Fingerprinting analysis of non-crystalline pharmaceutical compounds using high energy X-rays and the total scattering pair distribution function

Timur D. Davis

Submitted in partial fulfillment of the

requirements for the degree

of Doctor of Philosophy

in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2011

ABSTRACT

Fingerprinting analysis of non-crystalline pharmaceutical compounds using high energy X-rays and the total scattering pair distribution function

Timur D. Davis

In the development of new medicinal products, poor oral bioavailability, due to the low solubilities of many active pharmaceutical ingredients (APIs), is increasingly a barrier for treatments to be administered using tablet or capsule formulations and one of the main challenges facing the pharmaceutical industry. Non-crystalline phases such as the amorphous and nanostructured states can confer increased solubility to a drug, and therefore, have recently garnered a lot of interest from pharmaceutical researchers. However, little is known about local ordering in non-crystalline pharmaceuticals due to the lack of reliable experimental probes, hindering the clinical application of these compounds. The powerful tools of crystallography begin to lose their potency for structures on the nanoscale; conventional X-ray powder diffraction (XRPD) patterns become broad and featureless in these cases and are not useful for differentiating between different local molecular packing arrangements.

In this thesis, we introduce the use of high energy X-rays coupled with total scattering pair distribution function (TSPDF) and fingerprinting analysis to investigate the local structures of non-crystalline pharmaceutical compounds. The high energy X-rays allow us to experimentally collect diffuse scattering intensities, which contain information about a sample's local ordering, in addition to the Bragg scattering available in conventional XRPD

experiments, while the TSPDF allows us to view the intra- and inter-molecular correlations in real space.

The goal of this study was to address some fundamental problems involving fingerprinting non-crystalline APIs using TSPDF in order to lay the groundwork for the proper use of the technique by the pharmaceutical community. We achieved this by developing the methodology as well as the exploring the scientific implications. On the methodology side, we introduced PDFGetX3, a new software program for calculating TSPDFs that simplifies the procedure and reduces user interaction. We also set a baseline for the minimum X-ray energy that is needed for fingerprinting analysis, which had implications on the type of X-ray diffractometers that can be used. On the science side, we investigated the local structures of nanocrystalline and amorphous materials as well mixtures containing crystalline and amorphous phases. First, we identified a non-crystalline sample of the mood-stabilizing drug carbamazepine as a nanocrystalline version of one of its polymorphs. Next, we found that amorphous forms created by spray drying and cryomilling a proprietary compound have the same local structure. Finally, we quantified the phase fractions of polymorphic and amorphous components in a sample of the antibiotic sulfamerazine that was recrystallizing from a cryomilling-induced amorphous state.

Table of Contents

List of Figures

List of Tables

Acknowledgments

It is with an optimistic, but heavy heart that I write these passages. Optimistic because I am excited to see what the future will bring, but heavy because I know that I will never again experience the freedom - both intellectual and physical - that I have the past four years. I want to acknowledge all of the people who have, directly or indirectly, helped me get to where I am today.

First and foremost, I want to thank my advisor, Prof. Simon Billinge. I can say, without any qualification, that one of the very best decisions that I have ever made was to join his research group. A picture of him needs to be in the dictionary as the definition for the word 'mentor'. His guidance, patience, and insight have allowed me to do something that I never thought possible. Of course, I also want to acknowledge other members of the Billinge Group. Pavol Juhás, in addition to answering my often foolish questions and helping me countless times, saved my harddrive (with this thesis on it) when all hope seemed lost. Christopher Farrow taught me the basics of object oriented programming and offered piercing insight into the mathematical nature of the PDF. Emil Božin and Milinda Abeykoon not only helped to organize synchrotron experiments, but also sat with me at the beamline, making data collection less lonely and less frustrating. I give my sincere thanks for fruitful conversation and cooperation with other members of the Billinge Group past and present, including Yingrui Shang, Jiwu Liu, Wenduo Zhou, Peng Tian, Kirsten Jensen, Hrishi Tiwari, and Xiaohao Yang.

The experiments in this thesis were done at two synchrotrons, beamline ID-11-B at

the Advanced Photon Source at Argonne National Lab, which is supported by the US Department of Energy through grant No. W-31-109-ENG-38, and bemaline X7-B at the National Synchrotron Light Source, Brookhaven National Laboratory, which is supported by the US Department of Energy, Division of Materials Sciences and Division of Chemical Sciences, under contract No. DE-AC02-98CH10886. I am grateful to Peter Chupas, Jon Hanson, and their respective groups for their help in facilitating data collection.

I also want to thank all of my wonderful collaborators including Alastair J. Florence, Ryan Taylor, Kenneth Shankland, Matthew Johnson, and Milen Gateshki for providing samples and data. A special thanks to Matthew Johnson for inviting me to London and being a fantastic host.

I also want to acknowledge members of my thesis committee, Prof. James Im, Prof. Chris Marianetti, Prof. Sanat Kumar, and Prof. Xiangyun Qiu, for their guidance during my proposal and their help throughout this process.

I want to acknowledge all of the wonderful friends that I have made over the last four years. There are too many to list them all here, but in particular I want to mention Guoji Zhu for being a tremendous friend and confidant and leading me to fields that I had never considered, Kevin Tang for being my first friend in college and staying in touch after all these years, Arseniy Kukanov and Olga Kolesnikova for being an all around great friends, Dory Kramer for commiserating with me in class, studying with me for the quals, and letting me use his fume hood for sample preparation, Miki Hayano for letting me use equipment and dry ice, and Eugene Braude for always having time to talk with me. I also want to acknowledge Jonathan Stark and all members of EGSC for giving me something to do outside of research. I want to give a very special acknowledgement to Mi Wang. Her support has kept me on the right path for the last two years, she is one of the most special people in my life, and I am lucky to have her.

Finally, I want to thank my family and especially my parents. They have given me incredible love, guidance, support, and advice for many years. Their sacrifices for me will

always be remembered and everything that I have achieved and become, I owe to them.

To Maya and Peter

Part I

Main Text

Chapter 1

Introduction to the Pair Distribution Function

Our natural curiosity has motivated us, throughout the ages, to study the properties of the world, as the Roman poet Virgil (70 - 19 BCE) wrote, "Happy is he who gets to know the reasons for things". Until the late 19th century, light was used to probe the world around us - first via macroscopic observation, and later with the microscope [Hooke, 1665]. The optical microscope, though, is resolution limited. Since it uses visible light as a probe, it is unable to discern features that are smaller in size than the wavelength of light, roughly 400 nm. For this reason, it is difficult to overstate the importance of Röntgen's discovery in 1895 of the X-ray [Stanton, 1896].

X-rays, like visible light, are electromagnetic radiation. However, the wavelength of X-rays is several orders of magnitude shorter than visible light. Although X-ray wavelengths can be as long as 10 nm, the X-rays that are relevant to this thesis are in the range of 0.01-0.15 nm, or 0.1-1.5 Å [Guinier, 1963].

X-rays have become ubiquitous in modern society. Initially, they were used for medical applications such as imaging bone, owing to the fact that X-rays pass through skin and flesh but are absorbed by bones [Pasveer, 1989]. In additional to the medical appli-

cations, thanks to the pioneering work of Bragg [Bragg and Bragg, 1913], Ewald [Ewald, 1921], Laue [Friedrich *et al.*, 1913], Compton [Compton, 1923], and many others, X-ray crystallography has become an indispensable tool in the laboratory owing to the fact that X-ray wavelengths are roughly on the same scale the distances between atoms. These days, X-rays are frequently used to characterize inorganic compounds such as ceramics and metals [Warren, 1990], and organic compounds like small molecules [Lowes *et al.*, 1987], polymers [Cullity and Stock, 2001], and proteins [Frauenfelder *et al.*, 1979].

1.1 Experimental Basics

This thesis assumes that the reader has a basic understanding of the main ideas of X-ray diffraction and crystallography, such as Bragg's Law, reciprocal space, and the Ewald sphere. A full treatment of these concepts is available elsewhere [Schwartz and Cohen, 1987; Cullity and Stock, 2001; Warren, 1990]. We will instead focus on the foundations of the pair distribution function (PDF).

X-ray diffraction experiments can be performed on single crystal and powder samples. For our discussions here, we will focus on X-ray powder diffraction (XRPD). In a typical XRPD experiment, a beam of X-rays with a known energy and wavelength is directed at the sample, whose atoms scatter the beam at some scattering angle, 2θ. A detector measures the intensity of the scattered beam as a function of the scattering angle and the scientist uses these data to study the structure of the sample [Egami and Billinge, 2003]. The technique we discuss here is often referred to as *total scattering* to highlight the fact that the intensities collected at the detector include not just elastic Bragg scattering, but also elastic diffuse scattering and inelastic scattering and all of reciprocal space is measured [Egami and Billinge, 2003]. Diffuse scattering, which occurs in disordered materials where diffracted intensities can be observed in all directions [Dinnebier and Billinge, 2008], is particularly important in the study of amorphous and nanostructured pharmaceuticals and will be

discussed later.

For elastic scattering, if we define the incident wavevector as $\vec{\mathbf{k}}_i$ and the scattered wavevector as $\vec{\mathbf{k}}_f$, then the difference between the two is

$$\vec{\mathbf{Q}} = \vec{\mathbf{k}}_i - \vec{\mathbf{k}}_f \tag{1.1}$$

where $\vec{\mathbf{Q}}$ is the so-called momentum transfer or diffraction vector. Since momentum is conserved, $k = |\vec{\mathbf{k}}_i| = |\vec{\mathbf{k}}_f| = \frac{2\pi}{\lambda}$. Thus, Equation (1.1) becomes

$$Q = |\mathbf{Q}| = 2k\sin(\theta) = \frac{4\pi\sin(\theta)}{\lambda} \tag{1.2}$$

where θ is half of the scattering angle and λ is the wavelength of the X-rays, which is dependent on the energy of the X-rays used in the experiment and is a known quantity [Egami and Billinge, 2003].

Excellent formal derivations of the PDF from first principles are readily available [Farrow and Billinge, 2009; Egami and Billinge, 2003; Dinnebier and Billinge, 2008], Farrow in particular does a thorough analysis, and will not be re-derived in full detail here. However, Sections 1.3- 1.6 do provide a high level outline of the PDF derivation. First, though, we introduce the total scattering experiment.

1.1.1 Detectors

As mentioned previously, the detector measures the intensity of the scattered beam of X-rays as a function of the scattering angle. Generally, either 1-dimensional point detectors or 2-dimensional linear or image plate area detectors are used [Chupas *et al.*, 2003]. 1-D detectors are often used in laboratory-based X-ray diffractometers. Generally, they move along a fixed axis in an arc collecting data at pre-specified values of the scattering angle, 2θ. Since it is possible to define the values of 2θ at which measurements should occur, 1-D detectors can collect high resolution data. However, this also makes them slow. 2-D detectors are used at synchrotrons and in some lab-based diffractometers. 2-D linear detectors are detector

strips in an arc around the sample, while 2-D image plates are rectangular detectors are located perpendicular to the incident X-ray beam at some known distance away from the sample. Both types of 2-D detectors are able to catch X-rays from many scattering angles at once, thereby making collection time relatively quick [Chupas *et al.*, 2003]. However, 2-D detectors can be resolution limited depending on the pixel size of the image plate. The experiments discussed in this thesis, unless otherwise stated, were performed with 2-D area detectors.

In an XRPD experiment, the scattered X-rays appear on the 2-D detector as rings known as Debye-Scherrer rings. For an ideal powder, by which we mean the grains are sufficiently small relative to the X-ray beam and randomly oriented, the intensity around the ring is uniform [Dinnebier and Billinge, 2008]. A typical Debye-Scherrer ring pattern of a fine crystalline powder is in Figure 1.1. The distance of each ring from the origin can be related to the scattering angle, 2θ, because the distance from the sample to the detector is known. Software, such as Fit2D [Hammersley, 1998], can be used to integrate the intensities of the rings as a function of 2θ, as we see in Figure 1.2.

We note that in Figure 1.1 intensity of the rings (i.e. brightness) drops off with increasing 2θ. This is due to the behavior of the atomic scattering factor, $f(Q)$, and the Debye-Waller factor, both of which will be discussed later.

The XRPD pattern in Figure 1.2 was created from data collected using a 2-D image plate detector at a synchrotron, but a similar plot of intensity as a function of 2θ could be created using a 1-D detector.

1.2 Data Analysis

The first step to transforming an experimental dataset into a PDF is to convert from 2θ-space into Q-space [Egami and Billinge, 2003]. If we assume that the sample is isotropic, only the magnitude of Q, and not the direction, is important [Billinge, 2008a]. Since the

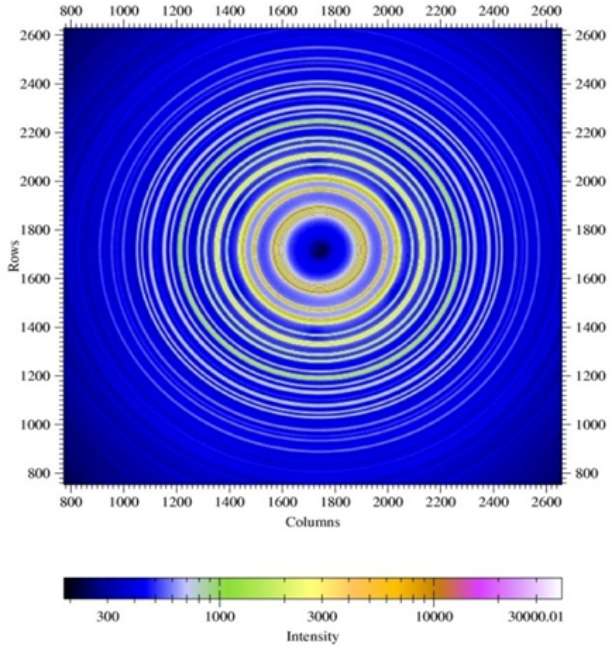

Figure 1.1: Debye-Scherrer ring pattern from a powder.

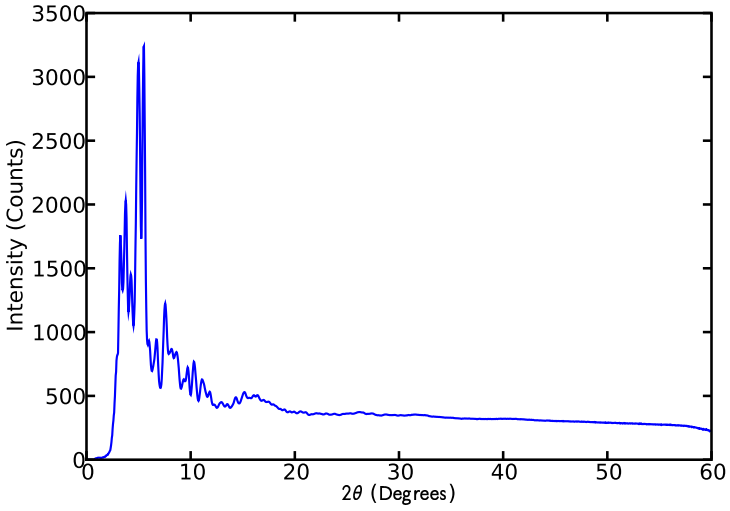

Figure 1.2: XRPD pattern from integrating rings of a Debye-Scherrer pattern.

wavelength of the X-rays is known, Equation (1.2) is used for the conversion, where θ is half the scattering angle, $\frac{2\theta}{2}$, and λ is the wavelength of the X-ray. Here we note that since $\sin(\theta) \leq 1$, the maximum Q accessible in an experiment strongly depends on λ. For example, a copper K_α diffractometer, which is widely used in the laboratory environment, has $\lambda = 1.54$ Å, so its available $Q_{max} \sim 8$ Å^{-1}. On the other hand, a synchrotron used for PDF analysis, which can produce X-rays with wavelength on the order of $\lambda \sim 0.1$ Å^{-1}, is capable of $Q_{max} \sim 45$ Å^{-1} [Petkov *et al.*, 1999; I.-K. Jeong *et al.*, 2001]. As we will see throughout this chapter and in Chapter 6, the Q_{max} is a very important parameter because it affects the resolution of the PDF.

The experimentally collected intensities, $I_{exp}(Q)$, like those in Figure 1.2, consist of several components [Egami and Billinge, 2003]:

$$I_{exp}(Q) = I_C(Q) + I_I(Q) + I_{MC}(Q) + I_{BG}(Q) \tag{1.3}$$

where $I_C(Q)$ is the coherent scattering intensity, $I_I(Q)$ is the incoherent scattering intensity, $I_{MC}(Q)$ is the multiple-scattering intensity, and $I_{BG}(Q)$ is the background scattering intensity [Egami and Billinge, 2003]. The incoherent scattering occurs when there is no phase relationship between the scattered waves, meaning that it does not contain structural information. Multiple scattering occurs when X-rays are re-scattered within the sample. Also, the sample container (i.e. a capillary), can contribute to the multiple scattering. Background scattering can be due to the sample container or the environment [Egami and Billinge, 2003; Chupas *et al.*, 2003].

Structural information is in the coherent scattering intensity, $I_C(Q)$. Therefore, before we can use experimental data for analysis, we must correct the $I_{exp}(Q)$ by removing incoherent scattering, multiple scattering, and background scattering components. The background scattering correction is performed by collecting data on only the sample container and then subtracting them from the total experimental data. For instance, if the sample was loaded in a capillary, then data would be collected on an empty capillary. Background data can

also be collected on the sample environment, (i.e. air). Depending on collection time and flux, the background data might need to be scaled prior to subtraction. Corrections for multiple scattering and incoherent scattering have typically been calculated by software such as PDFGetX2 [Qiu *et al.*, 2004]. These corrections will be discussed in greater detail in Chapter 3.

Sections 1.3- 1.6 offer an analytical derivation of the PDF starting from the definition of the coherent scattering cross section. Although the derivation is not necessary for an understanding of PDF, it does provide deep insight into the technique.

1.3 Coherent Scattering Intensity

First, we introduce the concept of the coherent scattering cross section, $\frac{d\sigma_C}{d\Omega}$ where σ_C is the coherent cross section and $d\Omega$ is a solid angle. The coherent scattering cross section, which has units of area, is the probability of a particle being scattered into some solid angle element $d\Omega$. It can be thought of as the area around a single atom that would scatter an incident X-ray photon with some given solid angle relative to the detector [Egami and Billinge, 2003]. In other words it is the probability per single atom that it will coherently scatter in some known direction.

After the coherent scattering intensity is corrected for absorption, polarization, and incident flux, which will be discussed later, we find that it is actually the same as the coherent scattering cross section [Egami and Billinge, 2003]. Therefore, we want to use the definition of the coherent scattering cross section above to find an expression for the coherent scattering intensity, $I_C(Q)$. A much more complete treatment of the following derivations is available in books [Egami and Billinge, 2003; Billinge, 2008a] and a paper [Farrow and Billinge, 2009], although here we try to explain important parts in an easy to understand, qualitative way.

First, we define the sample scattering amplitude

$$\Psi(\mathbf{Q}) = \frac{1}{\langle f(Q)\rangle}\sum_{\nu} f_{\nu}(Q)\mathrm{e}^{i\mathbf{Q}\mathbf{R}_{\nu}} \tag{1.4}$$

where $\Psi(\mathbf{Q})$ is the wavefunction of the scattered wave, $f(Q)$ is the atomic scattering factor for X-rays, i is the imaginary number, $\mathbf{Q}$ is the momentum transfer, $\mathbf{R}_{\nu}$ is the position of the ν-th atom, and we sum over all atoms.

From the first Born approximation [Egami and Billinge, 2003]

$$\frac{d\sigma_C}{d\Omega} = \langle V\rangle^2|\Psi(\mathbf{Q})|^2 \tag{1.5}$$

where $\langle V\rangle$ is the average scattering potential of an atom. In our case, $\langle V\rangle = \langle f(Q)\rangle$, where $f(Q)$ is the aforementioned atomic scattering factor and $\langle ...\rangle$ is a compositional average [Egami and Billinge, 2003]. We can imagine $\langle f(Q)\rangle$ and $\langle f(Q)\rangle^2$ to be the amplitude and the intensity of a solitary photon scattered by one atom in a material, respectively. In other words, the $\langle f(Q)\rangle^2$ is the average scattering power per atom [Billinge, 2008a]. The behavior and shape of the atomic scattering factor will be discussed in more depth later, but mathematically we can express the $\langle f(Q)\rangle^2$ as

$$\langle f(Q)\rangle^2 = \frac{1}{N^2}\sum_{i,j} f_j^* f_i \tag{1.6}$$

where $f*$ represents the complex conjugate of the atomic scattering factor, N is the total number of atoms, and we sum across all atom pairs. From here onward, we simplify the notation by not always mentioning the Q dependence of f, although it is obviously dependent on Q throughout the derivation.

When we square the wavefunction of the scattering amplitude and the atomic scattering factor, we have to take the complex conjugate. Therefore, substituting Equations (1.4)

and (1.6) into (1.5), we get

$$\frac{d\sigma_C}{d\Omega} = \left(\frac{1}{N^2}\sum_{i,j} f_j^* f_i\right)\left(\frac{1}{\frac{1}{N^2}\sum_{i,j} f_j^* f_i}\sum_{i,j} f_j^* f_i \mathrm{e}^{i\mathbf{Q}\cdot(\mathbf{R}_j-\mathbf{R}_i)}\right) \tag{1.7}$$

$$= \sum_{i,j} f_j^* f_i \mathrm{e}^{i\mathbf{Q}\cdot(\mathbf{R}_j-\mathbf{R}_i)} \tag{1.8}$$

As mentioned previously, the measured coherent scattering intensity, $I_C(Q)$ is related to the $\frac{d\sigma_C}{d\Omega}$ modified by the absorption and polarization factors, A and P respectively. We will talk about the corrections in detail in Chapter 3, but, for now we substitute into (1.8).

$$I_C(\mathbf{Q}) = AP\frac{d\sigma_C}{d\Omega} \tag{1.9}$$

$$I_C^{corr}(\mathbf{Q}) = \frac{I_C(\mathbf{Q})}{AP} \tag{1.10}$$

$$I_C^{corr}(\mathbf{Q}) = \sum_{i,j} f_j^* f_i \mathrm{e}^{i\mathbf{Q}\cdot(\mathbf{R}_j-\mathbf{R}_i)} \tag{1.11}$$

In Equation (1.10) we defined a new function, $I_C^{corr}(\mathbf{Q})$ which is the coherent scattering cross section corrected for the absorption and polarization factors, A and P, respectively. We defined this new function because we do not want to carry the A and P through the rest of the derivations and also to stay consistent with the notation introduced in Equation (1.3). We still need to keep in mind, though, throughout the rest of this derivation that the absorption and polarization corrections are important for obtaining a proper $I_C(\mathbf{Q})$.

Finally, we account for self-scattering where $i = j$ (i.e. the same atom), and so $\exp^{i\mathbf{Q}(\mathbf{R}_i-\mathbf{R}_i)} = 1$. Equation (1.11) becomes

$$I_C^{corr}(\mathbf{Q}) = \sum_i f_i^* f_i + \sum_{i\neq j} f_j^* f_i \mathrm{e}^{i\mathbf{Q}\cdot(\mathbf{R}_j-\mathbf{R}_i)} \tag{1.12}$$

$$= N\langle f^2\rangle + \sum_{i\neq j} f_j^* f_i \mathrm{e}^{i\mathbf{Q}\cdot(\mathbf{R}_j-\mathbf{R}_i)} \tag{1.13}$$

where N is the total number of atoms.

1.4 The Total Scattering Structure Function

We now have isolated the coherent scattering intensity, $I_C(Q)$, but we cannot use it directly to study structures. Instead, we prefer to work with a normalized version of the scattering intensity. We normalize it by dividing by the number of atoms, N, and the average scattering per atom, $\langle f \rangle^2$, this is important for the high-Q behavior of the structure function that we see in Section 1.4.1 [Billinge, 2008a; Farrow and Billinge, 2009].

Equation (1.13) becomes

$$\frac{I_C^{corr}(\mathbf{Q})}{N} = \langle f^2 \rangle + \frac{1}{N} \sum_{i \neq j} f_j^* f_i \mathrm{e}^{i\mathbf{Q}\cdot(\mathbf{R}_j - \mathbf{R}_i)} \tag{1.14}$$

$$\frac{I_C^{corr}(\mathbf{Q})}{N\langle f \rangle^2} = \frac{\langle f^2 \rangle}{\langle f \rangle^2} + \frac{1}{N\langle f \rangle^2} \sum_{i \neq j} f_j^* f_i \mathrm{e}^{i\mathbf{Q}\cdot(\mathbf{R}_j - \mathbf{R}_i)} \tag{1.15}$$

$$\frac{I_C^{corr}(\mathbf{Q})}{N\langle f \rangle^2} - \frac{\langle f^2 \rangle}{\langle f \rangle^2} = \frac{1}{N\langle f \rangle^2} \sum_{i \neq j} f_j^* f_i \mathrm{e}^{i\mathbf{Q}\cdot(\mathbf{R}_j - \mathbf{R}_i)} \tag{1.16}$$

We define the quantity in Equation (1.16) as $S(\mathbf{Q}) - 1$ where $S(\mathbf{Q})$ is the total scattering structure function, or just structure function for short [Egami and Billinge, 2003]. Substituting this into Equation (1.16)

$$S(\mathbf{Q}) - 1 = \frac{I_C^{corr}(\mathbf{Q})}{N\langle f \rangle^2} - \frac{\langle f^2 \rangle}{\langle f \rangle^2} \tag{1.17}$$

$$S(\mathbf{Q}) = \frac{I_C^{corr}(\mathbf{Q}) - N(\langle f^2 \rangle + \langle f \rangle^2)}{N\langle f \rangle^2} \tag{1.18}$$

Finally, we pull out the N by defining $I(\mathbf{Q}) = I_C^{corr}(\mathbf{Q})/N$, which also means that $I(\mathbf{Q}) = I_C(\mathbf{Q})/NAP$ [Farrow and Billinge, 2009], so equation (1.18) is simplified to

$$S(\mathbf{Q}) = \frac{I(\mathbf{Q}) - \langle f^2 \rangle + \langle f \rangle^2}{\langle f \rangle^2} \tag{1.19}$$

The $S(\mathbf{Q})$ is the normalized scattering intensity from a sample, it includes both Bragg and diffuse scattering [Egami and Billinge, 2003]. In general, we assume that scattering from

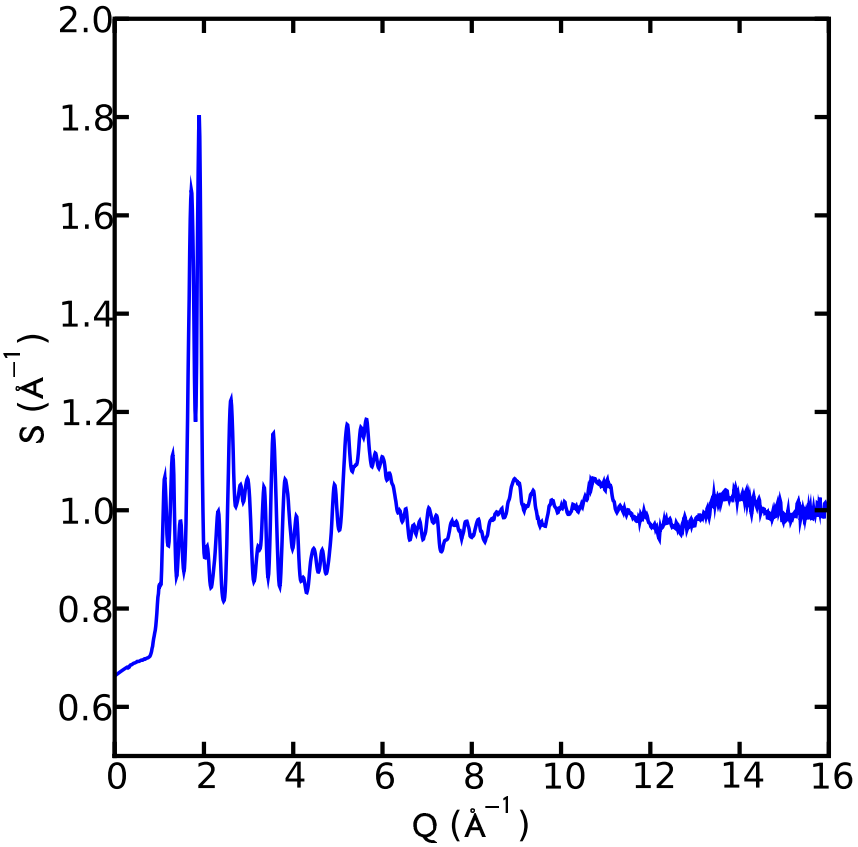

Figure 1.3: Example of a typical $S(Q)$. Created from the $I(Q)$ data in Figure 1.2. Note the Bragg scattering at low-to-mid Q and diffuse scattering at high Q.

a powder is isotropic, and depends on the magnitude of the wavevector, not the direction. To cement this point, we will no longer use the vector notation for Q. A typical-looking $S(Q)$, created from the $I(Q)$ data in Figure 1.2 is in Figure 1.3.

1.4.1 Properties of the Structure Function

The $S(Q)$ has a few interesting features that we would like to briefly touch on, using Figure 1.3 as a reference. First, we note that the function oscillates around 1, and, in fact, the average value $\langle S(Q) \rangle = 1$ [Billinge, 2008a]. The behavior of $S(Q)$ as $Q \to \infty$ is an

important phenomenon that will be discussed later. Also, we see the intensity of peaks decreases at higher values of Q. This is due to the Debye-Waller factor, which reduces intensities of Bragg scattering due the thermal motions of atoms [I.-K. Jeong *et al.*, 1999]. Finally, we note that the $S(Q)$ contains both Bragg intensities, in the form of sharp peaks in the Q-range up to about 12 Å^{-1} and broad, diffuse intensities in the Q-range 12-16 Å^{-1}. The Q_{max} value, 16 Å^{-1}, was not arbitrarily chosen. In fact, the Q_{max} is a very important parameter in data processing, which will be discussed in detail later. In general, we want to have a high Q_{max} [Dykhne *et al.*, 2011], which is why we prefer to use high-energy XRPD instruments such as synchrotrons. However, we need to be cognizant of the fact that noise in the data collection has a strong effect at high Q because the $\langle f \rangle^2$ (by which we divided the $I_C^{corr}(\mathbf{Q})$) amplifies oscillations in the signal because it approaches 0 at high Q. Therefore, long collection times or high X-ray flux might be necessary for good signal-to-noise ratio.

1.5 Reduced Structure Function

In practice, for reasons that will become obvious below in Section 1.6, we prefer to discuss total scattering data in terms of the so-called reduced structure function, $F(Q)$, as described in Equation (1.20):

$$F(Q) = Q(S(Q) - 1) \tag{1.20}$$

As we see in equation (1.20), the transformation between $S(Q)$ and $F(Q)$ is trivial, and indeed all of the structural information present in one is present in the other. Figure 1.4 contains a plot of the $F(Q)$ corresponding to the $S(Q)$ of the data in Figure 1.3.

As expected from Equation (1.20), the $F(Q)$ is centered around 0 and $\langle F(Q) \rangle = 0$.

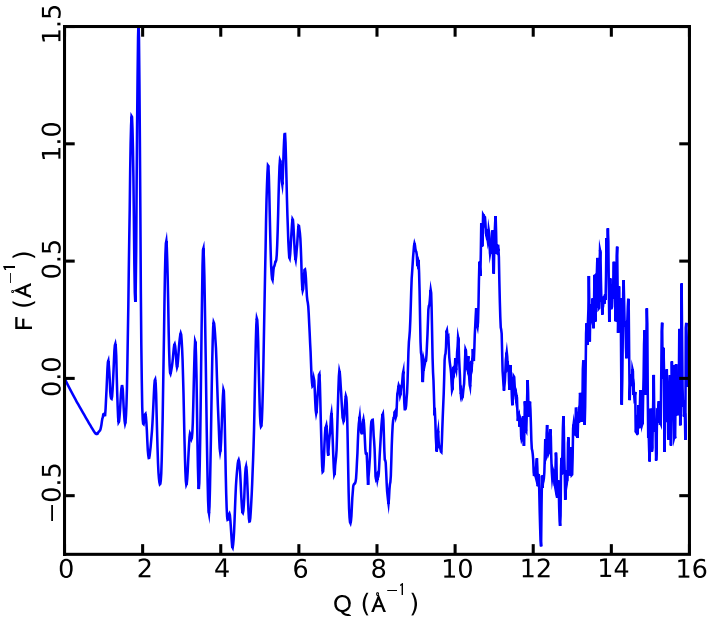

Figure 1.4: Example of a typical $F(Q)$. Created from the $S(Q)$ data in Figure 1.3.

1.6 Pair Distribution Function

Finally, we are ready to introduce three flavors of the pair distribution function - the atomic pair distribution function, $g(r)$, the reduced pair distribution function, $G(r)$, and the radial distribution function, $R(r)$. Luckily, the sine Fourier transform of the normalized scattered intensity, $S(Q)$, leads us directly to the all three functions [Egami and Billinge, 2003]. We use the sine transform because the $S(Q)$ is an even function [Farrow and Billinge, 2009].

$$G(r) = 4\pi r\rho_0(g(r) - 1) = \frac{2}{\pi}\int_0^\infty Q[S(Q) - 1]\sin(Qr)dQ \tag{1.21}$$

$$R(r) = 4\pi r^2\rho_0 g(r) \tag{1.22}$$

using the relationship in (1.20),

$$G(r) = \frac{2}{\pi}\int_0^\infty F(Q)\sin(Qr)dQ \tag{1.23}$$

where ρ_0 is the average number density of the material (i.e. the number of atoms per unit cell). It is easy to transform between the $g(r)$, $G(r)$, and $R(r)$ by simply adding and

multiplying constants.

From Equation (1.23), we instantly see the reasoning behind the introduction of the $F(Q)$ in Equation (1.20) - although the $F(Q)$ in and of itself does not provide any new structural information, it is nonetheless the function that we Fourier transform, so it makes sense in practice to work with it.

The $g(r)$ is, in effect, a histogram of the atom-atom distances in a material [Egami and Billinge, 2003]. It gives the probability of finding two atoms separated by some distance r inside of the sample. Each of the peaks in the PDF can be attributed to an atom-atom correlation, and the relative intensities of the peaks is related to how many such correlations are in the sample, meaning that taller peaks correspond to more atoms that distance apart in the sample.

We introduced the $R(r)$ in Equation (1.22) mostly for completeness and illustrative purposes. The RDF is rarely used in our analysis because it diverges like r^2, so the data are difficult to analyze when plotted [Egami and Billinge, 2003]. However, the RDF is the most physically intuitive of the correlation functions [Egami and Billinge, 2003] because it gives the number of atoms in an annulus of thickness dr at distance r from another atom, as shown in Figure 1.5, which illustrates an idealized $R(r)$ (right panel) for one atom at the origin of a two dimensional lattice (left panel, represented by a red 'x'). The closest atoms to the origin are the four atoms which are 2 Å away, as marked by the red circle in the left panel of Figure 1.5. The RDF, in response, has a peak at 2 Å. Going a bit further out, the next nearest neighbors also happen to be four atoms, now at a distance of $2\sqrt{2} \approx 2.8$ Å from the origin. We note that both peaks in the RDF have the same relative height, given that they correspond to the same number of atom pairs. Four more atoms are found 4 Å from the origin, and so we see another peak of the same height as the previous two. The next peak is at $2\sqrt{5} \approx 4.47$ Å from the origin. However, in this case, we note that the circle hits eight atoms, and, accordingly, the peak in the RDF is twice as tall as the previous peaks. This is an idealized case. In reality, the background of the RDF would be increasing

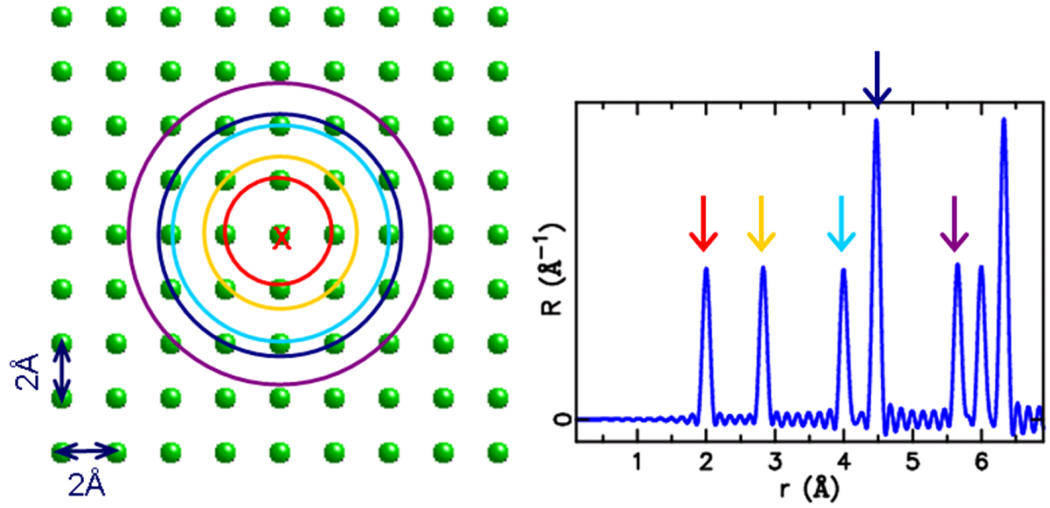

Figure 1.5: Idealized schematic of a radial distribution function. The rings in the left panel correspond to peaks at various r distances away from the origin in the right panel [Billinge, 2007].

as r^2 instead of being constant at 0.

The process of calculating partial PDFs or RDFs from atom-atom correlations, as above, is repeated for each atom in the structure [Billinge, 2007]. In this case, since all atoms are separated by exactly 2 Å, they would each provide the same partial PDF. However, in a real material atoms can have different neighborhoods. Also, if a sample contains multiple components, then peak height is also correlated to the scattering power of different species of atom [Billinge, 2007].

In Equation (1.21) we defined both the atomic pair distribution function, $g(r)$ and the reduced pair distribution function, $G(r)$. The two functions contain the same structural information about the compound, and are effectively different by a factor of $4\pi\rho_0$. From Equation (1.21) we see that, as $r \to 0$, $g(r) \to 0$ while $G(r) \to -4\pi\rho_0$ [Egami and Billinge, 2003]. In general, we find that, the $G(r)$ has a background of a straight line with slope $-4\pi\rho_0$ at low r. Although, in this sense, the $G(r)$ is less physically intuitive, it is still the PDF function that is the most widely used [Egami and Billinge, 2003] for two reasons. The

first reason, as per Equation (1.21), is that the $G(r)$ is the direct Fourier transformation of the $S(Q)$. No additional information is necessary to convert between the two functions. The number density, ρ_0, is needed in order to obtain the $g(r)$. Of course, even though this value could be calculated if the crystal structure is known, or even estimated from the $G(r)$, it is still much easier and more philosophically satisfying to do the transformation with no additional *a-priori* knowledge. The second reason, which is a little bit more subtle, is that in the $g(r)$ the amplitudes of oscillations containing structural information fall off like $1/r$ [Egami and Billinge, 2003]. This means that uncertainties in the low-r region are more heavily weighted than the high-r region. In the $G(r)$, the uncertainties in the data are constant in r, so all of the data are equally weighted. This result becomes very important when trying to fit a model to the PDF or fingerprint the PDF against other PDFs. For these reasons, from this point onward, when we discuss the PDF, we are exclusively referring to the $G(r)$.

At this point, we want to differentiate between normal PDFs and total scattering PDFs (TSPDFs). In principle, both functions have the same theoretical underpinning and are processed in the same manner with the TSPDF accessing a larger range of Q. However, as we will see in Chapters 5 and 6, a large Q-range, beyond what is accessible to copper-based laboratory diffractometers is needed in order to positively fingerprint pharmaceutical compounds. Therefore, for the rest of this thesis, we will specifically refer to PDFs made with $Q_{max} \geq 8.5$ Å^{-1} as TSPDFs.

A real TSPDF, from the experimental data in Figure 1.4 is in Figure 1.6. This TSPDF is actually of the over-the-counter drug Tylenol® (which mostly consists of the active ingredient acetaminophen) that had been ground up into a powder and measured at the synchrotron (we will discuss this TSPDF in great detail in Chapter 9). First, we note that this TSPDF of a 'real' compound is much more complicated than the idealized RDF in Figure 1.5 (of course it's also a slightly different function). In general, it is difficult to get an understanding for a whole structure by simply observing a PDF. However, we can

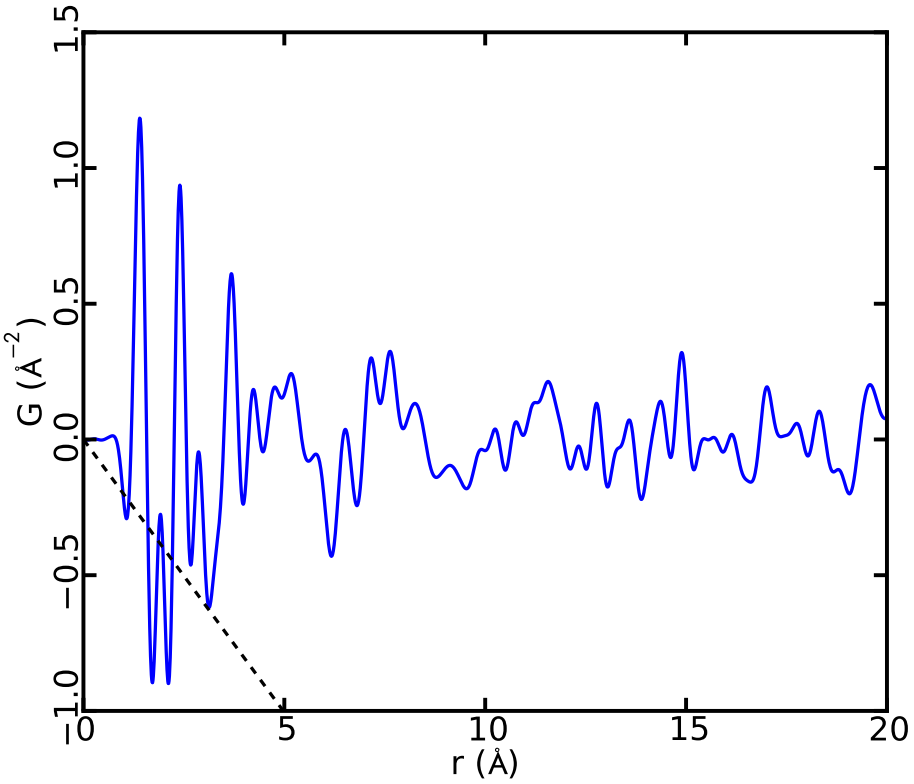

Figure 1.6: Example of a $G(r)$ of a pharmaceutical compound, in this case of Tylenol®. Created from the $F(Q)$ data in Figure 1.4. The dashed line represents the background with slope $-4\pi\rho_0$.

still visually observe some key features in Figure 1.6. For instance, the two most intense peaks are at 1.4 Å and 2.4 Å. These distances correspond to nearest and next-nearest neighbor carbon atoms in a carbon ring, and these features will be present in all carbon-based pharmaceutical compounds. As discussed earlier, we also note the linear baseline slope at low values of r as represented by the dashed line. The baseline should touch the bottoms of the first few peaks in the TSPDF. In this case, it slightly overshoots the bottom of the peak at 1.07 Å and just misses the bottom of the peak at 2.7 Å, but it is quite close. The peak at 1.9 Å, as we discuss below, is a termination ripple.

Unlike the Q-range and Q_{max} which have profound effects on the quality of a PDF, any range of r can be used, although the signal will naturally die out at some point due to the Debye-Waller factor. In Figure 1.6, we have set the r_{max} to be 20 Å, but this was an arbitrary decision and other values could have been used. In practice, though, since PDF if used to study local structure, r-ranges of 20 Å to 30 Å are used in the literature.

