Pharm Res (2011) 28:1041–1048
DOI 10.1007/s11095-010-0350-0

RESEARCH PAPER

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

Timur Dykhne • Ryan Taylor • Alastair Florence • Simon J. L. Billinge

Received: 27 September 2010 / Accepted: 10 December 2010 / Published online: 8 January 2011

ABSTRACT

Purpose To determine the optimal measurement strategy for fingerprinting condensed phases of pharmaceutical systems using atomic pair distribution functions (PDFs) obtained from data collected using several types of x-ray diffraction instruments.

Methods PDFs of crystalline and amorphous-phase molecular systems derived from data accessible to copper-, molybdenum-, and silver-anode laboratory sources were compared to one another and synchrotron data using qualitative and quantitative methods.

Results We find that reliable fingerprinting is still possible using silver and molybdenum laboratory sources, but data from copper anode laboratory sources are unreliable for fingerprinting, yielding ambiguous and potentially incorrect results.

Conclusion The ambiguities make data measured using low energy x-rays unsuitable for fingerprinting active pharmaceutical ingredients 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 for this application.

T. Dykhne (✉) · S. J. L. Billinge
Department of Applied Physics and Applied Mathematics
Columbia University
500 W. 120th St., Room 200 Mudd, MC 4701
New York, New York 10027, USA
e-mail: td2218@columbia.edu

R. Taylor · A. Florence
Solid-State Research Group
Strathclyde Institute of Pharmacy and Biomedical Sciences
University of Strathclyde
Glasgow, UK G4 ONR

S. J. L. Billinge
Condensed Matter Physics and Materials Science Department
Brookhaven National Laboratory
Upton, New York 11973, USA

KEY WORDS amorphous · fingerprinting · nanocrystalline · pair distribution function (PDF) · x-ray diffraction (XRD)

ABBREVIATIONS

API	active pharmaceutical ingredient
PDF	pair distribution function
XRD	x-ray diffraction
TS-PDF	total scattering pair distribution function
CALS	copper-anode laboratory source
MALS	molybdenum-anode laboratory source
SALS	silver-anode laboratory source

INTRODUCTION

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 (1). 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 diffraction (XRD) fingerprinting methods are invalid for amorphous forms due to the lack of long-range atomic structure (2). Recently, atomic pair distribution function (PDF) methods have been suggested as an alternative approach for fingerprinting amorphous APIs (2,3) 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.

PDF analysis of x-ray powder diffraction data is a powerful approach for studying amorphous (4) and nanostructured materials (5,6). 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 XRD instruments (7–10), and also PDFs obtained from powder data collected on a high energy synchrotron source (2).

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: $Q = 4\pi \sin\theta/\lambda$. 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 (5). This is readily done using available software (11).

It has been shown previously 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 (2). This approach, referred to as total scattering pair distribution function (TS-PDF) 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 (7–10). Here we evaluate the relative ability for fingerprinting the inherently low-Q CALS-derived PDFs and TS-PDFs 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 (10). 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 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 $Q = 30$–$40\,\text{Å}^{-1}$ (5).

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 $\sim 8\,\text{Å}^{-1}$, equivalent to 160° 2θ, whereas for molybdenum-anode laboratory sources (MALS) and silver-anode laboratory sources (SALS), it is $\sim 16\,\text{Å}^{-1}$ and $\sim 22\,\text{Å}^{-1}$, respectively. Q-ranges up to $\sim 45\,\text{Å}^{-1}$ are accessible when measuring using high energy synchrotron radiation (12). In this paper 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.

Theoretical Background

The reduced structure function $F(Q) = Q[S(Q) - 1]$ 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)* (5). 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 function *S(Q)* is known: *S(Q)* oscillates around unity, and as $Q \rightarrow \infty$, *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 (5), 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 40–50Å^{-1} in most cases (5). In practice, even when data are collected over a range limited to Q_{max}=15−20Å^{-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 (5). 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 [11] 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 according to

$$G(r) = 4\pi r[\rho(r) - \rho_0] = \frac{2}{\pi}\int_{Q_{min}}^{Q_{max}} Q[S(Q) - 1]sinQr\ dQ,$$

where $\rho(r)$ is the microscopic pair density, and ρ_0 is the average number density (13). 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)*: $G_{exp}(r) = \frac{2}{\pi}\int_{Q_{min}}^{Q_{max}} Q[S_{exp}(Q) - 1]sinQr\ dQ$. 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 this paper, 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.

EXPERIMENTAL PROCEDURE

Samples

TS-PDF 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) (14) and the α (15) and γ (16) forms of IND (IND-α and IND-γ). The molecules are shown in Fig. 1.

Details of the sample preparation are described elsewhere (2). 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.

