PDFgetX3: a rapid and highly automatable program for processing powder diffraction data into total scattering pair distribution functions

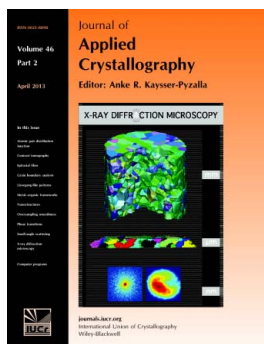**P. Juhás, T. Davis, C. L. Farrow and S. J. L. Billinge**

Journal of Applied Crystallography covers a wide range of crystallographic topics from the viewpoints of both techniques and theory. The journal presents papers on the application of crystallographic techniques and on the related apparatus and computer software. For many years, the *Journal of Applied Crystallography* has been the main vehicle for the publication of small-angle scattering papers and powder diffraction techniques. The journal is the primary place where crystallographic computer program information is published.

Crystallography Journals Online is available from **journals.iucr.org**

computer programs

PDFgetX3: a rapid and highly automatable program for processing powder diffraction data into total scattering pair distribution functions

P. Juhás,[a]‡ **T. Davis,**[a]‡ **C. L. Farrow**[a] **and S. J. L. Billinge**[a,b]*

[a]Department of Applied Physics and Applied Mathematics, Columbia University, New York, New York 10027, USA, and [b]Condensed Matter Physics and Materials Science Department, Brookhaven National Laboratory, Upton, New York 11973, USA. Correspondence e-mail: sb2896@columbia.edu

PDFgetX3 is a new software application for converting X-ray powder diffraction data to an atomic pair distribution function (PDF). *PDFgetX3* has been designed for ease of use, speed and automated operation. The software can readily process hundreds of X-ray patterns within a few seconds and is thus useful for high-throughput PDF studies that measure numerous data sets as a function of time, temperature or other environmental parameters. In comparison to the preceding programs, *PDFgetX3* requires fewer inputs and less user experience and it can be readily adopted by novice users. The live-plotting interactive feature allows the user to assess the effects of calculation parameters and select their optimum values. *PDFgetX3* uses an *ad hoc* data correction method, where the slowly changing structure-independent signal is filtered out to obtain coherent X-ray intensities that contain structure information. The output from *PDFgetX3* has been verified by processing experimental PDFs from inorganic, organic and nanosized samples and comparing them with their counterparts from a previous established software. In spite of the different algorithm, the obtained PDFs were nearly identical and yielded highly similar results when used in structure refinement. *PDFgetX3* is written in the Python language and features a well documented reusable code base. The software can be used either as a standalone application or as a library of PDF processing functions that can be called from other Python scripts. The software is free for open academic research but requires paid license for commercial use.

1. Introduction

With the recent increased interest in producing and exploiting nanostructured materials, it is necessary to expand the methods that go beyond crystallography (Billinge, 2010) for characterizing the atomic scale structure of such materials. In recent years, total scattering and atomic pair distribution function (PDF) analysis (Egami & Billinge, 2013) has emerged as a popular and powerful tool for this purpose (Billinge & Kanatzidis, 2004; Billinge, 2008; Young & Goodwin, 2011). To satisfy this demand a number of X-ray and neutron beamlines dedicated to or optimized for such measurements have emerged (Egami & Billinge, 2013), and manufacturers of laboratory X-ray sources are also beginning to market instruments for this kind of measurement. Especially with the use of two-dimensional detectors, modern beamlines are yielding total scattering data at unprecedented rates, allowing detailed parametric and time-resolved total scattering studies to be carried out in special environments (Chupas *et al.*, 2004, 2007; Jensen *et al.*, 2012; Redmond *et al.*, 2012). A bottleneck in further growth of the method is now the lack of robust and automatable software for creating PDFs from the raw data, currently a computationally and user-intensive process (Egami & Billinge, 2013).

This shortcoming can be illustrated by considering one of the most widely used software programs for this purpose, *PDFgetX2* (Qiu *et al.*, 2004). The program offers users a great deal of flexibility and control in choosing exactly which corrections to apply to X-ray scattering intensities in order to convert them to PDFs. However, because of the myriad of options available to users as well as the esoteric nature of many of the corrections (Egami & Billinge, 2013), PDF generation requires considerable user input and expertise in arcane details of the technique. Although the software has a graphical user interface, it is a time-consuming process to carry out the corrections, with many possibilities for input errors, and the process cannot be easily automated for high throughput of many data sets.

In this paper, we describe a new software program, *PDFgetX3*, which implements an *ad hoc* data reduction algorithm (Billinge & Farrow, 2012) that requires little user input, generates PDFs in a fraction of a second and can be straightforwardly automated to batch process thousands of PDFs. Here we show that in the physically relevant region of the PDF this software produces quantitatively accurate PDFs that are the same as those obtained using *PDFgetX2* for the considered cases, and which yield refined structural parameters that are also indistinguishable from those refined from *PDFgetX2*-determined PDFs.