Peak broadness also plays an important role in the PDF. Theoretically, if there was no atomic disorder, then peaks would simply be delta functions [Egami and Billinge, 2003]. However, atomic disorder in the form of the thermal motion of atoms and static displacement of atoms away from ideal lattice sites in the case of crystalline materials results in a distribution of atom-atom distances and broadens the peaks into Gaussian distributions [Billinge, 2008a]. As we will see later, the effects of thermal motion and static displacement become important when dealing with disordered materials such as amorphous pharmaceutical compounds.

Not every single oscillation in a PDF carries structural information. For instance, in Figure 1.6 we see a very small peak at 1.9 Å, between the first two tall peaks at 1.4 Å and 2.4 Å. This peak is most likely a termination ripple. Termination ripples are not noise. Rather, they arise from the limited range over which the Fourier transform is performed. In Equation (1.21), we see that the Fourier transformation is being performed from 0 to ∞. In reality, of course, this is not the case. We do the integration from Q_{min} to Q_{max}. Oftentimes, the $Q_{min} = 0$ (although experimental intensities will start at higher values because scattered intensities will be blocked by the beamstop at extremely low Q), but the Q_{max} is always finite. The consequence of doing the finite Fourier transform is that termination ripples appear in the PDF with wavelength $\sim 2\pi/Q_{max}$ [Egami and Billinge, 2003]. For example, the PDF in Figure 1.6 was created with $Q_{max} = 16.0$ Å^{-1}, so $2\pi/Q_{max} \sim 0.39$ Å, and indeed the width of the peak at 1.9 Å is ~ 0.4 Å. The effect of the termination ripples can be reduced in a few ways including increasing Q_{max}, which would require increasing experimental collection time (to improve statistics) [Chupas *et al.*, 2007], and reducing

experimental artifacts by properly applying data corrections and good instrument alignment [Laaziri *et al.*, 1999]. In fact, as we will see later, if too low of a Q_{max} is used, then it will be difficult to tell the difference between oscillations coming from the structure and those coming from termination ripples [Dykhne *et al.*, 2011].

We want to again emphasize here that the pair distribution function is in real space, r, not reciprocal space, Q. Although this statement is quite obvious, the implications of it are huge. No information is gained or lost in the Fourier transform, in fact we could work exclusively in reciprocal space and fit the $S(Q)$ to a model, but, the Fourier transform allows us to measure relative positions of atoms in a solid - both from the periodic structure in the form of Bragg peaks, and the local structure in the form of diffuse scattering - something that regular crystallographic methods do not allow. Taking this one step further, *ab initio* methods have been developed that seek to build structures purely from the information contained in the PDF [Juhás *et al.*, 2006]. Going beyond the science, there is just something satisfying that arises from being able to directly relate real space atom-atom distances to their structure.

1.7 Applying the PDF

1.7.1 Modeling

One of the best ways to extract structural information out of a PDF is to try to reproduce the PDF using a calculated structure model by refining parameters such as atom positions, lattice constants, and thermal factors [Egami and Billinge, 2003]. The structure models are typically calculated using powerful software tools such as PDFfit2 and PDFgui [Farrow *et al.*, 2007]. We demonstrate several structural refinements using PDFgui on a variety of compounds in Chapter 4.

Amorphous pharmaceuticals are difficult to model using existing software methods because they only contain short range structural information. New modeling tools, that allow

for structures with 'rigid' and 'floppy' components need to be developed to face the challenges presented by these materials. However, these tools are beyond the scope of the work presented here. For this reason, traditional PDF modeling will not be used in our analysis of PDFs of pharmaceutical compounds.

1.7.2 Fingerprinting

Since tools for modeling TSPDFs of pharmaceutical compounds have not fully been worked out, we instead rely on fingerprinting them against one another. Fingerprinting can be used to provide unambiguous structural identification of a compound or to highlight possible structural changes over time that may lead to product performance, e.g. due to crystallization of an amorphous pharmaceutical [Dykhne *et al.*, 2011]. In fact, as we will learn later, an important practical barrier to the development of amorphous pharmaceuticals in product development is a lack of reliable methods for fingerprinting them [Dykhne *et al.*, 2011].

We do fingerprinting both qualitatively and quantitatively. For the qualitative analysis we simply use a visual comparison of the plots of the PDFs. For the quantitative analysis we quantify the similarity of curves by using a home-written program that computes the Pearson product-momentum correlation [Myers and Well, 2010]

$$\mathbf{R} = \frac{1}{1-n}\sum_{i=0}^{n}\left(\frac{X_i - \overline{X}}{\sigma_x}\right)\left(\frac{Y_i - \overline{Y}}{\sigma_y}\right) \tag{1.24}$$

where $\overline{X}$ and σ_x are the mean and standard deviation of a data set, respectively, and n is the total number of data points in the data set. The Pearson correlation technique creates an $m \times m$ matrix that contains a correlation value $\mathbf{R}$ in the range -1 to 1 between each pair of m data-sets. The value 1 implies complete correlation, zero implies no correlation, and -1 implies an anti-correlation. The Pearson correlation technique is extremely powerful because it ignores absolute scaling, but is sensitive to relative scaling and slight shifts in peak positions. Empirically, we have found that correlations between two PDFs above 0.8

mean that they are highly correlated.

We study the correlations between PDFs in the range $r = 3.0 - 20$ Å. We chose this range because the very local structure (i.e. $r < 3.0$ Å) of all molecular samples is similar due to intra-molecular atom pairs, for example, consisting of nearest and next-nearest neighbor carbon-carbon bonds at 1.4 Å and 2.4 Å, respectively. Applying the correlation analysis to the entire data range does not change the result significantly but reduces the sensitivity to finding differences in molecular packing of the correlation analysis by including a range of r-that is highly similar regardless of the packing.

A sample table of Pearson correlations is in Table 6.1, reproduced from Chapter 6. This table shows correlations between three phases of the pharmaceutical compound carbamazepine (CBZ) and three phases of the compound indomethacin (IND). Crystalline phases of CBZ are CBZ-I and CBZ-III, and crystalline phases of indomethacin are α and γ. In both cases, the non-crystalline phases are denoted with an 'a'. From the table, we see very high correlations between CBZ-III and CBZ-a (0.881) and no high correlations between any other pair of compounds. As we will see in Chapters 5 and 6, this led us to correctly identify the non-crystalline phase structure to be similar to that of CBZ-III. In fact, this was the first time that TSPDF was used to successfully identify the structure of a non-crystalline phase of a pharmaceutical compound [Billinge *et al.*, 2010].

1.8 The Rest of the Thesis

In this chapter, we briefly highlighted a typical total scattering experiment, the derivation of the PDF from first principles, the properties of a PDF, and how we use PDF to study structures. In the next chapter we will introduce the importance of non-crystalline pharmaceutical compounds. In Chapter 3 we will study the data corrections in more detail. In Chapter 4 we will discuss a new algorithm for creating PDFs that allows for high-throughput data processing with little user interaction. In Chapters 5, 6, 7, 8, and 9 we will show ex-

Table 1.1: Summary of Pearson correlation coefficients between the PDFs of carbamazepine and indomethacin with Q_{max} = 20 Å^{-1}. Correlations higher than 0.8 are shown in bold (except when they are trivially unity).

	CBZ-III	CBZ-a	CBZ-I	IND-α	IND-a	IND-γ
$Q_{max} = 20$ Å^{-1}						
CBZ-III	1	**0.88121**	0.580032	0.36072	0.520868	0.535466
CBZ-a		1	0.721854	0.499347	0.692577	0.585051
CBZ-I			1	0.4143	0.607663	0.353945
IND-α				1	0.706309	0.477629
IND-a					1	0.648231
IND-γ						1

amples of TSPDFs being used to study non-crystalline and mixed phase pharmaceuticals and then we will conclude in Chapter 10. Appendix A has safety guidelines for handling pharmaceutical compounds and Appendix B is the manual for PDFGetX3 (introduced in Chapter 4).

Chapter 2

Introduction to Pharmaceuticals

Pharmaceutical drugs are, in the most broad sense, chemical compounds used to treat an illness. Treatment with drugs, also known as active pharmaceutical ingredients (APIs), is the most frequently used and usually most cost-effective preferred method of treating disease [Banker and Rhodes, 2002]. Although drugs can be administered into the body using a variety of routes (e.g. orally, intravenously, rectally, nasally, subcutaneously, etc. [Banker and Rhodes, 2002]), here we will focus on drugs that are delivered orally in solid dosage forms such as tablets.

Although the pharmaceutical industry spends more money on medical research than do the National Institutes of Health in the United States [Lexchin *et al.*, 2003], a major problem facing the industry is that it is becoming harder to discover new drugs that could be brought to market. In fact, even as R&D expenditures continue to increase dramatically, fewer new products enter the market [Gardner *et al.*, 2004b]. A variety of explanations have been proposed to explain this phenomenon such as: poorly understood *in vivo* physico-chemical properties of pharmaceutical materials, poor control over materials properties, and poor aqueous solubility and dissolution rate [Gardner *et al.*, 2004b].

In this chapter, and indeed the thesis, we focus on the last of those explanations - the problem of poor aqueous solubility, or, to be more precise, bioavailability of pharma-

ceutical products. We begin by introducing some basic information about pharmaceutical compounds.

2.1 Basics of Pharmaceutical Drugs

As mentioned earlier, we confine our discussion here to APIs that are delivered orally in solid dosage forms such as tablets, which are generally compacted crystalline powders. Crystalline forms are used largely for reasons of physico-chemical stability and processibility [Dykhne *et al.*, 2011]. As discussed in Chapter 1, crystalline compounds can be characterized and fingerprinted using the tools of crystallography such as XRPD.

2.1.1 XRPD Fingerprinting

A fingerprint of an API is a very powerful tool. In addition to being used at the formulation stage to identify the structure of a compound or highlight structural changes [Dykhne *et al.*, 2011], the fingerprint is also used for litigation purposes [Shankland *et al.*, 2004], and for quality control [Maurin *et al.*, 2008]. XRPD is a common technique for fingerprinting crystalline APIs because it is easy and quick [Byrn, 1982].

A detailed discussion of the XRPD fingerprinting process is in Chapter 1, but, in summary, XRPD patterns collected from samples are either visually compared to reference XRPD patterns (which are calculated from models or obtained from databases) by plotting them on top of one another and identifying similar features, or quantitatively compared using software such as PolySNAP [Barr *et al.*, 2004], which uses algorithms to determine the degree of similarity between two XRPD patterns [Billinge *et al.*, 2010].

2.1.2 Polymorphism

Fingerprints play a particularly important role in the formulation stage when dealing with drugs that exhibit polymorphism.

A polymorph is a unique crystal packing arrangement of a compound [Byrn, 1982]. When an API is synthesized, great care is taken to determine whether or not the compound can exist in polymorphic crystal forms and to characterize any such polymorphs [Byrn, 1982]. Knowledge of the polymorphs is important because different forms can have different thermodynamic properties such as melting points and solubilities [Hancock and Zografi, 1997].

Since polymorphs of the same compound have different crystal packing arrangements, they also have different XRPD patterns [Byrn, 1982] and PDFs [Dykhne *et al.*, 2011], as we see in Figure 2.1.

Figure 2.1 shows XRPD patterns and PDFs of two polymorphs of the API carbamazepine (CBZ), called Form I and Form III (or CBZ-I and CBZ-III for short). Panels **(a)** and **(b)** show the XRPD patterns of CBZ-I and CBZ-III, respectively, while panels **(c)** and **(d)** show the TSPDFs of the those polymorphs. We see that the XRPD patterns look completely different, while the TSPDFs share strong peaks at 1.4 Å and 2.4 Å which we've mentioned before correspond to carbon-carbon nearest- and next-nearest neighbor distances. We also see that the PDFs share peaks at 3.7 Å and 4.4 Å. These two peaks correspond to short range distances where the crystals structures are the same (i.e. intramolecular distances). Beyond 4.4 Å, the PDFs look completely different from one another (due to inter-molecular distances).

Fingerprints of all known polymorphs of a compound are needed when trying to fingerprint an unknown formulation (particularly when dealing with amorphous and nanostructured APIs, as we will discuss later). Databases such as the Cambridge Crystal Structure Database (CCSD) [Allen, 2002] contain structure models of many known polymorphs for many compounds. Using structural models for fingerprinting will be briefly touched on in Chapter 8.

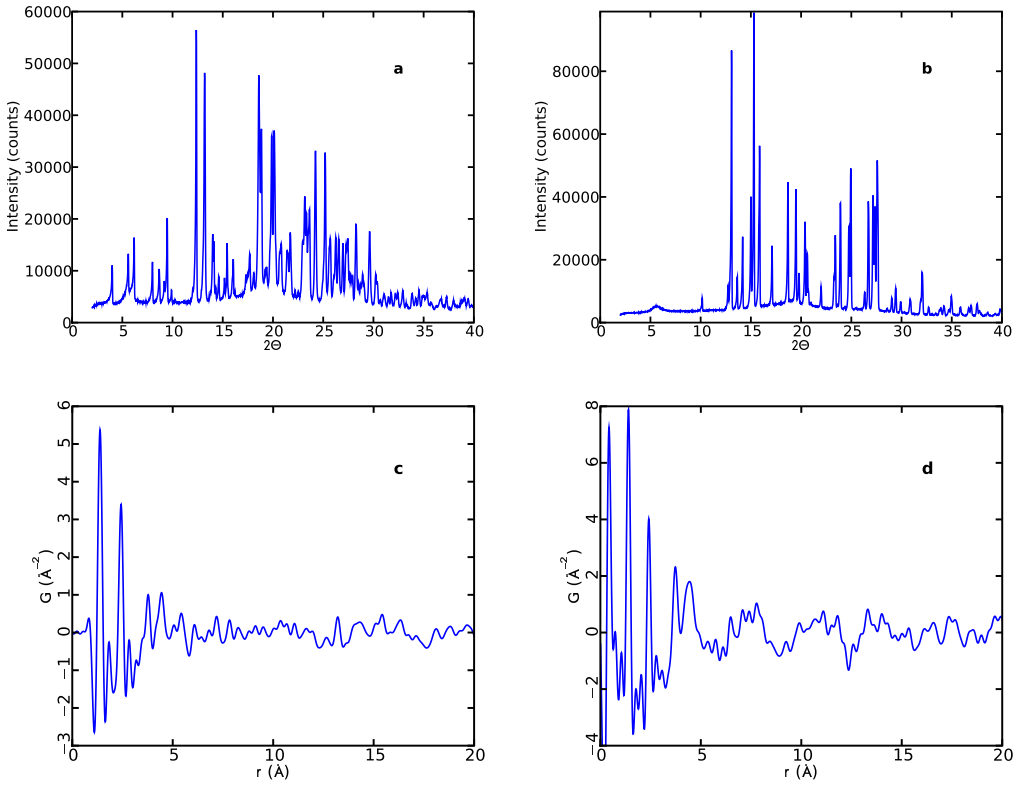

Figure 2.1: XRPD patters and PDFs of two polymorphs of the API carbamazepine (CBZ), CBZ-I and CBZ-III. Panels **(a)** and **(b)** contain the XRPD patterns of CBZ-I and CBZ-III, respectively. Panels **(c)** and **(d)** contain the TSPDFs of the two forms.

2.1.3 Excipients

A typical solid drug formulation that is delivered orally, i.e. a tablet, has two main components, the API, which is the focus of our discussion, and the excipient [Baldrick, 2000]. Excipients are the 'everything else', so to speak, that is inside of a tablet. Excipients are used for a variety of applications such as binders which hold the tablet together, coatings which protect the API from the environment, and fillers if an API needs to delivered in a small dosage [Baldrick, 2000; Rowe *et al.*, 2006].

Excipients tend to be polymeric compounds that lack long range structure and thus do not typically show up in the XRPD pattern [Hancock and Zografi, 1997]. In our investigations we tend to look at powders that contain purely APIs without any excipients for simplicity, and in this sense we ignore excipients for most of this thesis except at the end in Chapter 9 where we investigate some over-the-counter analgesic medications that contain both an API and excipient by fingerprinting a TSPDF of Tylenol® against TSPDFs of the three known polymorphs of its API, acetaminophen.

2.1.4 Bioavailability

Bioavailability is a pharmacokinetic property which measures the fraction of an administered API that reaches the circulatory system and then is not metabolized before it can exert its intended biological effect, in other words it is the degree to which drugs dissolve in their intended destination in the body [Byrn, 1982; Heaney, 2001]. Bioavailability is incredibly important because drug dissolution is a prerequisite to drug absorption and clinical response for almost all drugs given orally [Amidon *et al.*, 1995]. Obviously, a poor bioavailability means that a drug might not have any therapeutic effect.

Bioavailability can be measured using a variety of analytical methods including chemical methods such as chromatography and mass spectroscopy, biological methods such as those based on immunoassay procedures, and microbiological methods [Shah *et al.*, 1992].

A general approach is to first administer pharmacologically relevant dosages of an API intravenously via an IV infusion, then measure the concentration of the API in the blood at various time intervals using liquid chromatography mass spectroscopy (LC-MS). Next, after the API clears from the blood, the same dosage of the API is administered orally and measurements are made the same way using LC-MS at the same time points [Veber *et al.*, 2002]. Since the intravenously administered API is considered to be fully soluble, then the ratio of the blood concentration of the oral dosage to the intravenous dosage is the bioavailability. This value is typically plotted as a function of time [Veber *et al.*, 2002].

Although *in-vivo* bioavailability depends on many factors such as the fasted/fed state of the subject, gastric-emptying and intestinal transit, metabolism, etc. [Amidon *et al.*, 1995], we will focus on the aqueous solubility of the API as the main measurement of bioavailability.

In the development of new medicinal products, poor oral bioavailability, due to low solubility of many crystalline API candidates, is increasingly a potential barrier to enabling treatments to be administered using tablet or capsule formulations [Hancock and Parks, 2000]. One reason for the increasing occurrence of poor aqueous solubility is that new high-throughput screening techniques used for discovering new molecules often lead to large lipophilic candidates, which are poorly soluble in water [Gardner *et al.*, 2004a].

2.2 The Amorphous Phase

Compounds in the amorphous phase lack the three-dimensional long-range order that exists in crystalline materials [Hancock and Zografi, 1997]. In the case of amorphous pharmaceutical compounds, the positioning of molecules relative to one another is more random. Therefore, amorphous pharmaceuticals typically have short-range order (intra-molecular and extending to neighboring molecules), but no long-range order or well-defined molecular conformation if the constituent molecules are conformationally flexible [Yu, 2001].

Due to the lack of long-range order, amorphous systems are in a highly energetic state [Hancock and Parks, 2000]. This means that solubilities of amorphous phase compounds are generally expected to be higher than their crystalline versions. For this reason, the amorphous state is of significant interest within the pharmaceutical industry as a means to enhance the aqueous solubility of APIs [Hancock and Parks, 2000].

Several examples in the literature have shown that increases in solubility for various pharmaceuticals after they have been amorphized [Imaizumi *et al.*, 1980; Sato *et al.*, 1981] (please refer to [Hancock and Parks, 2000] for a more complete listing). An *in-vivo* study in rabbits found that amorphous indomethacin (IND) had a higher absorption rate over time than crystalline IND [Fukuoka *et al.*, 1987].

2.2.1 Amorphous Stability

The high potential energy of amorphous APIs also means that they are more unstable than their crystalline analogues and are prone to recrystallization [Yu, 2001]. This recrystallization can happen unexpectedly and with potentially disastrous results. For instance, the Pfizer drug atorvastatin (popularly known as Lipitor®) was formulated as an amorphous salt during the development phase, but then the salt crystallized during Phase III clinical trials. Additional studies had to be performed on the new crystalline compound because its properties had changed from the amorphous form that was originally developed [Gardner *et al.*, 2004b].

One interesting technique that has been proposed to stabilize amorphous phase APIs is by encapsulating them in polymer emulsions [Shaikh *et al.*, 2009; Kesisoglou *et al.*, 2007]. These stabilizers prevent them from aggregating [Kesisoglou *et al.*, 2007].

2.2.2 Breakdown of XRPD

Since traditional XRPD techniques rely on a sample having long range structure, the technique breaks down when it is applied to the amorphous phase [Hancock and Parks, 2000;

Dykhne *et al.*, 2011; Billinge *et al.*, 2010]. For example, Figure 2.2 shows the XRPD patterns collected from amorphous and crystalline phases of CBZ and IND. Amorphous CBZ and IND are in panels **(a)** and **(b)** while crystalline CBZ-I and IND-γ are in panels **(c)** and **(d)**. The amorphous forms of each compound were created by melt-quenching their respective crystalline phases. Unlike the XRPD patterns from crystalline data, the amorphous patterns have no unique, identifiable features. They cannot be used for fingerprinting or for extracting structural information.

The inability to characterize the local packing in amorphous compounds has long been a roadblock toward popularizing amorphous APIs and bringing them to market. However, as we will discuss throughout most of this thesis, the TSPDF method seems to offer a solution to this long-standing problem because of the local structural information available in the diffuse (non-Bragg) scattering.

2.2.3 Nanostructuring

Between the two extremes of completely amorphous and crystalline phases sit nanostructured APIs. Nanostructured compounds can be thought as being partially crystalline or partially amorphous and can be created by introducing some disorder into a crystalline compound using techniques such as milling or melt-quenching [Hancock and Parks, 2000]. These compounds are attractive because they offer increased bioavailability over purely crystalline compounds, and, unlike the purely amorphous compounds in Figure 2.2**(a)** and 2.2**(b)**, have some order. XRPD patterns of nanostructured compounds have some peaks, albeit they are broader and less intense than Bragg peaks [Hu *et al.*, 2002].

A bottleneck in the development of this class of APIs is that it is difficult to determine the degree to which a compound is nanostructured purely from XRPD patterns [Hu *et al.*, 2002]. However, as we will see in Chapter 5, TSPDFs can be used as a powerful tool to determine the domain size of the nanoparticles that constitute a nanostructured API.

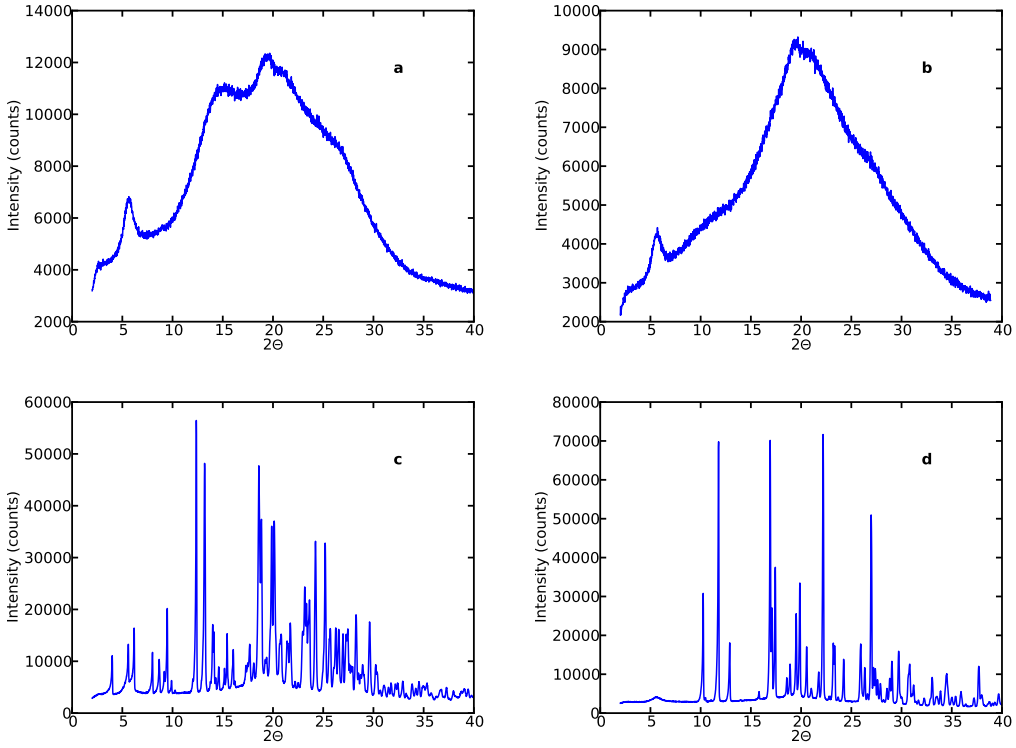

Figure 2.2: XRPD patters of amorphous and crystalline forms of carbamazepine (CBZ) and indomethacin (IND). Panels **(a)** and **(b)** contain the XRPD patterns of amorphous CBZ and IND, respectively. Panels **(c)** and **(d)** contain the XRPD patterns of the CBZ-I and IND-γ polymorphs. The amorphous forms of each sample were made by melt-quenching the respective crystalline forms.

Chapter 3

PDF Data Corrections

In Chapter 1, we made several references to various corrections applied to the experimental $I_{exp}(Q)$ during its transformation to the $S(Q)$. In this chapter, we discuss those corrections in more detail. We keep our discussion relatively qualitative and only show equations where analytical solutions exist and are relatively easily understood. Just as in Chapter 1, our goal here is not to exhaustively detail the step-by-step process of creating a PDF, but instead to provide insight into the process. Indeed, since these corrections are nowadays performed by software, it is more important to understand why these corrections are needed than their nitty-gritty details.

For this discussion, we re-introduce equations (1.3) and (1.9), now referred to as (3.1) and (3.2)

$$I_{exp}(Q) = I_C(Q) + I_I(Q) + I_{MC}(Q) + I_{BG}(Q) \tag{3.1}$$

$$I_C(Q) = AP\frac{d\sigma_C}{d\Omega} \tag{3.2}$$

From equations (3.1) and (3.2), there are five main corrections we consider, these corrections can be divided into two groups. Additive corrections (those in equation (3.1)),

such as incoherent scattering, multiple scattering, and background scattering, and multiplicative corrections (those in equation (3.2)) such as absorption and polarization. There are obviously more corrections that can be applied, such as oblique incidence, fluorescence, detector efficiency, incident flux, etc., but since the goal of this chapter is to be a survey, not a technical manual, they will not be considered here.

3.1 Multiplicative Corrections

3.1.1 Absorption

The absorption correction is used to account for the X-rays that are absorbed by the sample. To calculate the absorption correction, one needs to know the geometry of the experiment, the energy of the X-rays (this can be obtained given their wavelength with the relationship $E = \frac{hc}{\lambda}$ where E is the energy, h is Planck's constant and c is the speed of light in a vacuum), the mass absorption coefficient, $\mu(E)$ [Chantler, 1995], of the sample, the elements that make up the sample, its density, and its thickness.

Most necessary parameters for the absorption calculation are reasonably easy to obtain. The energy of the X-rays used in an experiment should be known ahead of time. The total mass absorption coefficient of a material is a linear addition of the absorption coefficients of each element weighted by the ratio of the elements in the compound. $\mu(E)$ of most elements has been tabulated as a function of X-ray energy in the literature [Chantler, 1995] and on the website of the National Institute of Standards and Technology (NIST) [Hubbell and Seltzer, 2004]. We usually express the mass absorption coefficient of each element divided by its density, $\mu(E)/\rho$. Luckily, the densities elements have also been tabulated [Klug and Alexander, 1974]. We multiply the total mass absorption coefficient by the thickness of the sample, t, to obtain the value μt, which is an important variable in the absorption calculation.

Finally, the geometry of the experiment also has an effect on the absorption. A more

detailed discussion can be found in [Egami and Billinge, 2003] and [Jeong *et al.*, 2001], but briefly flat plate geometries (both reflection and transmission) have analytic expressions for the absorption correction. The capillary geometry has a more complicated correction dependent on the diameter of the capillary.

The final absorption correction is tabulated by software using several complicated equations and then applied to the experimental data. In general, the magnitude of μt is inversely proportional to the magnitude of the correction, so a larger μt will mean that a smaller correction is necessary, and Q is proportional to the correction, so the absorption correction is bigger at larger values of Q.

3.1.2 Polarization Correction

The Thompson scattering process polarizes the scattered X-ray beam [Egami and Billinge, 2003] such that the electric vector lies in a direction perpendicular to the scattering plane. This affects the measured intensities as a function of the scattering angle [Egami and Billinge, 2003]. Unlike the other corrections described in this chapter, which rely on complicated functions with tabulated values, the polarization correction to account for this angle dependence can be expressed relatively easily analytically as a function of scattering angle and degree of polarization of the incident beam [Egami and Billinge, 2003], as below in Equation (3.3)

$$P = \frac{1 + A\cos^2 2\theta}{1 + A} \tag{3.3}$$

where A is a term that depends on the degree of polarization of the incident beam and 2θ is the scattering angle [Egami and Billinge, 2003]. A, not to be confused with the absorption correction in Equation (3.2), can take on values between 0 and 1.0, where $A = 0$ means that the incident beam is unpolarized, while $A = 1.0$ means that the incident beam is fully polarized.

Synchrotron radiation is theoretically fully plane polarized perpendicular to the scatter-

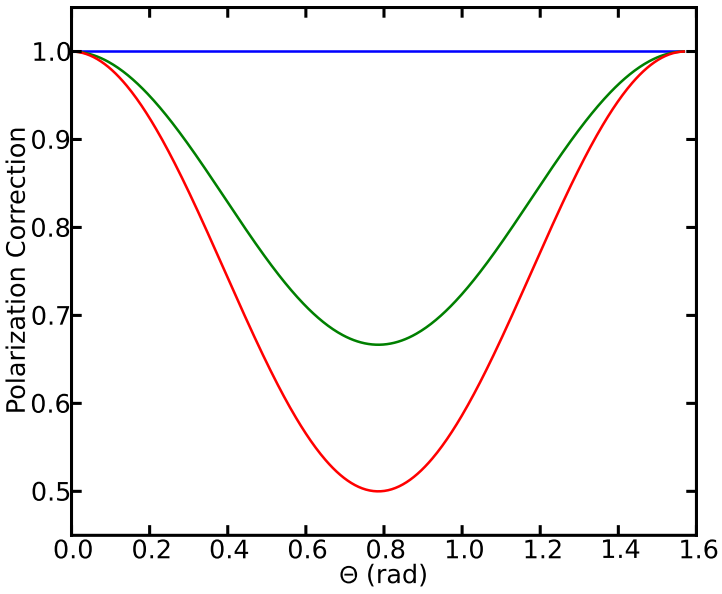

Figure 3.1: Plot of the polarization correction as a function of angle in radians. The blue curve corresponds to the incident beam being fully polarized perpendicular to the scattering plane ($A = 1$), the green curve to a half polarized incident beam ($A = 0.5$), and the red curve to an unpolarized incident beam ($A = 0$).

ing plane, and therefore does not need to be corrected for polarization. However, in reality, the beam is not completely polarized and some small corrections might be necessary [Egami and Billinge, 2003; Jeong *et al.*, 2001].

In Figure 3.1, we see the polarization correction plotted as a function of angle, θ, in radians with different values of A. The blue curve corresponds to a situation where $A = 1$, the green to when $A = 0.5$, and the red to $A = 0$. We see that when the beam is not fully polarized, the shape of the correction function is roughly the same: $P = 1$ at $\theta = 0$ and $\pi/2$ and reaches its minimum, $P = \frac{1}{1+A}$, at $\theta = \pi/4$. Overall, as A gets smaller, the magnitude of the polarization correction increases.

There are two special geometries that are worth a mention [Egami and Billinge, 2003].

In the first geometry, a monochromator crystal is placed in the unpolarized incident beam (i.e. placed in the beam before it hits the sample). In this case, $A = \cos^2 2\alpha$, where 2α is the scattering angle of the monochromator, in Equation (3.3). In the other geometry, the monochromator crystal is placed in the diffracted beam (i.e. after the beam hits the sample, this is also known as an analyzer). In this case, we use the equation

$$P = \frac{1 - \cos^2(2\alpha)\cos^2(2\theta)}{2} \tag{3.4}$$

instead of Equation (3.3).

3.2 Additive Corrections

3.2.1 Multiple Scattering

The multiple scattering contribution contains no usable structural inormation and must be removed from the measured intensity before we can obtain the $S(Q)$ [Egami and Billinge, 2003]. Multiple scattering occurs when the primary beam is scattered twice or more by the sample and container before reaching the detector [Egami and Billinge, 2003; Warren, 1990]. In our analysis, we consider only double scattering because it represents the majority of the contribution of multiple scattering to the data and higher orders are much more difficult to calculate [Egami and Billinge, 2003; Jeong *et al.*, 2001]. We also assume that the double differential cross section is isotropic, which is not the case in reality, however the assumption does simplify the math and has been found to work [Egami and Billinge, 2003]. We further simplify the calculation by separating out the incoherent scattering (i.e. Compton scattering), as done in equation (3.1) so that only completely elastic multiple scattering events need to be considered for this correction [I.-K. Jeong *et al.*, 2001].

The multiple scattering correction depend on sample thickness, transparency, absorption [Egami and Billinge, 2003] (a discussion of absorption is above in section 3.1.1), experiment geometry, and a complicated function that is dependent on Q, sample absorption,

thickness, density, and other parameters [Warren and Mozzi, 1966]. Luckily, values of this function have been tabulated so that they could be used in computer assisted calculations [C. W. Dwiggins, Jr and Park, 1971]. Multiple scattering contributions increase with sample thickness and with higher energy X-rays [I.-K. Jeong *et al.*, 2001] and therefore the correction is important in TSPDF experiments where X-ray energies are generally high and the data in the high-Q region contain important information [I.-K. Jeong *et al.*, 2001]. Finally, multiple scattering is far larger in transmission geometry than reflection geometry [Egami and Billinge, 2003].

3.2.2 Incoherent Scattering

Incoherently scattering waves do not interfere when diffracted off the sample (unlike coherent scattering where the waves interfere either constructively or destructively), meaning that there is no phase relationship between the waves [Egami and Billinge, 2003]. Since structural information in scattering comes from the interference effects of scattered waves, incoherently scattered waves do not contain any structural information [Egami and Billinge, 2003].

As incoherent scattering mainly consists of Compton scattering [Egami and Billinge, 2003], we devote our discussion exclusively to it. Compton scattering is an inelastic incoherent type of scattering that occurs when an incident X-ray photon collides when an electron in the sample and transfers some of its energy to the electron. The scattered photon thus has a different energy and wavelength than the incident [Egami and Billinge, 2003].

We correct for Compton scattering by subtracting Compton intensities from the data. These intensities have been calculated as a function of Q from tabulated data of the Compton scattering cross section [Balyuzi, 1975; Cromer and Mann, 1967]. A proper correction for Compton scattering is very important in high-energy total scattering analysis because the Compton signal can be larger than the coherent scattering intensity at high-Q [I.-K. Jeong *et al.*, 2001]. This is due to the fact that the elastic scattering cross section decreases as the

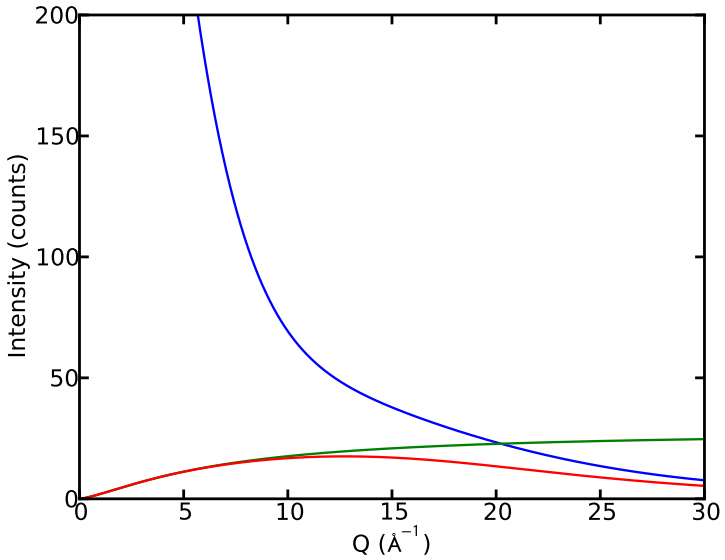

Figure 3.2: Compton correction for nickel. The blue line is the mean-square atomic scattering factor, $\langle f^2 \rangle$, the green line is the tabulated Compton scattering intensity, and the red line is the Compton modified by the Ruland attenuation function (0.01 width).

Compton scattering cross section increases [Egami and Billinge, 2003]. This also leads to a problem because, as we see in Figure 3.2, at high-Q values ($\sim$ 20 Å^{-1} in this case) subtracting the Compton scattering (green) from the experimental scattering intensities (blue, mean-square atomic scattering factor, $\langle f^2 \rangle$, is used as a proxy for the coherent scattering) will result in a negative intensity. We solve this problem by applying a smooth attenuation function to the Compton intensities as suggested by Ruland [Ruland, 1964] (red) and then subtracting the attenuated intensities from those that were collected experimentally.

3.2.3 Background Subtraction

Background scattering intensities come from scattering that is not done by the sample. For instance, the sample container (e.g. a capillary) contributes to the background scattering, as does air, and the instrument itself can have a background signal [Chupas *et al.*, 2003]. We

want to minimize the contribution of background scattering intensities to the total intensity. This can be done in two ways, first by optimizing the experiment and second by subtracting the background [Chupas *et al.*, 2003]. We optimize the experiment by, for example, using a poorly diffracting sample container (i.e. a Kapton® capillary rather than glass) and collecting a dark current on the detector between measurements. To properly subtract the background, we collect three data sets: (1) the sample and container together; (2) an empty container; and (3) the air [Toby and Egami, 1992]. The scattering the container and the air are then subtracted from the total measurement.

For optimal background subtraction, we want the collection time for all data sets to be the same. In reality, though, this rarely happens because samples are measured for different time periods depending on how well (or how poorly) they diffract, while a background measurement is done once. In this case, we simply scale the background measurements by multiplying them by a factor equal to the ratio of the collection times in order to even out the counts. For example, if the container was measured for 5 minutes and the sample was measured for 10 minutes, then the background data would be multiplied by 2 prior to being subtracted from the total intensity.

The contribution of the background intensities to the total intensities varies depending on how well the sample diffracts relative to the container. For example, Figure 3.3 shows two extreme cases. In panel **(a)**, the sample is nickel, a well-diffracting material with Bragg peaks that are several orders of magnitude higher than the background intensities. On the other hand, the sample in panel **(b)** is amorphous salbutamol, a much poorer diffracting material. We see that the background intensities are of the same order of magnitude as the total intensity.

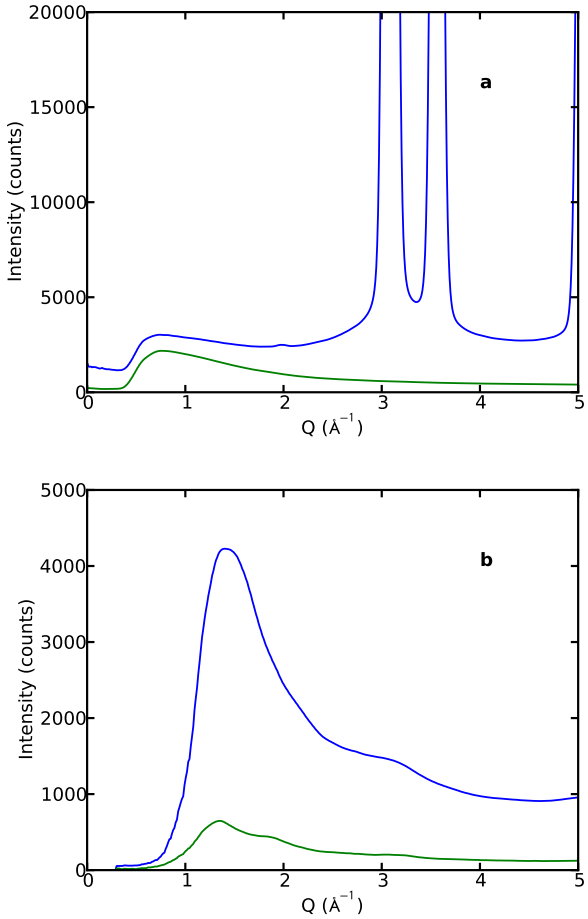

Figure 3.3: Examples of background intensity relative to total intensity, the samples were contained in Kapton® capillaries in both cases. **(a)** Background intensity (green) plotted with total intensity collected on a nickel sample (blue). The Bragg peaks from the nickel are several orders of magnitude higher than the background signal. **(b)** Background intensity (green) plotted with total intensity collected on amorphous salbutamol (blue). The background intensity is of the same order of magnitude as the total intensity.

Chapter 4

The PDFGetX3 Software

In this chapter, we introduce new software developed by members of the Billinge Group for processing PDFs, PDFGetX3. PDFGetX3 is a successor to PDFGetX2, the current gold-standard for making PDFs, which was also developed by the Billinge Group [Qiu *et al.*, 2004]. Although PDFGetX3 is able to process nearly all high energy PDF and TSPDF diffraction data, it is of particular importance to this thesis because it makes the PDF technique more accessible to researchers in the life sciences and the pharmaceutical industry and therefore fits in well with the theme of the thesis.

We first briefly discuss the motivation of why the program was developed and the need that it fulfills, then we devote most of the chapter to comparing PDFs made using PDFGetX3 to those made using PDFGetX2 (processed from the same raw data), and finally we briefly mention the novel parameter tuning feature of PDFGetX3.

4.1 Motivation for PDFGetX3

Although PDFGetX2 is a great software program, it has a few drawbacks. First, PDFGetX2 was written in the commercial IDL programming language and therefore requires the IDL Virtual Machine to be installed on a computer for it to be used, however, the latest version

of the IDL Virtual Machine does not support PDFGetX2 and a legacy version is difficult to find. Second, the program is quite complicated and sometimes frustrating for neophyte users because they need to manually enable, disable, and sometimes provide starting values for a whole host of data corrections (such as those described in Chapters 1 and 3) and parameters. Third, PDFGetX2 does not support any means of batch processing, and requires that each PDF be made individually. With the advent of the RAPDF method [Chupas *et al.*, 2003], time- or temperature-resolved measurements produce large numbers of data sets. Processing each data set individually has proven to be inefficient, even if most parameters across data sets remain the same.

PDFGetX3 was written with usability and high throughput in mind. First, the program was written in Python using only the Numpy library. This means that the program can be used on practically any system that has the freely available (BSD license) Python programming language installed. It has successfully been used on Windows Vista and Windows 7, MacOS, and Linux. Furthermore, PDFGetX3 removes frustration by minimizing the user interaction with the program. In order to process a data set, the user needs to provide, in a plain text configuration file, very basic information about the experiment such as the wavelength of the X-rays used, the chemical composition of the sample, and the range of Q over which to perform the Fourier transform as in Equation (1.21). The user does not need to enable or disable any corrections nor does he or she need to provide values for parameters used in corrections. Finally, PDFGetX3 allows the user to batch process data after he or she has set up the appropriate configuration file. For instance, if a user has made several hundred temperature-resolved measurements on the same material, then he or she will need to set up just one configuration file and then let PDFGetX3 sequentially process each data set. Most importantly, PDFGetX3 is fast. A data set is usually processed in a fraction of a second, so even several hundred measurements can be processed in just a few minutes.

More information on the program is in the PDFGetX3 manual in the Appendix of this thesis.

4.2 Comparison with PDFGetX2

The quality of the PDFs made with PDFGetX3 can be measured using qualitative and quantitative means. Qualitatively, the PDFs are plotted on top of those made with PDFGetX2 with an offset difference curve. Quantitatively, the same structure model can be fit to PDFs made with both programs and the refined parameters can be compared to one another. In this chapter, we do both to several classes of materials: Inorganic materials such as bulk nickel and barium titanate, nanostructured γ-alumina, and bulk and nanocrystalline cadmium selenide, as well as crystalline and nanostructured phases of the organic pharmaceutical carbamazepine. We choose these very different types of materials to show that PDFGetX3 is a robust program that can handle all sorts of high energy X-ray data.

