xINTERPDF User Guide 0.1.0

Chenyang Shi, Senior Scientist at AbbVie Inc

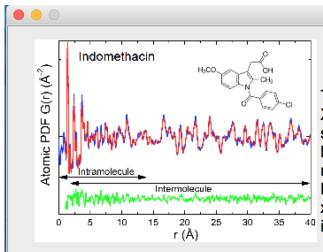


Overview
Installation
Technical Terms
Examples
Disclosure/Acknowledgement

Overview

xINTERPDF is written in Python 2.7. It utilizes Tkinter and matplotlib modules for creating a Graphical User Interface (GUI) and plot visualization. It builds on DiffPy-CMI package (http://www.diffpy.org/products/diffpycmi/index.html) for analysis of synchrotron/laboratory X-ray total scattering data collected for organic materials. The homepage for xINTERPDF is at https://github.com/curieshicy/xINTERPDF.

XINTERPDE

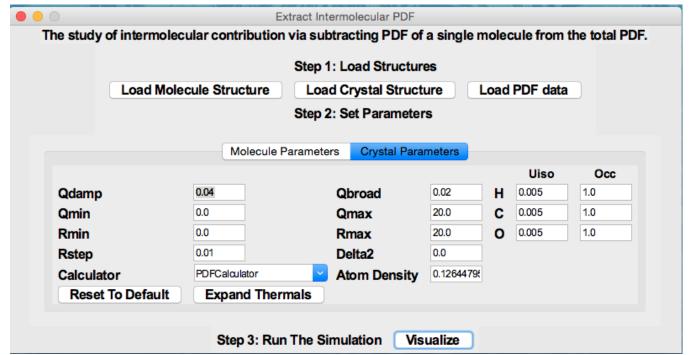


Welcome to use xINTERPDF program!

This program is designed to extract structural information from measured X-ray pair distribution function data for organic materials. Currently it supports (1) The study of intermolecular interaction (e.g. hydrogen bonds) by subtracting the scattering signal of a single molecule. (2) The PDF model fit of a crystalline organic X-ray PDF using the method proposed by Prill et al. (J. Appl. Cryst., 2015, 48, 171-178.). The homepage for xINTERPDF is at https://github.com/curieshicy/xINTERPDF. The program is developed by Chenyang Shi at AbbVie Inc.

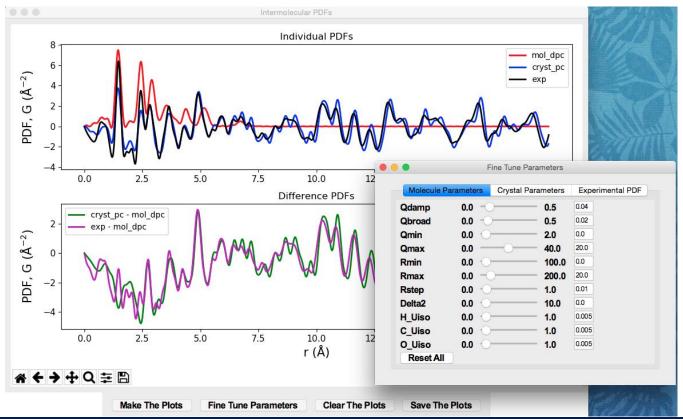
Usage 1: Extracting Intermolecular PDF

Users may use xINTERPDF to study the molecule-molecule interaction in organics. To do this, users first supply the structures for a single molecule (in xyz format) and a crystal (in cif format) and (optionally PDF data). Then in step 2, click <u>Expand Thermals</u> to bring up Uiso/Occ info. A variety of parameters for molecule and crystal will be specified by users. These include instrument parameters, thermal factors, occupancy and calculator to calculate PDF. In Step 3, hit <u>Visualize</u> to plot results.



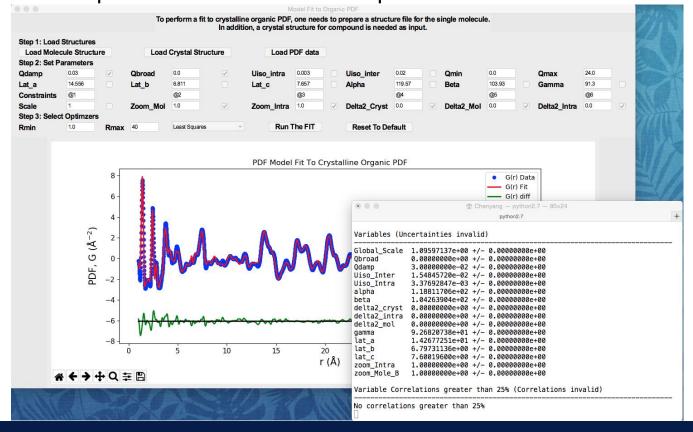
Usage 1: Extracting Intermolecular PDF

In the plot, the top panel shows the individual PDFs. In the example below, it shows the PDF (red) for molecule calculated by DebyePDFCalculator and PDF (blue) for a crystal using PDFCalculator. The experimental total PDF is shown in black. The bottom panel displays the theoretical/experimental intermolecular PDFs. Click <u>Fine Tune Parameters</u> to adjust the simulated plots. Hit <u>Save the Plots</u> to save raw data.