The intensities measured in a total scattering experiment, $I_m(Q)$, can be expressed as (Billinge & Farrow, 2012)

‡ These authors contributed equally to this work.

$$I_m(Q) = a(Q) I_c(Q) + b(Q), \tag{1}$$

where $I_c(Q)$ is the intensity from coherent scattering events, which contains all of the structural information about the sample, and $a(Q)$ and $b(Q)$ are multiplicative and additive corrections to the measured intensity, which do not contain structural information (Billinge & Farrow, 2012). Examples of the additive contributions are incoherent Compton scattering and background scattering from the sample container. Examples of the multiplicative contributions are sample self-absorption and polarization of the X-ray beam. The approach used by *PDFgetX2*, and other PDF data analysis programs, is to apply known corrections to $I_m(Q)$ to obtain the coherent scattering, $I_c(Q)$, which is transformed into the structure function, $S(Q)$ according to

$$S(Q) = \frac{I_c(Q) - \langle f(Q)^2 \rangle + \langle f(Q) \rangle^2}{\langle f(Q) \rangle^2}. \tag{2}$$

Here $f(Q)$ is the atomic scattering factor and the angle brackets indicate an average over all the atom types in the sample. For the neutron case the atomic scattering factors are replaced by coherent neutron scattering lengths, b, in this equation.

$S(Q)$ is Fourier transformed into the PDF, $G(r)$, according to (Farrow & Billinge, 2009)

$$G(r) = (2/\pi) \int_{Q_{min}}^{Q_{max}} Q[S(Q) - 1] \sin Qr \, dQ = (2/\pi) \int_{Q_{min}}^{Q_{max}} F(Q) \sin Qr \, dQ, \tag{3}$$

where the quantity $F(Q) = Q[S(Q) - 1]$ is the reduced structure function (Warren, 1990). The many corrections required are discussed in detail by Egami & Billinge (2013, ch. 5), including background subtraction, polarization, self-absorption, multiple scattering and Compton scattering, among many others, and these are implemented in *PDFgetX2* (Qiu *et al.*, 2004) and other similar programs (Petkov, 1989; Petkov & Danev, 1998; Jeong *et al.*, 2001; Soper & Barney, 2011).

It has recently been pointed out (Billinge & Farrow, 2012) that sufficient information is known about the general behavior of the correction terms in equation (1), and about the asymptotic behavior of the resulting $F(Q)$ function, that it may be possible to determine $a(Q)$ and $b(Q)$ through an *ad hoc* approach, where they are parameterized and the parameters are varied using a regression method in such a way as to yield an accurate $F(Q)$ function. Here we describe an algorithm for doing this, as well as a software implementation, *PDFgetX3*, and we show that, indeed, this approach yields PDFs that are not significantly different from those obtained using *PDFgetX2*. The program is fast, easy to use and highly automatable.

The method was developed initially for analyzing rapid acquisition PDF (RAPDF) data from two-dimensional detectors, though we show below that it is not limited to this application. It is assumed that the two-dimensional data have been correctly azimuthally integrated, and multiple frames summed or averaged, to obtain a one-dimensional intensity *versus* Q or intensity *versus* 2θ profile. A number of integration programs exist for this purpose, for example *Fit2D* (Hammersley *et al.*, 1996).

The algorithm (Billinge *et al.*, 2011) starts with raw intensity data measured *versus* scattering angle 2θ. At first, the angle is converted to scattering vector magnitude Q and the data are re-sampled to an equidistant Q grid, which is suitable for a fast Fourier transformation at a later step and also ensures constant weights in a Q-dependent fitting. Note that resampling introduces error correlations between points, which can be minimized if the data are azimuthally integrated from two dimensions directly onto a constant-Q grid (Yang & Billinge, 2013). The background intensities from an empty container are

then resampled to the same Q grid and subtracted from the sample data. This yields raw intensities from the specimen only, which are not normalized per incident intensity or per the number of scatterers. The structure function $S(Q)$ should oscillate around and then approach unity as Q tends to infinity, which in practice is about $Q = 25$ Å^{-1}. This means that the difference

$$S(Q) - 1 = I(Q)/\langle f \rangle^2 - \langle f^2 \rangle/\langle f \rangle^2 \tag{4}$$

must oscillate around zero and the normalized intensity $I(Q)/\langle f \rangle^2$ must be close to the normal scattering factor $\langle f^2 \rangle/\langle f \rangle^2$ for any Q. The raw sample intensities are therefore rescaled by a least-squares procedure to approach the normal scattering factor curve. A physically correct scattering function $S(Q)$ should also display proper asymptotic behavior as a derived function $F(Q) = Q[S(Q) - 1]$, which should oscillate around zero and approach it with increasing Q. The *PDFgetX3* algorithm is based on an assumption that the experimental function $S_m(Q)$ deviates from the correct value by a slowly changing additive factor $\beta_S(Q)$ such that

$$S_m(Q) - 1 = S(Q) - 1 + \beta_S(Q). \tag{5}$$

The derived function $F_m(Q)$ is then

$$F_m(Q) = Q[S(Q) - 1 + \beta_S(Q)] = F(Q) + Q\beta_S(Q). \tag{6}$$

Because the correct function $F(Q)$ oscillates around zero, the error term $\beta_S(Q)$ produces a slowly changing, Q-increasing background in $F_m(Q)$. The *PDFgetX3* algorithm estimates the background by modeling the $\beta_S(Q)$ function as an nth-degree polynomial $P_n(Q)$, which is then fitted as $QP_n(Q)$ to the $F_m(Q)$ function. This assumes that the changes in the actual zero-average $F(Q)$ signal are too rapid to influence a low-degree polynomial fit, which in effect yields the same result as if we were fitting the background term in $F_m(Q)$ alone. The corrected function $F_c(Q)$ is afterwards obtained by subtracting the polynomial fit:

$$F_c(Q) = F_m(Q) - QP_n(Q). \tag{7}$$

The function $F_c(Q)$ shows the correct asymptotic behavior, with $F \to 0$ for large Q values. Finally, the $F_c(Q)$ signal is converted to $G(r)$ using the fast Fourier transformation as per equation (3).