In all cases, PDFs from both programs are made from the same raw data and use the same input parameters (i.e. Q_{max}, X-ray wavelength, chemical composition, and container background). All data sets except the γ-Al_2O_3 were collected at high-energy synchrotron instruments, however, the synchrotron is not a requirement. PDFGetX3 can handle data from lab-based XRPD instruments as long as the Q-range is sufficiently large (more information in Chapter 6). For example, the γ-Al_2O_3 data were collected with a silver anode diffractometer.

The PDF data points do not have error bars because uncertainties are not measured when integrating around the Debye-Scherrer rings (Figure 1.1). Attempts are being made to develop the proper methodology to do this, but they are beyond the scope of this thesis. Instead, we follow the approach that has been used in the literature: We draw dashed horizontal lines that represent $\pm 2\sigma$ (standard deviations) in the difference curve [Proffen *et al.*, 2003; Peterson *et al.*, 2003]. Furthermore, again in following examples in the literature, the range in r space over which we calculate the difference curve and the standard deviations begins with the first nearest neighbor peak, and, in the case of nanoparticles, ends at the maximum size of the nanoparticle.

In general, as we will see in the following examples, we find that the PDFs made by the different programs look somewhat different from one another at r values lower than the first nearest neighbor peak. This is mostly due to differences in how the data are processed by different algorithms and improper applications of corrections. Luckily, any peaks in this range (i.e. before the nearest-neighbor peak) have no physical significance. The PDFs look almost exactly the same over the relevant range of data from the nearest neighbor peak and onward.

We rescale the PDFs by a constant such that the nearest neighbor peak is the same height between PDFs on the same plot. This is a standard procedure and has no effect on data quality (in fact, the scaling parameter is almost always used when comparing experimental data to models) [Peterson *et al.*, 2003]. The relative scaling of peaks to one another within the same PDF, which is preserved when scaling by a constant, is, however, important.

As we discussed in Chapter 1, we fit models to the PDFs by refining a variety of parameters such as lattice parameters, thermal factors, etc. using the program PDFgui [Farrow *et al.*, 2007]. When doing a typical refinement, our goal is to minimize the goodness of fit parameter Rw. Please refer to the PDFgui manual for more information on the various parameters and how refinements are done. In this chapter, though, we are not as much concerned with the perfect refinement as we are concerned with reproducible results. Therefore, in each case, we will compare the Rw value as well as the values of the refined parameters without going into the implications of the fit.

When we fit structure models to PDFs, we generally want to choose the largest range of r for the fit in order to ensure the most independent observations [Egami and Billinge, 2003] (unless we're interested in specific local ordering, as we are with γ-alumina). However, we avoid the low-r region (below the nearest-neighbor peak) where non-physical peaks might have a deleterious effect on the fit. As a final note, we will not fitting structure models to the pharmaceutical compounds in these examples. As discussed in Chapter 1, new modeling

tools that are beyond the scope of this dissertation need to be developed to do this properly.

4.2.1 Nickel and Barium Titanate

First we look at pure nickel (Ni) and barium titanate ($BaTiO_3$) in Figure 4.1. Both compounds diffract very well. Panel **(a)** shows the two PDFs of nickel plotted on top of one another. The PDF made with PDFGetX2 is in blue and the PDF made with PDFGetX3 is in green with $Q_{max} = 26.0$ Å^{-1} in both cases. The difference curve between the two PDFs is red. The area between the dashed horizontal lines represents $\pm 2\sigma$ in the difference curve. We see that there is almost no difference between the PDFs after the Ni-Ni nearest neighbor peak (at $r = 2.2$ Å). We see the same behavior in panel **(b)** with barium titanate. In general, it is quite easy to make PDFs of these compounds (there is a joke that the PDF of nickel can be made with a ruler and graph paper), but they do offer a spring board for us to look at more complicated compounds.

In Table 4.1, we see the parameters refined in fitting the nickel model [Wyckoff, 1967] to the PDFs and the Rw values from the fits. Not many parameters were refined in this fit because only one type of element is present, the lattice parameters are all equal ($a = b = c$), and there is only one isotropic thermal parameter ($U_{iso} = U_{11} = U_{22} = U_{33}$). Overall, we see very good agreement between most of the parameters and the Rw values. The Rw from PDFGetX2 is just slightly lower, but, realistically there is very little difference.

Table 4.2 shows the refined parameters for fitting $BaTiO_3$ PDFs to a model [Megaw, 1962]. We see that more parameters were refined for this material than for nickel, since all of the lattice parameters are no longer equal and neither are the thermal factors, plus a different set of thermal factors is needed for each element. Regardless, we see that the parameters agree very well with one another. The Rw from PDFGetX2 is a little bit lower.

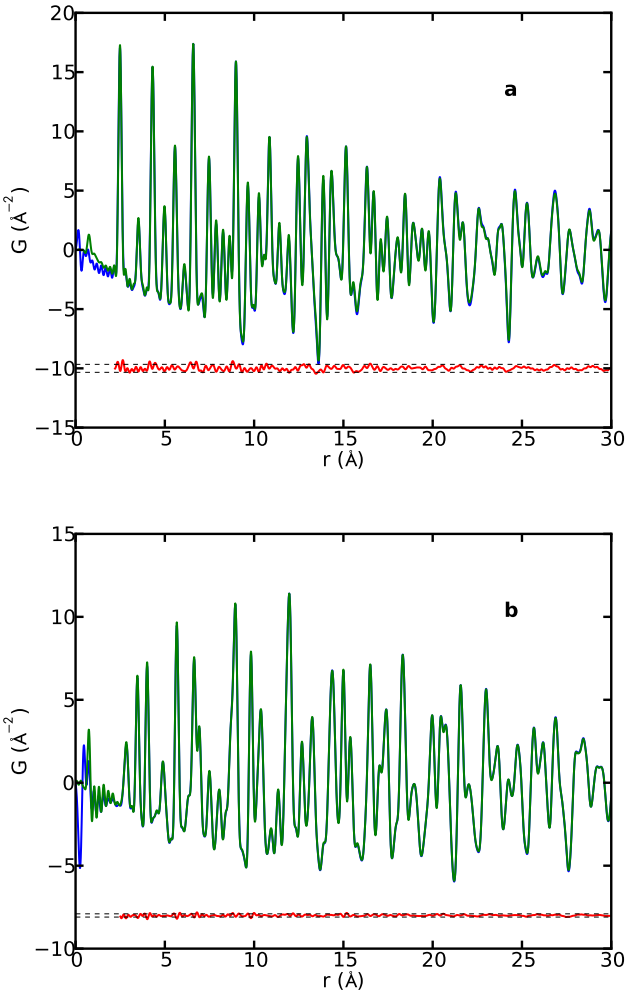

Figure 4.1: PDFs of **(a)** nickel and **(b)** barium titanate made with PDFGetX2 (blue) and PDFGetX3 (green) with $Q_{max} = 26.0$ Å^{-1} in both cases. Difference curve (offset) is in red. The dashed lines represent two standard deviations in the difference curve (r values below the nearest neighbor peaks were not included in the standard deviation calculation).

Table 4.1: Comparison of the parameters refined in fitting the Ni model to the PDFs.

Parameter	PDFGetX2	PDFGetX3
Qdamp (Å^{-1})	0.055	0.057
$a = b = c$ (Å)	3.524	3.524
δ_2 (Å^2)	2.525	2.705
U_{iso} (Å^2)	0.006	0.006
Rw	0.080	0.082

Table 4.2: Comparison of the parameters refined in fitting the $BaTiO_3$ model to the PDFs.

Parameter	PDFGetX2	PDFGetX3
Qdamp (Å^{-1})	0.048	0.049
$a = b$ (Å)	3.995	3.995
c (Å)	4.040	4.040
δ_2 (Å^2)	4.317	4.372
$U_{11\text{Ba}} = U_{22\text{Ba}}$ (Å^2)	0.005	0.005
$U_{33\text{Ba}}$ (Å^2)	0.005	0.004
$U_{11\text{Ti}} = U_{22\text{Ti}}$ (Å^2)	0.009	0.008
$U_{33\text{Ti}}$ (Å^2)	0.013	0.012
$U_{11\text{O}} = U_{22\text{O}}$ (Å^2)	0.011	0.010
$U_{33\text{O}}$ (Å^2)	0.092	0.094
Rw	0.118	0.121

Table 4.3: Comparison of the parameters refined in fitting the γ-Al_2O_3 model to the PDFs.

Parameter	PDFGetX2	PDFGetX3
Qdamp ($Å^{-1}$)	0.077	0.081
a (Å)	3.394	3.394
b (Å)	2.780	2.780
c (Å)	7.042	7.039
δ_2 ($Å^2$)	1.131	0.991
$U_{iso\mathrm{O}}$ ($Å^2$)	0.013	0.012
$U_{iso\mathrm{Al}}$ ($Å^2$)	0.015	0.014
Rw	0.164	0.166

4.2.2 Nanocrystalline γ-Alumina

Next, we investigate γ-alumina (Al_2O_3) using X-rays from a silver anode diffractometer ($\lambda = 0.56$ Å). The γ phase of Al_2O_3 has a local nanocrystalline structure that is different from the cubic and tetragonal structures of other alumina phases [Paglia *et al.*, 2006]. For this reason, a new structure model was developed for the local structure of γ-Al_2O_3 up to $r = 8$ Å (ICSD 173014) [Paglia *et al.*, 2006]. Figure 4.2 shows the PDFs of γ-Al_2O_3 made with PDFGetX2 and PDFGetX3. We see very good agreement between the PDFs. In fact, the PDFGetX3 PDF looks better at low r values.

Refined parameters are in Table 4.3. Unlike in previous cases where we tried to use a large r range for our refinement, in this case we refined only over the range $r = 1.5 - 8$ Å because the model only applies over this range. For this reason, we wanted to refine few parameters (this is why U_{iso} was used rather than separating out the thermal factors). Regardless, though, we see very good agreement between the fits.

As mentioned previously, the γ-Al_2O_3 data were measured using a silver anode lab

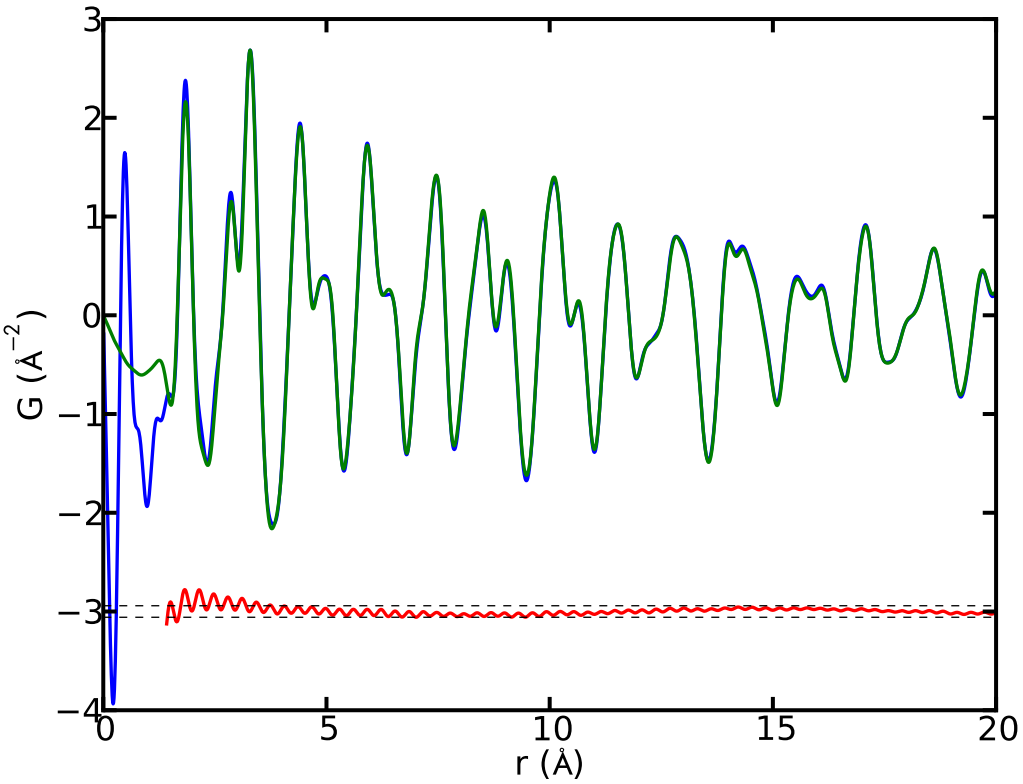

Figure 4.2: PDFs of γ-Al_2O_3 made with PDFGetX2 (blue) and PDFGetX3 (green) with $Q_{max} = 20.5$ Å^{-1} in both cases. Difference curve (offset) is in red. The dashed lines represent two standard deviations in the difference curve (r values below the nearest neighbor peaks were not included in the standard deviation calculation).

diffractometer, and not a synchrotron. From this example we see that PDFGetX3 is capable of processing data from lab-based instruments (with sufficient Q-range) as well as synchrotrons.

4.2.3 Cadmium Selenide Nanoparticles

We now turn our attention to a more complicated class of materials: nanoparticles. Generally, nanoparticles do not diffract as well as the bulk because they do not have long range order and smaller nanoparticles diffract worse than larger ones. In fact, it is quite difficult to extract quantitative structural information from nanoparticles using standard crystallography techniques [Billinge and Levin, 2007] (i.e. a copper-anode diffractometer).

In Figure 4.3 we look at PDFs of three samples of cadmium selenide (CdSe) taken from data that have previously been published by Masadeh *et al.* from the Billinge Group [Masadeh *et al.*, 2007]. The bulk CdSe in panel **(a)** is included for completeness. The nanoparticles in panels **(b)** and **(c)** were calculated to have diameters of 37 Å and 22 Å, respectively [Masadeh *et al.*, 2007].

We see that in all three panels of Figure 4.3, the PDFs made with the two programs are almost identical. The features are all reproduced beautifully by PDFGetX3, even though many more corrections needed to be applied to the nanoparticle data in PDFGetX2 than were needed for nickel or barium titanate. The low-r region looks a little bit different between the PDFs, especially as the size of the nanoparticles gets smaller, but we remember that this region contains no physical information. In fact, we might even argue that in panel **(c)**, the PDFGetX3 PDF looks cleaner than the PDFGetX2 PDF. At this point we again want to stress the fact that the PDFs in PDFGetX3 were made in a fraction of a second with little user interaction beyond inputting a few parameters.

Tables 4.4 and 4.5 contain the refined parameters for the CdSe samples compared to a model [Wyckoff, 1967]. The Qdamp that was refined for the bulk was fixed for the nanoparticles. We again see very good agreement between all parameters. Furthermore,

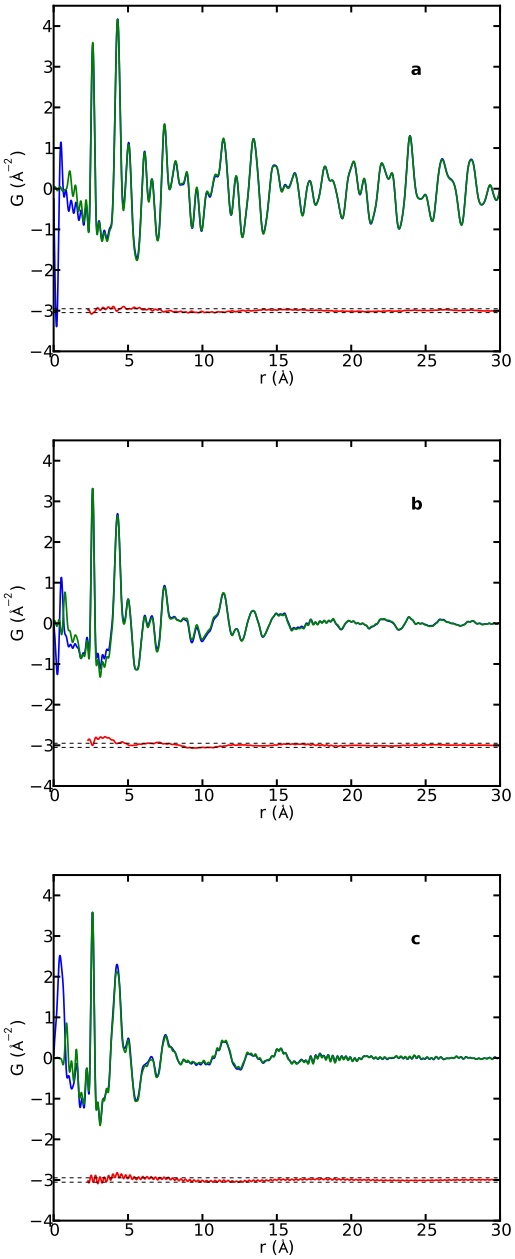

Figure 4.3: PDFs of **(a)** bulk CdSe and **(b)** 37 Å, and **(c)** 22 Å CdSe nanoparticles made with PDFGetX2 (blue) and PDFGetX3 (green) with Q_{max} = 18.0 Å^{-1} in all cases. Difference curve (offset) is in red. The dashed lines represent two standard deviations in the difference curve (r values below the nearest neighbor peaks were not included in the standard deviation calculation and, for the nanoparticle in panel **(c)**, r values larger than 22 Å were not included).

Table 4.4: Comparison of the parameters refined in fitting the CdSe model to the PDFs.

Parameter	PDFGetX2	PDFGetX3
Bulk		
Qdamp ($Å^{-1}$)	0.059	0.059
$a = b$ (Å)	4.299	4.299
c (Å)	7.011	7.011
δ_2 ($Å^2$)	3.210	3.260
$U_{11Cd} = U_{22Cd}$ ($Å^2$)	0.015	0.015
U_{33Cd} ($Å^2$)	0.014	0.014
$U_{11Se} = U_{22Se}$ ($Å^2$)	0.012	0.012
U_{33Se} ($Å^2$)	0.058	0.057
Rw	0.114	0.104

the Rw was lower for PDGetX3 in two out of three fits.

4.2.4 Drugs

The final class of materials that we want to look at are organic pharmaceutical compounds, since we will be discussing them exclusively in the rest of this thesis. These materials can be crystalline, as we see in Figure 4.4**(a)** and **(b)**, nanostructured as in Figure 4.4**(c)**, or amorphous (as we will see in Chapter 7). These materials tend to be made up of mostly light, organic elements such as hydrogen, carbon, and oxygen that do not diffract well, therefore even crystal phase pharmaceutical compounds require quite a bit of tinkering in PDFGetX2 to produce a good TSPDF.

In the examples here, we use three phases of the mood stabilizing drug carbamazepine (CBZ), crystalline CBZ-I and CBZ-III as well as nanostructured CBZ. The TSPDFs of these samples that were created with PDFGetX2 have been published in the literature [Billinge

Table 4.5: Comparison of the parameters refined in fitting the CdSe model to the PDFs.

Parameter	PDFGetX2	PDFGetX3
37 Å NP		
$a = b$ (Å)	4.299	4.296
c (Å)	7.007	7.007
δ_2 (Å^2)	4.661	4.736
$U_{11\mathrm{Cd}} = U_{22\mathrm{Cd}}$ (Å^2)	0.022	0.022
$U_{33\mathrm{Cd}}$ (Å^2)	0.003	0.004
$U_{11\mathrm{Se}} = U_{22\mathrm{Se}}$ (Å^2)	0.012	0.011
$U_{33\mathrm{Se}}$ (Å^2)	0.199	0.197
Particle diameter (Å)	36.38	35.33
Rw	0.194	0.173
22 Å NP		
$a = b$ (Å)	4.294	4.295
c (Å)	6.857	6.863
δ_2 (Å^2)	4.966	5.198
$U_{11\mathrm{Cd}} = U_{22\mathrm{Cd}}$ (Å^2)	0.043	0.042
$U_{33\mathrm{Cd}}$ (Å^2)	0.040	0.041
$U_{11\mathrm{Se}} = U_{22\mathrm{Se}}$ (Å^2)	0.020	0.020
$U_{33\mathrm{Se}}$ (Å^2)	0.233	0.221
Particle diameter (Å)	23.13	23.34
Rw	0.262	0.265

et al., 2010; Dykhne *et al.*, 2011] and will be discussed in great detail in Chapters 5 and 6 of this thesis.

Like the nanoparticles in Figure 4.3, the TSPDFs in Figure 4.4 made with PDFGetX2 have quite a bit of noise at low r. This is because those data were quite difficult to process and these were the best TSPDFs that could be made at the time of publication. We note that the TSPDFs made with PDFGetX3 appear to be quite clean at low r in comparison and require very little tinkering by the user.

As mentioned earlier, we did not fit these TSPDFs to models because new modeling tools need to be developed for this class of materials.

4.3 Tuning Feature

Although this chapter has just briefly touched the surface of the capabilities of PDFGetX3, one feature in particular deserves mention. The tuning feature allows the user to dynamically tune input parameters such as the Q_{max} and background scale and instantly see the effect that this will have on the PDF. This is possible because PDFgetX3 is able to process a data set so quickly that it can adjust parameters and calculate a PDF on the fly. This extremely powerful feature allows the user to interact with the data in order to best choose the parameters used in data processing.

More information on the program is available on the website of the Billinge Group (http://www.diffpy.org) and the manual is the Appendix of this thesis.

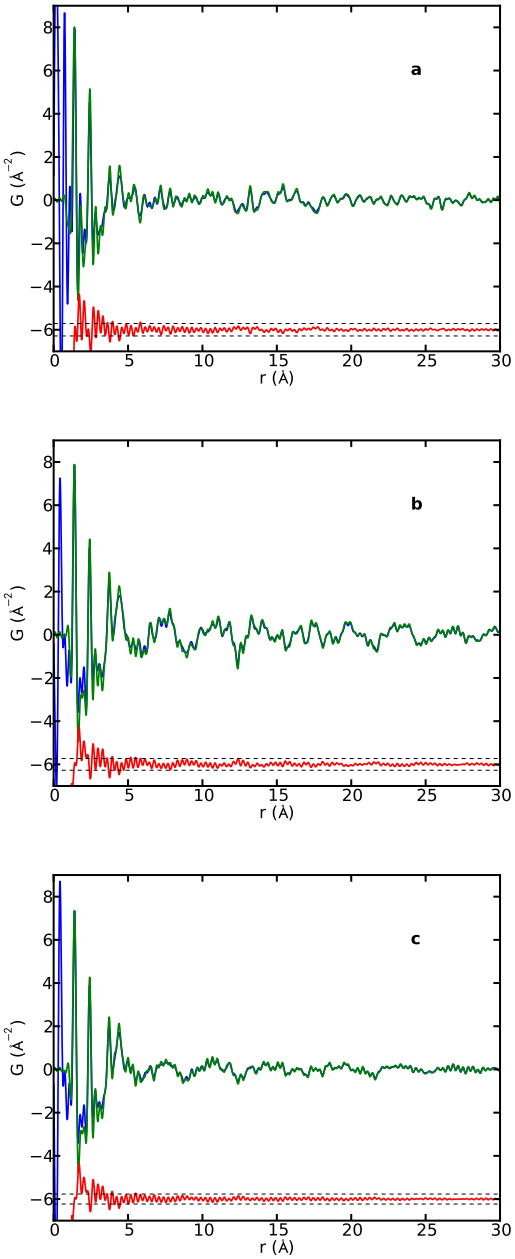

Figure 4.4: PDFs of **(a)** CBZ-I, **(b)** CBZ-III, and **(c)** nanostructured CBZ made with PDFGetX2 (blue) and PDFGetX3 (green) with $Q_{max} = 20.0$ Å^{-1} in all cases. Difference curve (offset) is in red. The dashed lines represent two standard deviations in the difference curve. (r values below the nearest neighbor peaks were not included in the standard deviation calculation).

Chapter 5

Characterization of Amorphous and Nanocrystalline Molecular Materials by Total Scattering

The work in this chapter was done in collaboration with Simon J. L. Billinge, Pavol Juhás, Emil Božin, Ryan Taylor, Alastair J. Florence, and Kenneth Shankland and published in the journal *CrystEngComm* [Billinge *et al.*, 2010].

This paper that this chapter is based on is a seminal work in fingerprinting crystalline and so-called 'X-ray amorphous' pharmaceutical compounds and a good demonstration of the principles discussed in the first three introductory chapters of the thesis. The use of high-energy X-ray total scattering coupled with pair distribution function analysis produces unique structural fingerprints from amorphous and nanostructured phases of the pharmaceuticals carbamazepine and indomethacin. The advantages of such facility-based experiments over laboratory-based ones are discussed and the technique is illustrated with the characterisation of a melt-quenched sample of carbamazepine as a nanocrystalline (4.5 nm domain diameter) version of Form III carbamazepine.

In this study, the commercial program PolySNAP (version 1.7.2 [Barr *et al.*, 2004]) was used to calculate correlations between PDFs because, at the time that this study was conducted, we had not yet developed our Pearson algorithm (Equation (1.24)) to do this. In principle, the PolySNAP correlations work in almost the same way as Pearson, but rely on a proprietary algorithm. A full set of Pearson correlations for the data in this chapter is in Chapter 1 in Table 6.1 and also in Chapter 6 where we discuss the effect of Q_{max} on fingerprinting.

5.1 Introduction

As we had discussed at length in Chapter 2, the majority of active pharmaceutical ingredients (APIs) are marketed as crystalline forms for reasons of stability. However, the formation, stability and performance of amorphous solids are also of significant interest within pharmaceutical research and development. Whilst the amorphous state can confer desirable properties to an API, such as increased aqueous solubility [Hancock and Parks, 2000], the inadvertent production of non-crystalline material during processing can also lead to uncontrolled variability in physical and chemical attributes. The potential for commercial exploitation of amorphous APIs is often complicated by their tendency to revert to a more thermodynamically favorable, and less soluble, crystalline state. Although the identification, characterization and quantification of amorphous pharmaceuticals has received considerable attention, little is known about local ordering in amorphous APIs due to the lack of reliable experimental probes. The powerful tools of crystallography begin to lose their power for structures on the nanoscale; conventional X-ray powder diffraction (XRPD) patterns become broad and featureless in these cases (Figure 5.1) and are not useful for differentiating between different local molecular packing arrangements [Billinge and Levin, 2007]. Accordingly, XRPD is generally used simply to identify such samples as non-crystalline (i.e. X-ray amorphous).

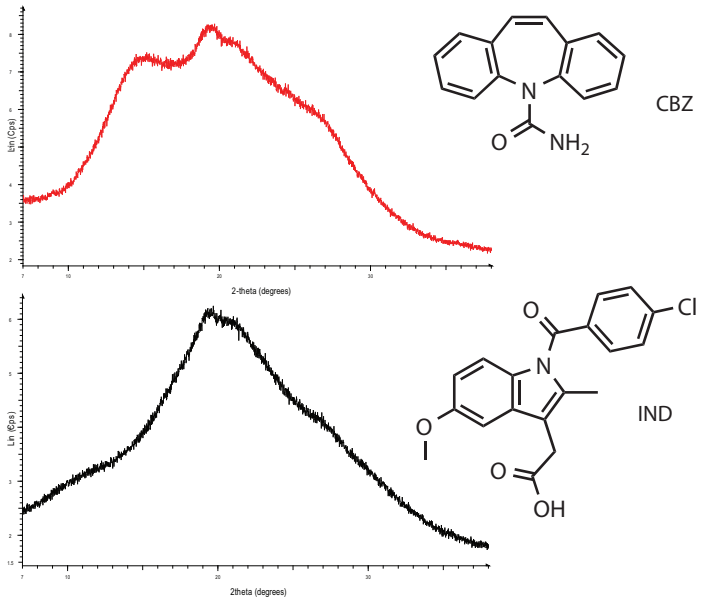

Figure 5.1: Molecular structures and laboratory Cu Kα_1 XRPD patterns for X-ray amorphous melt-quenched samples of CBZ (top) and IND (bottom).

It has recently been suggested that Fourier transforming conventional laboratory XRPD data (Typically Cu Kα radiation, $\lambda = 1.54$ Å; $2\theta_{max} = 60°$; $Q_{max} = 4.1$ Å^{-1}) to obtain the atomic pair distribution function (PDF) [Egami and Billinge, 2003; Warren, 1990] allows more structural information to be extracted [Bates *et al.*, 2006]. The PDF, $G(r)$, yields the probability of finding an atom at a distance r from any reference atom and so provides information on local structure in real space. However, this approach is intrinsically limited by the relatively low momentum transfer magnitude Q $(4\pi \sin(\theta)/\lambda)$ values typically accessible in the laboratory environment, resulting in a PDF of limited real-space resolution (proportional to $2\pi/Q_{max}$). For an accurate PDF across a wide range of r, data should be collected with low instrumental background and good counting statistics to high Q. These requirements can be met by combining high-energy (synchrotron) X-rays with imaging plate detectors.

In this work, a high-energy XRPD method known as total scattering [Egami and Billinge, 2003], coupled with Fourier transformation and PDF analysis, is applied to individual X-

ray amorphous samples of the anti-epileptic drug carbamazepine (CBZ; Figure 5.1) and the non-steroidal anti-inflammatory drug indomethacin (IND). This approach is referred to as the total scattering pair distribution function (TSPDF) method to differentiate it from the approach of obtaining the PDF from conventional laboratory XRPD data [Bates *et al.*, 2006]. The TSPDF method has been widely applied to inorganic materials to study amorphous structures [Wright, 1998] and more recently crystalline and nanocrystalline systems [Egami and Billinge, 2003; Billinge, 2008b] but its application to molecular systems has to date been very limited.

Up to this point in the thesis, we have been a little bit liberal in our use of the terms 'PDF' and 'TSPDF', since TSPDFs are a subset of the PDF umbrella and the underlying theory and corrections are the same for both. However, at this point, we find it instructive to refer exclusively to TSPDFs to highlight the requirement of high-energy X-rays and a large Q-range.

5.2 Experimental Procedure

Melt-quenched samples of CBZ and IND as well as polycrystalline samples of CBZ-I and CBZ-III [Grzesiak *et al.*, 2003] and α [Chen *et al.*, 2002] and γ [Cox and Manson, 2003] IND were measured with a lab-based diffractometer and a synchrotron. Melt-quenched samples were prepared by rapidly cooling molten compound in liquid N_2, then lightly grinding the compound, sieving it, and filling it into a 1 mm diameter Kapton® capillary. Crystalline samples were also ground, sieved, and filled into capillaries of the same type and diameter.

5.2.1 Lab Diffractometer

The laboratory data shown in Figure 5.1 were collected on a Bruker-AXS D8 diffractometer using capillary transmission geometry, primary monochromated Cu Kα_1 radiation (λ = 1.54056 Å) in the range 2-40° 2θ, 0.016° 2θ step size, 10 s per step, at 100 K.

5.2.2 Synchrotron

Data were collected at 100 K at beamline 11-ID-B at the Advanced Photon Source (APS) using the rapid acquisition PDF (RAPDF) method [Chupas *et al.*, 2003]. Samples were sealed in 1 mm diameter Kapton® capillaries and irradiated with X-rays of wavelength $\lambda = 0.1370$ Å. A large area 2D image plate detector (MAR345) was placed centered on and perpendicular to the incident beam 198 mm behind the sample. Sufficient statistics in the high Q-range were obtained using multiple exposures of the image plate, exposing for 300 s, between 5 and 8 times for each data point. The separate exposures were summed together before further processing, resulting in an integrated exposure time of 30 min per sample. 1D powder diffraction patterns were obtained by integrating around the Debye-Scherrer rings on the image plate images, correcting for beam polarization effects using the program Fit2D [Hammersley, 1998]. Data were collected to $Q_{max} = 31$ Å^{-1} and Fourier transformed using data to $Q_{max} = 20$ Å^{-1} to generate TSPDFs to $r_{max} = 30$ Å.

5.2.3 Total Scattering Data Processing

The short wavelength used (0.137 Å), combined with an appropriate data collection strategy enables data to be recorded over a sufficiently high Q-range to provide the necessary resolution in real-space for quantitative structural analysis to be attempted. In these data, a useable $Q_{max} = 20$ Å^{-1} was achieved, equating to a real-space resolution of $2\pi/Q_{max} = 0.31$ Å.

The data were corrected for multiple scattering, Compton scattering, and sample self-absorption, and normalized by the incident intensity and number of scatterers in the sample as described in Chapters 1 and 3. The total scattering reduced structure function, $F(Q)$ (Equation (1.20)), was calculated as per the steps in Chapter 1. The $F(Q)$ was Fourier transformed into the $G(r)$. All of these steps were carried out using the program PDFGetX2 [Qiu *et al.*, 2004].

5.3 Results

5.3.1 Carbamazepine

The total scattering data, presented as $F(Q)$, and the resultant TSPDFs for the amorphous and polycrystalline samples of CBZ are shown in Figure 5.2. The $F(Q)$ for the melt-quenched sample (Figure 5.2**(b)**), measured over a wide enough range of momentum transfer and properly normalized, is rich in information content compared to a conventional laboratory-based XRPD measurement (e.g. Figure 5.1). Close inspection shows no clear Bragg diffraction, confirming a lack of long-range order in the sample. The $F(Q)$ clearly distinguishes the polycrystalline samples CBZ-III and CBZ-I (Figure 5.2**(a)** and **(c)**, respectively) with the melt-quenched sample showing a closer resemblance to that of CBZ-III than CBZ-I.

Transforming $F(Q)$ to $G(r)$ (i.e. the TSPDF) allows interpretation and comparison to be carried out in real space. There is a striking resemblance between the TSPDF of CBZ-III and the melt-quenched sample. Full-profile comparisons of the TSPDFs in the range dominated by inter-molecular interactions, 3 - 20 Å, for the three samples using PolySNAP [Barr *et al.*, 2004] yielded a correlation co-efficient of 0.8389 for the melt-quenched and CBZ-III TSPDFs (perfect match = 1.0). The next closest similarity was observed for melt-quenched CBZ and Form I, but yielding a correlation coefficient of only 0.6124. The PolySNAP correlations between the three CBZ phases are in Table 5.1

Given such close agreement between the melt-quenched and Form III TSPDFs, the structural similarity between these samples was explored in more detail. The TSPDF of CBZ-III was modified by attenuating the TSPDF peaks in the high-r region using a spherical characteristic function to simulate the effects of reducing the range of structural coherence (or long range ordering) on the data, assuming spherical particles, using Equation (5.1) [Guinier, 1963], where r is the r-grid the TSPDF is on, d is the particle diameter, and Φ is a step function which is 0 when $r > d$. If the internal atomic arrangement of a nanocrystalline

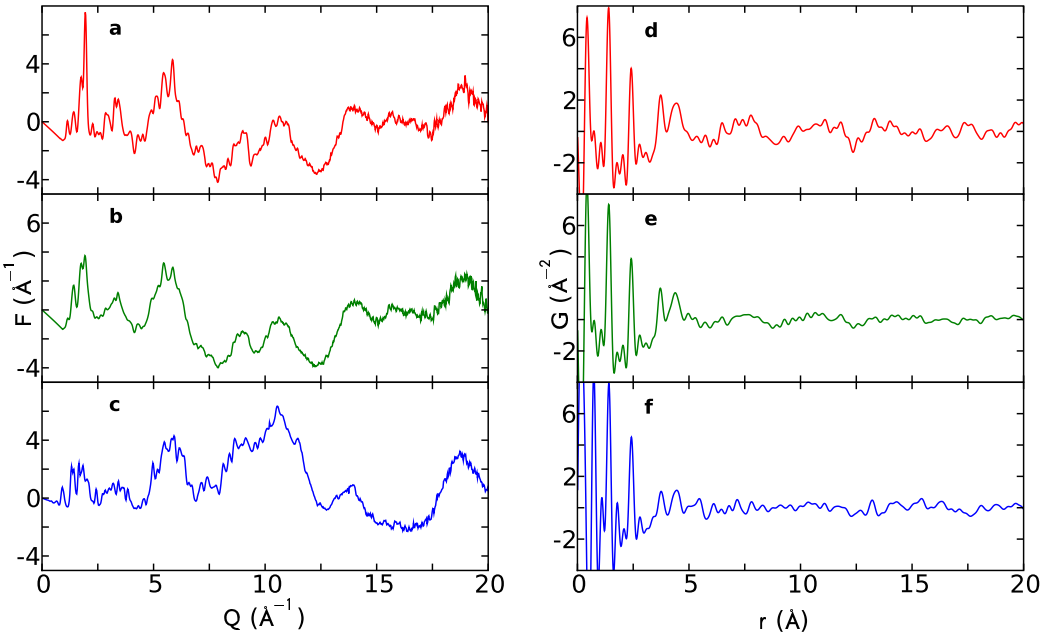

Figure 5.2: Total scattering diffraction patterns and TSPDFs of CBZ. Panels **(a)** and **(d)** correspond to CBZ-III, **(b)** and **(e)** to the melt-quenched sample and **(c)** and **(f)** to CBZ-I; **(a)**, **(b)**, **(c)** show the total scattering data in the form of $F(Q)$ whilst **(d)**, **(e)**, **(f)** are in the form of the TSPDF, $G(r)$.

Table 5.1: Summary of PolySNAP correlation coefficients between CBZ-I, CBZ-III, and melt-quenched (MQ) CBZ. Correlations higher than 0.8 are shown in bold (except when they are trivially unity).

	CBZ-I	CBZ-III	CBZ-MQ
CBZ-I	1	0.4467	0.6124
CBZ-III		1	**0.8389**
CBZ-MQ			1

domain resembles that of a bulk crystalline analog, its TSPDF resembles that of the bulk except that the amplitude of the TSPDF peaks is attenuated with increasing r due to the loss of far-neighbor correlations outside the particle. This can be modeled by multiplying the crystalline PDF with the auto-correlation of the shape function of the particle as in Equation (5.2).

$$f(r;d) = \left[1 - \frac{3r}{2d} + \frac{1}{2}\left(\frac{r}{d}\right)^3\right]\Phi(d-r) \tag{5.1}$$

$$G(r;d) = G(r) * f(r;d) \tag{5.2}$$

The overlay shown in Figure 5.3 was obtained using a nanocrystalline domain diameter of 4.5 nm. The excellent agreement between the attenuated TSPDF from CBZ-III and the melt-quenched CBZ TSPDF is definitive proof that the local packing in the melt-quenched sample is that of Form III with a range of structural coherence of 4.5 nm. It is interesting to ask whether the sample is made up of discrete 4.5 nm nanocrystallites of Form III or whether it is truly a homogeneous amorphous structure with short-range molecular CBZ-III-like packing. The data suggest the former since the sharpness of features in the TSPDFs is preserved with increasing r whilst their amplitude is simply reduced, which is not the behavior seen in truly amorphous samples. We thus conclude that the structure of the melt-quenched CBZ used in the measurement is actually nanocrystalline CBZ-III with an average particle diameter of 4.5 nm. Although the TSPDF of the melt-quenched sample is well explained by CBZ-III attenuated by the PDF characteristic function for a sphere, we cannot rule out that there is a dispersion of nanoparticle sizes centered around the value of 4.5 nm. For example, narrow dispersions with $\sim 10\%$ polydispersity are well explained using the characteristic function for a single sphere.

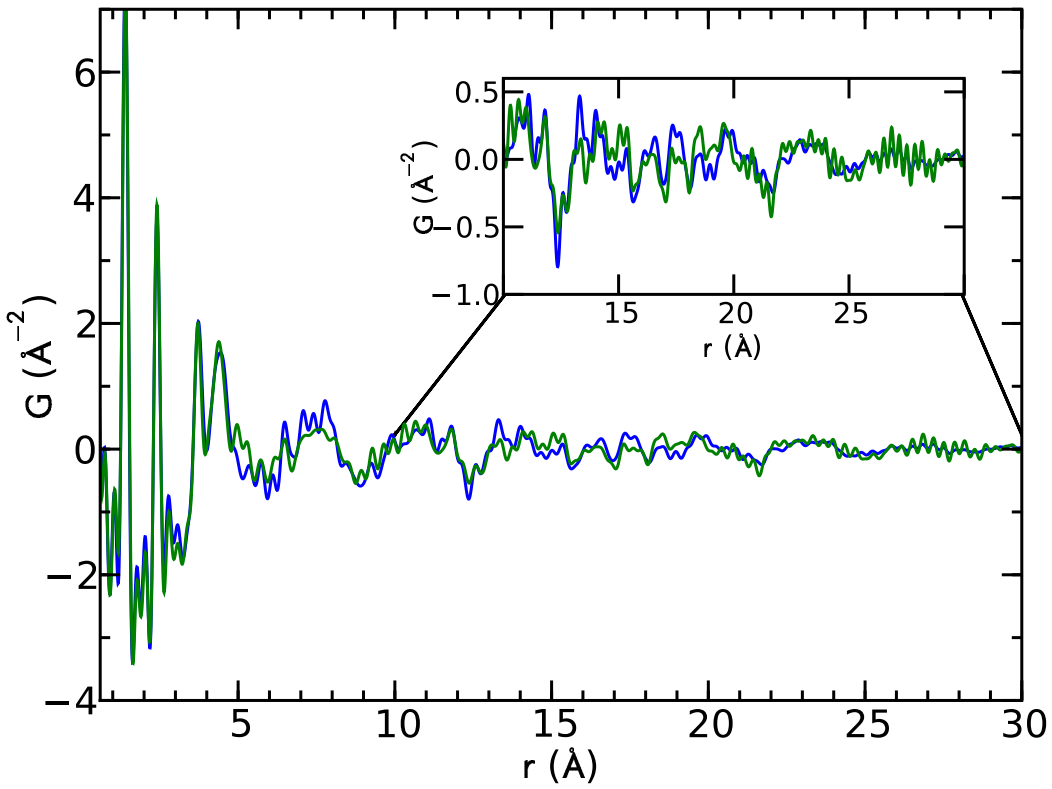

Figure 5.3: Comparison of TSPDF from the melt-quenched amorphous sample (green) and CBZ-III (blue), modified as if it were a 4.5 nm nanoparticle (see text for details). PolySNAP correlation coefficient 0.8601.

5.3.2 Indomethacin

A similar analysis has also been carried out on a melt-quenched sample of IND, and the results are shown in Figure 5.4. Again, $F(Q)$ shows the melt-quenched IND sample to be X-ray amorphous (no evident Bragg diffraction) and rich in structural information. From Table 5.2, the highest correlation coefficient from full-profile comparisons of the TSPDFs of melt-quenched, α and γ IND in PolySNAP was 0.6770, returned for the melt-quenched and α-IND phases. This is significantly lower than the highest value obtained for the CBZ TSPDF comparisons. All other coefficients were less than 0.5. Thus, the TSPDFs indicate that the local structure of the melt-quenched IND sample at 100 K is largely distinct from the α and γ crystalline forms. This contrasts with the suggestion based on crystallization and spectroscopic evidence that below T_g (315 K [Andronis and Zografi, 2000]) amorphous IND has a local structure, with dimeric hydrogen bonding, similar to the γ form [Shalaev and Zografi, 2002]. Linear combinations of the α and γ crystalline phases also do not give good agreement with the TSPDF from the melt-quenched sample. Further comparisons with the δ Form of IND were not possible at the time of writing as neither a crystal structure nor experimental TSPDF were available [Borka, 1974]. However, this result clearly shows that the TSPDF can readily characterize distinct local molecular packing arrangements in the amorphous IND sample. We note that, as with the CBZ, oscillations in the PDF are apparent over the whole r-range shown and clearly extend beyond 20 Å, which shows that the melt-quenched IND sample we studied is also nanocrystalline rather than truly amorphous.