Usage 2: Model fit to organic crystalline PDF

Another capability of xINTERPDF is to perform a PDF fit to organic crystalline PDF. In Step 1, users load in structure files and PDF data. In step 2, various parameters are fixed or allowed to vary. In Step 3, the range of the fit and optimizers are further determined by users. Hit <u>Run the FIT</u> to start the refinement. After the fit is complete, the fit is plotted and fit results are printed on the terminal.



Installation

xINTERPDF can be installed on Linux and macOS computers. The easiest way to install it is through **conda**. Here is an example of installing it on macOS 10.10.3.

(1) Download Anaconda Distribution for macOS at https://www.anaconda.com/download/?lang=en-us#macos. Select Python 2.7 version to install.

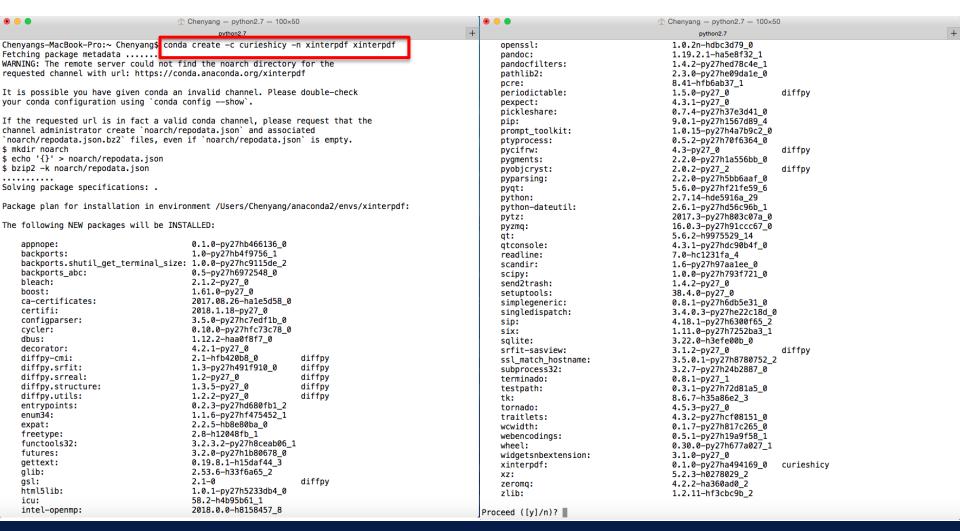
(2) Invoke a terminal, type conda config --get channels to check any channels that have been added. diffpy and xinterpdf are required. If you don't see both, type conda config --add channels diffpy and conda config --add channels xinterpdf

to add them.

```
♠ Chenyang — bash — 80×24
Last login: Fri Feb 2 14:06:01 on ttys000
Chenyangs-MacBook-Pro:~ Chenyang$ conda -V
conda 4.3.30
Chenyangs-MacBook-Pro:~ Chenyang$ conda env list
# conda environments:
                        /Users/Chenyang/anaconda2
root
Chenyangs-MacBook-Pro:~ Chenyang$ conda config --get channels
--add channels 'defaults'
                            # lowest priority
--add channels 'diffpy'
--add channels 'rdkit'
--add channels 'xinterpdf'
                             # highest priority
Chenyangs-MacBook-Pro: Chenyang$
```

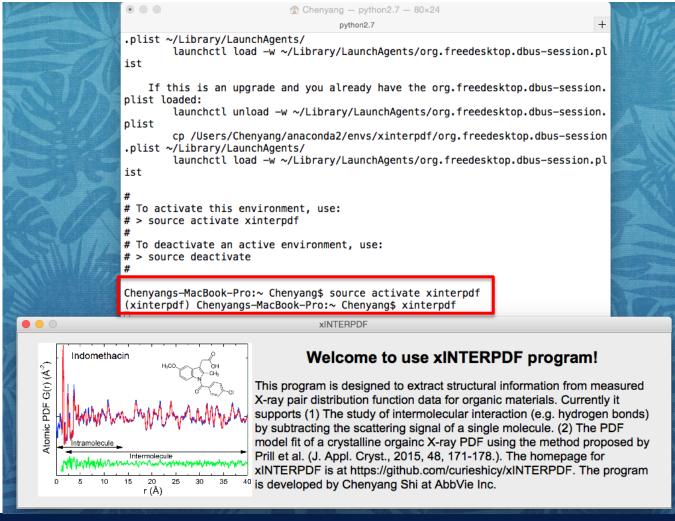
Installation

(3) Type conda create -c curieshicy -n xinterpdf xinterpdf to install it.