Since the fitted polynomial is an approximation to the actual error term $\beta_S(Q)$, the corrected function $F_c(Q)$ still deviates from the ideal F by

$$\Delta F(Q) = F_c(Q) - F(Q) = Q\beta_S(Q) - QP_n(Q), \tag{8}$$

and the difference introduces an error signal $\Delta G(r)$ in the obtained PDF. The function $QP_n(Q)$ is an $(n + 1)$th-degree polynomial approximation to the $Q\beta_S(Q)$ function on a fit interval running from zero to $Q_{maxinst}$; therefore we can assume that the $\Delta F(Q)$ difference has $(n + 1)$ roots that are approximately equidistant between $Q = 0$ Å^{-1} and $Q_{maxinst}$. The difference function $\Delta F(Q)$ switches between positive and negative values at each root, which roughly corresponds to oscillations with a half-period of $Q_{maxinst}/n$. Assuming this to be the maximum Q 'frequency' in the difference signal $\Delta F(Q)$, the Fourier transformation would introduce nonphysical signal $\Delta G(r)$ extending up to

$$r_{poly} = \pi n/Q_{maxinst}. \tag{9}$$

For typical RAPDF experimental data, the *PDFgetX3* program uses an eighth-degree polynomial correction with $Q_{maxinst} = 28$ Å^{-1}, which yields $r_{poly} = 0.9$ Å. Assuming there are no higher-frequency aberrations in the data themselves, the error signal $\Delta G(r)$ arising from the polynomial data correction is thus present only for lengths smaller than $r \simeq 0.9$ Å, *i.e.* in a region below the shortest bond lengths in

most materials. Furthermore, the polynomial fit cannot accidentally remove real structural signals from the experimental intensity provided the value of r_{poly} is chosen to be below the nearest neighbor bond distance in the material.

Under some experimental conditions, such as from lower-energy X-ray sources or where the experimental take-off angle is very limited, the instrument Q range $Q_{maxinst}$ is much smaller and may increase the error extent r_{poly} to physically meaningful distances. In such cases, the degree of the correction polynomial n needs to be reduced to avoid overcorrecting the measured data and to keep the value of r_{poly} small. The *PDFgetX3* procedure uses equation (9) in reverse, and for a fixed value of the error extent r_{poly} and instrument range $Q_{maxinst}$ it obtains the degree of the correction polynomial as

$$n_r = r_{poly}Q_{maxinst}/\pi. \quad (10)$$

This estimate of the polynomial degree n_r is almost never an integer, and rounding it to an integer would introduce abrupt changes in the PDF at the half-integer values. We would prefer the PDF to respond smoothly to the r_{poly} and $Q_{maxinst}$ parameters. To simulate a polynomial fit at an arbitrary floating-point degree, the correction polynomial is therefore refined twice, for an integer floor and ceiling of n_r, and the two fits are then averaged with the weights given by the distance of n_r from its integer bounds.
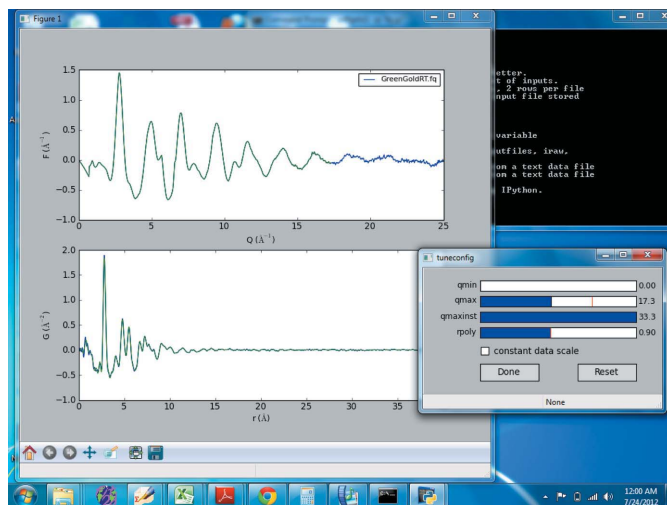
2. Program availability and operation

PDFgetX3 is written in the Python programming language (http://www.python.org). To run, it requires Python 2.6 or later with the NumPy (http://www.numpy.org/) and Matplotlib (http://matplotlib.org/) libraries installed (note that Python 3.0 is not currently supported). *PDFgetX3* has been tested on the Windows, Linux and Mac operating systems. Information on the installation and operation of the software can be found at http://www.diffpy.org. The command-line version is free for researchers conducting open academic research, but other uses require a paid license. A version with a

Figure 1
Screenshot of *PDFgetX3* in interactive mode. The user selected to plot $F(Q)$ and $G(r)$. These plots are updated in real time as the user uses the mouse to move the sliders. There are four sliders in this example, for Q_{min}, Q_{max}, $Q_{maxinst}$ and r_{poly}. The first two are self-explanatory. $Q_{maxinst}$ varies the range over which the correction polynomial is fitted and r_{poly} places an upper bound on the frequency of information that the *ad hoc* procedure can remove by fitting (for details, see Billinge & Farrow, 2012). If the user wishes to subtract a background signal the background scale will also appear as a slider option.

graphical user interface and an online version are also under development. Information can be found at http://www.diffpy.org.