5.4 Discussion

The key to obtaining useful TSPDF curves is not the use of synchrotron radiation *per se* but collecting data to high Q with good statistics. This is possible from laboratory based diffractometers that have Ag ($\lambda = 0.556$ Å) or Mo ($\lambda = 0.7107$ Å) sources, where the fact

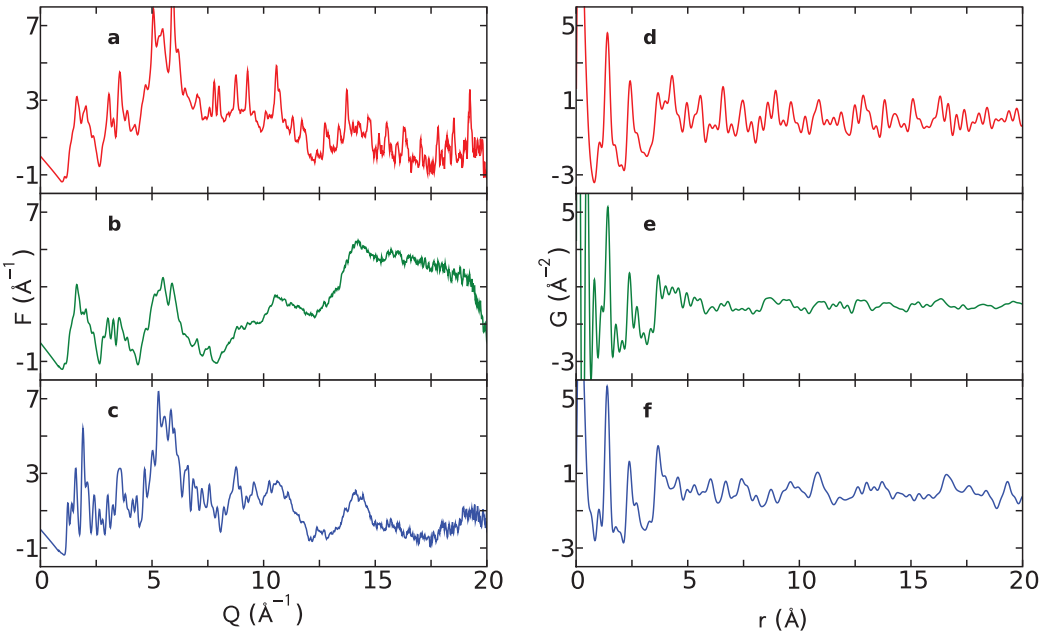

Figure 5.4: Total scattering diffraction patterns and TSPDFs of IND samples. **(a)** and **(d)** contain patterns from IND α, **(b)** and **(e)** to the melt-quenched sample and **(c)** and **(f)** IND γ. The column **(a)**, **(b)**, **(c)** shows the synchrotron total scattering data in the form of $F(Q)$ and the second column, **(d)**, **(e)**, **(f)**, contains the total scattering data in the form of the TSPDF, $G(r)$.

Table 5.2: Summary of PolySNAP correlation coefficients between IND α, IND γ, and melt-quenched IND. Correlations higher than 0.8 are shown in bold (except when they are trivially unity).

	IND α	IND γ	IND-MQ
IND α	1	0.4057	0.6770
IND γ		1	0.4886
IND-MQ			1

that λ is a factor of $\sim$ 2-3 times smaller than that of a Cu source means that higher Q values can be accessed for any given diffraction angle. The current data were Fourier transformed with a Q_{max} of 20 Å^{-1}, which is certainly accessible with a Ag lab diffractometer, although a suitably configured Mo instrument would offer significant practical advantages such as higher incident flux, increased X-ray scattering and higher detector efficiency. That said, synchrotron measurements are advantageous because the requisite statistics can be obtained over the whole Q-range in a short time (in this case 30 min) compared to many hours on a laboratory-based source. Future developments in high intensity, short wavelength laboratory X-ray sources will certainly help close this particular gap.

5.5 Conclusion

These results have a number of important implications. They show that TSPDF data can be used to unambiguously differentiate between different forms of amorphous or nanocrystalline molecular solids. As such, TSPDF is an approach that can take the fingerprinting role for amorphous pharmaceuticals that XRPD takes for polycrystalline pharmaceuticals. This opens the door to future studies exploring the effects of processing or storage on amorphous materials and of phase stability in molecular dispersions, for example. There can also be sufficient information in the TSPDF to enable the fitting of well-defined structural models for the molecular conformation and packing arrangements in amorphous and nanocrystalline samples. Clearly this would have particular application in the case of the melt-quenched IND TSPDF presented here, however the development of such models is beyond the scope of the current work. This capability offers the potential to revolutionize the study of amorphous samples, by illuminating the basic science underpinning the structure of non-crystalline molecular materials to add to the wealth of thermodynamic and spectroscopic literature available. Also, by tracking the evolution of structure of melt-quenched glasses for example, this tool may help identify new crystalline polymorphs *via* an amorphous or nanocrystalline

route.

As we hinted in the Discussion section, the work done in this chapter does raise the very important question of what Q-range is sufficient for fingerprinting nanocrystalline and amorphous pharmaceuticals. We have seen that a Q_{max} of 20 Å^{-1} is enough to produce TSPDFs with a good r-resolution, but this range is accessible only to synchrotron and Ag-anode instruments, both of which have their drawbacks. Synchrotron beamtime needs to be booked sometimes months in advance and is not always available for commercial use, and Ag-anode instruments have low intensities which could result in very long data collection times [Egami and Billinge, 2003]. On the other hand, measurements using a Cu-anode diffractometer would be ideal, particularly since this type of instrument is the cheapest and most common, but it can only reach a theoretical Q_{max} of 8 Å^{-1}. We answer this question in Chapter 6 by doing a systematic study of the effect of Q_{max} on the quality and reliability of PDFs used in fingerprinting pharmaceutical compounds.

Chapter 6

Data Requirements for the Reliable Use of Atomic Pair Distribution Functions in Amorphous Pharmaceutical Fingerprinting

The work in this chapter was done in collaboration with Ryan Taylor, Alastair Florence, and Simon J.L. Billinge and published in the journal *Pharmaceutical Research* [Dykhne *et al.*, 2011].

In Chapter 5, we investigated the effectiveness of the total scattering pair distribution function (TSPDF) method for fingerprinting disordered (i.e. nanostructured and amorphous) pharmaceutical compounds. We found that TSPDFs were able to identify that melt-quenched carbamazepine had the same structure as its Form III polymorph. However, we were left with the question of whether it would be possible to fingerprint pharmaceutical compounds using X-rays with lower energies than that of a synchrotron.

In response, we have carried out a systematic study comparing the efficacy of finger-

printing condensed phases of small molecule systems using atomic pair distribution functions (PDFs) obtained from X-ray diffraction data measured over different ranges of momentum transfer, Q. Data collected over a wide Q-range, for example from high energy X-ray synchrotron beamlines, contain more information and are expected to yield more reliable results. However, we find that reliable fingerprinting is still possible down to Q-ranges accessible to silver and molybdenum laboratory sources, but data from copper anode laboratory sources, which have the smallest Q-range, are unreliable for fingerprinting, yielding ambiguous and potentially incorrect results. These ambiguities make data measured over too narrow of a Q-range unsuitable for fingerprinting APIs and small molecule systems and, in general, copper anode diffractometers are undesirable for this purpose; however, laboratory X-ray sources with either Mo or Ag anodes are well suited to this application.

6.1 Introduction

As discussed in Chapters 2 and 5, the majority of drug compounds, or active pharmaceutical ingredients (APIs), are marketed in crystalline forms largely for reasons of physico-chemical stability and processibility. However, the amorphous state is also of significant interest within the pharmaceutical industry, particularly as a possible means to enhance aqueous solubility of APIs [Hancock and Parks, 2000]. In the development of new medicinal products, poor oral bioavailability due to low aqueous solubility/dissolution rates of many crystalline API candidates is increasingly a potential barrier to enabling treatments to be administered using tablet or capsule formulations. However, to ensure that all batches of manufactured dosage forms are bioequivalent, the selection of a robust solid form is necessary to ensure that the batch-to-batch variability in raw material properties is minimized and that the selected form is stable during all stages of processing, manufacture, and storage. An important practical barrier to the development of amorphous APIs in product development is the lack of reliable methods for fingerprinting amorphous APIs, for example to provide unambiguous

structural identification of raw material or to highlight possible structural changes over time that may lead to changes in product performance e.g. due to crystallization of the API. Whilst largely routine for crystalline solids, traditional X-ray powder diffraction (XRPD) fingerprinting methods are invalid for amorphous forms due to the lack of long range atomic structure [Billinge *et al.*, 2010]. Recently, atomic pair distribution function (PDF) methods have been suggested as an alternative approach for fingerprinting amorphous APIs [Bates *et al.*, 2006; Billinge *et al.*, 2010] and for providing valuable insight into the local packing of pharmaceuticals in non-crystalline forms. If these methods are validated as giving reliable fingerprints of APIs in the amorphous state, they could make a significant contribution to expediting the wider commercial exploitation of amorphous pharmaceuticals.

Although the PDF has historically been used on inorganic materials, it has recently received some attention from the pharmaceutical community and has been applied to molecular materials in crystalline and amorphous phases. There are examples in the literature of PDFs of molecular materials obtained from X-ray diffraction data from standard, widely available copper anode laboratory based XRPD instruments [Sheth *et al.*, 2004; Bates *et al.*, 2007; Sheth *et al.*, 2005; Newman *et al.*, 2008], and also TSPDFs obtained from powder data collected on a high energy synchrotron source [Billinge *et al.*, 2010].

The first step in a PDF measurement is to collect powder diffraction data with low noise over some range of momentum transfer, Q, where Q is related to the Bragg angle, θ, which is half the scattering angle, and the wavelength, λ of the incident radiation: $Q = 4\pi \sin\theta/\lambda$ (Equation (1.2)). The data are then corrected for experimental artifacts and normalized to obtain the reduced structure function, $F(Q)$, which can be Fourier transformed to get the PDF [Egami and Billinge, 2003]. This is readily done using available software [Qiu *et al.*, 2004].

We showed in Chapter 5 that powder diffraction data collected over a wide Q-range can successfully be used to differentiate the local structure of X-ray amorphous pharmaceutical forms (i.e. samples that do not give rise to measurable Bragg diffraction on a standard

laboratory diffractometer) of carbamazepine and indomethacin from their respective crystalline forms, both in Q-space (raw data) and in real-space as the PDF [Billinge *et al.*, 2010]. This approach, referred to as total scattering pair distribution function (TSPDF) analysis, is therefore a good candidate for fingerprinting amorphous small molecule materials. An alternative approach for fingerprinting amorphous forms within the laboratory environment is to use powder diffraction data from a conventional copper-anode laboratory source (CALS) to obtain PDFs [Sheth *et al.*, 2004; Bates *et al.*, 2007; Sheth *et al.*, 2005; Newman *et al.*, 2008]. Here we evaluate the relative ability for fingerprinting the inherently low-Q CALS derived PDFs and TSPDFs obtained with higher energy sources.

In general, if there is too little information in a diffraction data-set it will not be possible to use it for fingerprinting as it will not give a uniquely identifiable pattern arising directly from the molecular packing of the particular form. For example, it has been pointed out that, in contrast to crystalline forms, there is insufficient information in the diffraction pattern of amorphous APIs when measured in the low-Q region accessible from a CALS [Newman *et al.*, 2008]. Indeed, the absence of diffraction peaks in a CALS powder pattern is often used as the basis to identify the sample as amorphous or more specifically as 'X-ray amorphous' or 'diffraction amorphous'. This lack of information in a conventional CALS diffraction pattern collected from an amorphous API is the main origin of the difficulty in realizing reliable fingerprinting of non-crystalline APIs or solid dispersions of API. However, when measuring diffraction data over a wider Q-range, PDFs can be obtained that are more information rich, and therefore more capable of being applied successfully in sample fingerprinting. This argument breaks down at the point where the structure-containing signal in the diffraction data disappears altogether. At this point the resolution is sample-limited and there is no benefit to measure over a wider range of Q. In most inorganic materials, this occurs typically at values of $Q = 30 - 40$ Å^{-1} [Egami and Billinge, 2003].

To access higher Q-ranges it is necessary to use shorter wavelength X-rays. For example, the maximum Q-range accessible to a typical CALS instrument is $\sim$ 8 Å^{-1}, equivalent

to 160° 2θ, whereas for molybdenum-anode laboratory sources (MALS) and silver-anode laboratory sources (SALS), it is $\sim$ 16 Å^{-1} and $\sim$ 22 Å^{-1}, respectively. Q-ranges up to $\sim$ 45 Å^{-1} are accessible when measuring using high energy synchrotron radiation [Petkov *et al.*, 2000]. In this chapter we address the issue of the range of Q that is required for effective fingerprinting of amorphous APIs. We also investigate the value of Q where scattering from amorphous APIs becomes sample limited.

A second requirement for successful fingerprinting of any solid form is that for the data to be representative they must be reproducible and uniquely attributable to the phase being measured. As the PDF is derived from the raw data through a process of data corrections and normalization, any ambiguities in the data reduction may lead to variations in the resulting PDFs. Thus, a given data-set could potentially result in more than one PDF. Clearly in such instances, the measured PDF is not unique and is of no value for fingerprinting. As with powder diffraction for fingerprinting polycrystalline phases, any observed changes in the pattern should relate directly to changes in the structure of the sample alone.

We find that data collected over Q-ranges accessible from MALS and SALS instruments yield PDFs that give unique and reliable fingerprints. The situation becomes more ambiguous as the Q-range measured decreases and low-Q_{max} measurements from copper sources are less reliably unique and the information content is not, in general, adequate to provide a unique and therefore reliable fingerprint.

6.1.1 Theoretical Background

The reduced structure function $F(Q) = Q[S(Q) - 1]$ (Equation (1.20)) is obtained from raw powder diffraction data by correcting for experimental effects and normalizing the data, which is then Fourier transformed to obtain the PDF, $G(r)$ [Egami and Billinge, 2003] (Equation (1.23). In practice it is not possible to know *a-priori* the proper normalization factors, which depend on things such as detector efficiency, beam size and sample volume

and density that are not typically well known. However, the behavior of the structure function, $S(Q)$, is known: $S(Q)$ oscillates around unity and as $Q \to \infty$, $S(Q)$ asymptotically approaches unity. If we assume that additive corrections to the data (for example, background subtraction, multiple scattering, and Compton scattering) have been well accounted for, or are minimal, we can use the asymptotic behavior to ensure a correct normalization by scaling the data such that it oscillates about $\langle f(Q)\rangle^2$ and approaches it in the high-Q region [Egami and Billinge, 2003], as we explain below. In practice we don't measure $S(Q)$ to ∞ but to some finite value, Q_{max}, which is determined by the wavelength of the incident radiation and the highest scattering angle used. However, the signal in the diffraction data dies out with increasing Q and is gone by a Q-value of 40-50 Å^{-1} in most cases [Egami and Billinge, 2003]. In practice, even when data are collected over a range limited to $Q_{max} = 15 - 20$ Å^{-1} it is possible to observe the decay of the signal.

The decay of the signal at high-Q allows for a reliable normalization of the data. By definition, $S(Q)$ oscillates around unity and the decay means that $S(Q)$ asymptotically approaches unity with increasing Q [Egami and Billinge, 2003]. Since $S(Q) = I(Q)/\langle f(Q)\rangle^2$ we need to multiply the measured intensity by a factor such that, in the high-Q region where the signal has died, $I(Q) = \langle f(Q)\rangle^2$. We used the program PDFGetX2 [Qiu *et al.*, 2004] to carry out the corrections and the normalization. The data were corrected for background scattering, self-absorption, and incoherent Compton scattering then normalized for incident flux, number of scatterers, and atomic form factor squared, $\langle f(Q)\rangle^2$, to obtain the structure function, $S(Q)$.

In practice, PDFGetX2 assists the user with the corrections by optimizing the shape of the $S(Q)$ over a user-specified range of data, typically in the high-Q region. The program varies a series of physical parameters such that the $S(Q)$ oscillates around the $\langle f(Q)\rangle^2$ and approaches 1 as Q approaches Q_{max}. This is an effective way of determining the proper corrections when oscillations of the signal around $\langle f(Q)\rangle^2$ are small, i.e., in the high-Q region. However, at low-Q, the signal is still oscillating strongly. The approximation used

by the program no longer holds since the value of $S(Q_{max})$ depends on precisely where Q_{max} falls in the oscillating signal. It is no longer true that the correct shape of the $S(Q)$ is obtained by matching $I(Q)$ to $\langle f(Q)\rangle^2$ over some user-specified range of Q. Different $S(Q)$ shapes lead to a variability in the PDF obtained by Fourier transforming the data and it is therefore possible to obtain qualitatively dissimilar PDFs by selecting different ranges of data over which to run the optimization. When data are available extending to high-Q, the ambiguity in determining the correct shape of the $S(Q)$ is resolved and unique PDFs are obtained from a given diffraction pattern, as we show below.

The PDF, $G(r)$, is obtained from the $S(Q)$, by a Fourier transformation as per Equation 1.23. The Fourier transform is a linear transform and there is a unique relationship between $S(Q)$ and the resulting $G(r)$. No information is added or removed in the transform. However, in practice, what is transformed is the corrected, experimentally determined $S_{exp}(Q)$ which we assume to be very close to the actual $S(Q)$:

$$G_{exp}(r) = \frac{2}{\pi}\int_{Q_{min}}^{Q_{max}} Q[S_{exp}(Q) - 1]\sin Qr\, dQ. \tag{6.1}$$

There is still a unique relationship between $G_{exp}(r)$ and $S_{exp}(Q)$ if the transform is carried out over the same Q-range, but fingerprinting will be compromised if $S_{exp}(Q)$ itself is not sufficiently unique, i.e., deviations of $S_{exp}(Q)$ from the underlying $S(Q)$ of the material under study are small compared to differences between the $S(Q)$'s of different materials of interest. In the first part of this chapter, we test this assumption for data measured to low and high values of Q_{max}. We do this by obtaining different $S_{exp}(Q)$ functions from the same raw data using different, but reasonable, protocols for carrying out the PDF optimization in PDFgetX2. This is done by fixing Q_{max} but changing the range given to PDFgetX2 for the automatic PDF optimization. Since the exact value of Q_{max} is a somewhat arbitrarily chosen experimental parameter we also show the sensitivity of $S_{exp}(Q)$ to the choice of Q_{max} over some narrow but reasonable range of values for data with low- and high-Q_{max}'s.

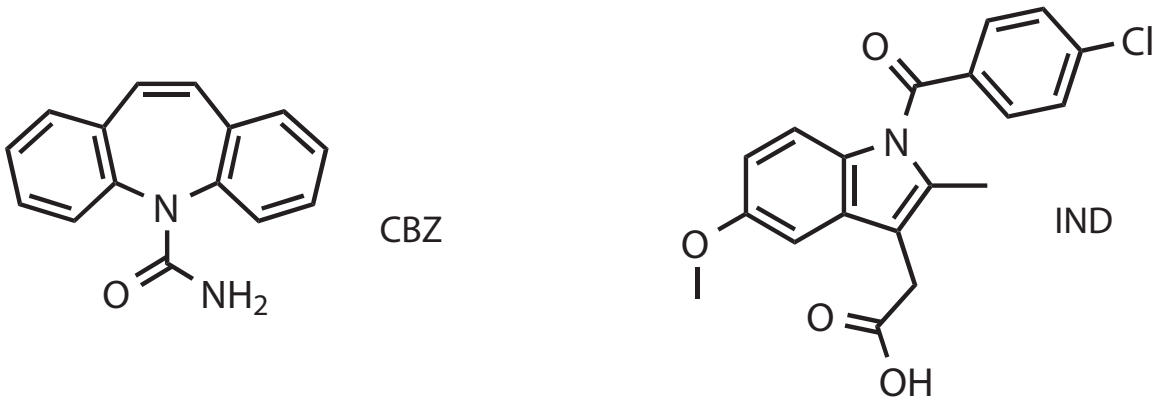

Figure 6.1: Molecular structures of CBZ (left) and IND (right).

6.2 Experimental Procedure

6.2.1 Samples

TSPDF data were collected from melt-quenched (amorphous) samples of carbamazepine (CBZ), an anti-epileptic drug, and indomethacin (IND), a non-steroidal anti-inflamatory drug (henceforth referred to as CBZ-a and IND-a, respectively), as well as polycrystalline samples of CBZ Form I and Form III (CBZ-I and CBZ-III) [Grzesiak *et al.*, 2003] and the α [Chen *et al.*, 2002] and γ [Cox and Manson, 2003] forms of IND (IND-α and IND-γ). The molecules are shown in Figure 6.2.1.

Details of the sample preparation are described in Chapter 5. Briefly, the amorphous samples were prepared using a melt-quenching method whereby molten compound was rapidly cooled in liquid N_2, lightly ground, sieved, and filled into a 1 mm diameter Kapton® capillary.

6.2.2 X-Ray Powder Diffraction Experiments

All samples were measured using a Bruker-AXS D8 diffractometer using capillary geometry with primary monochromated Cu-$K_{\alpha 1}$ radiation (λ = 1.54056 Å) and Lynxeye position

sensitive detector (PSD) in the range 2-40° 2θ (i.e. Q_{max} ca. 2.8 Å^{-1}) with step size 0.016° 2θ, 10 s per step at 100 K [Billinge *et al.*, 2010]. Although this is well below the potential Q_{max} accessible on the instrument, this reflects the typical data range collected for fingerprinting polycrystalline samples. Total scattering data were also collected at beamline 11-ID-B at the Advanced Photon Source (APS) at Argonne National Laboratory using the rapid acquisition PDF method [Chupas *et al.*, 2003]. A 2D image plate detector was placed perpendicular to a high energy X-ray beam ($\lambda = 0.137024$ Å) 198 mm behind the sample. Data were collected for 300 s and this was repeated between 5-8 times for a total collection of about 30 min for each data-set.

6.2.3 Fingerprinting

After PDFs have been obtained from the data using the methods discussed above, we test their efficacy at fingerprinting. This is done both qualitatively and quantitatively. For the qualitative analysis we simply use a visual comparison of the plots of the PDFs. For the quantitative analysis we use programs that quantify the similarity of curves. As discussed in Chapter 5, we use the commercially available PolySNAP program, which uses a modified version of the Spearman correlation parameter [Barr *et al.*, 2004]. In addition to PolySNAP we use a home-written program that computes the Pearson product-momentum correlation [Myers and Well, 2010], $\mathbf{R} = \frac{1}{1-n}\sum_{i=0}^{n}\left(\frac{X_i-\overline{X}}{\sigma_x}\right)\left(\frac{Y_i-\overline{Y}}{\sigma_y}\right)$ (Equation (1.24)), where $\overline{X}$ and σ_x are the mean and standard deviation of a data set, respectively. Both correlation techniques create an $m \times m$ matrix that contains a correlation value r in the range -1 to 1 between each pair of m data-sets. The value 1 implies complete correlation, zero implies no correlation, and -1 implies an anti-correlation. The correlation techniques are extremely powerful because they ignore absolute scaling, but are sensitive to relative scaling and slight shifts in peak positions.

We study the correlations between PDFs in the range $r = 3.0 - 20$ Å. We chose this range because the very local structure (i.e. $r < 3.0$ Å) of all molecular samples is similar due

to intra-molecular atom pairs, for example, consisting of nearest and next-nearest neighbor carbon-carbon bonds at 1.4 Å and 2.4 Å, respectively. Applying the correlation analysis to the entire data range does not change the result significantly but reduces the sensitivity to finding differences in molecular packing of the correlation analysis by including a range of r-that is highly similar regardless of the packing.

6.3 Results

6.3.1 Uniqueness Tests

First we consider TSPDF data from the synchrotron which has a $Q_{max} = 20.0$ Å^{-1}. Figure 2**(a)** shows 4 PDFs obtained from the same data-set but where the range of data used for the automated normalization in PDFgetX2 was from 62%, 71%, 80% and 92% of the Q_{max} to Q_{max}, respectively; e.g., in the first, 62%, data were normalized from 12.5 Å^{-1} to 20 Å^{-1}.

It is clear from the figure that the resulting PDFs are all highly similar. This is supported by the correlation analyses with the Pearson correlations for the PDFs all being larger than 0.99: the PDF generated is always unique regardless of reasonable variations in the selected processing parameters.

This is not true for the data with $Q_{max} = 2.8$ Å^{-1} from the CALS data, as is evident in Figure 2**(b)**. Again the 4 PDFs are all from the same data-set obtained by providing PDFgetX2 with the same relative Q-ranges as in Figure 2**(a)** for normalization. The four curves have peaks shifted and even features disappearing. For example, the peak at 12 Å is wholly absent in two of the four PDFs. Correspondingly, the Pearson correlations are as low as 0.73 between these PDFs.

6.3.2 Q_{max} Sensitivity

So far, we have looked exclusively at data collected from a synchrotron ($\lambda = 0.137$ Å) or a CALS not used to its full potential ($\lambda = 1.54$ Å, $2\theta_{max} = 40$ degrees). There are other

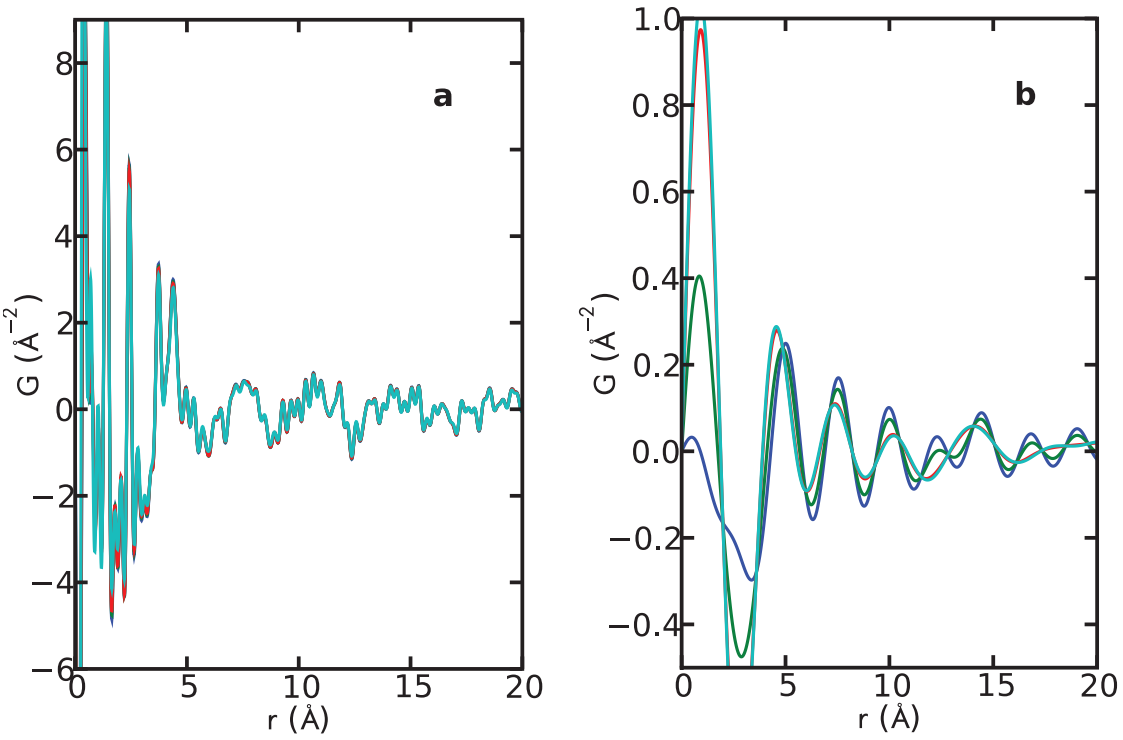

Figure 6.2: PDFs of amorphous-phase CBZ with **(a)** Q_{max} = 20.0 Å^{-1} and **(b)** Q_{max} = 2.8 Å^{-1} optimized over various ranges of Q. In panel **a**, all of the PDFs match up to each other perfectly, while in panel **b**, the PDFs exhibit variations in peak positions, relative intensities, and broadness.

X-ray sources between these two extremes. These include molybdenum anode laboratory sources (MALS), with a wavelength of 0.709 Å [Deslattes *et al.*, 2003] and silver anode laboratory sources (SALS) with a wavelength of 0.559 Å [Deslattes *et al.*, 2003]. We choose the maximum 2θ value of 160°, which corresponds to $Q_{max} = 8.0$ Å^{-1}, for the CALS, and a reasonable 2θ, for example 90 degrees, which corresponds to a limit where experiments may be completed in a short amount of time, for the MALS and SALS. The Q_{max} values that correspond to these conditions, with $2\theta_{max} = 90°$, are 12.5 Å^{-1} and 15.9 Å^{-1} for MALS and SALS, respectively.

To provide a uniform comparison of these instruments conditions, and in the absence of access to a complete set of suitable diffractometers, we have simulated data for the three instrument types using the data collected from the synchrotron only over the Q-range up to the relevant Q_{max} values given above for each anode type. In this test we found that it was important to renormalize the synchrotron data-sets for each value of Q_{max}. Although, in principle, the data should require the same normalization factor regardless of Q_{max}, in practice some deficiencies in the data corrections mean that the normalization needs to be tweaked for each Q_{max}. This underscores the importance of being able to carry out a reliable normalization, though we note that the fact of having the higher-Q data available allows us to make better corrections before the final step of treating the normalization.

Figure 3**(a)** shows the results for the crystalline Form-III CBZ, with the low angle $2\theta_{max} = 40$ degrees CALS data also shown for completeness.

It is clear from the figure that a Q_{max} of 12.5 Å^{-1}, accessible to a MALS, is sufficient to obtain virtually all of the features that distinguish the underlying molecular packing (i.e., in the region above $r = 3$ Å). However, although the general shape of the PDF is reproduced in the $Q_{max} = 8.0$ Å^{-1} data, there is a significant loss of information. For example, features such as the sharp peaks at 3.7 Å and 4.4 Å are completely lost. This suggests that, for differentiating the molecular packing in different forms of a typical API, an experiment with a Q_{max} of 12.5 Å^{-1} should be adequate. Such an experiment is easily carried out with a

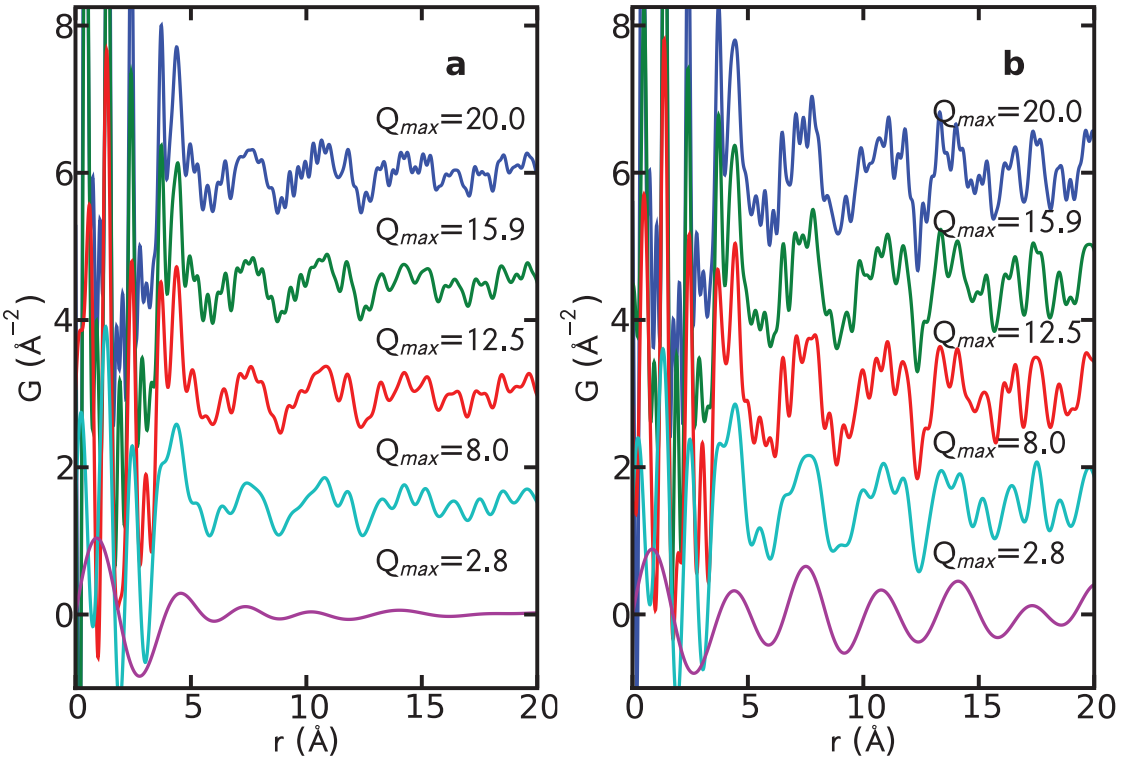

Figure 6.3: PDFs of **(a)** amorphous-phase CBZ and **(b)** Form III CBZ with Q_{max} values corresponding to, in order from top to bottom, a typical synchrotron setup, SALS, MALS, CALS ($2\theta_{max} = 90°$) and CALS ($2\theta_{max} = 40°$).

MALS, but not a CALS.

For the low-Q_{max} = 2.8 Å^{-1} CALS data, there is a dramatic loss of information and peak maxima in the curve do not even qualitatively correspond to the positions of sharp peaks in the higher resolution PDFs. Data collected over this range are clearly inadequate for fingerprinting.

It is evident from Figure 3**(b)** that the same conclusions can be drawn from the PDFs obtained from these amorphous APIs. They are considered to be X-ray amorphous in the sense that they have no sharp peaks in the powder diffraction measurement on a CALS but, as synchrotron based total scattering PDF measurements [Billinge *et al.*, 2010] show, they can have rather well defined local molecular packing. As with the crystalline Form-III we see that a Q_{max} of 12.5 Å^{-1} yields virtually all the fingerprinting information, but there is significant loss of information on reducing to Q_{max} = 8.0 Å^{-1}. We thus see that, as for the crystalline APIs, amorphous APIs can be fingerprinted using a practical measurement on a MALS.

We now turn to the quantitative correlation analysis using the Pearson correlation coefficients from these data, reproduced in Tables 6.1 and 6.2.

We see an overall trend that the Pearson correlation coefficients increase with decreasing Q_{max} for each of the same comparison-pairs (e.g., CBZ-III with CBZ-I or CBZ-III with CBZ-a). This reflects the fact that as the $G(r)$ (i.e. PDF) curves broaden due to the decreasing Q_{max}, the PDFs become less effective at differentiating between different underlying structures: PDF curves from dissimilar structures are less discriminating as Q_{max} decreases and the PDFs lose information content. However, this loss of information, and the consequent changes in the Pearson correlation coefficients, are quite small for $Q_{max} > 12.5$ Å^{-1}. By Q_{max} = 8.0 Å^{-1} the Pearson correlation coefficients have increased significantly for each comparison-pair, to the extent that it is now difficult to distinguish CBZ-a between CBZ-III and CBZ-I, whereas the synchrotron data clearly show that CBZ-a is made up of locally Form-III molecular packing [Billinge *et al.*, 2010]. We also see a false-positive for IND-a

Table 6.1: Summary of Pearson correlation coefficients between the PDFs shown in Fig. 6.3.2. Correlations higher than 0.8 are shown in bold (except when they are trivially unity).

	CBZ-III	CBZ-a	CBZ-I	IND-α	IND-a	IND-γ
$Q_{max} = 20$ Å^{-1}						
CBZ-III	1	**0.88121**	0.580032	0.36072	0.520868	0.535466
CBZ-a		1	0.721854	0.499347	0.692577	0.585051
CBZ-I			1	0.4143	0.607663	0.353945
IND-α				1	0.706309	0.477629
IND-a					1	0.648231
IND-γ						1
$Q_{max} = 15.9$ Å^{-1}						
CBZ-III	1	**0.88806**	0.587318	0.408013	0.540808	0.554994
CBZ-a		1	0.735184	0.528379	0.711449	0.603083
CBZ-I			1	0.461576	0.633917	0.371975
IND-α				1	0.747408	0.512554
IND-a					1	0.656109
IND-γ						1
$Q_{max} = 12.5$ Å^{-1}						
CBZ-III	1	**0.884121**	0.602841	0.414968	0.532329	0.552159
CBZ-a		1	0.743738	0.550489	0.694594	0.596248
CBZ-I			1	0.493308	0.642929	0.385381
IND-α				1	0.796457	0.546241
IND-a					1	0.648245
IND-γ						1

Table 6.2: Summary of Pearson correlation coefficients between the PDFs shown in Fig. 6.3.2. Correlations higher than 0.8 are shown in bold (except when they are trivially unity).

	CBZ-III	CBZ-a	CBZ-I	IND-α	IND-a	IND-γ
$Q_{max} = 8.0$ Å^{-1}						
CBZ-III	1	**0.896121**	0.609902	0.488752	0.541615	0.588639
CBZ-a		1	**0.803586**	0.611696	0.717735	0.620615
CBZ-I			1	0.567149	0.668514	0.403731
IND-α				1	**0.874273**	0.593747
IND-a					1	0.660049
IND-γ						1
$Q_{max} = 2.8$ Å^{-1}						
CBZ-III	1	0.592352	0.56047	0.242189	0.301645	0.315372
CBZ-a		1	0.738592	**0.815624**	**0.889357**	0.291114
CBZ-I			1	0.749409	0.688959	0.128833
IND-α				1	**0.95458**	0.25323
IND-a					1	0.324015
IND-γ						1

compared with IND-α that is not seen in the higher resolution high-Q_{max} data. Finally, at $Q_{max} = 2.8$ Å^{-1} the Pearson correlation don't provide any meaningful information. The correlation between CBZ-III and CBZ-a, which we expect to be high, is not significant and there are more false-positives than in any other case.

6.3.3 Information content of PDFs

Although a PDF can be plotted on an arbitrary grid, by convention, PDFs, including all of the PDFs shown above, are plotted on a relatively fine grid, with $\delta r \sim 0.01$ Å. However, it has recently been shown [Farrow *et al.*, 2011] that the ideal grid spacing for PDFs to maximize information content while minimizing correlations between data points is a grid spacing slightly less than the Nyquist-Shannon sampling frequency, i.e., $\delta r \leq \frac{\pi}{Q_{max}}$. Therefore, the number of independent data points in a PDF, N, is directly proportional to Q_{max}, since $N \sim \frac{r_{max}}{\delta r} = \frac{r_{max}(Q_{max})}{\pi}$. For example, the ideal grid spacings for $Q_{max} =$ 20.0 Å^{-1}, 8.0 Å^{-1}, and 2.8 Å^{-1} are 0.157 Å, 0.393 Å, and 1.122 Å, respectively, which means that the synchrotron PDF contains roughly 2.5 times more independent data-points than the 160° $2\theta_{max}$ CALS PDF, and 7.15 times as much as the 40° $2\theta_{max}$ CALS data. For illustration, Figure 4 contains the PDFs from Figure 3 on the Nyquist-Shannon grid.

We note how the CALS PDFs are much more coarse than those from the synchrotron data, reflecting their reduced information content and further confirming their limited value in amorphous API fingerprinting.

6.4 Conclusion

We have studied the efficacy of fingerprinting small molecular solids using atomic pair distribution functions (PDFs) obtained from data measured over different ranges of momentum transfer, Q. The chosen ranges corresponded to the maximum Q_{max} value accessible to copper-anode lab based XRPD devices, reasonable Q_{max} values accessible to molybdenum-

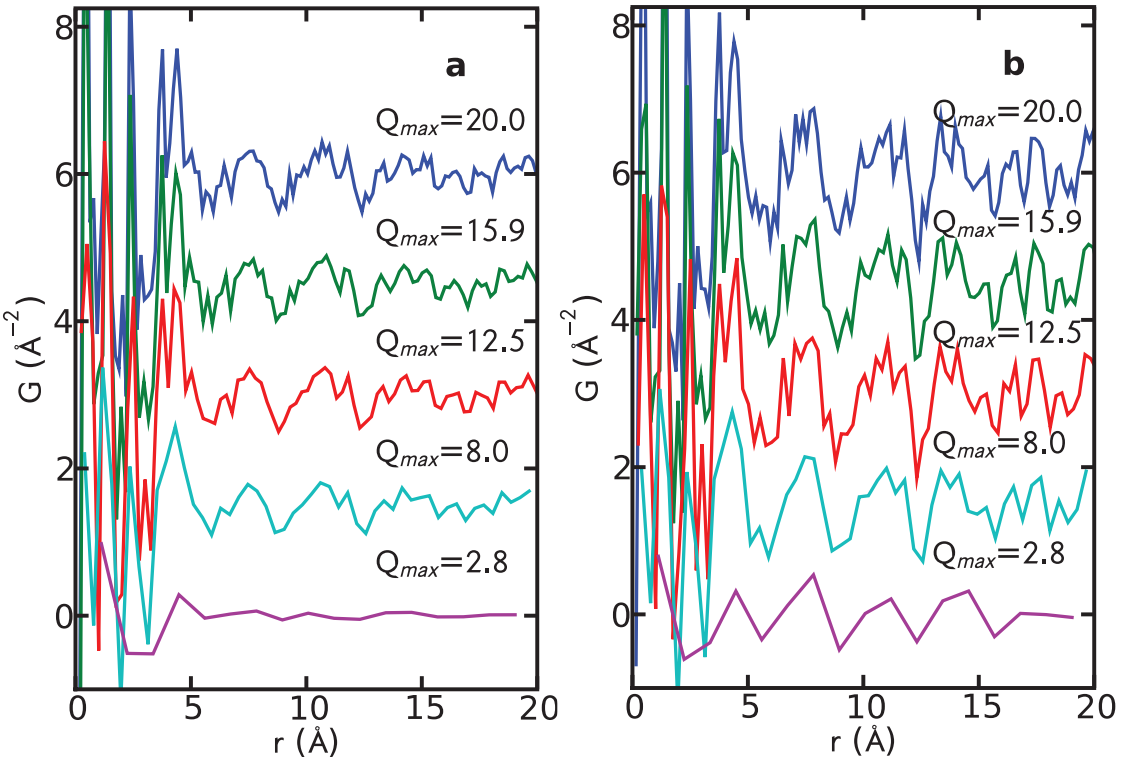

Figure 6.4: PDFs of **a** Amorphous CBZ and **b** beta phase CBZ from Figure 6.3.2 plotted on the Nyquist grid. The lower Q_{max} PDFs have a coarser grid.

and silver-anode devices, and a synchrotron beamline. We found that the ability to reliably fingerprint samples using PDF is closely related to the Q_{max} of the available data. As Q_{max} decreases, the PDFs become broader and lower resolution and the information content in the PDFs goes down. While the best results were offered by data collected at the synchrotron beamline, it was still possible to fingerprint with data in the Q regime attainable by silver and molybdenum lab-based instruments. However, fingerprinting failed with the low Q_{max} simulating copper anode lab-based instrument. Furthermore, we also found that data of a sufficiently high-Q could be robustly normalized resulting in a unique PDF. However, for low-Q_{max} data different normalization protocols for the same dataset resulted in a non-uniqueness in the resulting PDFs. This shows that data with a Q_{max} of 12.5 Å^{-1} or above should be used for effective fingerprinting of pharmaceuticals. Such data are easily obtained in a laboratory environment using a Mo or Ag anode X-ray source, or at a synchrotron source where higher quality data and shorter counting times are possible. We expect that other small molecular systems will have similar requirements.

Over the course of the past two chapters, we have laid the groundwork for the application of TSPDF to pharmaceutical compounds by first demonstrating a proof of concept and then investigating the data requirements for the technique. However, we have not yet explored a so-called truly amorphous compound, as the melt-quenched samples of CBZ and IND turned out to be nanocrystalline. For this reason, in Chapter 7 we investigate two truly amorphous compounds using TSPDF to determine whether we are able to divine out any structural information from the amorphous phase.

Chapter 7

Total scattering pair distribution function analysis of amorphous pharmaceuticals

The work in this chapter was done in collaboration with Simon J.L. Billinge and presented to the pharmaceutical industry.