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 ($\lambda = 1.54056$Å) and Lynxeye position sensitive detector (PSD) in the range 2–40° 2θ (i.e. Q_{max} ca. 2.8Å^{-1}) with step size 0.016° 2θ, 10 s per step at 100 K (2). 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 (17). 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 and 8 times for a total collection of about 30 min for each data-set.

Fingerprinting

After PDFs were obtained from the data using the methods discussed above, we tested their efficacy at fingerprinting. This was done both qualitatively and quantitatively. For the qualitative analysis, we simply used a visual comparison of the plots of the PDFs. For the quantitative analysis, we used programs that quantify the similarity of curves. In particular, we used the commercially available PolySNAP program, which uses a modified version of the Spearman correlation parameter (18). In addition to PolySNAP, we used a home-written program that computes the Pearson product-momentum correlation (19), $r = \frac{1}{1-n}\sum_{i=0}^{n}\left(\frac{X_i-\overline{X}}{\sigma_x}\right)\left(\frac{Y_i-\overline{Y}}{\sigma_y}\right)$, where $\overline{X}$ and σ_x are the mean and standard deviation of a data set, respectively. Both correlation techniques create an $n \times n$ matrix that contains a correlation value r in the range -1 to 1 between each pair of n 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 studied the correlations between PDFs 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 neighbor carbon-carbon bonds at 1.4Å and 2.4Å, 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.

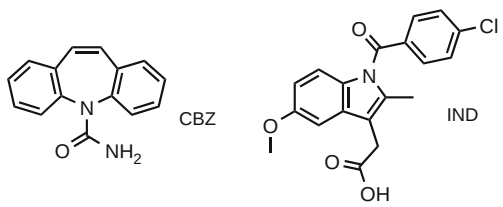

Fig. 1 Molecular structures of CBZ (*left*) and IND (*right*).

RESULTS

Uniqueness Tests

First we considered TS-PDF data from the synchrotron which has a $Q_{max} = 20.0$Å^{-1}. Figure 2(a) shows four 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 case, 62%, data were normalized from 12.5Å^{-1} to 20Å^{-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 Fig. 2(b). Again, the four PDFs are all from the same data-set obtained by providing PDFgetX2 with the same relative Q-ranges as in Fig. 2(a) for normalization. The four curves have peaks shifted and even features disappearing. For example, the peak at 12Å is wholly absent in two of the four PDFs. Correspondingly, the Pearson correlations are as low as 0.73 between these PDFs.

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 x-ray sources between these two extremes, including molybdenum anode laboratory sources (MALS), with a wavelength of 0.709Å (20), and silver anode laboratory sources

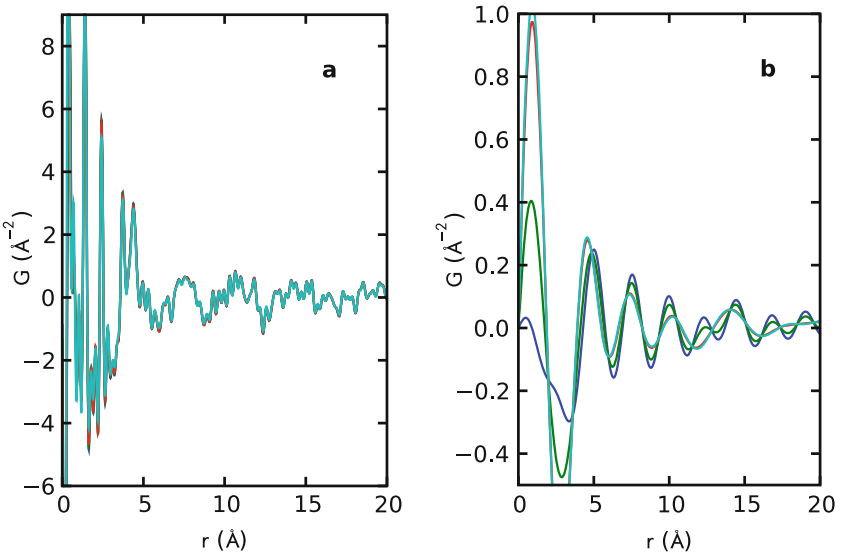

Fig. 2 PDFs of amorphous-phase CBZ with **(a)** $Q_{max}=20.0$ Å^{-1} and **(b)** $Q_{max}=2.8$Å^{-1} optimized over the same relative ranges of Q. In panel **(a)**, all of the PDFs match up to each other perfectly, while in panel **(b)**, the PDFs exhibit variations in peak positions, relative intensities, and broadness.

(SALS), with a wavelength of 0.559 Å (20). 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°, 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 12.5Å^{-1} and 15.9Å^{-1} for MALS and SALS, respectively.