Because the corrections are *ad hoc*, only minimal information needs to be supplied by the user, and this can be contained in a configuration file or specified as a command-line argument. In the current implementation the program reads data that are stored in a multi-column text file with the independent variable, Q or 2θ, in the first column and the measured intensity in the second. If the uncertainties on points in the data are known, these may be placed in subsequent columns. The filename for the input file, and a measured background file if one wants to subtract it, must be specified and if the independent variable is 2θ then an X-ray wavelength must also be specified. The approximate composition of the sample is also specified so that the $f(Q)$ averages may be computed accurately. A background scale parameter, and Q_{max} to be used in the Fourier transform, should be specified, though these have default values and the program works when they are not provided. The optimal values of some of these parameters may not be known *a priori*, and the program may be run in an interactive mode where various tuning parameters may be varied by sliding a slider with the resulting PDFs updating in real time in a plot window. In this way a user may quickly find the optimal Q_{max} and background scale values, which are fed back to the program. It takes only a few microseconds to complete the corrections on the raw data, and so the plots update in real time as the user adjusts the slider. Some other parameters may also be controlled by the user to obtain the desired output, as described in the manual.

The program has a powerful Python-based command-line interpreter capability, for example, allowing templates to be used for multiple files that have the same filename root but which differ in some way in the name, for example, by run number. This makes the automation of data reduction of many hundreds or thousands of data sets straightforward. The program is also written with a well documented application programming interface so that programmers can access the functionality of the engine within home-written Python scripts of arbitrary complexity. A screenshot of the program working in interactive mode is shown in Fig. 1.

3. Comparison of *PDFgetX3* and *PDFgetX2* PDFs

PDFs have been determined with *PDFgetX3* for a number of representative samples and compared with those determined from the same data using *PDFgetX2*. The resulting PDFs are compared by plotting them on top of each other. Where possible, structural models have been refined to both PDFs allowing a direct comparison of fit quality and the values of refined structural parameters from each PDF. The examples include 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 derived from the same raw data and, where appropriate, 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 data were collected at high-energy synchrotron instruments using the RAPDF mode (Chupas *et al.*, 2003), where data are collected on a two-dimensional detector and azimuthally integrated to obtain one-dimensional data sets; however, the synchrotron is not a requirement. *PDFgetX3* can handle
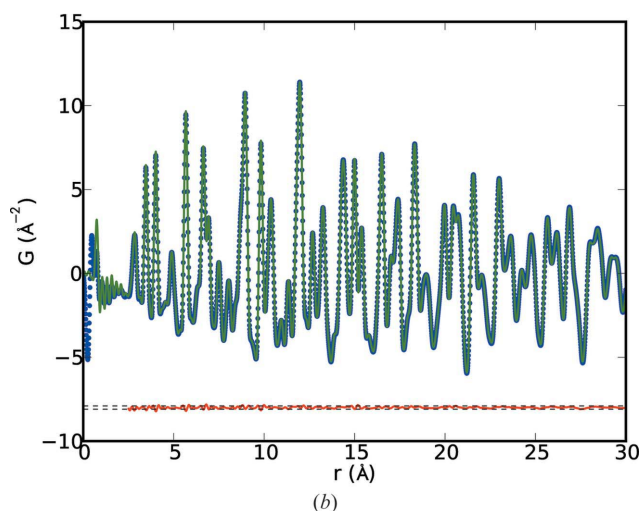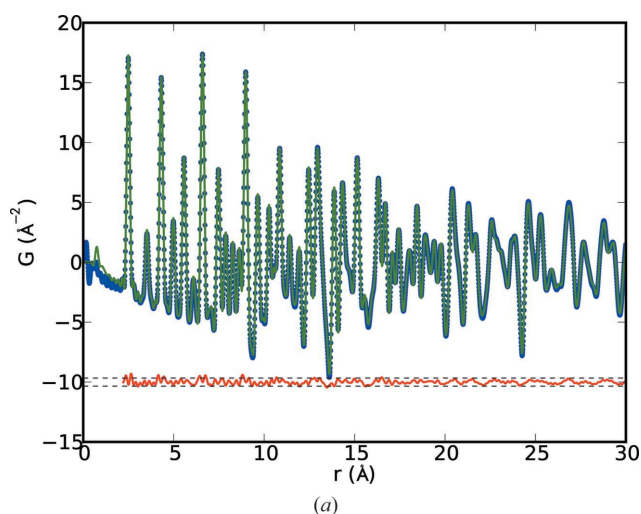
Table 1
Comparison of the parameters refined in fitting the Ni model to the PDFs.

Parameter	PDFgetX2	PDFgetX3
Q_{damp} (Å^{-1})	0.0554	0.0570
$a = b = c$ (Å)	3.5239	3.5237
δ_2 (Å^2)	2.52	2.71
U_{iso} (Å^2)	0.00612	0.00564
R_w	0.0796	0.0821

Table 2
Comparison of the parameters refined in fitting the $BaTiO_3$ model to the PDFs.