Installation

(4) Once the installation is complete. Type **source activate xinterpdf** to start the virtual environment and **xinterpdf** to invoke the main window of xINTERPDF.



Installation (alternative)

If conda install failed, alternatively, one may download the raw files (Logo.gif and cli.py) at

https://github.com/curieshicy/xINTERPDF/tree/master/Conda Build Recipe/xinter

<u>pdf</u>. To start the program, in a terminal, navigate the folder where you put both files, and type **python cli.py** to invoke the main window. Make sure you have installed Diffpy-CMI and matplotlib (2.0.2). Follow

http://www.diffpy.org/products/diffpycmi/index.html to install DiffPy-CMI. If you have conda, matplotlib can be installed by conda install matplotlib=2.0.2.

Technical Details

xINTERPDF uses diffpy.srreal.pdfcalculator module to calculate PDFs.

(http://www.diffpy.org/diffpy.srreal/api/diffpy.srreal.html?highlight=pdfcalculator#
diffpy.srreal.pdfcalculator.DebyePDFCalculator
PDF Calculator
DEDF Calculator<

PDF Calculator (PC)

Given a unit cell, and with periodic boundary conditions applied, the PDF for a crystalline material can be calculated via

$$G(r) = \frac{1}{Nr} \sum_{i} \sum_{j \neq i} \left\{ \frac{f_i f_j}{\langle f \rangle^2} \exp\left[\frac{-(r - r_{ij})^2}{\sigma_{ij}^2}\right] \right\} - 4\pi r \rho_0$$

Where σ_{ij} is defined by

$$\sigma_{ij} = \sigma'_{ij} \sqrt{1 - \frac{\delta_1}{r_{ij}} - \frac{\delta_2}{r_{ij}^2} + Q_{broad}^2 r_{ij}^2}$$

Here, f_i , f_j and <f> are X-ray form factor for species i and j, and average value weighted by concentration. N is the number of atoms in the unit cell. ρ_0 is the atomic density. σ'_{ij} is the root mean squared displacement coming from the atomic displacement parameters (ADP) tensors of the atom-pair. δ_1 and δ_2 are corrections that can be separately used to account for correlated atomic motion and Q_{broad} is an instrumental broadening factor coming from the finite Q resolution of the experiment.

Debye PDF Calculator (DPC)

DPC calculates PDF in Q-space. First F(Q) is calculated by Debye sum and then Fourier transformed to G(r)

$$F(Q) = \frac{1}{N\langle f(Q)\rangle^2} \sum_{i,j} f_i(Q) f_j(Q) \frac{\sin Q r_{ij}}{r_{ij}} \exp\left[-\frac{1}{2} \sigma_{ij}^2 Q^2\right]$$

$$F(Q) = Q[S(Q) - 1]$$

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

$$\sigma_{ij} = \sigma'_{ij} \sqrt{1 - \frac{\delta_1}{r_{ij}} - \frac{\delta_2}{r_{ij}^2} + Q_{broad}^2 r_{ij}^2}$$

Using DPC one can calculate PDF from any structure given known coordinates (typically in xyz format).

Explanation of parameters

 $\mathbf{Q}_{\mathsf{damp}}$: PDF Gaussian dampening envelope due to limited Q-resolution. Not applied when equal to zero. The Gaussian envelope is of the form

$$B(r) = e^{-\frac{(rQ_{damp})^2}{2}}$$

 $\mathbf{Q}_{\text{broad}}$: PDF peak broadening from increased intensity noise at high Q. Not applied when equal zero.

Delta 2: Coefficient for $(1/r^2)$ contribution to the peak sharpening.

\mathbf{Q}_{\min} :

- -PC: Lower bound of the experimental Q-range used. Affects the shape envelope.
- -DPC: Lower bound of the Q-grid for the calculated F(Q). Affects the shape envelope.

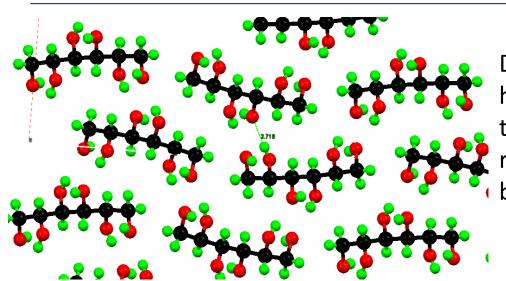
$\mathbf{Q}_{\mathsf{max}}$:

- -PC: Upper bound of the experimental Q-range used. Affects the termination ripples. Not used when zero.
- -DPC: Upper bound of the Q-grid for the calculated F(Q). Affects the termination ripples.