To provide a uniform comparison of these instrument 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 12.5Å^{-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

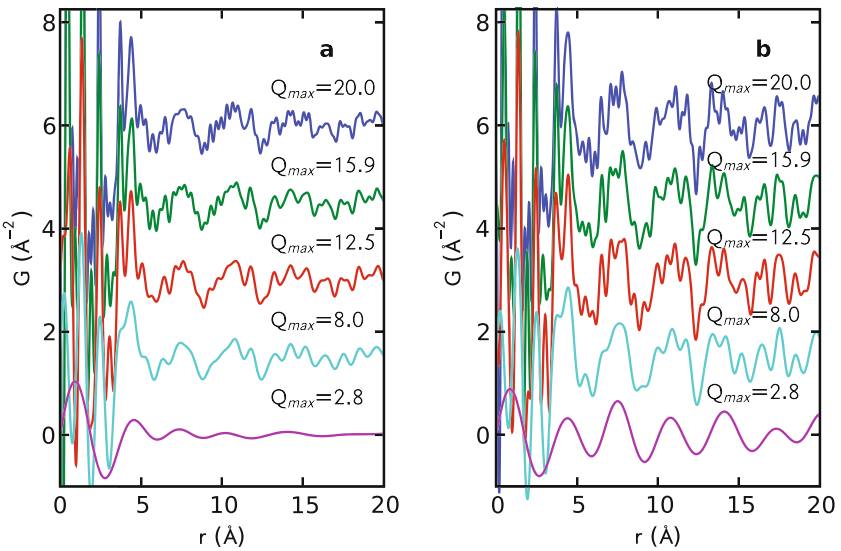

Fig. 3 PDFs of **(a)** amorphous-phase CBZ and **(b)** Form III CBZ with Q_{max} values corresponding to, in order from top to bottom, a typical synchrotron setup, SALS ($2\theta_{max}=90°$), MALS ($2\theta_{max}=90°$), CALS ($2\theta_{max}=160°$), and CALS ($2\theta_{max}=40°$).

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 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 Fig. 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 (2) 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 Table 1.

Table 1 Summary of Pearson Correlation Coefficients Between the PDFs Shown in Fig. 3. 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
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

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 $Q_{max}>12.5\text{Å}^{-1}$. By $Q_{max}=8.0\text{Å}^{-1}$, 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 (2). We also see a false-positive for IND-a compared with IND-α that is not seen in the higher resolution high-Q_{max} data. Finally, at $Q_{max}=2.8\text{Å}^{-1}$ the Pearson correlation does not 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.

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 $\delta r\sim 0.01\text{Å}$. However, it has recently been shown (Farrow et al., unpublished) 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., $\delta r\lesssim\frac{\pi}{Q_{max}}$. 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 $Q_{max}=20.0\text{Å}^{-1}$, 8.0Å^{-1}, and 2.8Å^{-1} are 0.157Å, 0.393Å, and 1.122Å, 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 many as the 40° $2\theta_{max}$ CALS data. For illustration, Fig. 4 contains the PDFs from Fig. 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.

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 XRD devices, reasonable Q_{max} values accessible to molybdenum-, 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

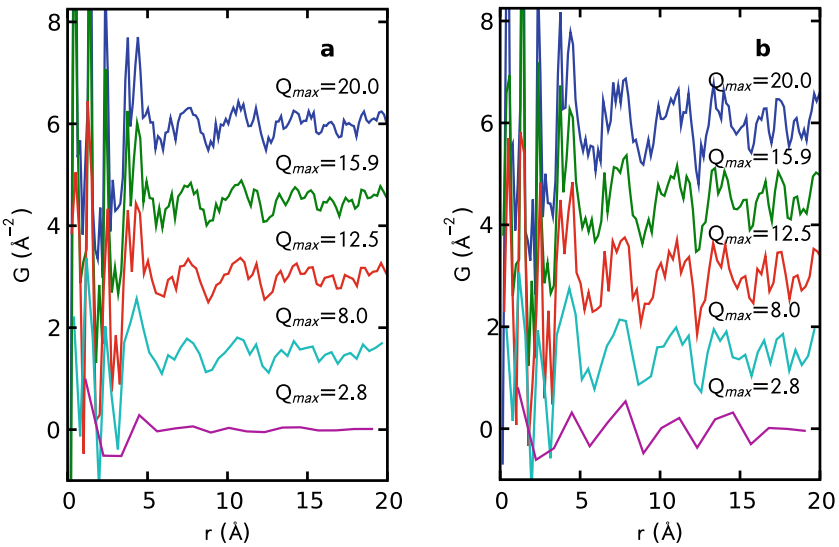

Fig. 4 PDFs of **(a)** Amorphous CBZ and **(b)** Form III CBZ from Fig. 3 plotted on the Nyquist grid. The lower Q_{max} PDFs have a coarser grid.