Parameter	PDFgetX2	PDFgetX3
Q_{damp} (Å^{-1})	0.0485	0.0491
$a = b$ (Å)	3.9952	3.9952
c (Å)	4.0399	4.0398
δ_2 (Å^2)	4.32	4.37
$U_{11,Ba} = U_{22,Ba}$ (Å^2)	0.00516	0.00494
$U_{33,Ba}$ (Å^2)	0.00454	0.00432
$U_{11,Ti} = U_{22,Ti}$ (Å^2)	0.00874	0.00839
$U_{33,Ti}$ (Å^2)	0.0125	0.0121
$U_{11,O} = U_{22,O}$ (Å^2)	0.0113	0.0103
$U_{33,O}$ (Å^2)	0.0927	0.0953
R_w	0.118	0.121

data from laboratory-based X-ray powder diffraction instruments and synchrotron data collected in point-by-point mode such as from high-resolution diffractometers like ID31 at ESRF. To show this the comparison is done for γ-Al_2O_3 data that were collected with a PANalytical laboratory-based silver anode diffractometer.

In general, as we will see in the following examples, we find that the PDFs derived by the different programs look somewhat different from one another at r values lower than r_{poly}. However, in the physically meaningful range beyond the first nearest neighbor peak the PDFs look almost exactly the same. In the plots the PDFs obtained by the different methods have been rescaled by a constant

Figure 2
PDFs of (*a*) nickel and (*b*) barium titanate produced by *PDFgetX2* (dots; blue in the electronic version of the journal) and *PDFgetX3* (line; green in the electronic version) with $Q_{max} = 26.0 \text{ Å}^{-1}$ in both cases. The difference curve (offset) is shown below (in red in the electronic version). 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).

such that the nearest neighbor peak is the same height for PDFs on the same plot. The *ad hoc* approach (Billinge & Farrow, 2012) does not result in absolutely normalized data and normalization must be carried out by other methods. A constant scale offset has been shown not to affect the structural information in the PDF (Peterson *et al.*, 2003) when it is modeled with a scale factor variable, since the relative scaling of peaks to one another within the same PDF is preserved.

Models were fitted to the PDFs by refining a variety of parameters as appropriate, such as lattice parameters, atomic positions and displacement parameters, using the program *PDFgui* (Farrow *et al.*, 2007). In each case, we compare the R_w value as well as the values of the refined parameters from the PDF obtained using *PDFgetX2* and *PDFgetX3*.

3.1. Nickel and barium titanate

First we look at pure nickel (Ni) and barium titanate ($BaTiO_3$) (Fig. 2).

Both compounds diffract strongly, making data corrections less challenging. Fig. 2(*a*) shows the two PDFs of nickel plotted on top of one another. In all the figures the PDF from *PDFgetX2* is shown as (blue) dots and the PDF from *PDFgetX3* as a (green) solid line (shown in color in the electronic version of the journal). Here $Q_{max} = 26.0 \text{ Å}^{-1}$ in both cases. The difference curve between the two PDFs is plotted offset below (in red) with dashed lines plotted at $\pm 2\sigma$ as guides to the eye; here σ is the standard deviation of the difference computed over the range above r_{poly}. We see only very small differences between the two PDFs after the Ni–Ni nearest neighbor peak (at $r = 2.2 \text{ Å}$). We see the same behavior in Fig. 2(*b*) with barium titanate.

The refined parameters from model fits are reproduced in Table 1. In the case of Ni only a few parameters may be varied because of the simplicity of the structure. Overall, we see very good agreement between most of the parameters and the R_w values. There are more structural parameters that may be varied in the $BaTiO_3$ case (Megaw, 1962), as reproduced in Table 2. The parameters still agree very well with one another and the quality of the fit as measured by R_w is the same. We do not report estimated standard deviations on the refined parameters since we do not have reliable error estimates for the data themselves. The enhancement of *PDFgetX3* to propagate uncertainties on the data and the problem of extracting reliable uncertainties on integrated powder data from two-dimensional integrating detectors are being addressed (Yang & Billinge, 2013), so we expect this problem to be resolved in the near future.

3.2. Nanocrystalline γ-alumina

Next, we investigate γ-alumina (Al_2O_3) using X-rays from a silver anode diffractometer ($\lambda = 0.56 \text{ Å}$). The γ phase of Al_2O_3 has a local

computer programs

Table 3
Comparison of the parameters refined in fitting the γ-Al$_2$O$_3$ model to the PDFs.

Parameter	PDFgetX2	PDFgetX3
Q_{damp} (Å^{-1})	0.0770	0.0808
a (Å)	3.3943	3.3941
b (Å)	2.7796	2.7802
c (Å)	7.0419	7.0395
δ_2 (Å^2)	1.13	0.991
$U_{\mathrm{iso,O}}$ (Å^2)	0.0126	0.0123
$U_{\mathrm{iso,Al}}$ (Å^2)	0.0148	0.0145
R_{w}	0.164	0.166