R_{min}: Lower bound of the r-grid for PDF calculation

 R_{max} : Upper bound of the r-grid for PDF calculation.

 $\mathbf{R}_{\mathsf{step}}$: Spacing in the calculated r-grid. r-values are at the multiples of rstep.



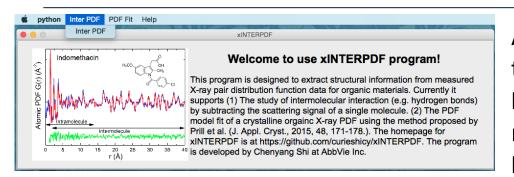
D-mannitol is a polyalcohol with rich hydrogen bonds. Based on its room temperature (283-303 K) structure as reported by Kim et al, the nearest O...O bond distance is ~2.72 Å.

Synchrotron X-ray total scattering was conducted on D-mannitol powder sample at 300 K. From a fit to PDF of cerium oxide, the instrumental resolution parameters are determined: $Q_{damp} = 0.02902 \text{ Å}^{-1}$, $Q_{broad} = 0.0004 \text{ Å}^{-1}$. A Q_{max} of 24 Å⁻¹ was used for Fourier transform. The software program PDFgetX2 was used.

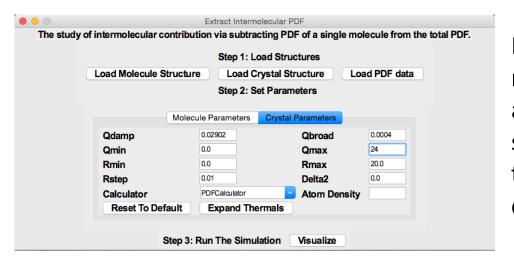
Examples files are available at

https://github.com/curieshicy/xINTERPDF/tree/master/Examples.

- H. S. Kim, G. A. Jeffrey and R. D. Rosenstein, Acta Cryst. B, 1968, 24, 1449.
- X. Qiu, J. W. Thompson and S. J. L. Billinge, J. Appl. Cryst. 2004, 37, 678.



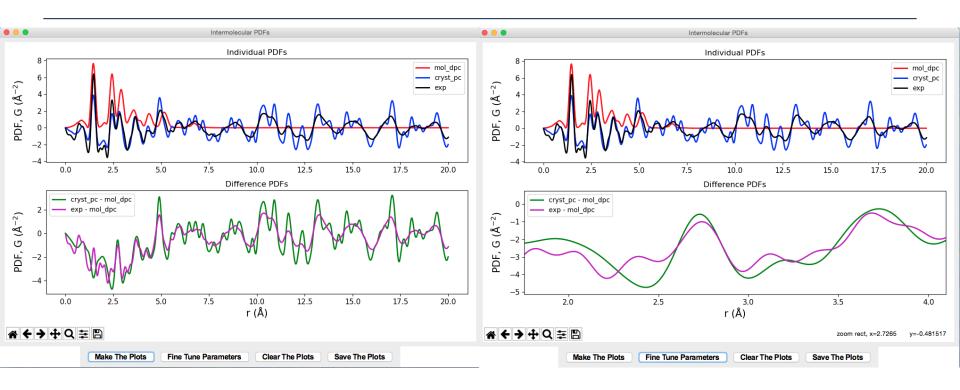
After invoke the main window in a terminal, select from drop-down menu, Inter PDF /Inter PDF, which brings up a new window titled Extract Intermolecular PDF.



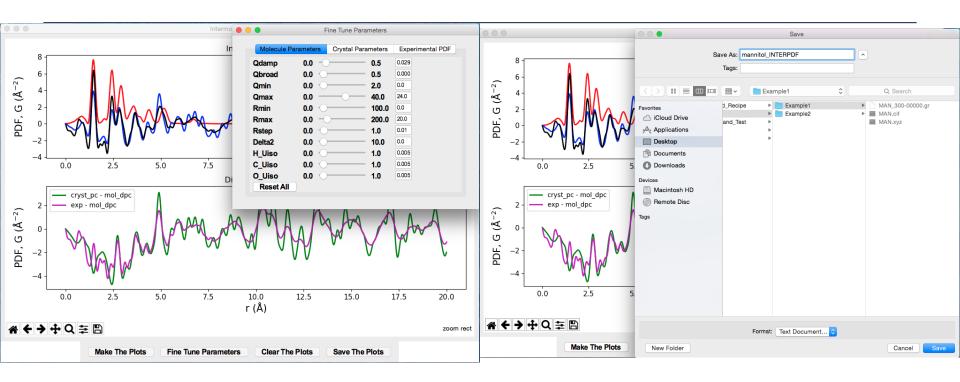
Following the Steps: first load in files for molecule (MAN.xyz), crystal (MAN.cif) and PDF data (MAN_300-00000.gr). In step 2, set Q_{damp} and Q_{broad} values to those determined from cerium oxide calibrant; set Q_{max} to 24 Å⁻¹.