In Chapters 5 and 6, we laid the fundamental groundwork in the technique of using TSPDFs to fingerprint small molecules (e.g. pharmaceutical compounds). However, so far, we have only looked at two compounds, carbamazepine and indomethacin, and, the X-ray amorphous melt-quenched samples of both turned out to be nanocrystalline rather than truly amorphous. To further test the TSPDF technique, in this chapter we attempt to fingerprint the crystalline and truly amorphous phases of two new compounds using TSPDF. We also investigate the effects that temperature, static disorder, and dynamic disorder have on the TSPDFs.

Furthermore, as we had discussed in Chapter 6, the high energy X-rays produced by a molybdenum anode laboratory diffractometer should theoretically provide for a sufficient

Q-range for fingerprinting with TSPDF. However, up to this point, all the data that we have analyzed were actually collected at a synchrotron. In this chapter, we investigate TSPDFs calculated from data collected using an actual Mo diffractometer compared to those collected using a synchrotron.

The analysis of individual TSPDFs is much more detailed in this chapter. While we have spared the reader the gory details in the past and have been content to describe TSPDFs in a general way, here we offer a much more detailed qualitative analysis by plotting several TSPDFs on the same plot and splitting the r range to highlight differences between areas of mostly intra-molecular and inter-molecular correlations. As part of this detailed discussion, we focus on specific features (i.e. peaks at certain r values) of the TSPDF and how they change between phases and at different temperatures. The byproduct of such an exhaustive study is that the material is quite dense compared to the previous chapters. However, this type of analysis is key to understanding the power of fingerprinting using TSPDF.

The two compounds presented in this chapter are of a proprietary nature and will henceforth be referred to simply as 'compound A' and 'compound B'.

7.1 Introduction

Little is known about the local ordering in amorphous and nanocrystalline APIs due to a lack of reliable experimental probes. Conventional X-ray powder diffraction (XRPD) patterns of amorphous compounds become broad and featureless due to a lack of long range atomic structure and are not useful for differentiating between different local molecular packing arrangements [Billinge *et al.*, 2010]. This behavior is illustrated in Figure 7.1 where we see that XRPD data collected from a crystalline compound (blue) has a unique structural fingerprint, while the data collected from an amorphous compound (green) offers no information that could be used in fingerprinting. This used to be the end of the story as far as fingerprinting amorphous phase compounds was concerned.

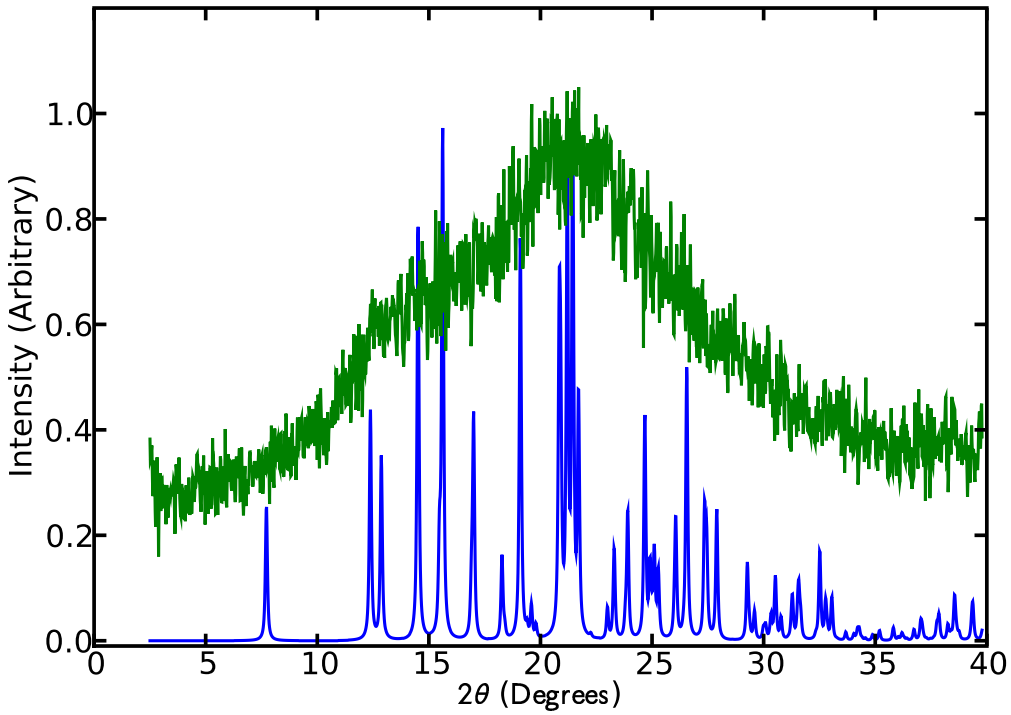

Figure 7.1: Typical XRPD patterns of an amorphous compound (green) and a crystalline compound (blue).

Recently, atomic pair distribution function (PDF) and total scattering atomic pair distribution function (TSPDF) methods have been suggested as an alternative approach for fingerprinting amorphous APIs [Bates *et al.*, 2006; Billinge *et al.*, 2010] and for providing valuable insight into the local packing of pharmaceuticals in non-crystalline forms. If these methods are validated as giving reliable fingerprints of APIs in the amorphous state, they could make a significant contribution to expediting the wider commercial exploitation of amorphous pharmaceuticals.

The PDF, $G(r)$, yields the probability of finding an atom at a distance r from any reference atoms and so provides information on local structure in real space. It is calculated by taking the sine Fourier transformation of the structure function, $S(Q)$, with the relationship in Equation (1.23). The $S(Q)$ is obtained by correcting X-ray diffraction data for self-absorption, multiple-scattering, incoherent scattering, etc. The details of these corrections and calculations are in Chapters 1 and 3 and they are applied to the data using special software, PDFGetX2 [Qiu *et al.*, 2004].

The information content, particularly the resolution in r, of a PDF depends on the information content of the $S(Q)$, which in-turn depends on the magnitude of Q_{max}, the momentum transfer, of the X-rays used. Q is related to the Bragg angle, θ, which is half the scattering angle, and the wavelength, λ, of the incident radiation with Equation (1.2). Generally, there are two ways to increase the Q_{max} that can be attained in an experiment. The first is to measure out to a higher $2\theta_{max}$, which has a hard limit, and the second is to increase the energy of the X-rays used, which is dependent on the source anode of the diffractometer.

As discussed in Chapter 6, the maximum possible Q_{max} values, typically measured in inverse angstroms, that are available to copper anode laboratory source (CALS), molybdenum anode laboratory source (MALS), and silver anode laboratory source (SALS) diffractometers are 8 Å^{-1}, 16 Å^{-1}, and 22 Å^{-1}, respectively. Synchrotrons are able to go as high as 40 Å^{-1} depending on the configuration. The analysis in Chapter 6 has shown that there is

not enough information in data collected on CALS diffractometers for proper PDF analysis. However, MALS and SALS diffractometers and synchrotrons are capable of producing sufficiently information-rich TSPDFs [Dykhne *et al.*, 2011].

After TSPDFs have been created, they can be 'fingerprinted' against one-another using qualitative and quantitative techniques. For qualitative analysis, we simply use a visual comparison of the plots of the PDFs [Billinge *et al.*, 2010; Dykhne *et al.*, 2011]. For quantitative analysis, we use a home-written program that computes the Pearson product-momentum correlation [Myers and Well, 2010; Dykhne *et al.*, 2011] to quantify the similarity of curves using Equation (1.24).

We study correlations between PDFs of crystalline compounds in the range $r = 3.0 - 20.0$ Å. We chose this range because the very local structure (i.e. $r < 3.0$ Å) of all molecular samples is similar due to intra-molecular atom pairs, for example consisting of nearest and next-nearest carbon-carbon bonds at 1.4 Å and 2.4 Å, respectively. Applying the correlation analysis to the entire data range does not change the result significantly but reduces the sensitivity to finding differences in molecular packing of the correlation analysis by including a range of r that is highly similar regardless of the packing.

In Chapter 5, we calculated the domain size of melt-quenched nanostructured carbamazepine (CBZ) using TSPDF [Billinge *et al.*, 2010]. By looking at how the ripples die out in a TSPDF we can tell the extent of structural coherence. However, it's very difficult to tell whether this is because of distinct physical grains of that size or structural coherence dying out due to disorder because the TSPDF dies out in the same way in both cases.

Here, we report the results of two TSPDF studies on two amorphous compounds, A and B. The goal of our investigation was to fingerprint and determine whether we could estimate the domain size or the range of structural coherence of the amorphous form of each compound. Also, for compound A, we wanted to determine, using PDF and fingerprinting, whether the amorphous forms created by using two methods - spray drying and cryomilling - are similar.

7.2 Experimental Methods

As mentioned earlier, two TSPDF studies were conducted. The first was to measure all the samples at a temperature of 100 K at a synchrotron. The second was to measure all the samples at room temperature, 300 K, using a MALS diffractometer. The measurements were performed at different temperatures to see what effect, if any, temperature has on the TSPDFs of APIs. In theory, there is less atomic motion at lower temperatures resulting in sharper TSPDF peaks in well ordered crystals at low-T. At higher temperatures, peaks broaden due to atomic motion and dynamic disorder in the compound. These measurements are not ideal, though, because they were conducted on different instruments with different experimental setups, which can have an effect on the resolution and broadening of peaks in a TSPDF. An ideal experiment would be to do the temperature comparison on the same instrument.

Data were collected on three crystalline forms and at least one amorphous form for each compound at both temperatures. There were several amorphous forms of Compound A that were synthesized using two different methods - cryomilling and spray drying.

7.3 Synchrotron

For the low temperature measurements at the synchrotron, amorphous samples were stored in a -80° C fridge at Columbia University, transported in dry ice, and kept in dry ice until they were needed for an experiment. Crystalline samples were kept in a regular refrigerator until the day of the experiment.

Experiments were performed at beamline X7B of the National Synchrotron Light Source (NSLS) at Brookhaven National Lab (BNL). Total scattering data were collected using the Rapid Acquisition PDF method [Chupas *et al.*, 2003]. A Perkin-Elmar 2D image plate detector was placed 114.9 mm behind the sample, perpendicular to a high energy X-ray beam with $\lambda = 0.3184$ Å. A spinner rotated the capillary holding the samples at 3 rpm.

Data were collected from the sample at at 100 K in frames of 30 s, repeated 60 to 90 times per sample. Measurements were integrated and normalized with respect to total collection time using the software Fit2D [Hammersley, 1998]. TSPDFs were calculated from the experimental data using the software PDFGetX2 [Qiu *et al.*, 2004]. All PDFs were made using $Q_{max} = 16.0$ Å^{-1} to match the maximum Q_{max} available to a MALS instrument.

7.4 Lab Diffractometer

The room temperature measurements were conducted on an in-house PANalytical MALS diffractometer with $\lambda = 0.709$ Å. These experiments were not conducted by members of the Billinge Group, hence we do not have access to the exact protocols used to measure the data. However, we calculated TSPDFs from the data measured by PANalytical using PDFGetX2 [Qiu *et al.*, 2004] with $Q_{max} = 16.0$ Å^{-1}, the maximum available to a MALS.

7.5 Results

7.5.1 Crystalline Compound B

As discussed earlier, conventional X-ray powder diffraction techniques are unable to probe amorphous phase compounds due to a lack of long range order in the compounds, as was illustrated in Figure 7.1. However, when X-ray diffraction data are collected on a sufficiently high energy instrument, such as a synchrotron, and then Fourier transformed into a TSPDF, real-space structural information becomes available.

Before we look at the amorphous phase, though, we first want to get a feeling for the features of a TSPDF by investigating typical TSPDFs of crystalline APIs. Figure 7.2, which shows the three polymorphs of compound B, is representative of how TSPDFs of crystalline compounds look.

In Figure 7.2**(a)**, we see that the TSPDF peaks up to about 3 Å in all of the crystalline

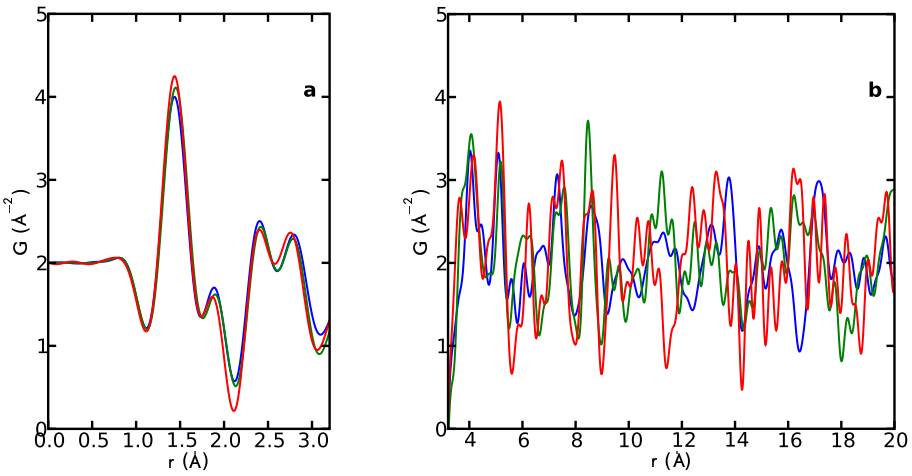

Figure 7.2: **(a)** The region between 0 Å and 3.2 Å and **(b)** the region between 3.2 Å and 20.0 Å of the TSPDFs of the three crystalline forms of compound B. Data collected at 100 K. Form α is in blue, form β is green, and form γ is in red. All three TSPDFs look the same in the low-r region, which corresponds to intra-molecular correlations, and are completely different when $r > 3.2$ Å, when inter-molecular correlations start to dominate. The TSPDFs in **(b)** have been offset and multiplied by 4.0 to fit on the axes.

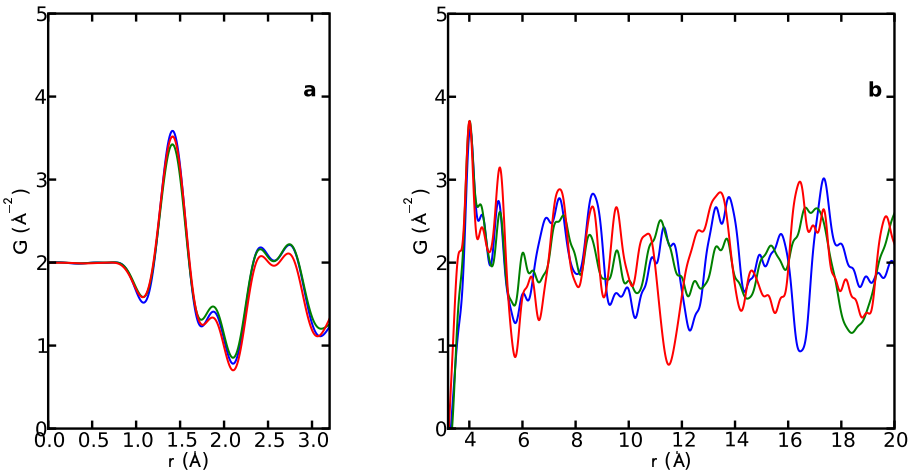

Figure 7.3: TSPDFs of the same formulations as Figure 7.2. Data gathered at 300 K on the MALS setup. The TSPDFs in **(b)** have been multiplied by 5.0 to fit on the axes.

forms look exactly the same. These peaks represent intra-molecular correlations from rigid sub-groups in the molecule such as nearest neighbor bonds and rings. Since the same molecule is present in all three forms of the compound, this region looks the same regardless of the form. However, as we see in Figure 7.2**(b)**, the TSPDFs start to look very different from one another at higher r. At this point, inter-molecular bonds start to dominate the TSPDF and, since all three forms of the compound have different molecular conformations and packing, the TSPDFs look different. The higher r region is what is used to differentiate the crystalline forms from each other via fingerprinting using the TSPDF.

The TSPDFs in Figure 7.2 are all from data collected at 100 K. As discussed in the Methods section, at this temperature there is less atomic motion than at room temperature. Thus we would expect sharper peaks in the TSPDFs. For comparison, we look at the TSPDFs of the crystalline compounds that were measured at room temperature on the MALS setup.

We see by comparing Figure 7.3**(a)** with Figure 7.2**(a)** that the low-r range looks exactly as it looked at low temperature, albeit with slightly broadened peaks. We also see that the

Table 7.1: Summary of Pearson correlation coefficients between the crystalline forms of compound B in the range 3.0-20.0 Å. Correlations higher than 0.8 are shown in bold (except when they are trivially unity).

	Form α	Form β	Form γ
Temperature: 100 K			
Form α	1	0.551	0.436
Form β		1	0.599
Form γ			1
Temperature: 300 K			
Form α	1	0.721	0.531
Form β		1	0.728
Form γ			1

peaks at the high-r range are in the same positions as they were at low temperature but are also broader. For instance, in Figure 7.2**(b)** the Form β (green) PDF has a triplet of peaks between 10 Å and 12 Å, while in Figure 7.3**(b)**, the same area looks like one broad peak with a shoulder. We see a similar situation in the range 12-14 Å in the Form β (red) TSPDFs. TSPDFs tend to broaden with increasing temperature due to dynamic disorder, meaning that the atoms tend to vibrate more at higher temperatures, and some larger scale librational motions can also set in.

The investigation we have done so far is an example of qualitative fingerprinting. Generally, this process involves plotting TSPDFs on top of one another and looking for similarities and differences between the various forms. For quantitative comparison, we use the Pearson product momentum correlation described in the Introduction. The best way to display the Pearson correlations is in a table as in Table 7.1. Empirically, we have found that correlations above 0.8 mean that the two TSPDFs are highly correlated.

We see from Table 7.1 that, as we would expect, the correlations between the crystalline phases are quite low. However, we also see that the correlations increase with higher temperature, since broad peaks tend to make PDFs look more similar to one another.

Overall, the above results show that PDF is a viable method for fingerprinting crystalline compounds as currently done using standard crystallographic techniques. However, unlike standard XRPD patterns, the PDF gives real space resolution to the features of a compound.

7.5.2 Amorphous Compound B

Now that we have become familiar with how a TSPDF looks, we introduce the amorphous form of compound B. We have four measured data sets - one at 100 K and three at 300 K. All four are plotted in Figure 7.4. The TSPDFs of the data measured at 300 K are blue, green, and red, while the TSPDF of the data measured at 100 K is light blue. The low temperature measurement is plotted with an offset in panel **(b)** for ease of viewing.

In Figure 7.4**(a)**, we see that the low-r region looks exactly the same for all four amorphous measurements and is also identical to all of the crystalline measurements. This is good news and as expected. From Figure 7.4**(b)**, we see some interesting results. First, we note that sharp features in the amorphous TSPDF disappear after about 10 Å. There is some additional signal in the form of broad features up to about 13.5 Å (18 Å at 100 K). After 13.5 Å the TSPDF goes flat, modulated by noise fluctuations. We can therefore estimate the range of structural coherence to be about 13.5 Å. This is the distance that we can see structural information away from any arbitrary atom. Most likely, this distance does not correspond to grain size, but more information about the compound would be needed to be sure.

We also note that most of the structural feature are reproduced in all the amorphous phase TSPDFs across temperatures. Structural features include the three sharp features at 4.0 Å, 4.5 Å, and 5.1 Å, the valley that drops down to about 5.5 Å, a peak at about 6.0 Å, and another peak at about 6.8 Å. However, there is one obvious feature that sticks out - we

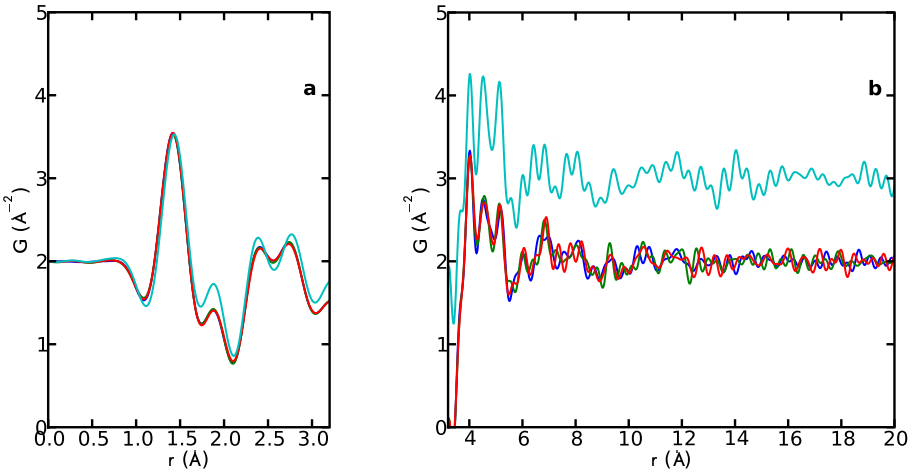

Figure 7.4: **(a)** The region between 0 Å and 3.2 Å and **(b)** the region between 3.2 Å and 20.0 Å of the TSPDFs of the amorphous form of compound B. The three measurements at room temperature are blue, green, and red, while the measurement at low temperature is light blue. The room temperature phases are plotted on top of one another and the low temperature phase is plotted with an offset for ease of viewing. The amorphous forms look exactly the same at low r, and then have identical, reproducible, structures up to about 7 Å. All structural information is lost by about 13.5 Å. The TSPDFs in **(b)** have been multiplied by 6.0 to fit on the axes.

see a sharp peak in the 100 K (light blue) PDF at about 6.4 Å which is not present in the other measurements. This feature could be investigated further, but is clearly particularly sensitive to temperature. However, this is beyond the scope of this study.

Figure 7.4 is important for two reasons. First, it corroborates that peaks in the TSPDF of the amorphous samples are real structural features, and not just random noise, since they are very closely reproduced across measurements that were taken at different times, on different machines, under different experimental conditions. Second, it allows us to estimate the range of structural coherence of the amorphous sample based on where there are no more structural features in the PDF. This method is not exact, but it does provide a good estimate.

Now that we have seen what crystalline and amorphous TSPDFs of compound B look like separately, we combine them by plotting them on top of each other in Figure 7.5. We plot the data up to 14 Å because there is no signal beyond that point.

Figure 7.5 shows all four measurements that were gathered at 100 K: three from crystalline samples and one from an amorphous sample. The three crystalline forms are plotted in different shades of blue and the amorphous is plotted in red. Visually we can fingerprint in two ways - first by determining which features in the PDF of the amorphous are in the crystal structures, and second, by the process of elimination, determining which features in the crystalline forms are not in the amorphous. The amorphous PDF shares some features with all three crystalline forms, but, it appears to share the most features with Form α, the light blue PDF. These features include a shoulder at 3.6 Å, peaks at 4.0 Å, 5.3 Å, 5.55 Å, and 6.05 Å, and a valley at 5.4 Å. This can lead us to believe that the local ordering of the amorphous phase is similar to that of Form α. On the other hand, we see that the sharp peak present in the Form γ TSPDF (dark blue) at 6.25 Å is not in the amorphous PDF. Also the shoulder at 5.8 Å in the Form γ corresponds to a trough in the amorphous phase. Finally, we also note that the peak present in Form α and the amorphous PDF at 4.0 Å is shifted to 4.16 Å in Form γ. These observations lead us to believe that the amorphous

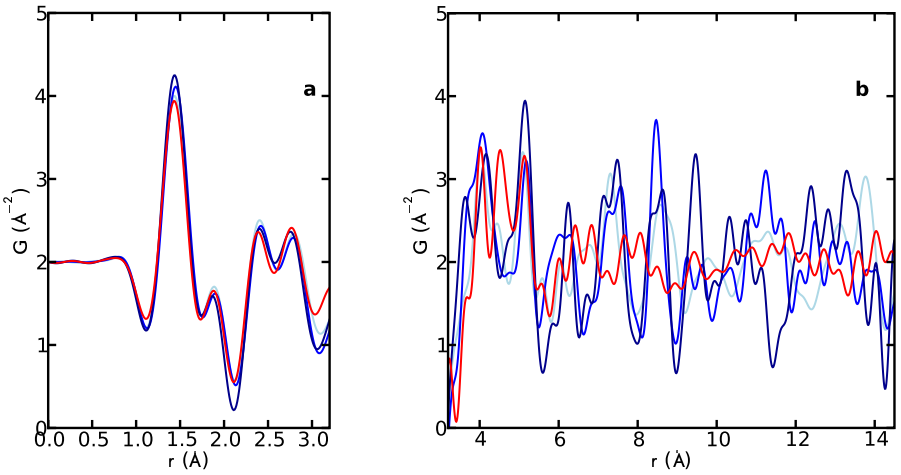

Figure 7.5: **(a)** The region between 0 Å and 3.2 Å and **(b)** the region between 3.2 Å and 14.0 Å of the PDFs of the three crystalline forms and the amorphous form of compound B collected at 100 K. Form α is light blue, Form β is medium blue, Form γ is dark blue, and the amorphous form is red. The amorphous form looks just like the crystalline forms at $r < 3.2$ Å. It loses sharp features at around 6 Å and all structural information at around 14 Å. The crystalline phase PDFs in **(b)** have been multiplied by 4.0 and the amorphous phase by 6.0 to fit on the axes.

Table 7.2: Summary of Pearson correlation coefficients between the TSPDFs shown in Fig. 7.5 in the range 3-7 Å. Correlations higher than 0.8 are shown in bold (except when they are trivially unity).

	Amorphous
Form α	**0.877**
Form β	0.727
Form γ	0.648

phase probably does not have the Form γ packing.

The most exciting result is possibly the observation of a very sharp peak in the amorphous phase at 4.5 Å that is not present in any of the crystalline forms. This is a surprising result. It is much easier to explain the loss of peaks on going from crystalline to amorphous due to disorder in the packing, but the appearance of a new sharp feature in the amorphous TSPDF means that there is a new, specific packing arrangement or conformation that is unique to the amorphous phase. Without the ability to model and refine the crystal structure in real space, we won't know what it is. This needs further development of the methodology, but the observation of a unique packing in the amorphous state now justifies such a development on the academic side. The possibility that it is an artifact in the data can be clearly ruled out as we have four independent measurements of the amorphous state, at two temperatures shown in Figure 7.4, and this feature is robust in all the measurements.

In addition to the qualitative observations above, we can also quantitatively compare the TSPDFs to one another using the Pearson product momentum correlations discussed earlier. We look at the correlations between each of the crystalline forms and the amorphous form in the range 3-7 Å because this is the range where we see sharp features in the amorphous phase and so this region is the only one where it makes sense to compare the crystalline and amorphous patterns.

Table 7.2 confirms our observations. The highest correlations occur between the amorphous phase and Form α, while the lowest correlations are between the amorphous and Form γ. In general, Pearson correlations are used as a complementary tool to visually comparing TSPDFs in fingerprinting. From experience, we have found Pearson correlations greater than 0.8 signify a strong correlation between two sets of data, but this is not a strict rule.

Next, we compare the crystalline and amorphous data collected at room temperature. Although three amorphous data sets were collected, they are sufficiently similar that one representative data set can be used in this analysis. Furthermore, as we have already determined that the low-r range will look the same, only the high-r range plot is shown below in Figure 7.6.

We see that all of the TSPDFs in Figure 7.6 look more similar to one another than in the low temperature case. This is not surprising as thermal disorder decreases the sharpness of features in the crystalline TSPDFs making them look more like amorphous TSPDFs which have been broadened due to static disorder. The first sharp peak at roughly 4.0 Å, which was slightly shifted in position in the three low temperature crystal forms, is now in the same exact spot for all TSPDFs in the room temperature case. Similarly, we see that the trough at around 5.7 Å exists in all measurements. Overall, the amorphous phase seems to look the most like Form α or Form β (light blue and medium blue, respectively). In addition to the two sharp peaks at 4.0 Å and 5.1 Å, they also share a peak at 6.0 Å. However, while the amorphous phase and Form α both have a peak at 6.8 Å, Form β does not have any distinguishable features. As with the 100 K temperature data, Form γ looks the least like the amorphous TSPDF.

We again note that the sharp peak in the amorphous phase at 4.5 Å still does not correspond to any of the three phases. Since this peak is reproducible between the amorphous data collected on different instruments at different temperatures, it is likely that the peak contains real structural information. The fact that the peak does not exist in any of the

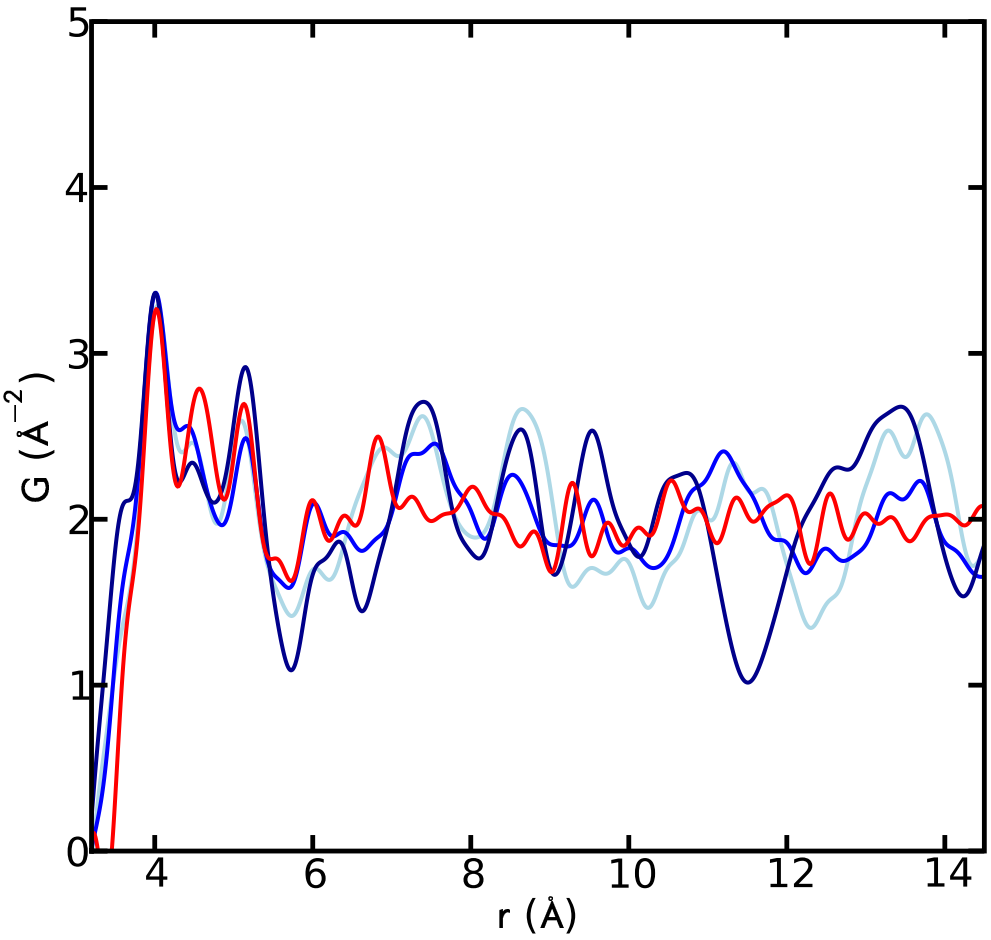

Figure 7.6: The region between 3.2 Å and 14.0 Å of the TSPDFs of the three crystalline forms and the amorphous form of compound B collected at room temperature. The three crystalline forms are in different shades of blue while the amorphous form is in red. The crystalline phase TSPDFs have been multiplied by 4.0 and the amorphous phase by 6.0 to fit on the axes.

Table 7.3: Summary of Pearson correlation coefficients between the TSPDFs shown in Fig. 7.6 in the range 3-7 Å. Correlations higher than 0.8 are shown in bold (except when they are trivially unity).

	Amorphous
Form α	**0.945**
Form β	**0.938**
Form γ	**0.804**

three crystalline forms at both temperatures means that it warrants further investigation.

We see in Table 7.3 that the correlations between the crystalline and the amorphous forms are all substantially higher than they were at low temperature. Furthermore, while Table 7.2 showed that the amorphous was far more correlated to Form α than the other two forms, in Table 7.3 we see that the amorphous is just slightly more correlated to Form α than Form β. This can be partially explained by the broadening that occurs in the TSPDFs at higher temperatures. The fact that has remained constant is that the amorphous form is the least correlated to Form γ. Further investigation is necessary to determine which is the ideal temperature for performing these types of measurements.

We can draw several conclusions from our analysis of compound B. First, we have successfully fingerprinted the three crystalline forms of the compound against one another and demonstrated the effects that temperature has on TSPDFs of APIs. Next, we have demonstrated that the TSPDFs of the amorphous form of compound B are reproducible and contain structural features. Using these PDFs we were able to estimate the range of structural coherence to be roughly 13.5 Å. Finally, when combining the crystalline and amorphous TSPDFs, we saw that the amorphous form has packing most similar to that of Form α. However, due to the sharp peak at 4.5 Å that is not present in any crystalline form, we are unable to make any conclusions about the amorphous form packing and can

only recommend further studies. In particular, temperature dependence studies should be performed on the same instrument to determine the best temperature to collect data without having to account for resolution effects such as broadening and damping that occur naturally when comparing data collected on different instruments. Also, these additional studies should explore why the peak in the amorphous samples at 4.5 Å is not present in any crystalline form.

7.5.3 Crystalline Compound A

For this compound, we have data collected from each of the three crystalline forms plus data collected from amorphous forms that were created using two different methods, cryomilling and spray drying.

First, we fingerprint the crystalline forms against one another, as we did for compound B. The plots will not be shown here, but the Pearson correlations in Table 7.4 demonstrate that, as expected, the three crystalline forms are quite different from one another at both temperatures.

7.5.4 Amorphous Compound A

Similar to our investigation above with compound B, we first want to see if there are any structural differences among the amorphous forms and see if we can determine the structurally coherent domain. We do this by plotting all of the collected amorphous data.

Figure 7.7**(a)** contains all of the TSPDFs from the amorphous samples plotted on top of one another in the low-r region. The three samples measured at 100 K were: one cryomilled (light blue), and two spray dried (medium blue and dark blue). The two 300 K measurements were of cryomilled (light green) and spray dried (dark green) samples. We instantly see that the measurements do not perfectly agree with one another in the low-r region as they had with compound B, but the general low-r structure is the same. We see the expected C-C peaks at 1.4 Å and 2.4 Å and also a peak at 1.9 Å. Within each temperature

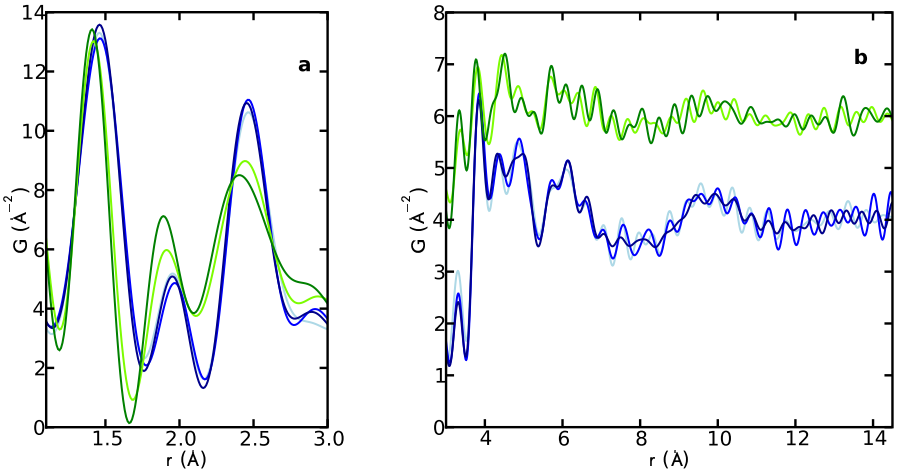

Figure 7.7: **(a)** The low-r regions of three TSPDFs of amorphous compound A collected at 100 K in shades of blue and the two PDFs collected at 300 K in shades of green plotted on top of each other. **(b)** The high-r regions of the same TSPDFs from panel **(a)** plotted with an offset. The 100 K data (blue) have been multiplied by 5.0 to fit on the same axes with the 300 K data.

Table 7.4: Summary of Pearson correlation coefficients between the crystalline forms of compound A in the range 3.0-20.0 Å. Correlations higher than 0.8 are shown in bold (except when they are trivially unity).

	Form α	Form β	Form γ
Temperature: 100 K			
Form α	1	0.375	0.372
Form β		1	0.465
Form γ			1
Temperature: 300 K			
Form α	1	0.397	0.554
Form β		1	0.559
Form γ			1

grouping the TSPDFs are almost the same (i.e. all three blue TSPDFs look just like one another) and we just see a slight peak shift across temperatures by 0.05 Å or less. The peak at 2.4 Å is a great example of peak broadening with increasing temperature. At 100 K, it is sharp and tall, but then it broadens out and becomes shorter at room temperature.

We note that the 100 K TSPDFs in Figure 7.7**(b)** match up very well, even though the samples were processed in different ways. We see peaks with structural information at 3.3 Å, 3.8 Å, 4.33 Å, 5.7 Å, and 6.1 Å. Two of the forms have strong peaks at 4.85 Å while a third has a peak-shaped hump in the area. We also see some broad structural information in the form of a broad feature between 8.5 Å and 11.5 Å, but no structural information in $r > 11.5$ Å. This is an important result because it implies that both amorphous forms are the same, even though they were created using different processing methods. The Pearson correlations among the three phases in the range 3.0-7.0 Å are roughly 0.99. We also estimate the range of structural coherence to be roughly 11.5 Å because there is no signal

beyond that point. This again probably does not correspond to a grain size and is instead the distance that we can see away from an arbitrary atom.

The 300 K TSPDFs in Figure 7.7**(b)** also match up well to each other. We see peaks at 3.35 Å, 3.8 Å, 4.5 Å, 4.9 Å, 5.7 Å, and 6.06 Å. In some instances the peaks don't exactly line up with each other as they do at 100 K, but these discrepancies are generally quite small, less than 0.1 Å, and can probably be chalked up to the dynamic disorder introduced at higher temperatures. The Pearson correlation between the two TSPDFs in the range 3-7 Å is 0.912, which is lower than what it was at low temperature, but still very high. Finally, we can again estimate the range of structural coherence to be roughly 11.5 Å.

When we compare the TSPDFs across temperatures, we see that many peaks are roughly in the same positions, such as those at 3.3 Å, 3.8 Å, 5.7 Å, and 6.1 Å. Many of the peaks are slightly shifted, but these shifts are less than 0.1 Å and are expected. A very interesting difference is that the 300 K TSPDFs have a sharp peak at roughly 4.5 Å, while the 100 K TSPDFs have peaks on either side of it. Most likely, the increase in atomic motion that results from the higher temperature caused the two peaks to combine into one big one.

Now that we have looked exclusively at the amorphous data and concluded that there is no difference between the various forms, we can compare them to the crystalline phases. We begin by looking at the low temperature data in Figure 7.8.

We plot Figure 7.8 between 3 Å and 12 Å because there are no differences in structure at the low-r range and there is no structural information after about 11.5 Å.

All three forms share a peak with the amorphous form at 3.8 Å and at 4.3 Å. However, Form γ (dark blue) reproduces these peaks the best. Furthermore, there is a hump between 4.7 Å and 5.1 Å that the amorphous form shares with Form γ but not the other two forms. From this, we can conclude that the amorphous form looks the most like Form γ. However, we also note that the peak in the amorphous phase at 6.1 Å does not have an analog in any of the crystalline forms. In fact, they all have troughs at that point. We also see that Form β looks the least like the amorphous form, particularly because it has a very strong

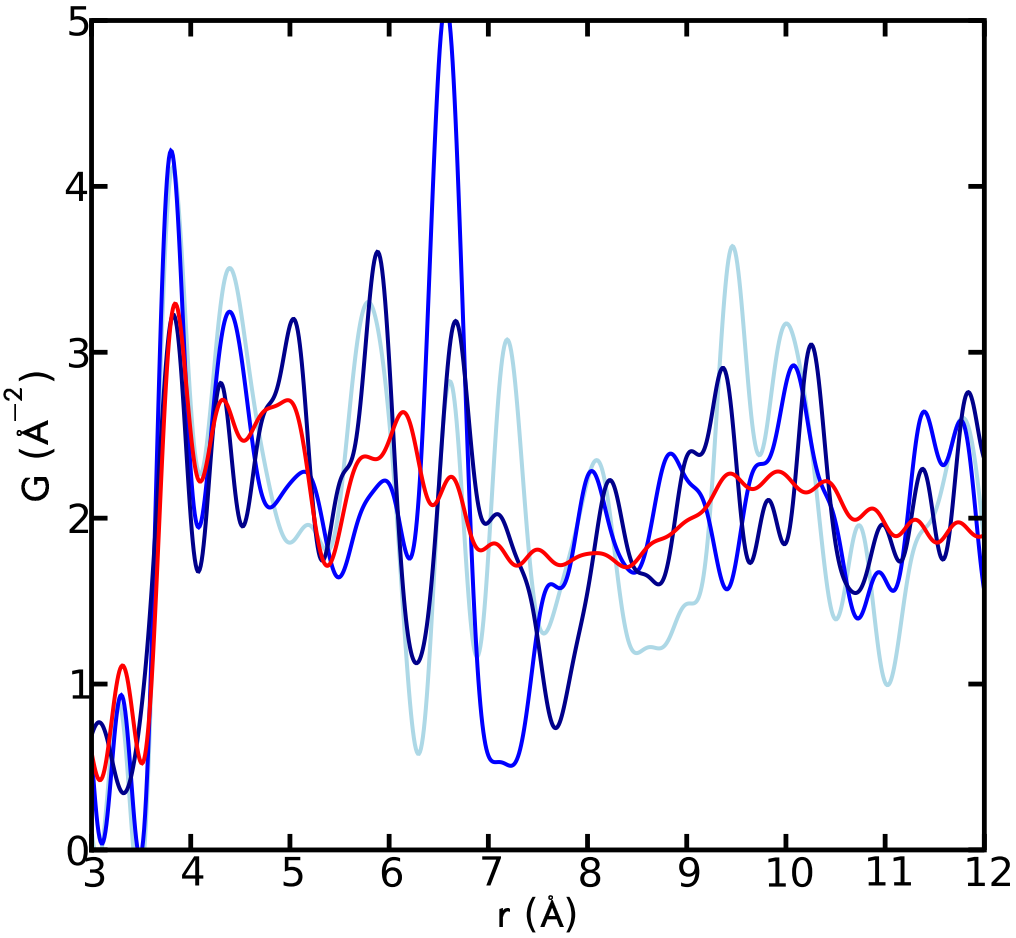

Figure 7.8: TSPDFs of the amorphous form of compound A (red) and three forms of the crystalline compound, Form α (light blue), Form β (medium blue), Form γ (dark blue) collected at 100 K plotted between 3.0 Å and 12 Å. TSPDFs have been multiplied by 3.0 to fit on the plot.

Table 7.5: Summary of Pearson correlation coefficients between the TSPDFs shown in Fig. 7.8 in the range 3-7 Å. Correlations higher than 0.8 are shown in bold (except when they are trivially unity).

	Amorphous
Form α	0.794
Form β	0.696
Form γ	0.759

features at 6.57 Å that is not in the amorphous.

From Table 7.5 we see, as expected, that the lowest correlation is with Form β. However, we also see that the Pearson factor is comparable between Forms α and γ, though the qualitative analysis indicated that Form γ looked the most like the amorphous form. The Pearson analysis does not yield a clear winner here, probably due to the sharpness of the peaks in the crystalline phases compared to the amorphous case.

We can now do the same analysis on the data collected at room temperature.

All phases have peaks at 3.8 Å, Forms β and γ reproduce the peak at 4.4 Å, while the Form α peak is shifted to 4.2 Å. Interestingly, the peak in the amorphous phase at 4.85 Å is replicated perfectly in Form γ, but in none of the other crystalline forms. Furthermore, the rise in Form γ at 5.55 Å and drop at 6.1 Å is imitated well in the amorphous phase. In fact, almost all features of Form γ are represented in the amorphous phase except a trough in at 6.4 Å, which coincides with a peak in the amorphous. In particular, the peak in the amorphous at 6.1 Å which had no analog at low temperature now looks like it is part of a Form γ peak.