nanocrystalline structure that is different from the longer-range structures (Paglia *et al.*, 2006). For this reason, a new structure model was developed for the local structure of γ-Al$_2$O$_3$ up to $r = 8$ Å (Paglia *et al.*, 2006; ICSD 173014; Inorganic Crystal Structure Database, FIZ Karlsruhe, Germany, and NIST, Gaithersburg, MD, USA). Fig. 3 shows the PDFs of γ-Al$_2$O$_3$ from *PDFgetX2* and *PDFgetX3* over this region. We see very good agreement between the PDFs. In fact, the *PDFgetX3* PDF looks better at low r values.
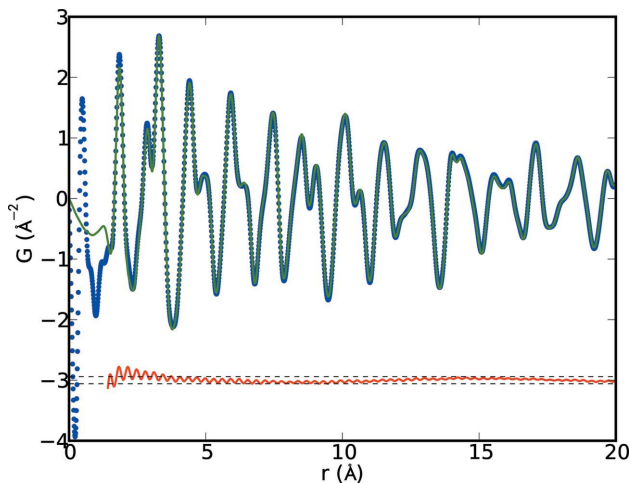
Refined parameters are given in Table 3. Unlike in previous cases where we tried to use a large r range for our refinement, in this case we refined only over the range $r = 1.5$–8 Å because the model only applies over this range. For this reason, we wanted to refine few parameters (this is why U_{iso} was used rather than anisotropic displacement parameters). Regardless, we see very good agreement between the fit results.

3.3. Cadmium selenide nanoparticles

We now turn our attention to a more challenging class of materials: nanoparticles which tend to be weakly scattering and more disordered. In Fig. 4 we show PDFs of three samples of cadmium selenide (CdSe) taken from data published by Masadeh *et al.* (2007). The bulk CdSe in Fig. 4(*a*) is included for completeness. The nanoparticles in Figs. 4(*b*) and 4(*c*) were calculated to have diameters of 37 and 22 Å, respectively (Masadeh *et al.*, 2007). We see that in all three panels of Fig. 4 the PDFs from the two programs are almost identical. It was challenging to obtain the PDFs from *PDFgetX2*, requiring considerable care, user intervention and parameter tuning. In the case of

Figure 3
PDFs of γ-Al$_2$O$_3$ produced by *PDFgetX2* (dots; blue in the electronic version of the journal) and *PDFgetX3* (line; green in the electronic version) with $Q_{\mathrm{max}} = 20.5$ Å^{-1} in both cases. The difference curve (offset) is shown below (in red in the electronic version). 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).

Figure 4
PDFs of (*a*) bulk CdSe and (*b*) 37 Å and (*c*) 22 Å CdSe nanoparticles produced by *PDFgetX2* (dots; blue in the electronic version of the journal) and *PDFgetX3* (line; green in the electronic version) with $Q_{\mathrm{max}} = 18.0$ Å^{-1} in all cases. The difference curve (offset) is shown below (in red in the electronic version). 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
Comparison of the parameters refined in fitting the CdSe wurtzite model to the PDFs.

| Parameter	PDFgetX2	PDFgetX3
Bulk		
Q_{damp} (Å^{-1})	0.0593	0.0599
$a = b$ (Å)	4.2996	4.2996
c (Å)	7.0112	7.0113
δ_2 (Å^2)	3.21	3.26
$U_{11,Cd} = U_{22,Cd}$ (Å^2)	0.0156	0.0155
$U_{33,Cd}$ (Å^2)	0.0143	0.0141
$U_{11,Se} = U_{22,Se}$ (Å^2)	0.0129	0.0128
$U_{33,Se}$ (Å^2)	0.0581	0.0575
R_w	0.114	0.104
37 Å nanoparticle		
$a = b$ (Å)	4.2956	4.2961
c (Å)	7.0068	7.0075
δ_2 (Å^2)	4.66	4.74
$U_{11,Cd} = U_{22,Cd}$ (Å^2)	0.0225	0.0221
$U_{33,Cd}$ (Å^2)	0.0302	0.0302
$U_{11,Se} = U_{22,Se}$ (Å^2)	0.0120	0.0118
$U_{33,Se}$ (Å^2)	0.199	0.194
Particle diameter (Å)	36.39	35.34
R_w	0.194	0.173
22 Å nanoparticle		
$a = b$ (Å)	4.2940	4.2948
c (Å)	6.8567	6.8633
δ_2 (Å^2)	4.97	5.20
$U_{11,Cd} = U_{22,Cd}$ (Å^2)	0.0433	0.0415
$U_{33,Cd}$ (Å^2)	0.0403	0.0409
$U_{11,Se} = U_{22,Se}$ (Å^2)	0.0199	0.0203
$U_{33,Se}$ (Å^2)	0.233	0.221
Particle diameter (Å)	23.13	23.35
R_w	0.262	0.265

PDFgetX3 the PDFs shown were produced with no more effort than the Ni and BaTiO$_3$ PDFs shown above. The low-*r* region looks slightly different in the two 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 Fig. 4(*c*) the *PDFgetX3* PDF looks cleaner than the *PDFgetX2* PDF.