For molecule DebyePDFCalculator is used; while for crystal, PDFCalculator is used. Click **Expand Thermals** for both molecule and crystal. Leave values at default.

In step 3, click **Visualize** to see the plots.



Click **Make The Plots**, the theoretical PDFs for molecule and crystal, together with experimental PDF are plotted in the top panel. In the bottom panel, the difference PDFs from experiment and theory are compared. Using magnifying tool in the embedded navigation toolbar, one can zoom into the peak around 2.7 Å. Hovering the mouse in the region, the peak positions for O...O is ~2.72 Å and ~2.74 Å, respectively.

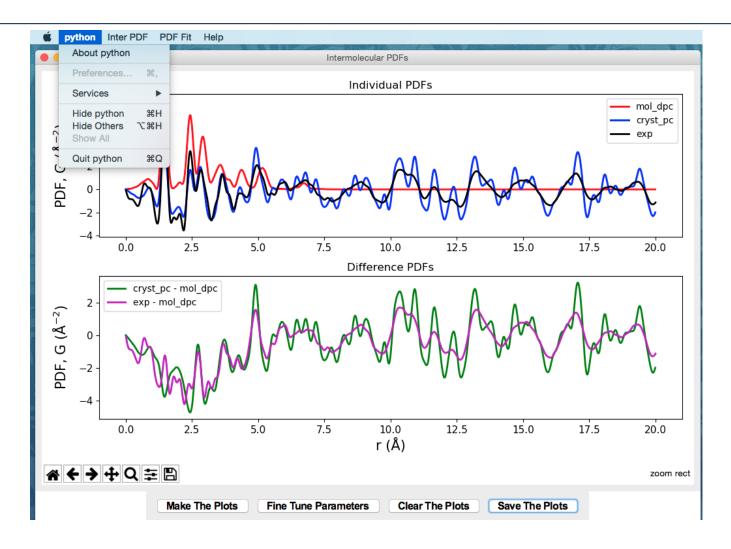


Hit **Fine Tune Parameters** to bring about a window for fine tuning parameters. It has three tabs for molecule, crystal and experiment data, respectively. Users can either drag the scale bar or type a number in the entry box and hit Enter. The plots will update in real time.