A table of the Pearson correlations is in Table 7.6. Note that, since the two amorphous forms are relatively similar to one another, but are not exactly the same, as they were at low temperature, Pearson correlations for both forms are in the table, even though only one

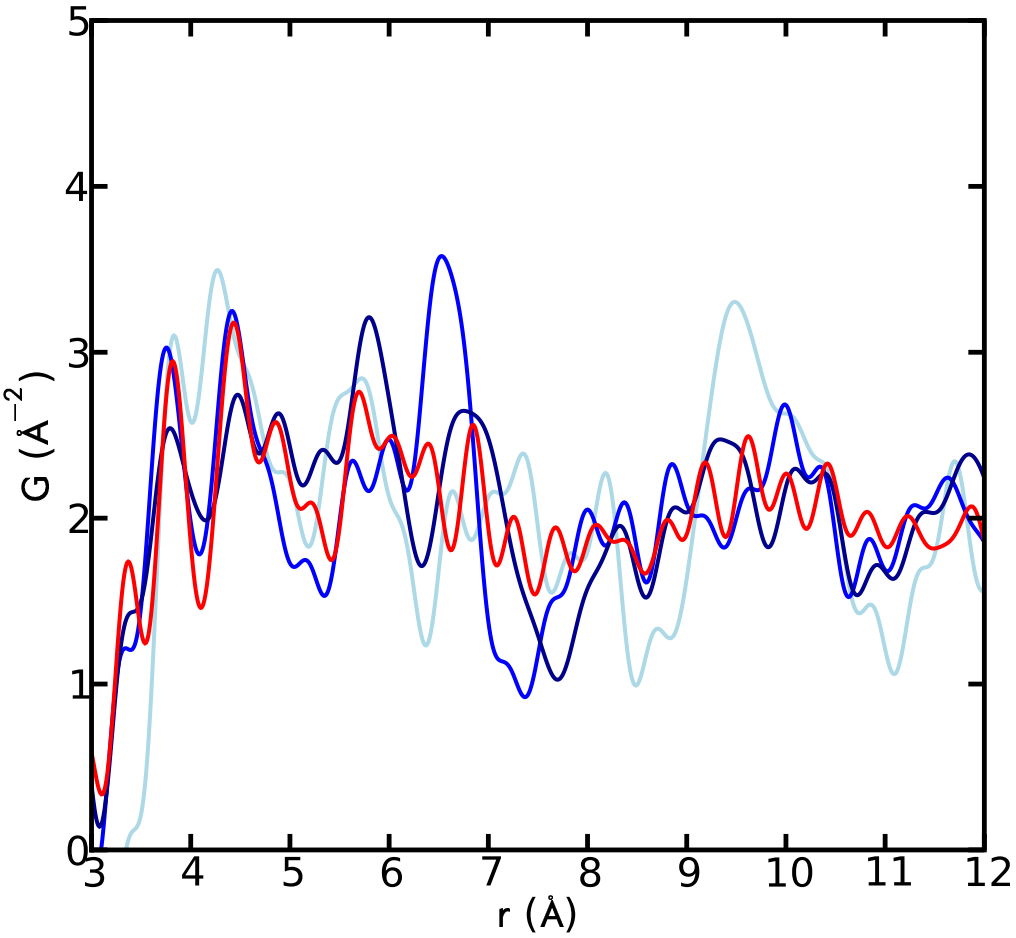

Figure 7.9: PDFs of the amorphous form of compound A (red) and three forms of the crystalline compound, Form α (light blue), Form β (medium blue), Form γ (dark blue) collected at 300 K plotted between 3.0 Å and 12 Å. The crystalline PDFs have been multiplied by 3.0 to fit on the same axes as the amorphous.

Table 7.6: Summary of Pearson correlation coefficients between the PDFs shown in Fig. 7.9. Correlations higher than 0.8 are shown in bold (except when they are trivially unity).

	Cryomilled	Spray Dried
Form α	0.758	0.733
Form β	0.737	0.739
Form γ	**0.835**	**0.820**

form was used for the visual analysis for simplicity sake.

First, we see in Table 7.6 that the correlations between the two amorphous phases and the crystalline phases are roughly the same as each other. This lends further credence to our hypothesis regarding the amorphous compound being the same regardless of how it was produced. We also see that, in the 300 K data, Form γ clearly has the highest Pearson correlation with the amorphous forms with the others significantly lower. Further systematic studies need to be performed to study the relationship between fingerprinting efficacy of amorphous TSPDFs and temperature dependence of the data collection.

We have three results from our analysis of this compound. First, we have shown conclusively that there is no difference in the packing of the two different amorphous forms. This can have big implications if one form is is easier or cheaper to process. Second, we have estimated the range of structural coherence of the amorphous compound to be about 11.5 Å at room temperature by determining the point at which we no longer have signal in the PDF. Third, we believe that the amorphous phase has similar packing to that of Form γ.

7.6 Conclusion

In this chapter we have demonstrated the power of total scattering pair distribution function analysis to study amorphous phase pharmaceutical compounds. Whereas regular crystallography and powder X-ray diffraction do not give us any information at all about the structure and local molecular packing of non-crystalline molecules, the TSPDF gives us relatively sharp peaks with structural information up to about 7 Å and broad features as high as 12-14 Å. There are several results to this investigation: First, we have shown that the amorphous phases that result from cryomilling and spray drying compound A are the same. Second, we estimated the ranges of structural coherence for the amorphous phase of each compound, 13.5 Å for compound B and 11.5 Å for compound A purely by visually inspecting the amorphous TSPDFs. This is different from the domain size that we calculated for CBZ in a previous paper. While the CBZ sample was most likely nanostructured, and thus had some ordering on the scale of a few nanometers, the compounds studied here are truly amorphous. Third, we have concluded that the packing in the amorphous form of compound B is distinct from any of the crystalline forms, and also cannot be explained as a linear combination of those forms. The packing in the amorphous state is distinct in some regard. In compound A the amorphous packing is similar to that of Form γ. Finally, we found that temperature effects play an important role in the TSPDF of the amorphous and crystalline phase compounds and so they need to be studied systematically on a single instrument where all parameters except temperature are kept constant.

Additionally, we have proven the results of Chapter 6 by using data collected on a MALS for pharmaceutical fingerprinting. This is an important result because of the advantages that MALS has over SALS and synchrotron instruments (higher intensity than SALS and more accessibility than the synchrotron).

Our results across Chapters 5, 6, and 7 have shown that TSPDF with fingerprinting can be used to identify local structural information of crystalline, amorphous, and nanocrys-

talline pharmaceutical compounds. However, we have yet to investigate compounds that contain polymorphic or amorphous impurities or are mixtures of several phases. These mixtures can develop, for example, from the application of processing techniques to reduce particle size, such as milling. In Chapter 8, we investigate whether TSPDF and fingerprinting could be used to quantify the mixing fraction of crystalline and amorphous components in a recrystallizing compound.

Chapter 8

Phase quantification at the nanoscale using the total scattering pair distribution function (TSPDF) method: recrystallization of cryomilled sulfamerazine

The work in this chapter was done in collaboration with Matthew Johnson and Simon J.L. Billinge.

In previous chapters, we dealt exclusively with single phase APIs, whether they be crystalline, nanocrystalline, or amorphous. In this chapter, we make the next logical step and investigate a mixture that contains crystalline and amorphous components using TSPDF and fingerprinting.

One of the most challenging problems facing the pharmaceutical industry is to identify and quantify the phase fractions in mixed phase samples that contain crystalline, nanocrys-

talline, and amorphous components. Here we demonstrate an approach that accomplishes this using high energy X-rays coupled with TSPDF analysis by studying samples of sulfamerazine, a sulfonamide antibiotic drug, recrystallizing from a cryomilled-induced amorphous state. Samples milled under different conditions were shown to have significantly different phase compositions. The initial amorphous state was obtained by cryomilling the stable Form 1 polymorph. This was then aged at low temperature to initiate controlled recrystallization. We show that depending on the milling and aging protocol we see a mixture of amorphous material with the metastable Form 2 polymorph. A minority of Form 1 is also observed. We describe the approach that allowed us to quantify the phase fractions despite the majority of the sample lacking crystalline order.

8.1 Introduction

Milling is a common method for reducing the particle size of a drug during manufacture. Particle size reduction is desirable, for example, for a particular drug delivery method such as for inhaled drugs that must access certain parts of the lung [Alagusundaram *et al.*, 2010], to improve processability [John, 2009], or in the case of poorly soluble drugs, to improve bioavailability [Gowthamarajan and Singh, 2010]. Milling of the drug particle can lead to a polymorphic and/or amorphous transformation. The creation of a disordered state as a result of the breakdown of long range order, increases the molecular mobility and the likelihood of spontaneous crystallization [Hancock and Zografi, 1997; Wildfong *et al.*, 2006]. As milling is commonly used to reduce the particle size of manufactured drugs, the output material must be tested for form and, where required, amorphous or polymorphic impurities prior to release. Tests for impurities are conducted using differential scanning calorimetry (DSC) and X-ray powder diffraction (XRPD), although it is difficult to quantify the level of amorphous content using XRPD.

It is typical to call any material that produces a broad halo in an XRPD pattern 'X-ray

amorphous', but this may not completely be the case. The material may also contain some order in the form of nano-particles [Billinge *et al.*, 2010] or a crystalline component. Techniques to probe the structure of amorphous compounds or to identify the amorphous phase fraction are limited. One technique that has been shown to probe amorphous structure is atomic pair distribution function (PDF) analysis. This technique, which has typically been applied to inorganic materials [Egami and Billinge, 2003; Billinge and Kanatzidis, 2004; Billinge, 2008b], has been suggested as an approach for fingerprinting amorphous APIs [Bates *et al.*, 2007] and investigating the local packing of pharmaceuticals in the disordered solid state [Bates *et al.*, 2007]. We now know that the results of these early studies are in question because the d-spacing range (Q range, where Q is the magnitude of the scattering vector [Egami and Billinge, 2003]) over which the data were measured was not sufficient [Dykhne *et al.*, 2011], but PDFs obtained from data of sufficient quality and Q-range are often rich in information and can reliably be used for fingerprinting [Dykhne *et al.*, 2011] and even studying the structure and size of nanocrystalline forms of API's [Billinge *et al.*, 2010]. To differentiate the earlier low-Q PDF measurements from the later high-Q measurements, we refer to the latter as total scattering PDFs (TSPDF). TSPDFs provide local information on a compound in the angstrom to nanometer scale, potentially unlocking the hidden world below the XRPD amorphous halo.

Here we apply the TSPDF technique for the first time to the phase quantification of amorphous and polymorphic forms of pharmaceutical compounds. Sulfamerazine, a sulfonamide drug widely used with other antibiotics [Maren, 1976], was chosen as the test compound for this study as it was known to convert upon hard milling [Zhang *et al.*, 2002] from Form 1 to Form 2 in a short time scale. Sulfamerazine has three known polymorphic forms: Form 1 [Caira and Mohamed, 1992], Form 2 [Acharya and Kuchela, 1982], and Form 3 [Hossain, 2006]. Using TSPDF we demonstrate that it is possible to extract, quantitatively, the phase composition of the samples as they recrystallize from the cryomilled form. This was made more difficult as we didn't have available all the crystalline forms or a purely

amorphous sample as controls. Instead, we use a sample that has mostly recrystallized into Form 2 as a substitute for Form 2 and a TSPDF calculated from a single molecule of the known crystal structure as a substitute for the amorphous. This demonstration increases the potential applications of TSPDF analysis for studying amorphous and mixed phase pharmaceutical compounds.

8.2 Experimental Methods

8.2.1 Sample Preparation and characterization

Two batches of sulfamerazine Form 1 (Sigma Aldrich 58876-50G) were used as input material for a Retsch cryomill. The two production batches are different as the weight of sample used in the cryomilling was varied to determine potential process yield.

For the first batch, 600 mg of sulfamerazine was cryomilled in a 25 ml jar, milling for 9 cycles of 3 min followed by 3 min cool-down (54 min total). A sample was taken for XRPD using a CALS on a silicon wafer with the diffractometer in reflection geometry. Data collection was over 10 minutes. The results showed that this material was x-ray-amorphous after cryomilling as seen by the green curve in Figure 8.1**(b)**.

For the second batch, 1500 mg of sulfamerazine was cryomilled in a 50 ml jar, milling for 9 cycles of 3 minutes, followed by 3 min cool-down. The jar was allowed to warm to room temperature before opening (about 2 hours) and then tested by XRPD, similar to the previous batch. The diffraction pattern indicated that the sample was still crystalline, so the cryomilling procedure was repeated for a second time (108 min total). After this second attempt, the XRPD results showed the material to be partially crystalline, but the diffraction trace did not match the input reference pattern indicating a polymorphic transformation had occurred during the milling (green curve, Figure 8.1**(c)**).

Both batches were transferred from the jars to glass bottles and stored for about one month in a freezer and were again characterized using XRPD before being brought to the

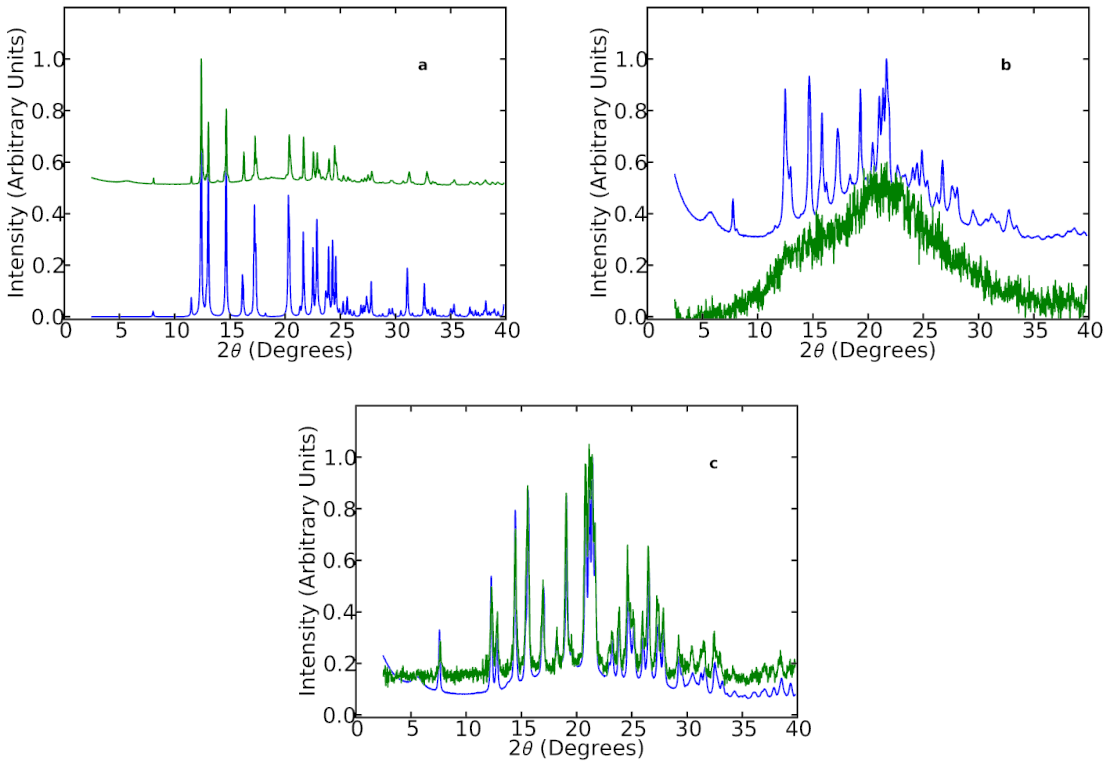

Figure 8.1: **(a)** Experimental XRPD pattern of F1 (green) plotted with a Form 1 pattern simulated from the structure model obtained from the CSD (blue). **(b)** XRPD pattern of C1 collected directly after cryomilling (green) and 1 month later (blue). **(c)** XRPD pattern of C2 collected directly after cryomilling (green) and 1 month later (blue).

synchrotron (blue curves in Figure 8.1**(b)** and **(c)**). A pure Form 1 sample was measured as a control (Figure 8.1**(a)**).

Henceforth, we will refer to data collected from the reference Form 1 sulfamerazine standard as 'F1', the first cryomilled (small) batch as 'C1', and the second cryomilled (large) batch as 'C2'.

8.2.2 Synchrotron Experiment

Powder samples were packed into Kapton® capillaries with 1 mm diameter and brought to the National Synchrotron Light Source (NSLS) at Brookhaven National Laboratory (BNL) in dry ice and kept in dry ice in the laboratory until they were put into the X-ray beam. On the diffractometer the capillary was immediately cooled to 100 K in a flowing stream of cold N_2 gas using an Oxford cryostream® cooler.

Total scattering data were collected at beamline X-7B at the NSLS using the rapid acquisition PDF method [Chupas *et al.*, 2003]. A Perkin-Elmar 2D image plate detector was placed perpendicular to the X-ray beam ($\lambda = 0.3184$ Å) 114.9 mm behind the sample. A spinner rotated the capillary holding the sample at 3 rpm. Data were collected for 30 s at 100 K and this was repeated between 60 and 90 times per sample, measurements were all integrated and normalized with respect to total collection time using the software Fit2D [Hammersley, 1998].

8.2.3 TSPDF Generation

Two programs were used to calculate PDFs. The first, PDFGetX2 [Qiu *et al.*, 2004], was used to calculate TSPDFs from experimental data. This program applies corrections to the data and Fourier transforms it from Q-space to real space. The second, PDFgui [Farrow *et al.*, 2007], was used to calculate PDFs from crystal structure files that are available in databases such as the Cambridge Structural Database (CSD) [Allen, 2002]. Occasionally, these crystal structures do not have information on the isotropic thermal parameters, U_{iso},

of a compound, which are necessary to calculate a PDF. In these cases a value of $U_{iso} = 0.01$ Å^2 was used. This value is appropriate for atomic vibrations in most materials at moderate temperatures comparable to room temperature.

8.3 Results

8.3.1 XRPD analysis

First, we visually compare the XRPD data in Figure 8.1. As expected, from panel **(a)** we see that the measured F1 diffraction pattern clearly corresponds very well with that calculated from the Form 1 structure model. Next, we consider the CuK$_\alpha$ diffraction patterns of C1 and C2 samples before and after aging for 1 month. Panel **(b)** shows us that C1 was X-ray amorphous right after the cryomilling, but transformed into a more crystalline phase after a month. However, this phase is not fully crystalline as evidenced by the broad halo in the middle of the plot. Finally, from panel **(c)**, we see that C2 was mostly crystalline after cryomilling and remained in the same phase over the course of a month, although it does appear to have a small amorphous halo.

We can use XRPD to fingerprint the crystalline component of the aged C1 and C2 samples. This comparison is shown in Figure 8.2.

Careful examination of Figure 8.2**(e)**, compared to panels **(d)** and **(f)**, clearly identifies C2 as being in Form 2. However, as evidenced by Figure 8.2**(a)** and 8.2**(b)**, we are unable to determine which of the crystalline forms C1 corresponds to, as the XRPD pattern contains features of both Form 1 and Form 2. There is no apparent intensity at the positions of strong peaks in the Form 3 diffraction pattern (8.2**(c)**) ruling this out as a constituent.

To supplement the visual comparison in Figure 8.2, we carry out a quantitative analysis of the data using Pearson correlation analysis [Dykhne *et al.*, 2011]. Since the XRPD data are in Q-space, not real space, the Pearson correlation is calculated over the entire range of data. The results are in Table 8.1.

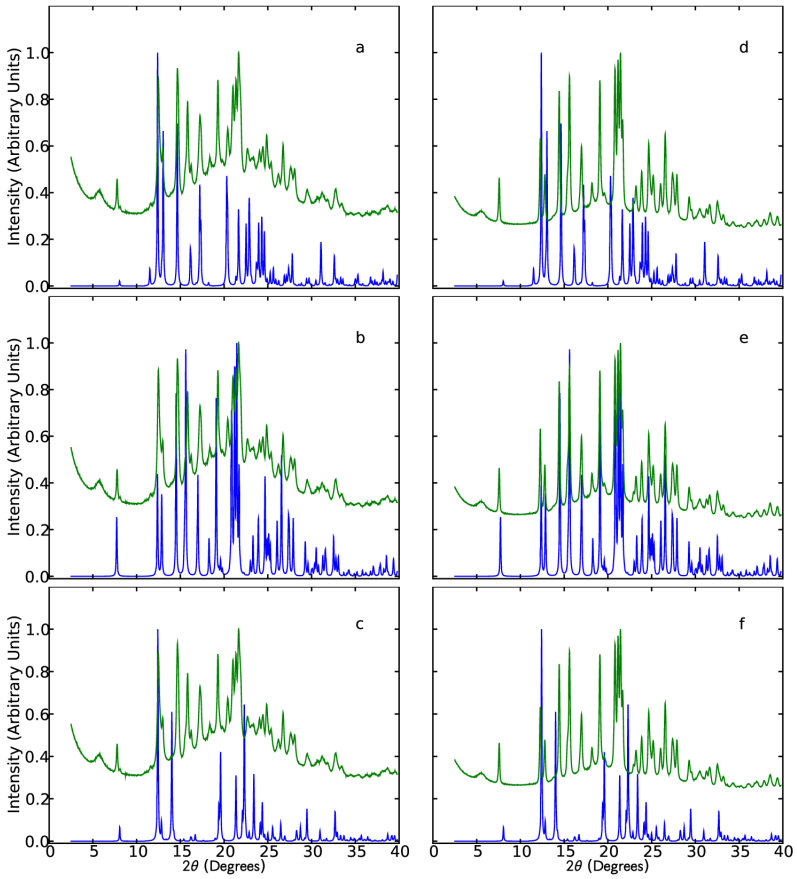

Figure 8.2: **(a)-(c)** XRPD data of C1 (green) plotted with calculated XRPD patterns of Form 1, Form 2, and Form 3 in order from top to bottom (blue). **(d)-(f)** XRPD data of C2 (green) plotted with calculated XRPD patterns of Form 1, Form 2, and Form 3 in order from top to bottom (blue). Panel **(e)** identifies C2 as having mostly the Form 2 structure, while panels **(a)** and **(b)** show C1 contains features of both Form 1 and Form 2.

Table 8.1: Pearson correlation coefficients between C1 and C2 and the crystalline Forms 1, 2, and 3. Coefficients higher than 0.8 are shown in bold (except when they are trivially unity).

	C1	C2	Form 1	Form 2	Form 3
C1	1	0.626	0.439	0.459	0.257
C2		1	0.095	**0.843**	0.083
Form 1			1	0.145	0.338
Form 2				1	0.118
Form 3					1

The Pearson correlation coefficients in Table 8.1 corroborate our assertion that C2 is strongly correlated with Form 2, but not correlated to Form 1 or Form 3. On the other hand, C1 does not correlate strongly with any of the crystalline forms. The correlation with Forms 1 and 2 are roughly similar, while the correlation with Form 3 is lower.

We try to gain more insight into the C1 structure using TSPDF analysis.

8.3.2 Experimental TSPDF Analysis

Figure 8.3 shows TSPDFs of F1, C1, and C2 plotted with an offset and scaled to highlight the high-r region. It shows that C1 and C2 clearly are a different phase than F1, even though they were created by cryomilling batches of F1. Furthermore, the TSPDFs of the cryomilled samples show us that they look quite similar to each another, with the C1 phase having some peaks in the same positions as those of C2, but with lower amplitudes at higher r. Pearson correlation coefficients calculated in the range 3-20 Å can be used to quantify the relationship between the TSPDFs. The Pearson correlation coefficients in Table 8.2 verify the visual observations that there is a higher similarity between the two cryomilled samples than with the precursor phase F1.

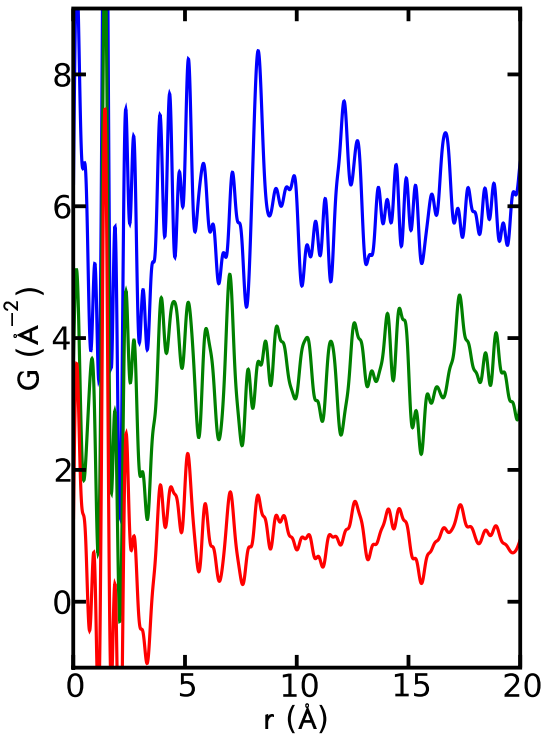

Figure 8.3: TSPDFs of the three samples of sulfamerazine. From top to bottom: crystalline F1, cryomilled C2, cryomilled C1. The TSPDFs of C1 and C2 look different from F1 and similar to one another.

Table 8.2: Pearson correlation coefficients between the TSPDFs of F1, C1, and C2 in the range 3-20 Å. Coefficients higher than 0.8 are shown in bold (except when they are trivially unity).

	F1	C1	C2
F1	1	0.668	0.340
C1		1	**0.892**
C2			1

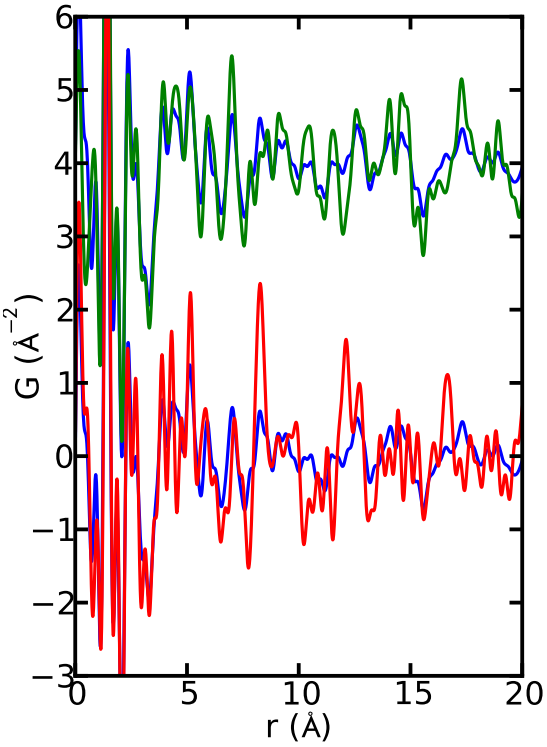

Figure 8.4: Top: C1 in blue and C2 in green plotted on top of one another (Pearson correlation is **0.892**). Bottom: C1 in blue and F1 in red plotted on top of one another (Pearson correlation is **0.668**). The axis have been scaled to focus on the high-r region.

8.3.3 Mixing Fraction Measurement

From the top of Figure 8.4, where we visually compare C1 and C2 by plotting them on top of one another, we see that the two TSPDFs share many similar features. However, the C2 peaks have a higher amplitude than those of C1. We also see that some features in C1, such as the peak at around 8.2 Å, are not reproduced in C2. From the bottom part of Figure 8.4 we see that there is less agreement between C1 and F1 than there was with C2. However, some features do match up. For instance, the aforementioned C1 sharp peak at 8.2 Å corresponds well to an F1 peak in the same position.

The information in Figures 8.1, 8.2, and 8.4, suggests that C1 is a mixture of Form 1, Form 2, and amorphous components. To test this theory, in Figure 8.5 panel **(a)** we plot the TSPDF curves of C1 (blue) and C2 (green) and a difference curve (red) with C2 scaled such that the values at high-r match up well. We notice that the difference curve looks a lot like F1. In panel **(b)** we plot the difference curve from panel **(a)** (red) along with F1 (indigo) scaled such that its high-r values match that of the difference curve between C1 and C2. Interestingly, we see fantastic agreement with F1. We subtract F1 and plot the difference curve in panel **(c)** (orange). It appears that the intramolecular carbon-carbon first and next nearest neighbor peaks are apparent in this curve. To test this idea, we plot the PDF of a single sulfamerazine molecule on top of it (black). Apart from the noise inherent in taking the second differential, this curve agrees rather well, especially in the low-r region. If the amorphous state is lacking in significant short range order, the PDF of the amorphous form will resemble the PDF of the single molecule, as we see here.

The analysis in Figure 8.5 gives us an idea for quantifying phase fractions including an amorphous phase fraction in C1. We have experimental Form 1 data, but we do not have pure experimental Form 2 data. However, from Figure 8.2**(e)** and Table 8.1 we know that C2 has mostly converted to Form 2 so we use it as a substitute for Form 2, although we keep in mind that there appears to be a little bit of an amorphous halo in C2 as well. We don't have pure amorphous data, but, as we have demonstrated in Figure 8.5, we can model the local structure of an amorphous phase by calculating a PDF of a single molecule of Form 1 sulfamerazine from a known structure (CSD refcode: SLFNMA02) using SrFit, a home written Debye PDF calculator. The idea here being that the PDF of a single molecule contains strong peaks from nearby intra-molecular atoms similar to the behavior exhibited in Figure 8.5**(c)**. We choose Form 1 as the base molecule because this is the original form of the compound. We will refer to this calculated PDF as α.

When a PDF is calculated by software, whether it be from experimental data, as is the case with C1, C2, and F1, or from a model, as we do with α, it is scaled by the software.

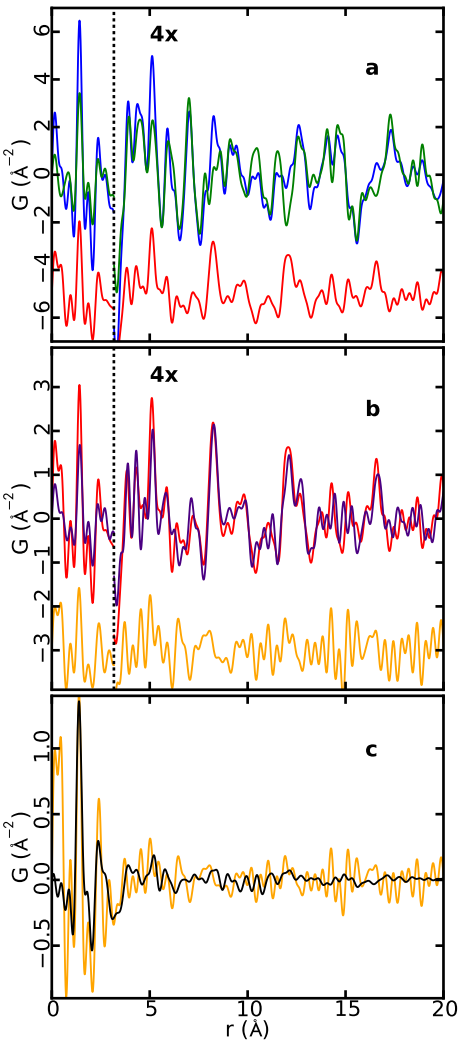

Figure 8.5: **(a)** C1 (blue) with 33% of C2 (green), the difference is in red. **(b)** The difference curve from panel **(a)** (red) with 12% of F1 (indigo), the difference is in orange. **(c)** The difference curve from panel **(b)** (orange) with 55% of the F1 model (black). In panels **(a)** and **(b)**, the TSPDFs to the right of the vertical line at $r = 3.17$ Å have been multiplied by 4.0 to highlight the data in the high-r region.

This scaling factor is dependent on the algorithm used in the calculation, so for instance even two PDFs calculated from the same experimental data using different corrections can be on a different scale. This is not a problem for regular fingerprinting, as we did above, where the curves are scaled for convenient comparison, for example, so that the tallest peaks match. However, this becomes a problem when quantifying mixing fractions because the scale of each PDF in the mixture affects its contribution to the whole.

We must place the TSPDFs of the components of the mixture (C2, F1, and α) onto an absolute scale. This was done such that the integrated areas of the first carbon-carbon peak in the radial distribution functions (RDFs) are the same and the same as that for C1. The RDF gives the number of atoms in an annulus of thickness dr at distance r from another atom [Egami and Billinge, 2003], which in our case means that the integrated area of the first peak is proportional to the coordination number of carbon, which should be the same in all our RDFs.

The RDF, $R(r)$ is calculated from the PDF by

$$R(r) = rG(r) + 4\pi r^2 \rho_0,$$

where ρ_0 is the average number density.

After we have found the appropriate scaling factors from the RDF, we apply an optimization routine that varies the mixing parameters of the three components:

$$xF1 + yC2 + z\alpha$$

$$x + y + z = 1$$

where x, y, and z are the mixing fractions of F1, C2, and α respectively. We set the Pearson correlation between the mixture and C1 as our target function to maximize.

The mixing fractions that produced the highest Pearson correlation (**0.9765**) are: 12% F1, 33% C2, and 55% α. This means that the normalized TSPDFs of the three components combined in those fractions add up to the TSPDF of C1.

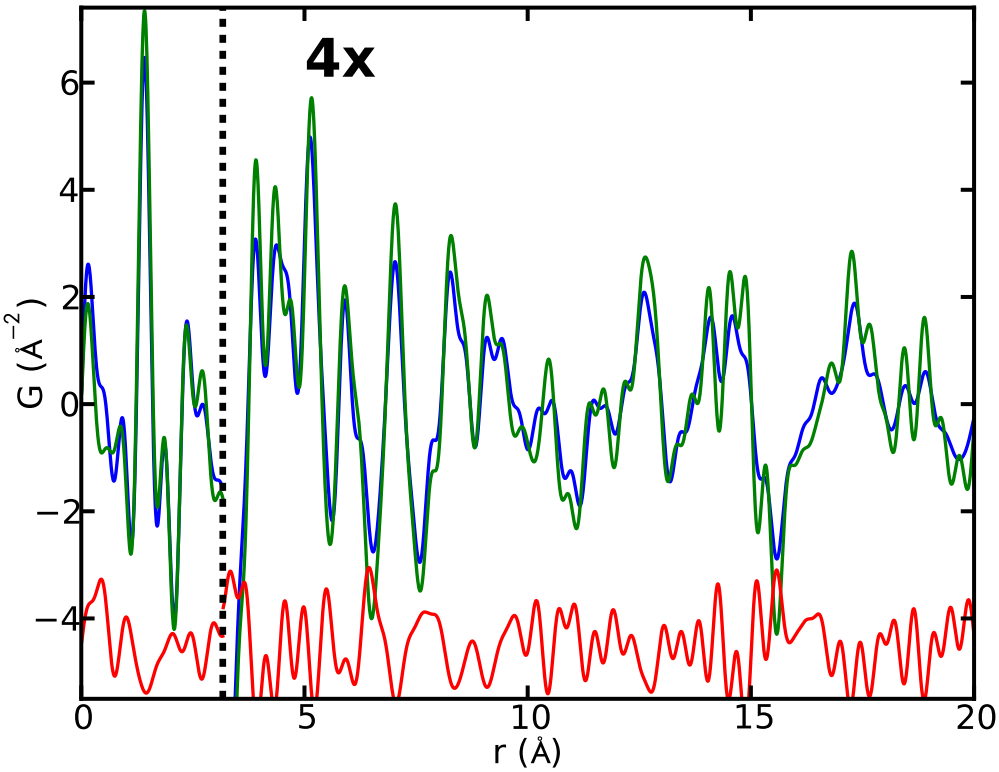

Figure 8.6: Fingerprinting of C1 (blue) with the calculated TSPDF of the mixture containing 33% C2, 12% F1, and 55% α (green) (Pearson correlation is **0.9765**). The TSPDFs to the right of the vertical line at $r = 3.17$ Å have been multiplied by 4.0 to highlight the data in the high-r region.

Figure 8.6 contains C1 plotted with the total TSPDF of the mixture. We see that almost all features of C1 are reproduced beautifully in the mixture.

8.4 Discussion

While both C1 and C2 predominantly crystallized as Form 2, C2 is more completely recrystallized than C1, which has a large amorphous component. We hypothesize that this is due to the effects of temperature. C2 was allowed to warm to room temperature after

cryomilling, while C1 was not. Therefore, it rapidly recrystallized at a temperature close to room temperature. On the other hand, low temperature seems to inhibit recrystallization because C1 was kept at low temperature for a month but still maintained an amorphous component. The recrystallizing state in C1 predominantly consists of a non-crystalline phase, but with a significant proportion of Form 1 and Form 2. The growth of Form 2 from the amorphous state has been seen before [Zhang *et al.*, 2002] though is not necessarily expected since the stable form, and the form that was cryomilled to create the precursor, is Form 1.

One of the major advantages of using the TSPDF is that with it we can quantify the amorphous phase component in a mixture of amorphous and crystalline material because both the crystalline and amorphous components contribute intensity to the carbon-carbon nearest neighbor peaks. This means that after we subtract the crystalline components, which we do by scaling them at high-r, as demonstrated in Figure 8.5, any intensity left in the peaks is due to contributions from the amorphous.

One may, at first, find it curious that the TSPDFs of C1 and C2 match up so well, as we see in Figure 8.4, even though C1 is only a third Form 2, while C2 is mostly Form 2. This occurs because the amorphous content is concentrated in the low-r, intramolecular, peaks. This is incidentally the same reason that allows us to model the amorphous contribution to the mixture using a calculated PDF of a single molecule. When we get out to high-r, intermolecular, peaks then the Form 2 content of C1 is nearly 3 times that of Form 1, and so we naturally observe good agreement with C2. Similarly, despite the C1 sample being more than half amorphous, the TSPDF technique is clearly able to quantify minority crystalline components at the 10% level.

These results help cement TSPDF as an accessible technique that can be used for effective fingerprinting of pharmaceutical APIs.

8.5 Conclusion

Using high energy X-rays coupled with TSPDF analysis, we quantified the phase fractions of crystalline and amorphous components in a mixed phase sample of sulfamerazine recrystallizing from a cryomilled-induced amorphous state. Even though the mixture consisted of a majority of an amorphous component (55%), the TSPDF was sensitive to the phase fractions of two crystalline components as well. These results demonstrate that TSPDF is an important tool in tackling some of the major challenges facing the pharmaceutical industry.

In the next chapter, we again demonstrate the quantification of mixing fractions in an API, only this time we know the expected mixing fractions ahead of time, so that we can compare them to those calculated with TSPDFs.

Chapter 9

Fingerprinting Analgesics with TSPDF

In Chapter 8 we found the mixing fraction of polymorphic and amorphous components in a recrystallizing compound using TSPDF. However, no other techniques were used to verify the mixing fractions, and they were not known *a priori*, so an argument might be made that we can't be sure of the results of a TSPDF study.

Additionally, in previous chapters we have measured samples of powders that contain only the API. However, as we discussed in Chapter 2, actual pharmaceutical tablets for oral ingestion are usually a mixture of two components - the API and the excipient. For fingerprinting with TSPDF to be a useful technique for quality control and litigation purposes, the API needs to be properly identified when it is in a mixed phase with excipient.

To address the above questions, in this chapter we do two simple investigation on the over-the-counter (OTC) analgesics (pain killers), Tylenol® and Sudafed®. First, we fingerprint the TSPDF of an *a priori* known mixture of two components with the TSPDFs of the constituent components and second we fingerprint the TSPDF of Tylenol against three polymorphs of its API.

9.1 Sample Preparation and Measurement

The main API in Tylenol is acetaminophen [de Villiers *et al.*, 2998] (it is also known as paracetamol), a non-steroidal anti-inflammatory drug, and the main API in Sudafed is phenylephrine [Troup and Mitchner, 1964], a decongestant. Tablets of both compounds were purchased at a local drug store.

Three samples were prepared. A sample of pure Tylenol was prepared by grinding up a tablet using a mortar and pestle into a fine powder and then filling the powder into a Kapton® capillary 1 mm in diameter. A pure Sudafed sample was prepared the same way. A mixture was prepared by grinding together equal parts Tylenol and Sudafed (roughly 300 mg each) with a mortar and pestle and then filling a capillary with the mixed powder.

Samples were measured at beamline X7B at the NSLS using X-rays with $\lambda = 0.3184$ Å. These data were collected at the same time as the sulfamerazine data, so the full experimental details are in Chapter 8 and will not be reiterated here. TSPDFs were made with PDFGetX3 using the methods outlined in Chapters 1-4. They were compared to PDFs made with PDFGetX2 using the methods in Chapter 4 to verify proper data processing.

Samples of the three polymorphs of acetaminophen [Perrin *et al.*, 2009] were procured by our collaborators at PANalytical and measured on their silver anode diffractometer with $\lambda = 0.56$ Å. TSPDFs of all the data were calculated using PDFGetX3 and with $Q_{max} = 16$ Å^{-1}.

9.2 Results

9.2.1 Mixture Comparison

Figure 9.1 shows the TSPDFs of the three samples, with the mixture of Tylenol and Sudafed in blue, pure Tylenol in green, and pure Sudafed in red. We see that the mixture PDF shares features with both components. Table 9.1 shows the Pearson correlations between the three

Table 9.1: Summary of Pearson correlation coefficients between the PDFs shown in Fig. 9.1. Correlations higher than 0.8 are shown in bold (except when they are trivially unity).

	Mixture	Tylenol	Sudafed
Mixture	1	**0.882**	0.780
Tylenol		1	0.414
Sudafed			1

TSPDFs. As expected, we see low correlations between Tylenol and Sudafed and pretty high correlations between each component and the mixture. The correlation between Tylenol and mixture is above our threshold for significance, and the correlation between Sudafed and the mixture is just below.

We use a similar technique to optimize the mixing fraction of Tylenol and Sudafed in the mixture as we used in Chapter 8 to optimize the polymorphic and amorphous mixing fractions. The main difference being that we now have two components rather than three. Briefly, we first put all the TSPDFs onto the same absolute scale such that the integrated areas of the first carbon-carbon peaks in the radial distribution functions (RDFs, Equation (1.22)) are the same, then we then apply an optimization routine that varies the mixing parameters of the two components:

$$xT + yS \tag{9.1}$$

$$x + y = 1 \tag{9.2}$$

where T and S are the TSPDFs and x and y are the mixing fractions of Tylenol and Sudafed respectively. We set the Pearson correlation between the Tylenol-Sudafed TSPDF and the mixture as our target function to maximize.