Table 4 contains the refined parameters for the CdSe samples compared to a model based on wurtzite (Wyckoff, 1967). Again we see very good agreement between all parameters determined from the *PDFgetX2* and *PDFgetX3* PDFs, and the residual, R_w, is comparable in the two pairs of PDFs.

3.4. Pharmaceuticals

The final class of materials that we tested are organic pharmaceutical compounds. These materials can be crystalline, as we see in Figs. 5(*a*) and 5(*b*), nanostructured as in Fig. 5(*c*), or amorphous. These materials tend to have relatively complicated crystal structures that are made up of mostly light organic elements such as hydrogen, carbon, nitrogen and oxygen. These are weak X-ray scatterers whose electrons can delocalize with respect to the atomic nucleus when engaged in chemical bonds, and in addition the organic molecules may have 'floppy' parts, for example, with torsion degrees of freedom. As a result, the structure signal in the X-ray data from organic materials is relatively weak and even crystal phase pharmaceutical compounds require quite a bit of tinkering in *PDFgetX2* to produce a good PDF.

In the examples here we consider three polymorphs of the drug carbamazepine (CBZ): crystalline CBZ form I and form III as well as a melt-quenched carbamazepine that turned out to be nanocrystalline (Billinge *et al.*, 2010; Dykhne *et al.*, 2011). As with the nanoparticles in Fig. 4, the PDFs in Fig. 5 derived by *PDFgetX2* have relatively large fluctuations from imperfect corrections at low *r*. This is common for weakly scattering samples. However, these were the best PDFs that
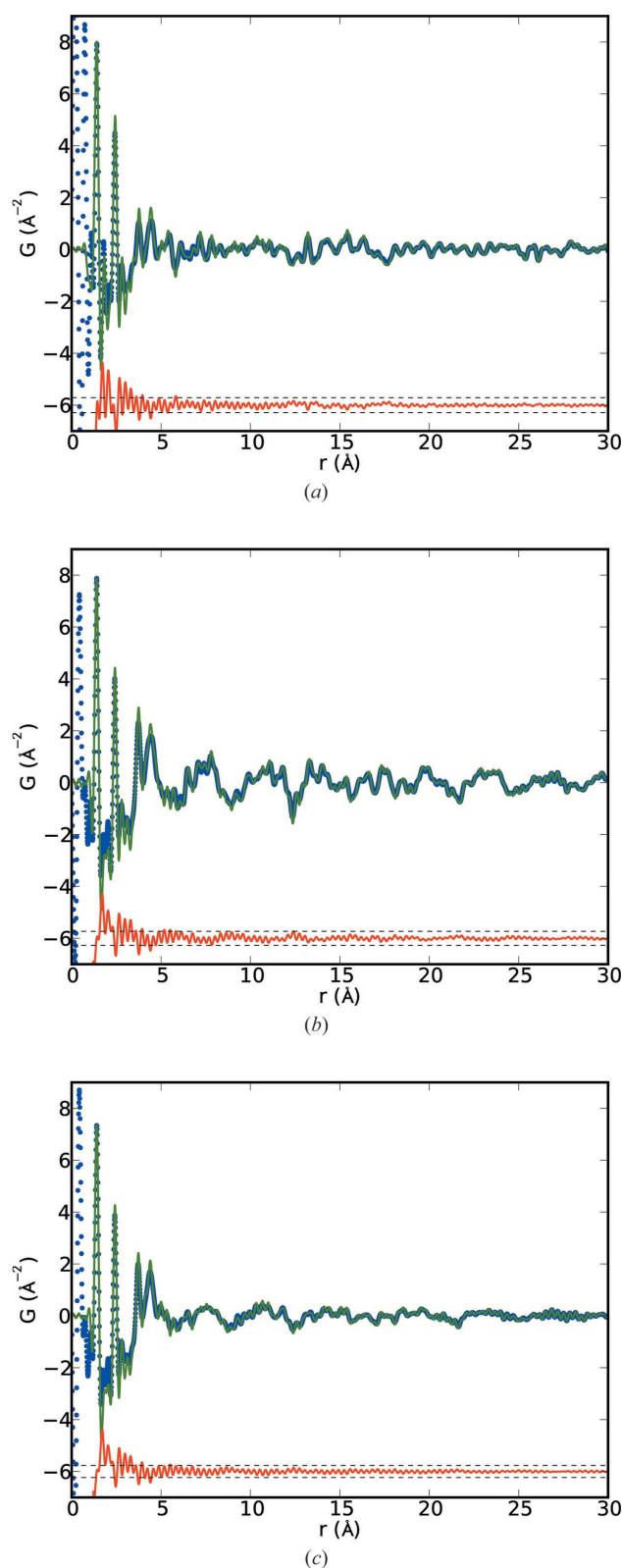
Figure 5
PDFs of (*a*) CBZ-I, (*b*) CBZ-III and (*c*) nanostructured CBZ produced by *PDFgetX2* (dots; blue in the electronic version of the journal) and *PDFgetX3* (line; green in the electronic version) with $Q_{max} = 20.0$ Å^{-1} in all cases. The difference curve (offset) is shown below (in red in the electronic version). 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).

could be obtained using *PDFgetX2* at the time of publication. The PDFs from *PDFgetX3* are highly similar in the physically meaningful region above the nearest neighbor separation (C—C bond at 1.4 Å), with the added benefit that they appear to be cleaner in the unphysical low-*r* region. This is an advantage because termination ripples from large features in the unphysical region may propagate into the physically meaningful region of the PDF.