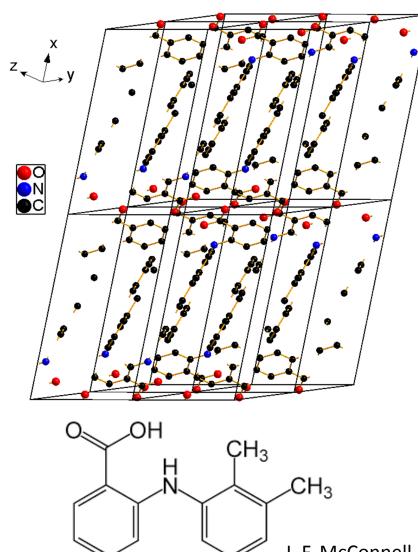
Click Clear the Plots to erase the plots. Click Save The Plots to save all raw data.

```
*Here are the raw data for the plotting the curves.********************
       *You have chosen to use DeybePDFCalculator for molecule; PDFCalculator for crystal.*******
       You also have scaled the measured PDF by a factor of 1.000 *********************
       From left to right, the data correspond to (1)r (2)m (3)c (4)e (5)c-m (6)e-m *************
0000000000021e-02 6.212443191915230020e-03 -1.059388955578956237e-02 -1.84141500000000135e-01 -1.680633274770479413e-02 -1.903539431919152314e-01
2.00000000000000042e-02 1.222590932327609001e-02 -2.149921887044312432e-02 -2.726172000000000040e-01 -3.372512819371921433e-02 -2.848431093232761113e-01
2.9999999999999889e-02 1.786282325375071478e-02 -3.300927702376095507e-02 -3.569217000000000084e-01 -5.087210027751166985e-02
4.00000000000000083e-02 2.296872986922959825e-02 -4.538209911720556683e-02 -4.3589549999999913e-01 -6.835082898643515814e-02
      00000000000278e-02 2.742156663333214159e-02 -5.882539791214071861e-02 -5.0856389999999577e-01 -8.624696454547285673e-02 -5.359854666333321305e-01
5.999999999999778e-02 3.113916665704996936e-02 -7.348440170292774853e-02 -5.7416659999999714e-01 -1.046235683599777144e-01 -6.053057666570499373e-01
7.00000000000000666e-02\ 3.408136856581576307e-02\ -8.943330360119053357e-02\ -6.32179200000000522e-01\ -1.235146721670062897e-01\ -6.662605685658158361e-01
8.0000000000000167e-02\ 3.623690446715328861e-02\ -1.066708208278051434e-01\ -6.8232300000000128e-01\ -1.429077252949584320e-01\ -7.185599044671533431e-01
1.199999999999996e-01 3.925448026295294546e-02 -1.859566010169879380e-01 -8.07011200000000396e-01 -2.252110812799408834e-01 -8.462656802629530128e-01
      000000000044e-01 3.942080670153031569e-02 -2.072085892115682615e-01 -8.21750300000000446e-01 -2.466293959130985702e-01 -8.611711067015304089e-01
      000000000133e-01 3.976139948555254544e-02 -2.284398564666100251e-01 -8.31504800000000439e-01 -2.682012559521625983e-01 -8.712661994855526171e-01
       999999999944e-01 4.051142979510184011e-02 -2.493352073592800811e-01 -8.37254000000000537e-01 -2.898466371543819142e-01 -8.777654297951018592e-01
1.6000000000000033e-01 4.186383830073221041e-02 -2.695979199506459700e-01 -8.4001979999999832e-01 -3.114617582513781735e-01 -8.818836383007322421e-01
        00000000122e-01 4.399199010123152481e-02 -2.889671668736874532e-01 -8.40827699999999557e-01 -3.329591569749189572e-01
1.9000000000000022e-01\ 5.108940734433653486e-02\ -3.242509790985316220e-01\ -8.40486500000000252e-01\ -3.753403864428681569e-01\ -8.915759073443365601e-01
2.0000000000000111e-01 5.620585249988695187e-02 -3.399467522291140242e-01 -8.4111559999999633e-01 -3.961526047290009900e-01 -8.973214524998869290e-01
```

In the file that saved, the header explains the details of each column of data saved. In this example, it describes the Calculators used for simulating PDFs for molecule and crystal, and the scale factor used (in **Fine Tune Parameters** tab) for experimental PDF. Each column, from left to right, corresponds to, respectively, (1)the radial distance, (2) theoretical PDF for molecule, (3) theoretical PDF for crystal, (4) experimental PDF, (5) theoretical intermolecular PDF and (6) experimental intermolecular PDF.

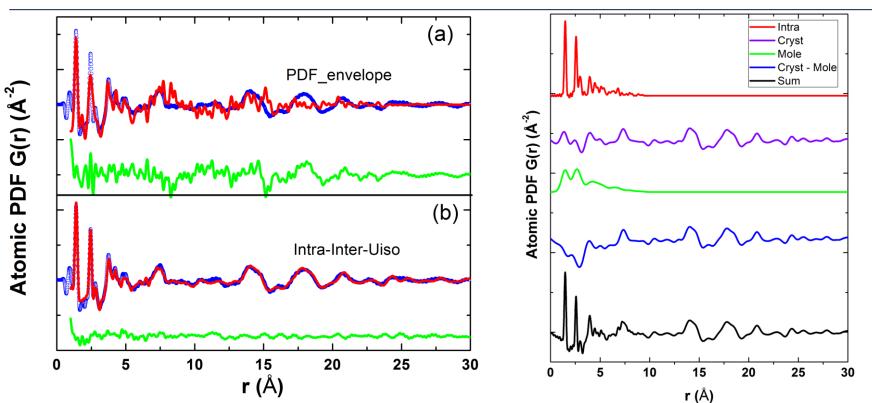


In the dropdown menu, click **Python**---**Quit Python** to quit.



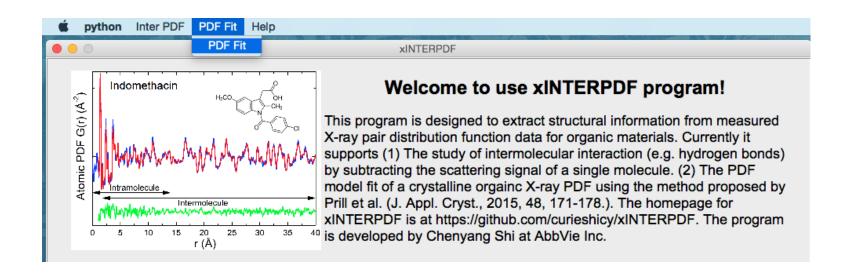
Mefenamic acid (MEF) is a nonsteroidal antiinflammatory and analgesic drug used to treat mild pain, especially menstrual cramps. The crystal structure of mefenamic acid (Form I) was solved by McConnell and Company in 1976. It has a triclinic structure with a space group $\bar{P}1$.