As we see in Figure 9.2, the highest Pearson correlation (**0.993**) occurs with the mixing fraction with 49% Tylenol and 51% Sudafed. We see that the fit is nearly perfect. Even

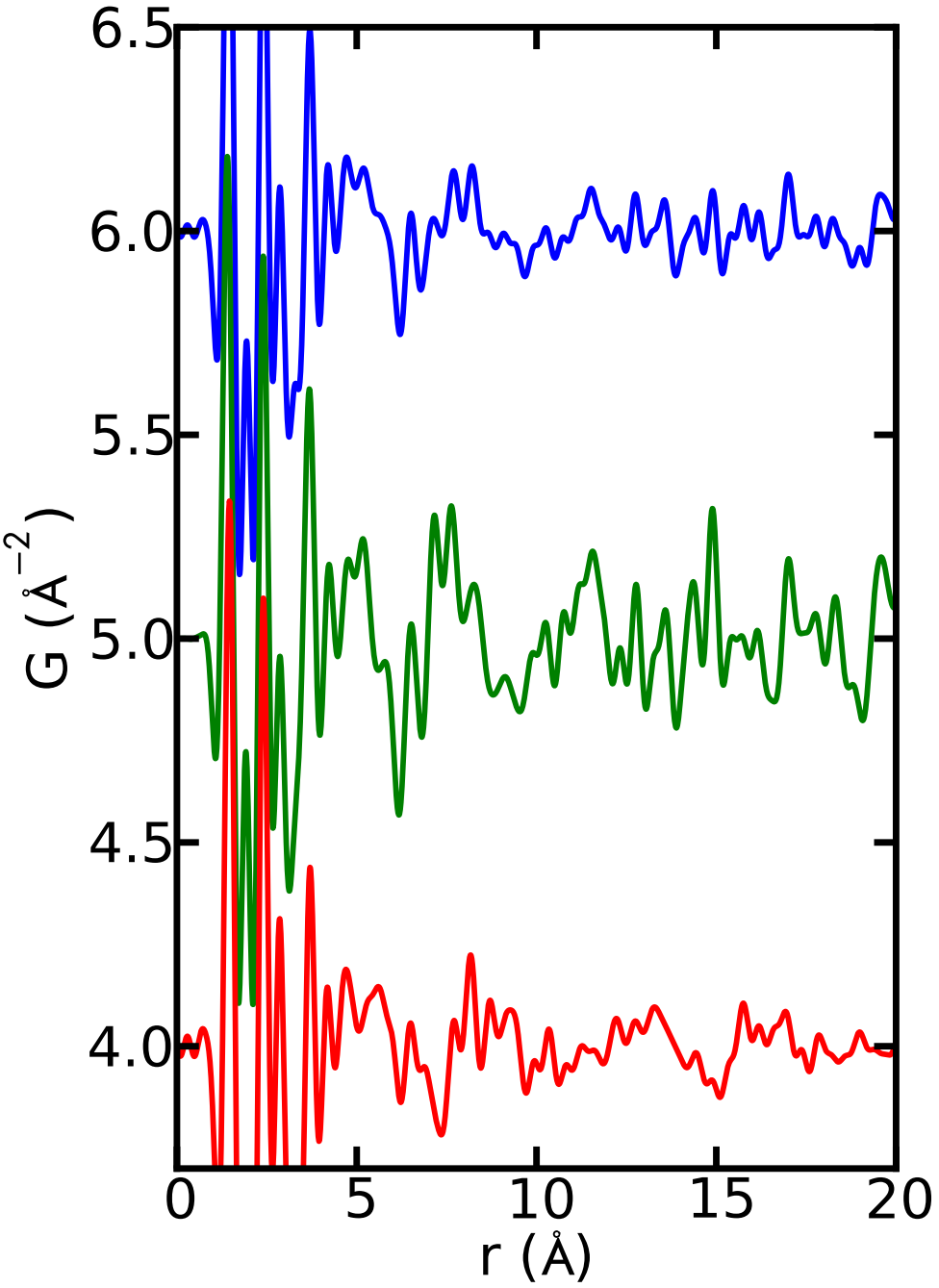

Figure 9.1: TSPDFs of the Tylenol-Sudafed mixture (blue), Tylenol (green), and Sudafed (red). The y-axis has been scaled to highlight the high-r region where the PDFs are different.

magnified 4x, the difference curve is very tiny. The mixture of 49-51 Tylenol-Sudafed is almost exactly the 50-50 we were expecting.

Overall, we have shown with this very simple example that TSPDF coupled with fingerprinting using Pearson correlations is a powerful way to identify the proportions of components in a mixture of APIs. The results here act as a 'sanity check' of sorts for the results in Chapter 8, showing that the mixing fractions that we quantify with TSPDF are correct.

9.2.2 Fingerprinting APIs and Excipient

Now, we turn our attention to the issue of identifying an API that is mixed with excipient using fingerprinting. This has implications in quality control (i.e. quick identification of a pill on a production line) and for litigation purposes (i.e. a company is sued for using a patent protected API with a different excipient composition).

To test the TSPDF's ability to do this, we fingerprint Tylenol against the three polymorphs of its constituent API, acetaminophen. By weight a Tylenol pill is roughly 80% API (500 mg of API per pill, total weight of the pill is roughly 625 mg), so we would expect it to fingerprint well. In Figure 9.3, we see the TSPDF of Tylenol (blue) plotted with the three crystalline forms of acetaminophen. We instantly see that Tylenol looks very much like Form I, not like the other two forms. The TSPDFs of the Form II and Form III APIs are, in general, a little bit noisy, but these factors are outside of our control (Form III in particular is quite hard to synthesize [Burley *et al.*, 2008], so the sample was very small). The fact that there is enough signal for us to fingerprint shows the robustness of the technique. The Pearson correlations are in Table 9.2 where we clearly see that the correlations between Tylenol and Form I are much higher than the correlations between any other forms.

We have shown that TSPDF can be used to fingerprint the API in a pill containing an API and excipient mixture, at least when the API is a large fraction of the mixture. Future investigations where the API accounts for a smaller weight fraction of a pill will need to

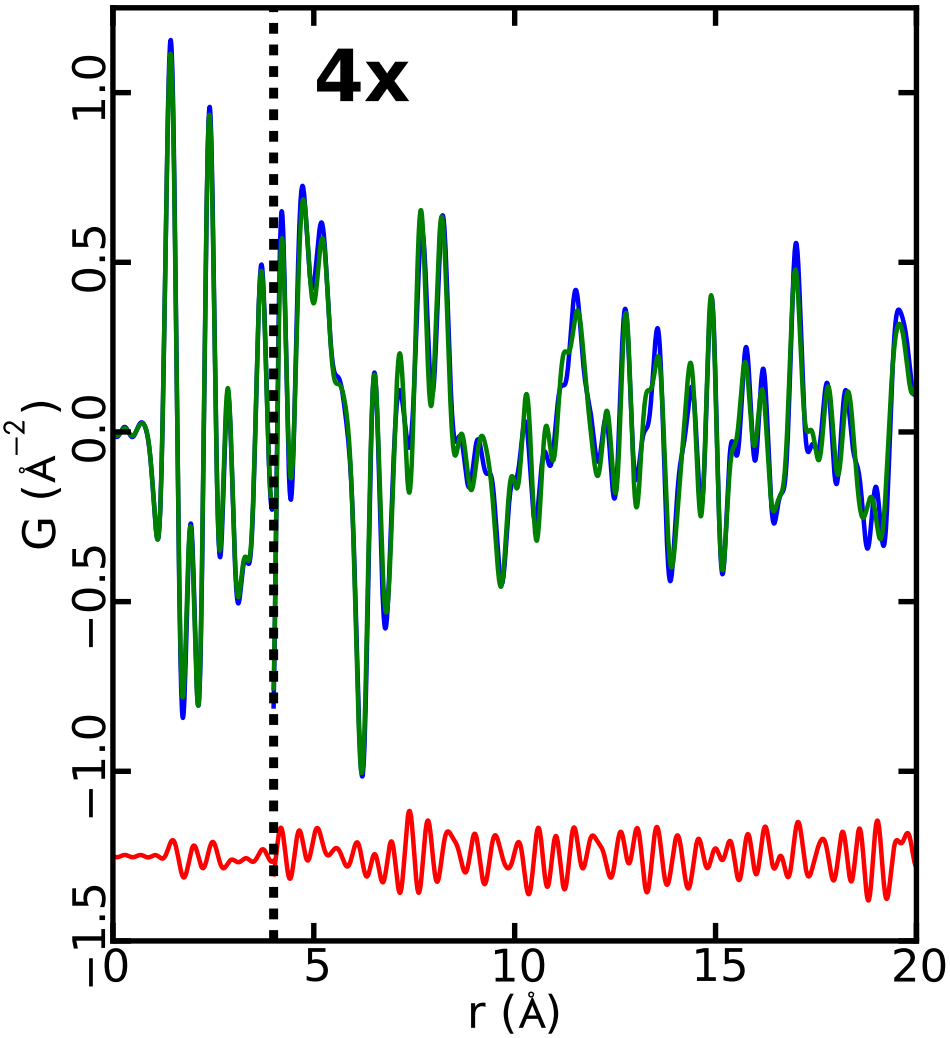

Figure 9.2: TSPDFs of the Tylenol-Sudafed mixture (blue) and the calculated mixture of 49% Tylenol and 51% Sudafed (green) with an offset (red). The Pearson correlation is **0.993**. The TSPDFs have been multiplied by a factor of 4.0 for r values greater than 4 Å(dashed line).

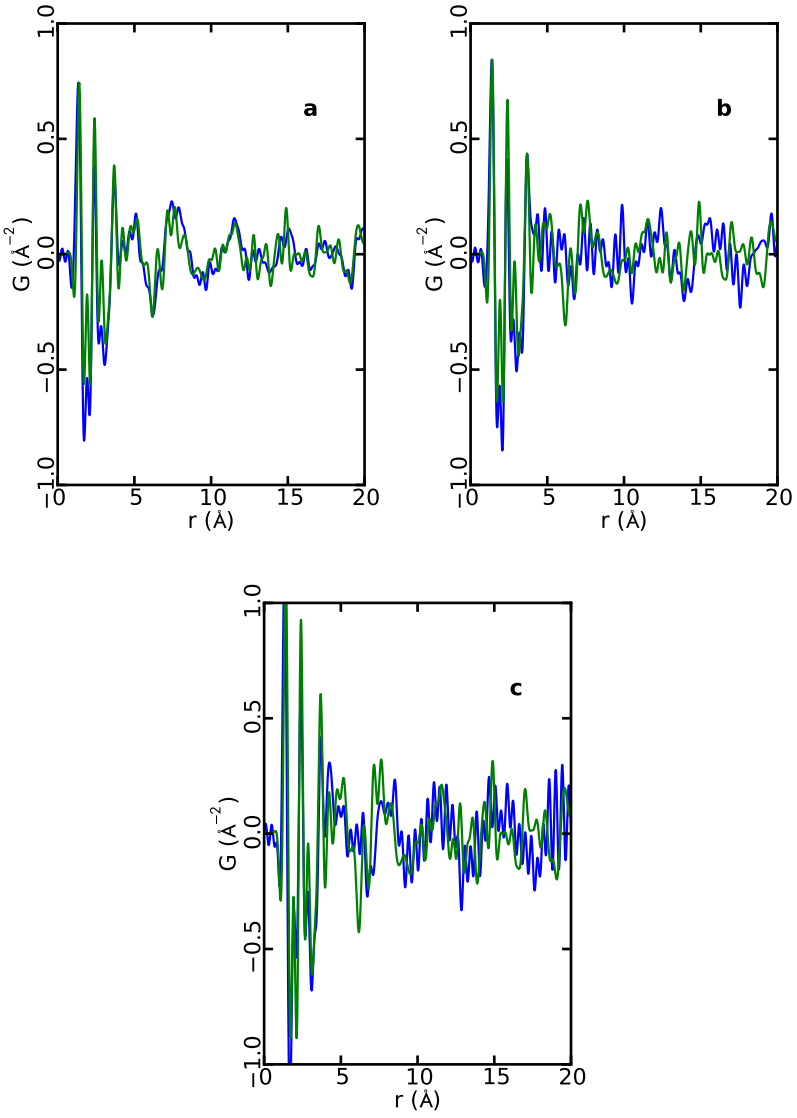

Figure 9.3: TSPDFs of Tylenol (blue) plotted on top of TSPDFs of **(a)** Form I, **(b)** Form II, and **(c)** Form III (all in green), $Q_{max} = 16$ Å^{-1} in all cases. Tylenol clearly looks very much like Form I, and not like the other forms.

Table 9.2: Summary of Pearson correlation coefficients between the PDFs shown in Fig. 9.3. Correlations higher than 0.8 are shown in bold (except when they are trivially unity).

	Tylenol	Form I	Form II	Form III
Tylenol	1	**0.856**	0.484	0.517
Form I		1	0.504	0.594
Form II			1	0.630
Form III			1	1

be performed. If the excipient consists of mostly polymers, then, in theory, the polymers should not produce a signal beyond the first few peaks and the TSPDF could still be used to identify the API.

9.3 Conclusion

The two results in this chapter are reasonably basic, but far reaching. First, were able to quantify the mixing fractions of two over-the-counter pain medications. We found taht the mixing fractions calculated from the TSPDFs were almost exactly the expected mixing fractions from sample preparation. This result is complementary to the results in Chapter 8 and acts as a case in point demonstration of the technique. Second, we fingerprinted the API in a pill containing an API-excipient mixture. This result demonstrates that TSPDF could be a useful technique for quality control and litigation purposes.

Chapter 10

Conclusions

A major challenge facing the pharmaceutical industry is that it is becoming more difficult to bring new drugs to market because many candidate compounds exhibit poor aqueous solubility. One of the proposed solutions to increase solubility is to formulate the compounds in a non-crystalline state. However, traditional X-ray powder diffraction (XRPD) techniques that are relied upon by the industry begin to lose their effectiveness when disorder is introduced because diffraction patterns become broad and featureless.

In this thesis, we attempted to show that high energy X-rays coupled with the total scattering pair distribution function (TSPDF) and fingerprinting analysis could be used to identify the local structure of non-crystalline pharmaceutical materials when XRPD methods break down. Our hypothesis was that the larger range of momentum transfer, Q, afforded by high energy X-rays would let us collect diffuse scattering intensities, which contain information about a sample's local structure, and the TSPDF would let us work in real-space where we could view intra- and inter-molecular correlations. As we have seen over the course of this thesis, our hypothesis proved successful and we were able to make important contributions to the field.

In our first study, we successfully used synchrotron X-rays to fingerprint three phases, two crystalline and one nanostructured, of two pharmaceutical compounds, carbamazepine

(CBZ) and indomethacin (IND). We also found that the disordered form of CBZ was actually the nanostructured Form III polymorph with a 4.5 Å diameter. When we first embarked on this project, we were not sure what to expect, as such systematic studies of active pharmaceutical ingredients (APIs) have not previously been attempted using TSPDF. Fortuantely, the investigation was successful and became a proof of concept that showed that useful information that is not available via traditional XRPD could be obtained from the TSPDF. One of the key insights was that fingerprinting could be done via direct comparison, meaning that no structural refinement was necessary to identify the nanostructured phase. This made the technique much more attractive and accessible to the pharmaceutical industry.

The results of the first study were encouraging, but they also introduced several issues that needed to be addressed before the technique could take off among pharmaceutical researchers. The first issue is that we had not defined exactly what we meant by 'high energy' X-rays and what sort of accessible Q range was needed for TSPDF investigations of pharmaceuticals. We showed that synchrotron radiation was energetic enough to get the job done, and we showed that a copper-anode diffractometer not used to its full potential was not, but there is a spectrum of X-ray diffractometers in between. To answer this question, we took the next logical step in developing the methodology by investigating exactly the data requirements for the reliable use of TSPDFs in fingerprinting, since not all researchers, particularly not those in industry, have unlimited and immediate access to a synchrotron. We found that molybdenum- and silver-anode diffractometers can be used to collect reliable total scattering data, however copper-anode diffractometers, even when used to their maximum potential, will not produce reliable results.

The other issue that our initial study raised was that, at the time, there were no automated methods of quickly generating many TSPDFs and comparing them to one another. This was another barrier to the adoption of the technique. Since high-throughput scanning techniques used in drug discovery produce a lot of data, manually making TSPDFs one a time and plotting them two at a time for comparison were not efficient approaches.

In response, we developed PDFGetX3, a new program for processing total scattering data that requires very little user interaction and is capable of automatically batch processing data sets. We also introduced the use of Pearson correlations to quantitatively compare TSPDFs. The Pearson correlations allow the researcher to focus his or her attention on data that display strange behavior (i.e. high correlations when they should be low or vice versa) instead of plotting every data set.

In our initial studies and development of the methodology, we had only analyzed nanostructured compounds. Luckily, though, our initial work attracted the interest of the pharmaceutical industry and we were able to obtain two proprietary compounds, which we called Compound A and Compound B in this thesis, in a truly amorphous form for investigation. We found that, unlike the nanostructured APIs which had attenuated signal out to high r, TSPDFs of truly amorphous compounds give sharp peaks with structural information up to about 7 Å and broad features up to 12-14 Å. However, this signal still provided more information than what could be obtained from traditional crystallographic techniques. Using fingerprinting, we discovered that the amorphous phases that result from cryomilling and spray drying Compound A are the same. This bit of information could have important implications on the compound if one of the amorphization techniques is easier and/or cheaper than the other. We also found that the local packing of the amorphous form of Compound A is similar that of the γ polymorph, while the local packing of the amorphous form of Compound B is distinct. Although the analysis was not quite as straightforward with these materials as it had been with CBZ and IND, our results definitely added valuable insight into the local structure of these compounds and allowed us to continue to develop the technique.

After fingerprinting Compounds A and B, we noticed that we had exclusively concerned ourselves with fingerprinting single phase APIs. However, the pharmaceutical industry was very concerned with being able to quantify mixing fractions of crystalline and non-crystalline components in APIs that had undergone a particle size reduction process such

as cryomilling. Therefore, it seemed logical for us to attempt to fingerprint a sample of sulfamerazine that was recrystallizing from a cryomilling-induced amorphous state. Using TSPDF, fingerprinting, and an in-house optimization code, we found that the recrystallizing sample was a mixture of 33% C2 (which mostly consisted of Form 2), 12% Form 1, and 55% non-crystalline components. This was an incredibly exciting result because it showed that TSPDF can be used to fingerprint complex mixtures consisting of several polymorphs of a compound in addition to the amorphous. Furthermore, the results showed that TSPDF was able to quantify a minority crystalline component at roughly the 12% level in the presence of a majority non-crystalline phase. Finally, we demonstrated that the calculated TSPDF of a single molecule of sulfamerazine could be used to quantify the amorphous mixing fraction. The results of this investigation are probably the most exciting results of the thesis. Our key insight in the study was the realization that the integrated area of the first carbon-carbon peak in the radial distribution function (RDF) of each data set could be used to normalize each TSPDF onto an absolute scale. Without that insight, we would have been unable to calculate the correct mixing fractions.

We followed our study of the quantification of mixing fractions of recrystallizing sulfamerazine by looking at a much simpler mixture of two over-the-counter pain medications as a case in point. We found that the mixing fractions calculated from the TSPDFs coincided almost exactly with the expected mixing fractions from sample preparation. Although there were no non-crystalline components in play this time, the result did act as a 'sanity check' of sorts to prove that our algorithm for calculating mixing fractions was working as expected.

Our final result is that we were able to fingerprint the correct polymorphic form of the API acetaminophen in a Tylenol that contained the API and excipient. This result is interesting for two reasons. First, the fingerprinting worked even though the data were collected on two types of instruments (Tylenol data collected on a synchrotron, acetaminophen data on a silver-anode diffractometer), which shows fingerprinting TSPDFs is insensitive to the instrument used. Second, the result demonstrates that TSPDF fingerprinting might be-

come a useful technique for quality control (for example, to test pills during manufacture) or litigation (for example, to show that a certain compound is infringing on the patent of a different compound).

Overall, all of the studies that we've done have definitely shown that our technique works well and that there is a lot of potential for some very interesting and fruitful collaboration between academia and the pharmaceutical industry in bringing a new class of APIs to the market. We have received a lot of interest from the pharmaceutical industry to continue developing the technique and also to investigate compounds that are of interest to them.

10.1 Future Work

Although we have laid the foundation for fingerprinting using TSPDF, there is still a lot of work that needs to be done to expand the technique. For instance, as we observed in Chapter 7, temperature has an effect on the dynamic disorder of the TSPDFs, which translated to peaks being broader at higher temperatures. However, the exact behavior of the effect of temperature changes on the TSPDFs of crystalline, nanostructured, and amorphous APIs needs to be further examined. One way to do this would be to use temperature-resolved TSPDF studies by collecting RAPDF data from a sample as it was warmed up or cooled down over some pre-defined temperature range. This type of analysis could produce dozens of data sets that need to be analyzed using tools like PDFGetX3 and Pearson correlations to investigate how the local structure evolves with temperature.

Another area where further investigation is needed is in fingerprinting of the API in an API-excipient mixture. We successfully fingerprinted acetaminophen with Tylenol, but Tylenol is roughly 80% acetaminophen by weight. Other APIs need to be investigated where the API to excipient ratio is lower. It might be that below a certain threshold in the API-excipient ratio, the signal of the API will no longer be distinguished from the noise of the excipient. On the other hand, though, many excipients are polymers and might not

produce signal in the TSPDF beyond the first few peaks, so the high-r region could still be used for fingerprinting.

In addition to API-excipient mixes, API-polymer mixes might also prove to be a fertile ground for TSPDF research. Since amorphous phase APIs are metastable, there is always a risk that they will revert to a thermodynamically favorable (and less soluble) crystalline state. Therefore, there is tremendous interest in developing a reliable means of stabilizing non-crystalline APIs. One suggested route to achieving this goal is through formulating the API in a dispersion in a water-soluble polymer. However, it is difficult to understand the structural composition of the dispersions [Craig, 2002] and some characterization techniques have failed. Similar techniques to what we have employed in Chapters 8 and 9 may be used in this investigation.

Finally, as we have mentioned several times in this thesis, there currently do not exist TSPDF tools designed specifically to model the structures of disordered APIs. For TSPDF to grow from a fingerprinting technique to an inexorable tool of the pharmaceutical industry, these modeling paradigms need to be developed. One idea for doing this is to allow users to label certain functional groups of atoms, such as a carbon ring, as 'rigid', so as to constrain certain parameters (i.e. atom positions) of the entire group, and to label other atoms, such as a chain of carbons, as 'floppy', meaning that they are able to move with fewer constraints. We hope that further developments in the modeling methodology will bring TSPDF to the forefront similarly to how XRPD evolved from a fingerprinting technique to a robust tool of structural investigation.

Part II

Appendices

Appendix A

Safety Guide for Handling Pharmaceuticals

This is a guide I wrote for handling the pharmaceutical samples that we have worked with throughout this thesis. Having it in the thesis is useful for completeness and also as a reference for future students who wish to continue this work.

A.1 Introduction

Now that the Billinge Group has started to collaborate with pharmaceutical researchers and industry, safety precautions have become a top priority. This document is a guide, a summary, and a task list for how to handle pharmaceutical samples from the moment that they arrive until they are disposed of.

A.2 Preparation

When you find out that a pharmaceutical sample is on its way from a collaborator, the first thing to do is check the MSDS form. You can either ask the collaborator to send you the MSDS form or check out a chemical website such as Sigma Aldrich's MSDS database

(http://www.sigmaaldrich.com/safety-center.html) for information. In particular, look at sections 2 and 8. Section 2 identifies the potential hazards of the material (e.g. inhaled, swallowed, exposed to skin, etc.) while section 8 lists how to protect oneself from hazards. MSDS forms will give you an indication of how to safely approach samples. Also, make sure to ask collaborators what additional precautions are needed.

Once the hazards are known, the next step is to check to make sure that there is ample space in the fridge for the sample(s). Storing pharmaceutical compounds in the fridge is particularly important for amorphous phase materials because it might slow down their transformation into a crystalline phase. At the time of writing, we do not yet have a fridge, so we share a fridge with the Herman Group in Mudd Room 1141. We intend to procure a fridge soon.

A.3 When Samples First Arrive

Note: Always wear gloves when handling samples, no exception.

Samples will most likely arrive in a box with the MSDS forms attached. Check the MSDS forms to ensure that you received the right materials. If you have received the wrong materials, contact the collaborator who sent them to figure out what's wrong. Collaborators might decide to send the MSDS forms by email rather than packing them with the samples. In this case, print out a copy of each MSDS form and store it near the samples.

The samples should arrive packed in sealed, labeled vials. While wearing nitrile gloves, check that the labels on the vials match with the MSDS forms and what is expected from the collaborator. If the vials are not labeled, contact the source of the samples and do not open them.

Check how much sample you have received. Very small amounts of sample are needed for synchrotron experiments since the sample will be loaded into kapton capillaries that

are roughly 3-4 mm long with a 1 mm diameter. If you have received a large amount of each sample (i.e. several grams), consider sub-sampling the material into smaller vials and sending the rest back to the collaborator. Since pharmaceutical materials need to be handled with extra care and are harder to dispose of, it's easier for us to just send extra material back to the source than worry about it ourselves. Please note that you shouldn't sub-sample too little of the material because you will need it to load the capillaries, after all, and it could be a difficult process.

Put all of the relevant sample information into your lab notebook and also on the wiki for your project.

Once you have completed the checking and labeling, put the samples into secure containers, for example by double- or triple-bagging the samples, and then put the containers into the fridge.

A.4 Filling Capillaries

The samples all need to be in sealed kapton capillaries before they can be taken to the synchrotron. Kapton capillaries can be obtained from the Cole-Parmer website under the name "polyimide tubing". Typically, we use tubing with 0.0395" (1.0 mm) inner diameter and 0.0435" outer diameter (1.1 mm).

All work that is done with the samples should be performed in a functioning fume hood. Before filling a capillary, we need to prepare the fume hood. To do this:

- Make sure you have the necessary supplies. These are:

    1. The sample powder (in a vial)
    2. A capillary cut to be roughly 1.5 inches long
    3. Clay
    4. Mortar and pestle

5. Ethanol (or another solvent)
6. Regular white sheets of paper
7. Weighting paper
8. Kimwipes (or other laboratory cleaning tissues)
9. Plastic bags for disposal
10. Long, thin pieces of wire that can be used as plungers to push the powders into the capillaries

- Cover the area where you intend to work in the fume hood with white sheets of paper. This will facilitate cleanup later.
- Clean the mortar and pestle and wire plunger with the ethanol.
- Make sure you have some Kimwipes on hand to clean up spills.

To fill a capillary:

- Seal one side of the capillary with modeling clay.
- Pour some of the sample powder into the mortar. Crush it with the pestle. It is important to have a fine powder for the experiment or the data might be spotty. If you have a sieve, try to sieve the powder as well.
- There are two methods you could use to load the powder into the capillary. The first is to try to scoop up the powder from the mortar into the capillary. Run the capillary through a small pile of powder making a scooping motion, then tap the capillary (while its in a vertical position) against a horizontal surface to get the powder to settle. This method is a bit messy, so be sure you don't get powder all over the place. The alternative is to use a funnel of some sort to try to funnel the powder into the capillary. In either case, you will have to use the wire plunger to pack the sample into the capillary so that it's tightly backed.

- Once you're done loading the sample, seal the other side of the capillary using modeling clay. Put the capillary into its own plastic bag and put that bag into the same bag as the sample vial. Clean everything up. Put used Kimwipes, weighting paper, and the paper that you lined the bottom of the fume hood with into the bag that was set aside for disposal. Label the bag using the environmental safety labels from Columbia facilities and then contact Environmental Health & Safety (EHS) to pick up and dispose of the waste.

A.5 Bringing the Sample to the Beamline

Generally, we want to bring only the sealed capillary to the beamline. This way, we can ensure that potentially toxic samples are never spilled at the experiment and no waste management / clean-up is necessary. It is also much easier for us to take the capillaries back home.

In preparation for travel, the capillaries should be in small, labeled plastic bags (I use snack size Ziploc bags). These small bags should be in larger, labeled plastic bags (I use either quart-sized or gallon-sized Ziploc bags). All of these bags should be in a cardboard box. The MSDS forms should also be close at hand for all materials. The box holding the samples should also include things that might be useful at the beamline such as some extra capillaries, clay, and wire plunger.

The experimental box should be in a piece of luggage or a bag. If you are flying, check in the luggage with the samples, rather than carrying it on.

A.6 Disposing of Samples

After all relevant experiments are complete and you are sure that you will not to use certain samples again, then pack the samples into plastic bags that are labeled using Columbia's environmental safety labels, and then contact EHS to pick up and dispose of the waste.

A.7 General Tips

This section is a summary of a conversation with a researcher at a pharmaceutical firm about the best way to handle pharmaceutical samples.

- Standard protection when handling toxic pharmaceutical materials includes: fume hood (to be used in all cases), glasses, disposable lab coat, gloves (to be used in all cases), and possibly a dust mask as well (if the sample is known to form dust clouds).
- A fume hood is necessary because many samples are toxic if inhaled.
- In case of skin or eye irritation (e.g. allergic reaction), flush with copious water and make sure to contact the proper safety people.
- In case of spill, use water to clean up the spilled material, collect it, and then dispose of it as above.
- Section 8 of the safety data sheet contains information on exposure control. Always read this when you get a new sample shipped to you.
- Don't try to put too much of a sample into capillary at once.
- Use nitrile gloves to handle substances. These substances are dry, so there will be no solvent interaction with gloves.
- Clean up each sample and put on a fresh pair of gloves before handling the next sample.

Appendix B

PDFGetX3 Documentation

This appendix contains the documentation for PDFGetX3, current as of June 1, 2011. This documentation is provided in the thesis a supplement to Chapter 4, which introduces the PDFGetX3 software. This version of the documentation omits the installation instructions. Please refer to the http://www.diffpy.org website for information about the software and the latest version of the manual. Please note that PDFGetO is not PDFGetX3. PDFGetO is a web-based front-end for the PDGetX3 engine, and therefore information about input parameters and command line arguments do not apply.

B.1 Introduction

PDFGetX3 represents the next step in the creation of high quality PDFs that can be used for productive research purposes. Just by inputting a few parameters into a user-friendly configuration file, a user will be able to automatically and rapidly create and plot PDFs using PDFGetX3.

This code was written by members of the Billinge Group at Columbia University including: Timur Dykhne, Pavol Juhas, Christopher Farrow, Simon Billinge.

B.2 Quick-Start Guide

More information on how the program works is in the *Using PDFGetX3* section, the *Examples* section contains instructions on how to run the examples included with the program, and the *Tutorial* section contains a step-by-step guide for processing a data set and changing input parameters.

To run PDFGetX3, input data need to be in the proper format. Please refer to the section *Input Files* for more information. The minimum configuration parameters needed for data processing are:

- The wavelength of the X-rays in the experiment;
- The chemical composition of the material;
- The Q_{max};
- The format of the input data (2θ, Q (Å^{-1}) or Q(nm^{-1}))

For more information on the configuration parameters or configuration file, refer to the *Using PDFGetX3* and the *Configuration and Command Line Options* sections.

Once we have the input parameters and an input data file, we run PDFGetX3 at the command line with:

```
>> rapdfgetx [DATA-FILE-NAME] -b BACKGROUND-FILE-NAME -c CONFIG-FILE-
-NAME
```

where `[DATA-FILE-NAME]` is the name of the file containing the data we want to process, `[BACKGROUND-FILE-NAME]` is the name of the file containing intensities collected from the background (if available), and `[CONFIG-FILE-NAME]` is the name of the file containing the configuration information. Note that if the configuration file has the default name `rapdfgetx.cfg` and is in the same folder as the data, then it will be detected automatically without the `-c` option. Also note that if the background file name is in the configuration file, then it does not need to be defined at the command line.

B.3 Input Files

PDFGetX3 accepts inputs in the form of two-column data where the first column contains values expressed in terms of 2θ in degreees, Q (momentum transfer), in nm^{-1}, or Q in Å^{-1} and the second column contains intensities. The user defines the type of data in the first column using the **dataformat** parameter in the configuration file or at the command line.

PDFGetX3 is generally smart about ignoring metadata in the header of a data file, but it will get confused if the data file is extremely peculiar. For instance, if one of the data columns has letters in addition to numbers. If PDFGetX3 is having trouble loading a data file, then we should open it in a text editor to verify that it contains legible data.

B.4 Using PDFGetX3

Using PDFGetX3 is simple and straightforward. First, we set up a configuration file with all of the necessary parameters defined, then we run the program at the command line, and finally we plot the output. We iterate these steps until we are satisfied with the results.

B.4.1 Configuration

The first thing that we need is a configuration file. The configuration file contains the necessary parameters that are needed to produce PDFs using PDFGetX3.

Note: Output files from PDFGetX2 and PDFGetX3 can also be used as configuration files. This is useful if we have pre-made PDFs but are not sure how they were made.

B.4.1.1 Creating a Configuration File

First, we check if our data came with a configuration file (usually files with a `.cfg` extension). If we don't have a configuration file, then we can easily create a template of a configuration file by running the following at the command prompt in Windows or terminal window in Linux:

```
>> rapdfgetx --template > config.cfg
```

Where `config.cfg` is what we want to call our configuration file. This command will output a configuration file named `config.cfg` that contains a template that can be modified to suit our needs. The file name `config.cfg` was chosen here as an example. Any file name can be used.

The config file can be opened using any text editor, such as WordPad or Vim. Once opened, a typical config file will look like Figure B.1. Any text in the configuration file that begins with a hash mark (`#`) will be ignored.

The order in which PDFGetX3 gathers configuration information is as follows: first it searches the users home directory for a file named `.rapdfgetx.cfg`, then it searches the working directory for a file named `.rapdfgetx.cfg`, and finally it searches the working directory for a file named `rapdfgetx.cfg`. If any of these files are found and they contain enough configuration information for PDFGetX3 to run, then the user will not need to manually define a configuration file at the command line using the `-c` option.

B.4.1.2 Sections of the Configuration File

The user can add sections to configuration files with configurations that are specific to a certain compound or sample. For instance, if the working directory has data collected from two different compounds, for example NaCl and $BaTiO_3$, then the main `rapdfgetx.cfg` configuration file can have two sections, `[NaCl]` and `[BaTiO3]`, that contain configuration information specific to just those compounds (e.g. composition, Q_{max}, and wavelength), and get the rest of the information from the `[DEFAULTS]` section. The user can then call each section using the `--section` option at the command line (please refer to the *Configuration and Command Line Options* section of this appendix).

Using this feature, it is possible to have one main configuration file that contains information about many different samples, rather than many configuration files that are specifically

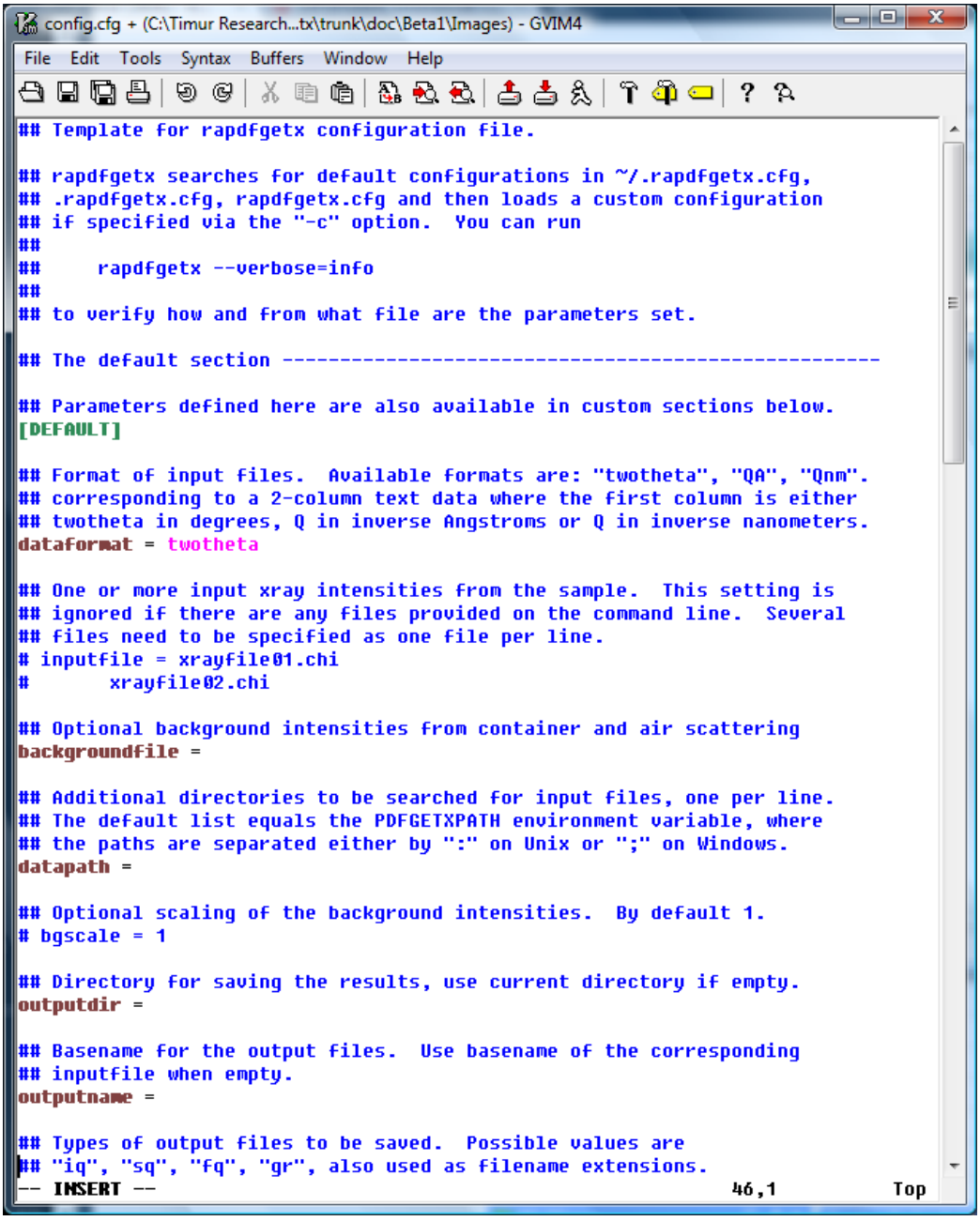

Figure B.1: Template of a configuration file.

made for each sample.

B.4.1.3 Options in the Configuration File

There are many options that can be turned on and off in the configuration file, the details of which are in the auto-generated template and in the *Configuration and Command Line Options* section. Options are enabled by setting them equal to some value in the configuration file. For instance, to set the wavelength to 0.709 Å, we would add or modify the line `wavelength=0.709` in the configuration file. The options that are needed for PDFGetX3 to run are: **dataformat**, **wavelength**, **composition**. Additionally **outputtype** is necessary for the software to output a file.

B.4.2 Running PDFGetX3

Once we have adjusted parameters in the configuration file, we are ready to run the program. To do so, we open a terminal window (in Linux) or command prompt (in Windows), navigate to a folder containing the raw data files that we want to process, the background intensities (optional), and the configuration file, and execute rapdfgetx.py with the following command:

```
>> rapdfgetx [DATA-FILE-NAME] -b [BACKGROUND-FILE-NAME] -c [CONFIG-
-FILE-NAME]
```

Note: If **outputtype** is not defined in the configuration file, then we will also need to add `outputtypes=x` or `-t=x` where `x` is the type of file that we want to output (the four options are: 'gr', 'fq', 'sq', or 'iq') or else there will be no output even if the program runs successfully.

B.4.3 Plotting

More information on exactly what is contained in files output by PDFGetX3 is in the *Output Files* section. Once we have an output (for example, a $G(r)$ or $F(Q)$) we naturally want to

plot it. We can use the `plot` argument at the command line when running PDFGetX3 to plot data right after they are processed. Additionally, if we want to plot an output without running PDFGetX3 (and thereby reprocessing data), we use the `plotdata` program. It can be run from the command line as follows:

```
>> plotdata [DATA-FILE-NAME] [DATA-FILE-NAME-2] ...
```

Where `[DATA-FILE-NAME]` corresponds to the PDFs that we want to plot. The plotdata script can accept multiple inputs and plot them all on top of one another. The plot window allows us to zoom in and out on the data and save an image of the plot contents. However, it does not allow for any sort of data manipulation. To manipulate the data, we need to load them into a data manipulation/plotting program, such as Excel, Origin, Kuplot, or Python (with the Matplotlib library).

B.5 Output Files

PDFGetX3 can output up to four different data file types:

- I(Q): This file consists of the intensities after the data have been converted to Q-space, if necessary, and the background has been subtracted.
- S(Q): This file contains the structure function.
- F(Q): This file contains the reduced structure function.
- G(r): This is the PDF.

We can choose which files to output either in the configuration file via the **outputtype** parameter, or at the command line with the `--outputtypes` or `-t` options.

The the structure of each type of output file is roughly the same. The header contains metadata such as the version of the software and the configuration information while the body of the file contain two columns of x-y data. Figure B.2 shows a typical PDF output file.

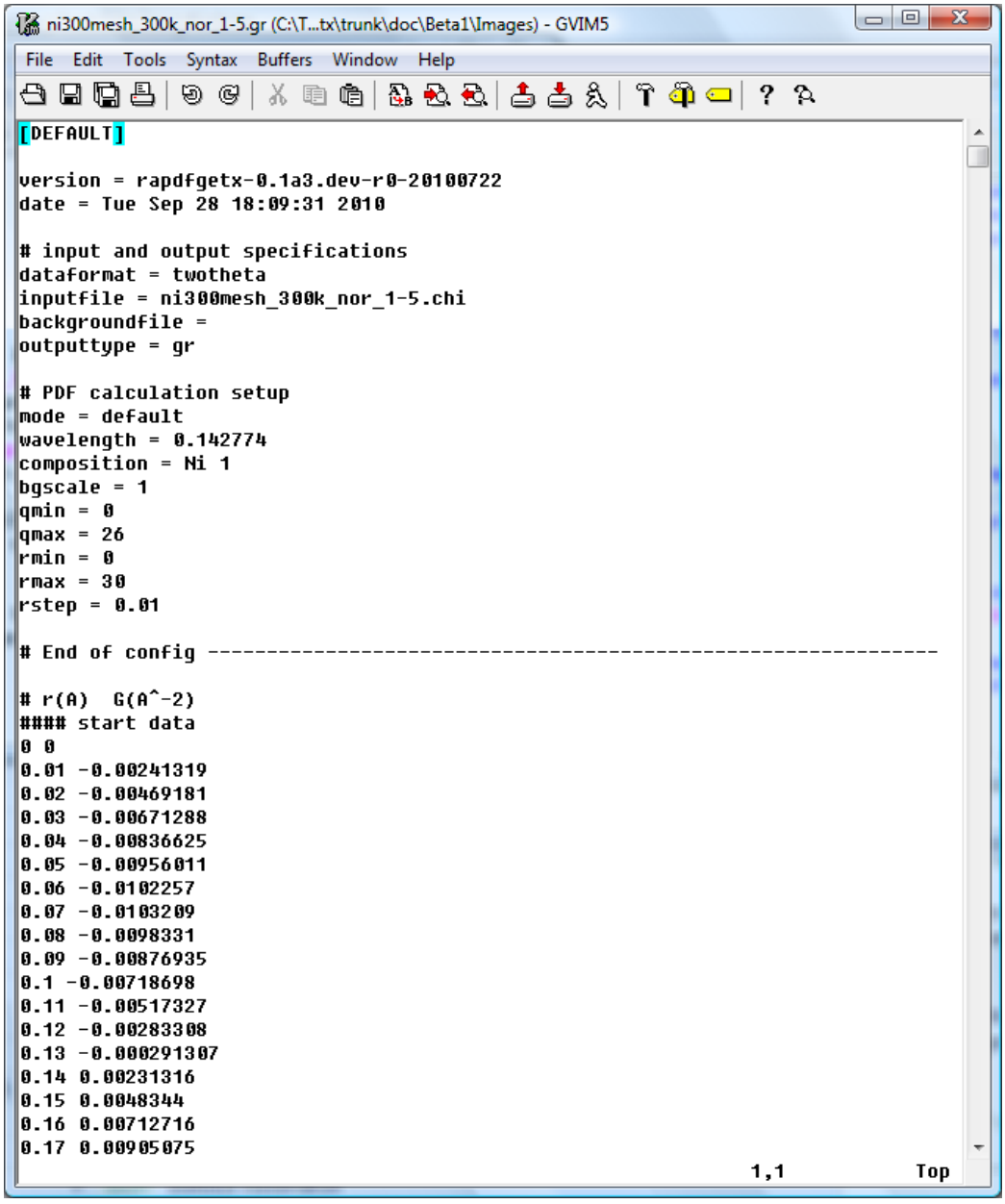

Figure B.2: A `.gr` ($G(r)$) file output by PDFGetX3.

B.6 Examples

PDFGetX3 comes with several examples in the `examples` directory of the installation. Each example includes the following files:

- A `.chi` file which contains the raw intensities plotted versus 2θ collected from a synchrotron beamline. Some examples also include a background data file (also in `.chi` format) while others do not.
- A `rapdfgetx.cfg` configuration file.
- A PDF (file with `.gr` extension) with `rapdfgetx` in its name and a PDF with `pdfgetx2` in its name. These two files are PDFs made with PDFGetX3 and PDFGetX2, respectively, from the same experimental data.