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 data-set resulted in a non-uniqueness in the resulting PDFs. This shows that data with a Q_{max} of 12.5Å^{-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.

ACKNOWLEDGMENTS

We would like to thank Pavol Juhás, Emil Božin, and Christopher Farrow for their help in collecting and analyzing the data. This work is supported through the US National Science Foundation through Grant DMR-0702940. The Advanced Photon Source is supported by the U.S. DOE, Office of Science, Office of Basic Energy Sciences, under contract No. W-31-109-Eng-38.

REFERENCES

1. Hancock BC, Parks M. What is the solubility advantage for amorphous pharmaceuticals? Pharm Res. 2000;17(4):397–404.
2. Billinge SJL, Dykhne T, Juhás P, Božin E, Taylor R, Florence AJ, *et al.* Characterisation of amorphous and nanocrystalline molecular materials by total scattering. CrystEngComm. 2010;12(5):1366–8.
3. Bates S, Zografi G, Engers D, Morris K, Crowley K, Newman A. Analysis of amorphous and nanocrystalline solids from their x-ray diffraction patterns. Pharm Res. 2006;23(10):2333–49.
4. Wright AC. Diffraction studies of glass structure: the first 70 years. Glass Phys Chem. 1998;24:148–79.
5. Egami T, Billinge SJL. Underneath the Bragg peaks: structural analysis of complex materials. Oxford: Pergamon; 2003.
6. Billinge SJL. In: Dinnebier RE, Billinge SJL, editors. Local structure from total scattering and atomic pair distribution function (pdf) analysis. Powder diffraction: theory and practice. London: Royal Society of Chemistry; 2008. p. 464–93.
7. Sheth AR, Bates S, Muller FX, Grant DJW. Polymorphism in piroxicam. Cryst Growth Des. 2004;4(6):1091–8.
8. Bates S, Kelly RC, Ivanisevic I, Schields P, Zografi G, Newman AW. Assessment of defects and amorphous structure produced in raffinose pentahydrate upon dehydration. J Pharm Sci. 2007;96(5):1418–33.
9. Sheth AR, Bates S, Muller FX, Grant DJW. Local structure in amorphous phases of piroxicam from powder x-ray diffractometry. Cryst Growth Des. 2005;5(2):571–8.
10. Newman A, Engers D, Bates S, Ivanisevic I, Kelly RC, Zografi G. Characterization of amorphous API:polymer mixtures using x-ray powder diffraction. J Pharm Sci. 2008;97(11):4840–56.
11. Qiu X, Thompson JW, Billinge SJL. PDFgetX2: a GUI driven program to obtain the pair distribution function from x-ray powder diffraction data. J Appl Crystallogr. 2004;37:678.
12. Petkov V, Billinge SJL, Shastri SD, Himmel B. Polyhedral units and network connectivity in calcium aluminosilicate glasses from high energy x-ray diffraction. Phys Rev Lett. 2000;85:3436.
13. Farrow CL, Billinge SJL. Relationship between the atomic pair distribution function and small angle scattering: implications for modeling of nanoparticles. Acta Crystallogr A. 2009;65(3):232–9.
14. Grzesiak AL, Lang M, Kim K, Matzger AJ. Comparison of the four anhydrous polymorphs of carbamazepine and the crystal structure of form i. J Pharm Sci. 2003;92(11):2260–71.
15. Chen X, Morris KR, Griesser UJ, Byrn SR, Stowell JG. Reactivity differences of indomethacin solid forms with ammonia gas. J Am Chem Soc. 2002;124(50):15012–9.
16. Cox PJ, Manson PL. Gamma-indomethacin at 120 k. Acta Crystallogr E. 2003;E59(7):o986–8.
17. Chupas PJ, Qiu X, Hanson JC, Lee PL, Grey CP, Billinge SJL. Rapid acquisition pair distribution function analysis (RA-PDF). J Appl Crystallogr. 2003;36:1342–7.
18. Barr G, Dong W, Gilmore CJ. Polysnap: a computer program for analysing high-throughput powder diffraction data. J Appl Crystallogr. 2004;37(4):658–64.
19. Myers JL, Well AD. Research design and statistical analysis. 3rd ed. Hillsdale: Lawrence Erlbaum Associates; 2010.
20. Deslattes RD, Ernest J, Kessler G, Indelicato P, de Billy L, Lindroth E, *et al.* X-ray transition energies: new approach to a comprehensive evaluation. Rev Mod Phys. 2003;75(1):35–99.