We did not fit these PDFs to models because new modeling tools need to be developed for this class of materials.

4. Summary

We have described and demonstrated an implementation of the *ad hoc* data reduction protocol described by Billinge & Farrow (2012) in a new Python-based software program *PDFgetX3*. PDFs obtained using this method have been compared with PDFs obtained using *PDFgetX2*, an established program for producing PDFs, and are found to be highly similar. Models fitted to the *PDFgetX2* and *PDFgetX3* PDFs yield refined parameters that are correspondingly similar. The program has been tested on a range of samples, from strongly scattering inorganic crystalline powders such as nickel and $BaTiO_3$ to weakly scattering low atomic number pharmaceutical compounds. The program is easy to use compared to *PDFgetX2* and rapid, giving PDFs in real time as parameters such as background scale or Q_{max} are varied. The program should be applicable to most PDF studies (though it does not yield data on an absolute scale), but will prove to be especially useful for high-throughput studies such as parametric or time-resolved experiments. More information about the program is available at http://www.diffpy.org.

References

Billinge, S. J. L. (2010). *Physics*, **3**, 25.
Billinge, S. J. L., Dykhne, T., Juhás, P., Božin, E., Taylor, R., Florence, A. J. & Shankland, K. (2010). *CrystEngComm*, **12**, 1366–1368.
Billinge, S. J. L. & Farrow, C. L. (2012). *arXiv*:1211.4284 [cond-mat.mtrl-sci].
Billinge, S. J. L., Farrow, C. & Juhás, P. (2011). Provisional Patent 61 563 258.
Billinge, S. J. L. & Kanatzidis, M. G. (2004). *Chem. Commun.* pp. 749–760.
Chupas, P. J., Chapman, K. W. & Lee, P. L. (2007). *J. Appl. Cryst.* **40**, 463–470.
Chupas, P. J., Chaudhuri, S., Hanson, J. C., Qiu, X., Lee, P. L., Shastri, S. D., Billinge, S. J. & Grey, C. P. (2004). *J. Am. Chem. Soc.* **126**, 4756–4757.
Chupas, P. J., Qiu, X., Hanson, J. C., Lee, P. L., Grey, C. P. & Billinge, S. J. L. (2003). *J. Appl. Cryst.* **36**, 1342–1347.
Dykhne, T., Taylor, R., Florence, A. & Billinge, S. J. (2011). *Pharm. Res.* **28**, 1041–1048.
Egami, T. & Billinge, S. J. L. (2013). *Underneath the Bragg Peaks: Structural Analysis of Complex Materials*, 2nd ed. Amsterdam: Elsevier.
Farrow, C. L. & Billinge, S. J. L. (2009). *Acta Cryst.* A**65**, 232–239.
Farrow, C. L., Juhás, P., Liu, J., Bryndin, D., Božin, E. S., Bloch, J., Proffen, T. & Billinge, S. J. L. (2007). *J. Phys. Condens. Mater.* **19**, 335219.
Hammersley, A. P., Svenson, S. O., Hanfland, M. & Hauserman, D. (1996). *High Pressure Res.* **14**, 235–248.
Jensen, K. M., Christensen, M., Juhas, P., Tyrsted, C., Bøjesen, E. D., Lock, N., Billinge, S. J. & Iversen, B. B. (2012). *J. Am. Chem. Soc.* **134**, 6785–6792.
Jeong, I.-K., Thompson, J., Proffen, Th., Turner, A. M. P. & Billinge, S. J. L. (2001). *J. Appl. Cryst.* **34**, 536.
Masadeh, A. S., Božin, E. S., Farrow, C. L., Paglia, G., Juhás, P., Karkamkar, A., Kanatzidis, M. G. & Billinge, S. J. L. (2007). *Phys. Rev. B*, **76**, 115413.
Megaw, H. D. (1962). *Acta Cryst.* **15**, 972–973.
Paglia, G., Božin, E. S. & Billinge, S. J. L. (2006). *Chem. Mater.* **18**, 3242–3248.
Peterson, P. F., Božin, E. S., Proffen, Th. & Billinge, S. J. L. (2003). *J. Appl. Cryst.* **36**, 53–64.
Petkov, V. (1989). *J. Appl. Cryst.* **22**, 387–389.
Petkov, V. & Danev, R. (1998). *J. Appl. Cryst.* **31**, 609–619.
Qiu, X., Thompson, J. W. & Billinge, S. J. L. (2004). *J. Appl. Cryst.* **37**, 678.
Redmond, E. L., Setzler, B. P., Juhas, P., Billinge, S. J. L. & Fuller, T. F. (2012). *Electrochem. Solid State*, **15**, B72–B74.
Soper, A. K. & Barney, E. R. (2011). *J. Appl. Cryst.* **44**, 714–726.
Warren, B. E. (1990). *X-ray Diffraction.* New York: Dover.
Wyckoff, R. W. G. (1967). *Crystal Structures*, Vol. 1, 2nd ed. New York: Wiley.
Yang, X. & Billinge, S. J. L. (2013). *J. Appl. Cryst.* Submitted.
Young, C. A. & Goodwin, A. L. (2011). *J. Mater. Chem.* **21**, 6464–6476.