The following compounds are included as example:

- BaTiO3
- CeO2
- Menthol
- Ni
- Pt
- Si90Ni10

Note that several of the example compounds were discussed at length (with images and model refinement results) in Chapter 4.

B.7 Tutorial

In this tutorial we will make and compare several PDFs of nickel by adding or removing the background, varying the Q_{max}, and tuning parameters in interactive mode.

B.7.1 Getting Started

First we navigate over to the `examples` folder in our PDFGetX3 directory and then to the `Ni` folder. The file named `ni300mesh_300k_nor_1-5.chi` contains the raw data collected from the beamline. We can open the file in a text editor to verify its contents. The first column contains two-theta values and the second column contains intensities. The file `kapton_bgrd_300k_nor_2-3.chi` contains the measured background intensities, in this case an empty Kapton® capillary.

In addition to the data files, we also see the configuration file `rapdfgetx.cfg`, which contains all of the configuration parameters for the experiment.

Finally, there are two more files in the folder. One has `pdfgetx2` and the other has `rapdfgetx` in the file name. Both have `.gr` extensions. These are sample PDFs that we can compare our output to. The file with `pdfgetx2` in the title was created using the PDFGetX2 software, while the file with `rapdfgetx` in the title was created using PDFGetX3.

B.7.2 Setting Up a Configuration File

The configuration file, `rapdfgetx.cfg` is already set up to make a PDF of this particular nickel data set. If we run the program using the configuration as it is now, we will get a PDF with the background subtracted and a Q_{max} of 26 Å^{-1}. Before we do that, though, we should make a copy of the configuration file and give it a different name, for example `config.cfg`.

We will modify the contents of this file to see what effects Q_{max} and background subtraction have on the resulting PDF. Modifying a config file is easy. We just need to change the parameters using a text editor.

In this new configuration file, put a hash mark, '#', in front of the line `backgroundfile = kapton_bgrd_300k_nor_2-3.chi`. The hash mark comments out that line of text so that it will be ignored by the program. PDFGetX3 will now no longer do a background subtraction

when processing this data set if this new configuration file is used.

B.7.3 Running PDFGetX3

Now lets run the program. Open a terminal or command prompt and navigate over to the folder where the data files and configuration file are stored. Run the program with:

```
>> rapdfgetx ni300mesh_300k_nor_1-5.chi -c config.cfg
```

This will create a PDF with the filename `ni300mesh_300k_nor_1-5.gr`. The default behavior of PDFGetX3 is to always create output files with the same file name root as the data files. This can be changed using the **outputname** parameter in the configuration file or the `--outputname` argument at the commandline. If a file with the name already exists in the folder, then PDFGetX3 will give a warning and not overwrite the existing files. To force PDFGetX3 to overwrite the files, use the `--force` argument at the command line.

Note that if the configuration file was given a different name, then we substitute that name for `config.cfg`.

B.7.4 Plotting the Output

Now that we have made a new PDF, lets plot it and the PDF with the background subtracted on top of one another. To plot, we use the `plotdata` program as follows.

```
>> plotdata ni300mesh_300k_nor_1-5.gr ni300mesh_300k_nor_1-5-rapdfgetx.gr
```

Where `ni300mesh_300k_nor_1-5.gr` can be replaced with the file name of the PDF that has just been generated, if an alternative name was chosen, and `ni300mesh_300k_nor_-1-5-rapdfgetx.gr` is the name of the pre-made PDF from the examples folder.

The output will look like Figure B.3. We see that the background subtraction does not have a large effect on the PDF. This is because the Ni sample diffracts more strongly than the capillary that the sample is contained in.

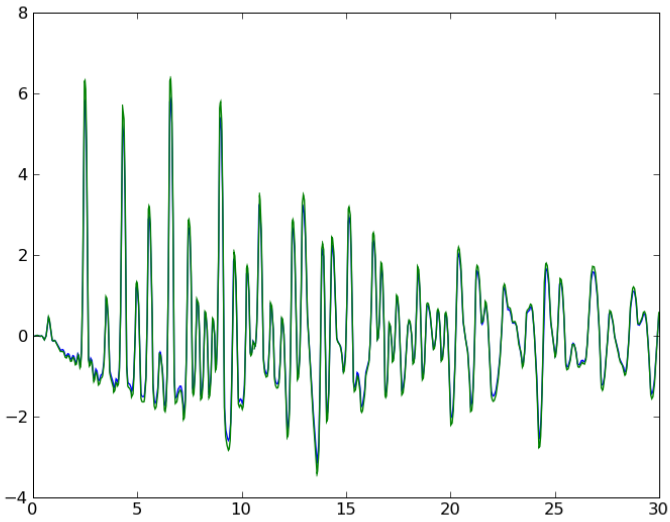

Figure B.3: PDFs of nickel with and without background subtraction.

B.7.5 Changing Q_{max}

Now, we want to see what effect adjusting the Q_{max} will have on the PDF. We again open the configuration file with a text editor. First, we reintroduce the background subtraction by removing the hash mark (#) in front of the line beginning with **backgroundfile**. Then we go to the line that says **qmax = 26.0** and change this value to 16.0, so the line should read **qmax = 16.0**. Save your changes.

Now, we run the program as we had done previously:

```
>> rapdfgetx ni300mesh_300k_nor_1-5.chi -c config.cfg --force
```

Note that we have now introduced the command line option `--force`. The reason for using the force command is because PDFGetX3 does not automatically overwrite existing files. Therefore, if we had not used the force command, the program would respond by telling you that the output file already exists.

After we run the program with the `--force` command, PDFGetX3 outputs a new PDF with the new Q_{max}. To determine what effect a change in Q_{max} has on the PDF, we plot

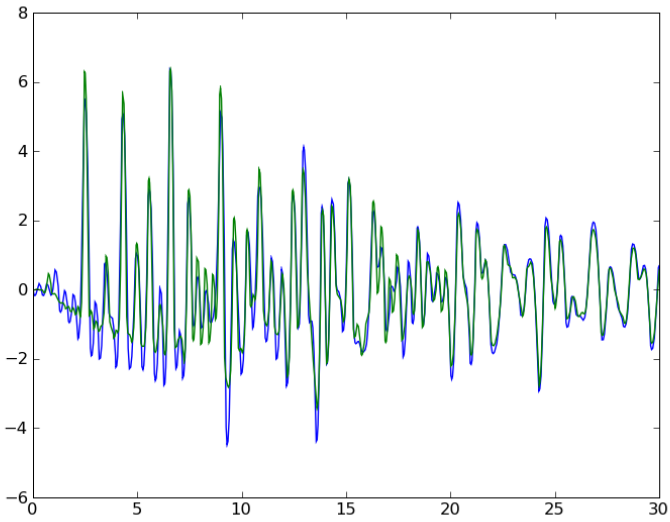

Figure B.4: PDFs of nickel with different Q_{max} values.

PDFs on top of one another, as before:

```
>> plotdata ni300mesh_300k_nor_1-5.gr ni300mesh_300k_nor_1-5-rapdfgetx.gr
```

The output should look like Figure B.4. Note that the lowering Q_{max} affects the peak intensity.

B.7.6 Tuning Parameters in Interactive Mode

One of the most powerful features of PDFGetX3 is the ability to tune various parameters in interactive mode. In the previous section, we compared two PDFs with different Q_{max} values. But, what if we want to see what effect an entire range of Q_{max} can have on a PDF and we don't want to run the program a bunch of times? To do this, we tune parameters in interactive mode.

First, to enter interactive mode, we run PDFGetX3 with the `-i` command line option. For example, to enter interactive mode with the Ni data and the `rapdfgetx.cfg` configuration file, we enter interactive mode with:

```
>> rapdfgetx ni300mesh_300k_nor_1-5.chi -i --force
```

This launches us into the interactive Python session with the PDF data pre-loaded.

Now, for this example, lets say we want to see what effect a change in Q_{max} has on the PDF and on the reduced structure factor, $F(Q)$. To do this, first set which variables we wanted to plot by typing:

```
>> config.plot=["fq","gr"]
```

In the interactive session. Next, we define the parameter that we want to tune by typing:

```
>> tuneconfig(qmax=(10,20))
```

In this case, we want to tune Q_{max} between 10.0 Å^{-1} and 20.0 Å^{-1}.

Now, a plot window should appear with the reduced structure function, $F(Q)$, in the top half, and the PDF, $G(r)$, in the bottom half. As we move the mouse through the plot window from left to right, the Q_{max} changes. The title line of the plot shows us the current value of Q_{max}. When we click the mouse in the window, we set the Q_{max} as that value. Figure B.6 shows the tuning window.

We might also be interested in having a baseline PDF to visually compare our tuned PDF to. For example, to test the effects of varying Q_{max}, we can compare a tuned PDF to the PDF with Q_{max} of 26 Å^{-1} that we made earlier in this tutorial. To do this, we simply run PDFGetX3 with the `--plot` command line option:

```
>> rapdfgetx ni300mesh_300k_nor_1-5.chi -i --force --plot="fq,gr"
```

This will put us into interactive mode with a plot window already open. The $F(Q)$ and $G(r)$ in this plot window were made with the parameters preset in the configuration file. Now, we can run `tuneconfig` again to see how a PDF made with a lower Q_{max} compares to the one with a Q_{max} of 26 Å^{-1} by typing:

```
>> tuneconfig(qmax=(10,25))
```

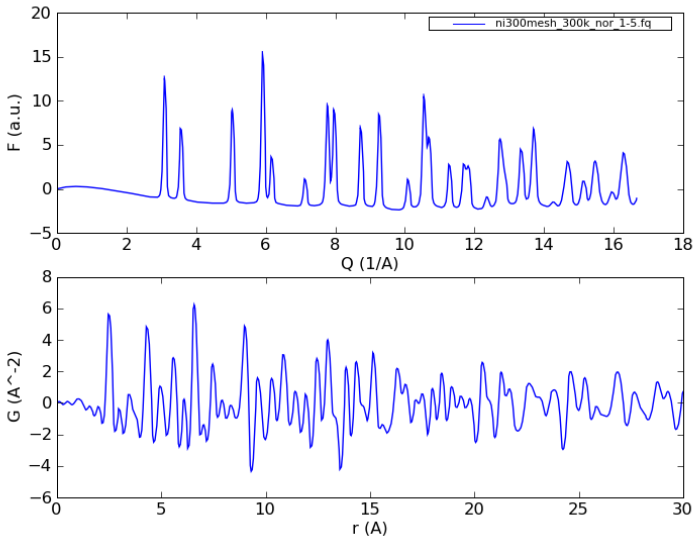

Figure B.5: Tuning the Q_{max} of nickel. Here it is at around 17 Å^{-1}.

Now, when we move the mouse on the plot, we see that the blue $F(Q)$ and $G(r)$ correspond to the original PDF while the green $F(Q)$ and $G(r)$ are being tuned. The plot should look like Figure B.6.

This tutorial has gone through the basic steps necessary to create good PDFs using PDFGetX3. We can refer to the *Configuration and Command Line Options* section for a list all of the available options.

B.8 Configuration and Command Line Options

PDFGetX3 is very flexible in allowing the user to customize how he or she want the program to run. Tables B.1, B.2, B.3, and B.4 show the options that are available in the configuration file and command line, what argument they take, and what they do. Note that the command line options that are more than a letter long have two dashes (`--`), while those that are a letter long have one dash (`-`).

Notes from the tables:

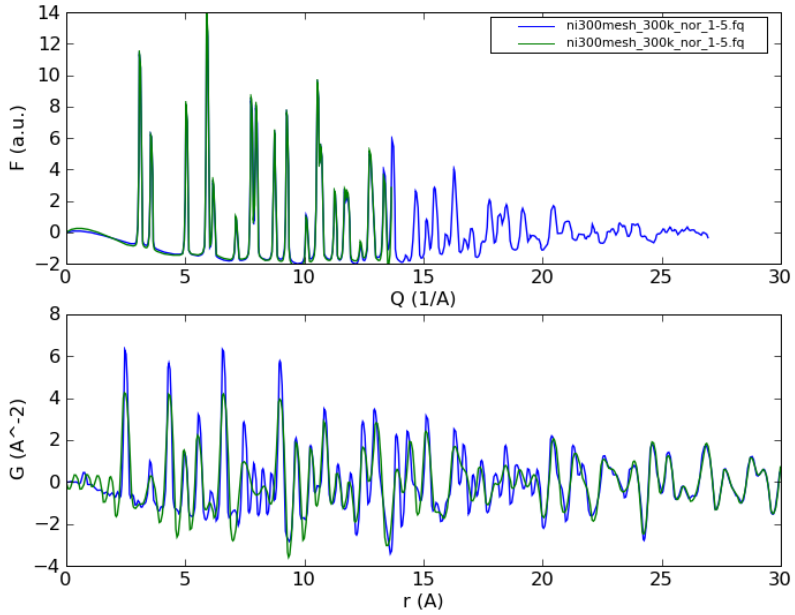

Figure B.6: Comparing PDFs of nickel by tuning the Q_{max}. The reference (blue) has 26 Å^{-1}, while the tuned PDF (green) is at around 14 Å^{-1}.

Table B.1: List of parameters that can be defined in the configuration file or at the command line, the type of argument that they take, and what they do.

Config File	Command Line	Argument (1)	What it Does
dataformat	`--format` or `-f`	'twotheta', 'Qnm', 'QA'	Tells PDFGetX3 the format the data collected using an X-ray instrument are in. twotheta means the that intensities are plotted versus two-theta in degrees. Qnm means that the intensities are plotted versus Q in inverse nanometers. QA means that the intensities are plotted versus Q in inverse angstroms. This option is required!
backgroundfile	`--background` or `-b`	filename	Optional background file name, by default none
outputtype	`--outputtypes` or `-t`	'iq', 'sq', 'fq', 'gr'	Results to be saved, by default none. iq is the intensity after background subtraction and normalization onto a regular grid in Q-space; sq is the structure factor; fq is the reduced structure factor; and gr is the PDF. Note more than one argument is accepted.
wavelength	`--wavelength` or `-w`	number (Å)	X-ray wavelength in angstroms (required)
plot	`--plot` or `-p`	'iq', 'sq', 'fq', 'gr'	Outputs to be plotted after the processing is complete, none by default

Table B.2: Continuation of the list of parameters that can be defined in the configuration file or at the command line, the type of argument that they take, and what they do.

Config File	Command Line	Argument (1)	What it Does
composition	`--composition`	Chemical composition	Chemical composition of the sample **(2)**. This option is required!
bgscale	`--bgscale`	number	Factor by which to scale the background data, by default 1
qmin	`--qmin`	number (1/Å)	Minimum Q value to use in Fourier transform in inverse angstroms, by default 0
qmax	`--qmax`	number (1/Å)	Maximum Q value to use in Fourier transform in inverse angstroms. Required!
rmin	`--rmin`	number (Å)	Low limit of r-grid in angstroms, by default 0
rmax	`--rmax`	number (Å)	Upper limit of r-grid in angstroms, by default 30
rstep	`--rstep`	number (Å)	r-grid step size in angstroms, by default 0.01
inputfile		filename	Names of files to be processed, one per line. Ignored if any inputs are given at command line
datapath	`--datapath`	directory name	Directories to search for files in addition to the current directory

Table B.3: Continuation of the list of parameters that can be defined in the configuration file or at the command line, the type of argument that they take, and what they do.

Config File	Command Line	Argument **(1)**	What it Does
outputdir	`--directory` or `-d`	directory name	Directory for output files. Current directory used by default
outputname	`--outputname` or `-o`	filename	Base name (before extension) for output files. Uses same base name as input files by default
interact	`--interact` or `-i`	'yes' in config file **(3)**	Starts an interactive Python session after processing data
verbose	`--verbose`	'error', 'warning', 'info', 'debug'	How much information the program should reports
N/A	`--config` or `-c`	config file name	Select the configuration file to use, rapdfgetx.cfg is used by default
force	`--force`	'yes' in config file	Overwrite output files, if they exist. By default PDFGetX3 does not overwrite files
N/A	`--help` or `-h`		Lists all available command line arguments and what they do
N/A	`--version` or `-V`		Displays the program version
N/A	`--template`		Displays the template for a configuration file
N/A	`--section` or `-s`	Section name	Select the active section in the config file **(4)**

Table B.4: Continuation of the list of parameters that can be defined in the configuration file or at the command line, the type of argument that they take, and what they do.

Config File	Command Line	Argument **(1)**	What it Does
N/A	`--mode`	'default', 'chris1005', 'timur', 'electron', 'pj1010', 'pj1010-electron'	Choose the algorithm that the PDF calculator uses. By default, the default mode is used
N/A	`--pattern`	see description in notes **(5)**	Process arguments in file name as a pattern **(5)**
N/A	`--list`		List all input files, useful when creating a pattern

1. Argument: The argument is what we set the option parameters to. Some parameters accept one or more of a specific set of arguments. For instance, dataformat can be one of 'twotheta', 'Qnm', or 'QA'. In other cases, parameters accept strings such as a file name or numbers. Finally, for some command line options, no argument is necessary.

2. Composition: This is a string containing letters and numbers. Supported formats are 'PbTi0.5Zr0.5O3', 'Pb Ti0.5 Zr0.5 O3' or 'Pb 1 Ti 0.5 Zr 0.5 O 3'. Space characters are ignored, unit counts can be omitted, but it is important to use proper case for atom symbols. Elements can repeat in the formula (e.g., 'CH3 CH3'), but parenthesis are not supported.

3. Interact: The interactive mode lets the user enter an interactive Python session with the variables from processing the PDF already loaded. Help can be found at the interactive Python prompt. If you enable the interactive session from the command line, then no argument is needed. To enable it from the configuration file, you must make it true as follows: **interact = yes**.

4. Section: Configuration files can have sections devoted to different materials in addition to the `[DEFAULT]` section. For example, there could be a `[NaCl]` section that contains parameters specific to that material (e.g. wavelength, Q_{max}, composition, etc).

5. Pattern: The pattern option lets the user process many sets of data sequentially without having to manually type the name of each data file. Using pattern allows the user to define a rule by which PDFGetX3 will process data sets. For instance, if we have datasets named test_1.chi, test_2.chi, test_3.chi in the same folder, they can all be processed using the following command line argument:

```
>> rapdfgetx --pattern "test_<1-3>"
```

The available options for patterns are:

- `^start` match start only at the beginning of the string
- `end$` match end only at the end of string
- `<7>` match number 7 preceded by any number of leading zeros
- `<1-34>` match an integer range. The matched number may have one or more leading zeros
- `<7->` match an integer greater or equal 7 allowing leading zeros
- `<->` match any integer

Part III

Bibliography

Bibliography

[Acharya and Kuchela, 1982] K. Ravindra Acharya and K. N. Kuchela. Crystal structure of sulfamerazine. *J. Crystallogr. Spectrosc. Res.*, 12(4):369–376, 1982.

[Alagusundaram *et al.*, 2010] M. Alagusundaram, N. Deepthi, S. Ramkanth, S. Angalaparameswari, T.S. Mohamed Saleem, K. Gnanaprakash, V.S. Thiruvengadarajan, and C. Madhusudhana Chetty. Dry powder inhalers - an overview. *Int. J. Res. Pharm. Sci.*, 1(1):34–42, 2010.

[Allen, 2002] F.H. Allen. The cambridge structural database: a quarter of a million crystal structures and rising. *Acta Crystallogr. B*, B58(1):380–388, 2002.

[Amidon *et al.*, 1995] Gordon L. Amidon, Hans Lennernas, Vinod P. Shah, and John Crison. A theoretical basis for a biopharmaceutic drug classification: the correlation of in vitro drug product dissolution and in vivo bioavailability. *Pharmaceut. Res.*, 12(3):413–420, 1995.

[Andronis and Zografi, 2000] Vlassios Andronis and George Zografi. Crystal nucleation and growth of indomethacin polymorphs from the amorphous state. *J. Non-Cryst. Solids*, 271(3):236–248, 2000.

[Baldrick, 2000] Paul Baldrick. Pharmaceutical excipient development: the need for preclinical guidance. *Regul. Toxicol. Pharm.*, 32(2):210–218, 2000.

[Balyuzi, 1975] H. H. M. Balyuzi. Analytic approximations to incoherently scattered x-ray intensities. *Acta Crystallogr. A*, A31(5):600–2, 1975.

[Banker and Rhodes, 2002] Gilbert S. Banker and Christopher T. Rhodes. *Modern Pharmaceutics*. Marcel Dekker, New York, 4th edition, 2002.

[Barr *et al.*, 2004] Gordon Barr, Wei Dong, and Christopher J. Gilmore. Polysnap: a computer program for analysing high-throughput powder diffraction data. *J. Appl. Crystallogr.*, 37(4):658–664, 2004.

[Bates *et al.*, 2006] S. Bates, G. Zografi, D. Engers, K. Morris, K. Crowley, and A. Newman. Analysis of amorphous and nanocrystalline solids from their x-ray diffraction patterns. 23(10):2333–2349, 2006.

[Bates *et al.*, 2007] S. Bates, R. C. Kelly, I. Ivanisevic, P. Schields, G. Zografi, and A. W. Newman. Assessment of defects and amorphous structure produced in raffinose pentahydrate upon dehydration. *J. Pharm. Sci.*, 96(5):1418–1433, 2007.

[Billinge and Kanatzidis, 2004] S. J. L. Billinge and M. G. Kanatzidis. Beyond crystallography: the study of disorder, nanocrystallinity and crystallographically challenged materials. *Chem. Commun.*, 2004:749–760, 2004.

[Billinge and Levin, 2007] S. J. L. Billinge and I. Levin. The problem with determining atomic structure at the nanoscale. *Science*, 316:561–565, 2007.

[Billinge *et al.*, 2010] Simon J. L. Billinge, Timur Dykhne, Pavol Juhás, Emil Božin, Ryan Taylor, Alastair J. Florence, and Kenneth Shankland. Characterisation of amorphous and nanocrystalline molecular materials by total scattering. *CrystEngComm*, 12(5):1366–1368, 2010.

[Billinge, 2007] S. J. L. Billinge. Nanostructure studied using the atomic pair distribution function. *Z. Kristallogr. Suppl.*, 26:17–26, 2007.

[Billinge, 2008a] S. J. L. Billinge. Local structure from total scattering and atomic pair distribution function (pdf) analysis. In Robert E. Dinnebier and Simon J. L. Billinge, editors, *Powder diffraction: theory and practice*, pages 464 – 493, London, England, 2008. Royal Society of Chemistry.

[Billinge, 2008b] Simon J. L. Billinge. Nanoscale structural order from the atomic pair distribution function (PDF): There's plenty of room in the middle. *J. Solid State Chem.*, 181:1698–1703, 2008.

[Borka, 1974] L Borka. The polymorphism of indomethacin. new modifications, their melting behavior and solubility. *Acta. Pharm. Suec.*, 11(3):295–303, 1974.

[Bragg and Bragg, 1913] W. H. Bragg and W. L. Bragg. The reflection of x-rays by crystals. *Proc. R. Soc. Lond. A*, 88:428–438, 1913.

[Burley *et al.*, 2008] Jonathan C. Burley, Melinda J. Duer, Robin S. Stein, and Ranko M. Vrcelj. Enforcing ostwald's rule of stages: Isolation of paracetamol forms iii and ii. *Eur. J. Pharm. Sci.*, 31(5):271–276, 2008.

[Byrn, 1982] Stephen R. Byrn. *Solid state chemistry of drugs.* Academic Press, New York, 1982.

[C. W. Dwiggins, Jr and Park, 1971] C. W. Dwiggins, Jr and D. A. Park. Calculation of the intensity of secondary scattering of x-rays by non-crystalline materials. *Acta Crystallogr. A*, 27:264, 1971.

[Caira and Mohamed, 1992] Mino R. Caira and Riana Mohamed. Positive identification of two orthorhombic polymorphs of sulfamerazine (c11h12n4o2s), their thermal analyses and structural comparison. *Acta Crystallogr. B*, 48(4):492–498, 1992.

[Chantler, 1995] C. T. Chantler. Theoretical form factor, attenuation, and scattering tabulation for z=1-92 from e=1-10 ev to e=0.4-1.0 mev. *J. Phys. Chem. Ref. Data*, 24:71, 1995.

[Chen *et al.*, 2002] Xiaoming Chen, Kenneth R. Morris, Ulrich J. Griesser, Stephen R. Byrn, and Joseph G. Stowell. Reactivity differences of indomethacin solid forms with ammonia gas. *J. Am. Chem. Soc.*, 124(50):15012–15019, 2002.

[Chupas *et al.*, 2003] Peter J. Chupas, Xiangyun Qiu, J. C. Hanson, P. L. Lee, Clare P. Grey, and Simon J. L. Billinge. Rapid acquisition pair distribution function analysis (RA-PDF). *J. Appl. Crystallogr.*, 36:1342–1347, 2003.

[Chupas *et al.*, 2007] P. J. Chupas, K. W. Chapman, and P. L. Lee. Applications of an amorphous silicon-based area detector for high-resolution, high-sensitivity and fast time-resolved pair distribution function measurements. *J. Appl. Crystallogr.*, 40:463–470, 2007.

[Compton, 1923] Arthur H. Compton. A quantum theory of the scattering of x-rays by light elements. *Phys. Rev. Lett.*, 21(5):483–502, 1923.

[Cox and Manson, 2003] Philip J. Cox and Pamela L. Manson. Gamma-indomethacin at 120 k. *Acta Crystallogr. E*, E59(7):o986–o988, 2003.

[Craig, 2002] Duncan Q. M. Craig. The mechanisms of drug release from solid dispersions in water-soluble polymers. *Int. J. Pharm.*, 231(2):131–144, 2002.

[Cromer and Mann, 1967] Don T. Cromer and Joseph B. Mann. Compton scattering factors for spherically symmetric free atoms. *J. Chem. Phys.*, 47(6):1892–1893, 1967.

[Cullity and Stock, 2001] B. D. Cullity and S. R. Stock. *Elements of X-Ray Diffraction.* Prentice Hall, 3rd edition, 2001.

[de Villiers *et al.*, 2998] Melgardt M. de Villiers, Dale Eric Wurster, Jakkie G. Van der Watt, and Amol Ketkar. X-ray powder diffraction determination of the relative amount of

crystalline acetaminophen in solid dispersions with polyvinylpyrrolidone. *Int. J. Pharm.*, 163(1-2):219–224, 2998.

[Deslattes *et al.*, 2003] Richard D. Deslattes, Jr. Ernest G. Kessler, P. Indelicato, L. de Billy, E. Lindroth, and J. Anton. X-ray transition energies: new approach to a comprehensive evaluation. *Rev. Mod. Phys.*, 75(1):35–99, 2003.

[Dinnebier and Billinge, 2008] Robert E. Dinnebier and Simon J. L. Billinge, editors. *Powder diffraction: theory and practice*. Royal Society of Chemistry, London, England, 2008.

[Dykhne *et al.*, 2011] Timur Dykhne, Ryan Taylor, Alastair Florence, and Simon J. L. Billinge. Data requirements for the reliable use of atomic pair distribution functions in amorphous pharmaceutical fingerprinting. *Pharmaceut. Res.*, 28:1041–1048, 2011.

[Egami and Billinge, 2003] T. Egami and S. J. L. Billinge. *Underneath the Bragg peaks: structural analysis of complex materials*. Pergamon Press, Elsevier, Oxford, England, 2003.

[Ewald, 1921] P. P. Ewald. Die berechnung optischer und elektrostatisher gitterpotentiale. *Ann. Phys.-Berlin*, 64:253–287, 1921.

[Farrow and Billinge, 2009] C. L. Farrow and S. J. L. Billinge. Relationship between the atomic pair distribution function and small angle scattering: implications for modeling of nanoparticles. *Acta Crystallogr. A*, 65(3):232–239, 2009.

[Farrow *et al.*, 2007] C. L. Farrow, P. Juhás, Jiwu Liu, D. Bryndin, E. S. Božin, J. Bloch, Th. Proffen, and S. J. L. Billinge. PDFfit2 and PDFgui: Computer programs for studying nanostructure in crystals. *J. Phys: Condens. Mat.*, 19:335219, 2007.

[Farrow *et al.*, 2011] Christopher L. Farrow, Margaret Shaw, Hyun-Jeong Kim, Pavol Juhás, and Simon J. L. Billinge. The nyquist-shannon sampling theorem and the atomic pair distribution function. *arXiv*, 2011. arXiv:1104.0874.

[Frauenfelder *et al.*, 1979] Hans Frauenfelder, Gregory A Petsko, and Demetrius Tsernoglou. Temperature-dependent x-ray diffraction as a probe of protein structural dynamics. *Nature*, 280:558–563, 1979.

[Friedrich *et al.*, 1913] W. Friedrich, P. Knipping, and M. Laue. Interference appearances in x-rays. *Ann. Phys.-Berlin*, 41:971–988, 1913. reprinted from *Sitzungsber. K. Bayer. Akad. Wiss.* pp. 303-322, (1912).

[Fukuoka *et al.*, 1987] Eihei Fukuoka, Midori Makita, and Shigeo Yamamura. Glassy state of pharmaceuticals. ii. bioinequivalence of glassy and crystalline indomethacin. *Chem. Pharm. Bull*, 35(7):2943–2948, 1987.

[Gardner *et al.*, 2004a] Colin R. Gardner, Orn Almarsson, Hongming Chen, Sherry Morissette, Matt Peterson, Zhong Zhang, Szu Wang, Anthony Lemmo, Javier Gonzalez-Zugasti, Julie Monagle, Joseph Marchionna, Steve Ellis, Chris McNulty, Alasdair Johnson, Doug Levinson, and Michael Cima. Application of high throughput technologies to drug substance and drug product development. *Comput. Chem. Eng.*, 28(6):943–953, 2004.

[Gardner *et al.*, 2004b] Colin R. Gardner, Christopher T. Walsh, and Orn Almarsson. Drugs as materials: valuing physical form in drug discovery. *Nat. Rev. Drug Discovery*, 3:926–934, 2004.

[Gowthamarajan and Singh, 2010] K. Gowthamarajan and Sachin Kumar Singh. Dissolution testing for poorly soluble drugs: a continuing perspective. *Dissolution Technologies*, 17(3):24–32, 2010.

[Grzesiak *et al.*, 2003] Adam L. Grzesiak, Meidong Lang, Kibum Kim, and Adam J. Matzger. Comparison of the four anhydrous polymorphs of carbamazepine and the crystal structure of form i. *J. Pharm. Sci.*, 92(11):2260–2271, 2003.

[Guinier, 1963] A. Guinier. *X-ray Diffraction in Crystals, Imperfect Crystals, and Amorphous Bodies.* San Francisco, W.H. Freeman, San Francisco, 1963.

[Hammersley, 1998] A. P. Hammersley. Fit2d v9.129 reference manual v3.1. ESRF Internal Report ESRF98HA01T, 1998.

[Hancock and Parks, 2000] Bruno C. Hancock and Michael Parks. What is the solubility advantage for amorphous pharmaceuticals? 17(4):397–404, 2000.

[Hancock and Zografi, 1997] Bruno C. Hancock and George Zografi. Characteristics and significance of the amorphous state in pharmaceutical systems. *J. Pharm. Sci.*, 86(1):1–12, 1997.

[Heaney, 2001] Robert P. Heaney. Factors influencing the measurement of bioavailability, taking calcium as a model. *J. Nutr.*, 131(4):1344S–1348S, 2001.

[Hooke, 1665] Robert Hooke. *Micrographia: or some physiological descriptions of minute bodies made by magnifying glasses with observations and inquiries thereupon.* J. Martyn and J. Allestry, London, 1665.

[Hossain, 2006] G. M. Golzar Hossain. A new polymorph of sulfamerazine. *Acta Crystallogr. E*, 62(6):o2166–o2167, 2006.

[Hu *et al.*, 2002] Jiahui Hu, True L. Rogers, Judith Brown, Tim Young, Keith P. Johnston, and Robert O. Williams III. Improvement of dissolution rates of poorly water soluble apis using novel spray freezing into liquid technology. *Pharmaceut. Res.*, 19(9):1278–1284, 2002.

[Hubbell and Seltzer, 2004] J.H. Hubbell and S.M. Seltzer. Tables of x-ray mass attenuation coefficients and mass energy-absorption coefficients (version 1.4) http://www.nist.gov/pml/data/xraycoef. *Natl. Inst. Stan. Technol.*, 2004.

[I.-K. Jeong *et al.*, 1999] I.-K. Jeong, Th. Proffen, F. Mohiuddin-Jacobs, and S. J. L. Billinge. Measuring correlated atomic motion using x-ray diffraction. *J. Phys. Chem. A*, 103:921–924, 1999.

[I.-K. Jeong *et al.*, 2001] I.-K. Jeong, F. Mohiuddin-Jacobs, V. Petkov, S. J. L. Billinge, and S. Kycia. Local structure study of $in_xga_{1-x}as$ semiconductor alloys using high energy synchrotron x-ray diffraction. *Phys. Rev. B*, 63:205202, 2001.

[Imaizumi *et al.*, 1980] Hiroyuki Imaizumi, Naoki Nambu, and Tsuneji Nagai. Stability and several physical properties of amorphous and crystalline forms of indomethacin. *Chem. Pharm. Bull*, 28(9):2565–2569, 1980.

[Jeong *et al.*, 2001] I.-K. Jeong, J. Thompson, A. M. P. Turner, and S. J. L. Billinge. Pdfgetx: a program for determining the atomic pair distribution function from x-ray powder diffraction data. *J. Appl. Crystallogr.*, 34:536, 2001.

[John, 2009] Edgar John. How to set specifications for the particle size distribution of a drug substance? *Am. Pharm. Rev.*, 12(3):74–77, 2009.

[Juhás *et al.*, 2006] P. Juhás, D. M. Cherba, P. M. Duxbury, W. F. Punch, and S. J. L. Billinge. Ab initio determination of solid-state nanostructure. *Nature*, 440(7084):655–658, 2006.

[Kesisoglou *et al.*, 2007] Filippos Kesisoglou, Santipharp Panmai, and Yunhui Wu. Nanosizing - oral formulation development and biopharmaceutical evaluation. *Adv. Drug Deliver Rev.*, 59(7):631–644, 2007.

[Klug and Alexander, 1974] H. P. Klug and L. E. Alexander. *X-ray diffraction procedures for polycrystalline and amorphous materials.* Wiley, New York, 2nd edition, 1974.

[Laaziri *et al.*, 1999] K. Laaziri, S. Kycia, S. Roorda, M. Chicoine, J. L. Robertson, J. Wang, and S. C. Moss. High resolution radial distribution function of pure amorphous silicon. *Phys. Rev. Lett.*, 82:3460, 1999.

[Lexchin *et al.*, 2003] Joel Lexchin, Lisa A. Bero, Benjamin Djulbegovic, and Otavio Clark. Pharmaceutical industry sponsorship and research outcome and quality: systematic review. *BMJ*, 326(7400):1–10, 2003.

[Lowes *et al.*, 1987] Mariette M. J. Lowes, Mino R. Caira, Antonie P. Lotter, and Jakkie G. Der Van Watt. Physicochemical properties and x-ray structural studies of the trigonal polymorph of carbamazepine. *J. Pharm. Sci.*, 76(9):744–752, 1987.

[Maren, 1976] Thomas H. Maren. Relations between structure and biological activity of sulfonamides. *Annu. Rev. Pharmacol. Toxicol.*, 16:309–327, 1976.

[Masadeh *et al.*, 2007] A. S. Masadeh, E. S. Božin, C. L. Farrow, G. Paglia, P. Juhás, A. Karkamkar, M. G. Kanatzidis, and S. J. L. Billinge. Quantitative size-dependent structure and strain determination of CdSe nanoparticles using atomic pair distribution function analysis. *Phys. Rev. B*, 76:115413, 2007.

[Maurin *et al.*, 2008] Jan K. Maurin, Franciszek Plucinski, Aleksander P. Mazurek, and Zbigniew Fijalek. The usefulness of simple x-ray powder diffraction analysis for counterfeit control - the viagra® example. *J. Pharmaceut. Biomed.*, 43(4):1514–1518, 2008.

[Megaw, 1962] H. D. Megaw. Refinement of structure of batio3 and other ferroelectrics. *Acta Crystallogr.*, 15:972–973, 1962.

[Myers and Well, 2010] Jerome L. Myers and Arnold D. Well. *Research Design and Statistical Analysis*. Hillsdale: Lawrence Erlbaum Associates, 3 edition, 2010.

[Newman *et al.*, 2008] A. Newman, D. Engers, S. Bates, I. Ivanisevic, R. C. Kelly, and G. Zografi. Characterization of amorphous API:polymer mixtures using x-ray powder diffraction. *J. Pharm. Sci.*, 97(11):4840–4856, 2008.

[Paglia *et al.*, 2006] G. Paglia, E. S. Božin, and S. J. L. Billinge. A novel fine-scale nanostructure in γ-Al_2O_3. *Chem. Mater.*, 18:3242–3248, 2006.

[Pasveer, 1989] Bernike Pasveer. Knowledge of shadows: the introduction of x-ray images in medicine. *Sociol. Health Ill.*, 11(4):360–381, 1989.

[Perrin *et al.*, 2009] Marc-Antoine Perrin, Marcus A. Neumann, Hagit Elmaleh, and Lionel Zaske. Crystal structure determination of the elusive paracetamol form iii. *Chem. Commun.*, (22):3181–3183, 2009.

[Peterson *et al.*, 2003] P. F. Peterson, E. S. Božin, Th. Proffen, and S. J. L. Billinge. Improved measures of quality for atomic pair distribution functions. *J. Appl. Crystallogr.*, 36:53, 2003.

[Petkov *et al.*, 1999] V. Petkov, I. K. Jeong, J. S. Chung, M. F. Thorpe, S. Kycia, and S. J. L. Billinge. High real-space resolution measurement of the local structure of $ga_{1-x}in_xas$ using x-ray diffraction. *Phys. Rev. Lett.*, 83(20):4089–4092, 1999.

[Petkov *et al.*, 2000] V. Petkov, S. J. L. Billinge, S. D. Shastri, and B. Himmel. Polyhedral units and network connectivity in calcium aluminosilicate glasses from high energy x-ray diffraction. *Phys. Rev. Lett.*, 85:3436, 2000.

[Proffen *et al.*, 2003] Th. Proffen, S. J. L. Billinge, T. Egami, and D. Louca. Structural analysis of complex materials using the atomic pair distribution function - a practical guide. *Z. Kristallogr.*, 218:132–143, 2003.

[Qiu *et al.*, 2004] Xiangyun Qiu, Jeroen W. Thompson, and Simon J. L. Billinge. PDFgetX2: a GUI driven program to obtain the pair distribution function from x-ray powder diffraction data. *J. Appl. Crystallogr.*, 37:678, 2004.

[Rowe *et al.*, 2006] Raymond C Rowe, Paul J Sheskey, and Sian C Owen. *Handbook of Pharmaceutical Excipients*. Pharmaceutical Press, London, Chicago, 5th edition, 2006.

[Ruland, 1964] W. Ruland. The separation of coherent and incoherent compton x-ray scattering. *Brit. J. Appl. Phys.*, 15:1301, 1964.

[Sato *et al.*, 1981] Toyomi Sato, Akira Okada, Keiji Sekiguchi, and Yasuyuki Tsuda. Difference in physico-pharmaceutical properties between crystalline and noncrystalline 9,3"-diacetylmidecamycin. *Chem. Pharm. Bull*, 29(9):2675–2682, 1981.

[Schwartz and Cohen, 1987] L. H. Schwartz and J. B. Cohen. *Diffraction from materials*. Springer-Verlag, New York, 1987.

[Shah *et al.*, 1992] Vinod P. Shah, Kamal K. Midha, Shrikant Dighe, Iain J. McGilveray, Jerome P. Skelly, Avraham Yacobi, Thomas Layloff, C.T. Viswanathan, C. Edgar Cook, R.D. McDowall, Kenneth A. Pittman, and Sidney Spector. Analytical methods validation: bioavailability, bioequivalence and pharmacokinetic studies. *Pharmaceut. Res.*, 9(4):588–592, 1992.

[Shaikh *et al.*, 2009] J. Shaikh, D.D. Ankola, V. Beniwal, D. Singh, and M.N.V. Ravi Kumar. Nanoparticle encapsulation improves oral bioavailability of curcumin by at least 9-fold when compared to curcumin administered with piperine as absorption enhancer. *Eur. J. Pharm. Sci.*, 37(3):223–230, 2009.

[Shalaev and Zografi, 2002] Evgenyi Shalaev and George Zografi. The concept of 'structure' in amorphous solids from the perspective of the pharmaceutical sciences. In Harry Levine,

editor, *Amorphous Food and Pharmaceutical Systems*, pages 11–30, Cambridge, 2002. The Royal Society of Chemistry.

[Shankland *et al.*, 2004] Kenneth Shankland, Anders J. Markvardsen, and William I.F. David. Powder diffraction based structural studies of pharmaceuticals. *Z. Kristallogr. Suppl.*, 219(12):857–865, 2004.

[Sheth *et al.*, 2004] A. R. Sheth, S. Bates, F. X. Muller, and D. J. W. Grant. Polymorphism in piroxicam. *Cryst. Growth Des.*, 4(6):1091–1098, 2004.

[Sheth *et al.*, 2005] A. R. Sheth, S. Bates, F. X. Muller, and D. J. W. Grant. Local structure in amorphous phases of piroxicam from powder x-ray diffractometry. *Cryst. Growth Des.*, 5(2):571–578, 2005.

[Stanton, 1896] Arthur Stanton. Wilhelm conrad roentgen on a new kind of rays: translation of a paper read before the wuerzburg physical and medical society, 1895. *Nature*, 53:274–276, 1896.

[Toby and Egami, 1992] B. H. Toby and T. Egami. Accuracy of pair distribution function analysis applied to crystalline and noncrystalline materials. *Acta Crystallogr. A*, 48(3):336–46, 1992.

[Troup and Mitchner, 1964] A.E. Troup and H. Mitchner. Degradation of phenylephrine hydrochloride in tablet formulations containing aspirin. *J. Pharm. Sci.*, 53(4):375–379, 1964.

[Veber *et al.*, 2002] Daniel F. Veber, Stephen R. Johnson, Hung-Yuan Cheng, Brian R. Smith, Keith W. Ward, and Kenneth D. Kopple. Molecular properties that influence the oral bioavailability of drug candidates. *J. Med. Chem.*, 45(12):2615–2623, 2002.

[Warren and Mozzi, 1966] B.E. Warren and R.L. Mozzi. Multiple scattering of x-rays by amorphous samples. *Acta Crystallogr.*, 21:459–461, 1966.

[Warren, 1990] B. E. Warren. *X-ray Diffraction*. Dover, New York, 1990.

[Wildfong *et al.*, 2006] Peter L.D. Wildfong, Bruno C. Hancock, Michael D. Moore, and Kenneth R. Morris. Towards an understanding of the structurally based potential for mechanically activated disordering of small molecule organic crystals. *J. Pharm. Sci.*, 95(12):2645–2656, 2006.

[Wright, 1998] A. C. Wright. Diffraction studies of glass structure: the first 70 years. *Glass. Phys. Chem.*, 24:148–179, 1998.

[Wyckoff, 1967] R. W. G. Wyckoff. *Crystal Structures*, volume 1. Wiley, New York, 2nd edition, 1967.

[Yu, 2001] Lian Yu. Amorphous pharmaceutical solids: preparation, characterization and stabilization. *Adv. Drug Deliver Rev.*, 48(1):27–42, 2001.

[Zhang *et al.*, 2002] Geoff G.Z. Zhang, Chonghui Gu, Mark T. Zell, R. Todd Burkhardt, Eric J. Munson, and David J.W. Grant. Crystallization and transitions of sulfamerazine polymorphs. *J. Pharm. Sci.*, 91(4):1089–1100, 2002.