Synchrotron X-ray total scattering was conducted on D-mannitol powder sample at 300 K. From a fit to PDF of cerium oxide, the instrumental resolution parameters are determined: $Q_{damp} = 0.02902 \ \text{Å}^{-1}$, $Q_{broad} = 0.0004 \ \text{Å}^{-1}$. A Q_{max} of 24 Å^{-1} was used for Fourier transform. The software program PDFgetX2 was used.

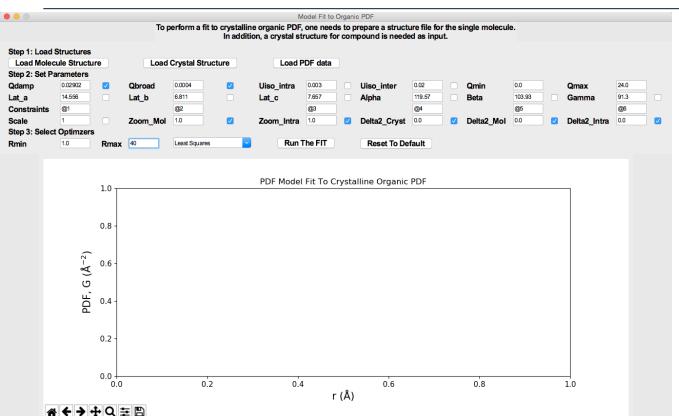


As Prill et al. reported, the PDF of organics cannot be modelled well by expanding unit cell as typically done for inorganic materials (left figure a), because the intermolecular forces are weaker than intramolecular ones. Instead, a model differentiating both intra- and intermolecular contributions is more appropriate. As shown in the breakdown of a total PDF for MEF (black curve in the right figure), it has contributions from intra- (red) and inter-molecules (blue), each with a distinct U_{iso} thermal factor.

D. Prill, P. Juhás, M. U. Schmidt and S. J. L. Billinge, J. Appl. Cryst. 2015, 48, 171-178.



To perform a PDF fit, in dropdown menu, click **PDF Fit/PDF Fit** to start the GUI window.

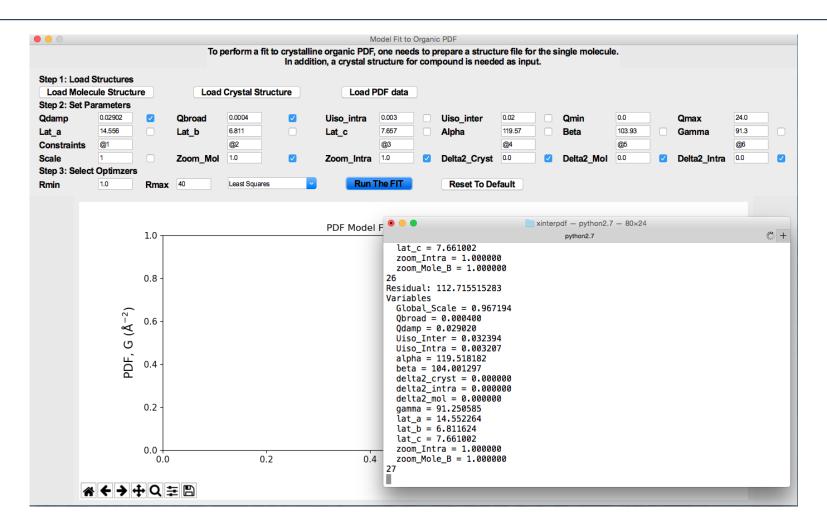


Following the steps, first load in structure files and PDF data (MEF.xyz, MEF.cif and MEF_300-00000.gr). In Step 2, specify the parameters. If a check box is marked, that parameter is fixed during the fit; otherwise is allowed to vary.

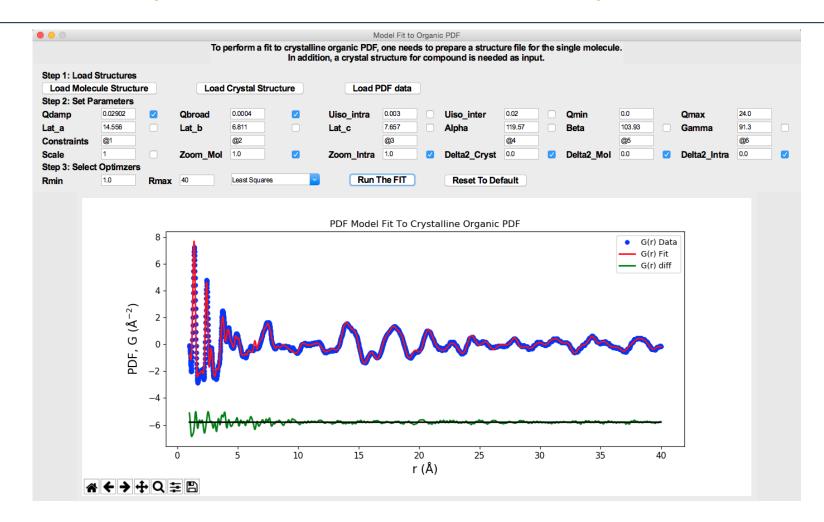
Below lattice parameters, there is a row for constraining them. For example, if it is a cubic structure, users can type @1 for Lat_a, Lat_b and Lat_c.

For a tetragonal structure, @1 for Lat_a and Lat_b, @2 for Lat_c. Similar syntax is applied to constrain the angles (Alpha, Beta and Gamma). Since MEF is triclinic, we need six different variables for lattice parameters. Zoom_Mol and Zoom_Intra give the possibilities to expand or shrink the molecule isotropically. The three delta 2 values can be further freed to explain for the r-dependent peak width (i.e. correlated motions). Refer to the right figure in Slide 22, for the meaning of Mol and Intra.

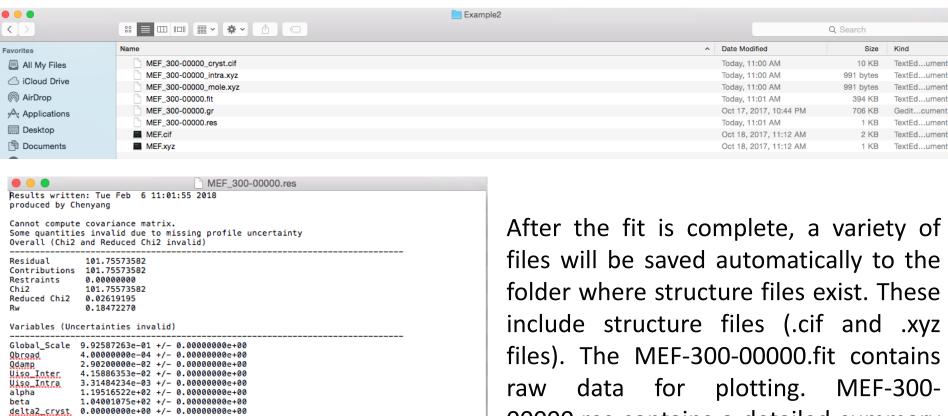
In Step 3, select the fit range and optimizer. Hit Run The Fit to start the run.



When the fit starts running, the terminal will update fit parameters in real-time.



When the fit is complete, the fit result is plotted. Blue circles, red and green curves, correspond to measured, calculated and difference PDF, respectively.



folder where structure files exist. These include structure files (.cif and .xyz files). The MEF-300-00000.fit contains data for plotting. MEF-300-00000.res contains a detailed summary of fit results. See the screenshot shown to the left.

delta2_mol

zoom_Intra

gamma

lat_a

lat_b

delta2_intra 0.00000000e+00 +/- 0.00000000e+00

zoom Mole B 1.00000000e+00 +/- 0.00000000e+00

No correlations greater than 25%

0.00000000e+00 +/- 0.0000000e+00

9.12728874e+01 +/- 0.00000000e+00

1.45539596e+01 +/- 0.00000000e+00

6.81128742e+00 +/- 0.00000000e+00

7.66290270e+00 +/- 0.00000000e+00 1.00000000e+00 +/- 0.00000000e+00

Variable Correlations greater than 25% (Correlations invalid)

Disclosure/Acknowledgement

AUTHOR INFORMATION

Corresponding Author

*E-mail: chenyang.shi@abbvie.com. Phone: +1 847-935-3328.

ORCID

Chenyang Shi: 0000-0002-2291-7325

Notes

C.S. is the employee of AbbVie and may own AbbVie stock. The design, study conduct, and financial support for this research were provided by AbbVie. AbbVie participated in the interpretation of data, review, and approval of the publication. The author declares no competing financial interest.

ACKNOWLEDGMENTS

C.S. thanks various individuals from Drug Product Development (Research and Development) at AbbVie for their helpful insights and discussions. C.S. thanks Drs. Karena Chapman, Kevin Beyer and Olaf Borkiewicz at beamline 11-ID-B, Argonne National Lab for assists of collecting PDF data. C.S. appreciates suggestions from Dr. Pavol Juhas, Prof. Lian Yu and Prof. Kirsten Jensen.