xINTERPDF User Guide v.0.1.0

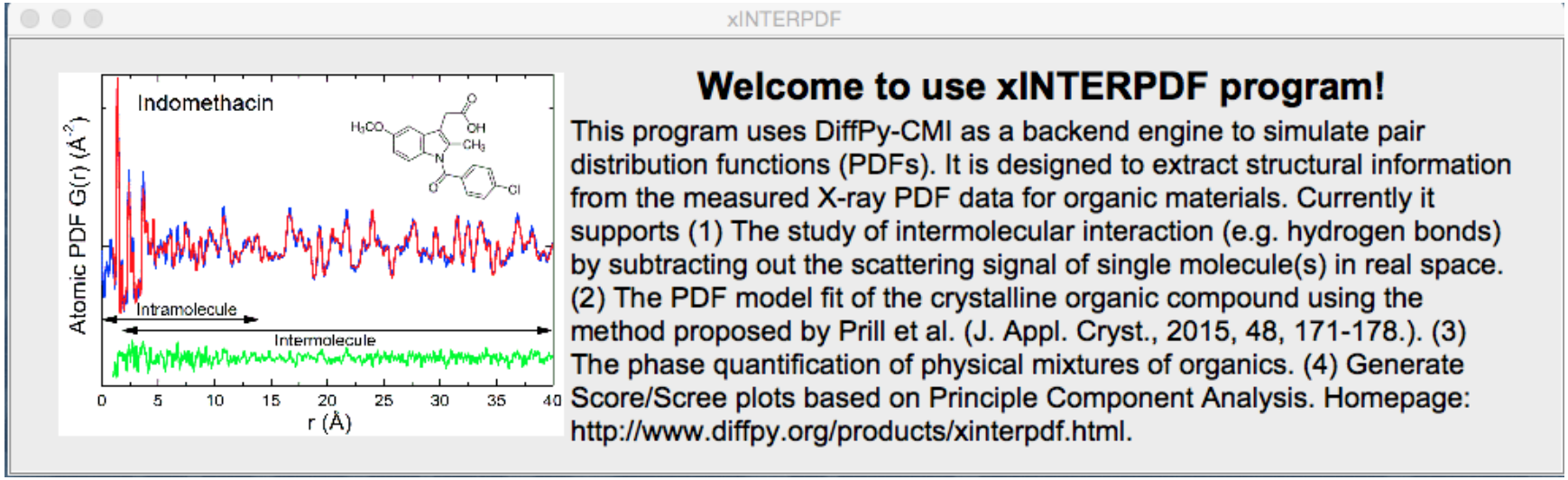
Chenyang Shi, Senior Scientist at AbbVie Inc.

Updated June 15th, 2018

A video demo about installation and usage is available at https://www.youtube.com/watch?v=lAFZ5VYEH1g

Overview

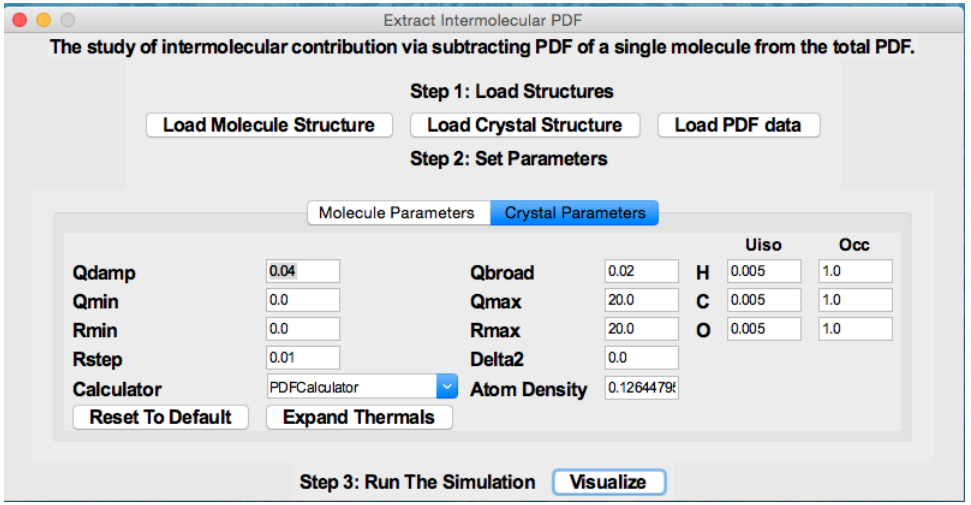
xINTERPDF is written in Python 2.7. It utilizes Tkinter and matplotlib modules for creating a Graphical User Interface (GUI) and plot visualization, and NumPy, SciPy and Scikit-learn (i.e. PCA feature) for scientific calculations. It builds on DiffPy-CMI package (http://www.diffpy.org/products/diffpycmi/index.html) for analysis of synchrotron/laboratory X-ray total scattering data collected for organic materials. The homepage for xINTERPDF is at https://www.diffpy.org/products/xinterpdf.html and its GitHub page is at https://github.com/curieshicy/xINTERPDF.



Usage 1: Extracting Intermolecular PDF

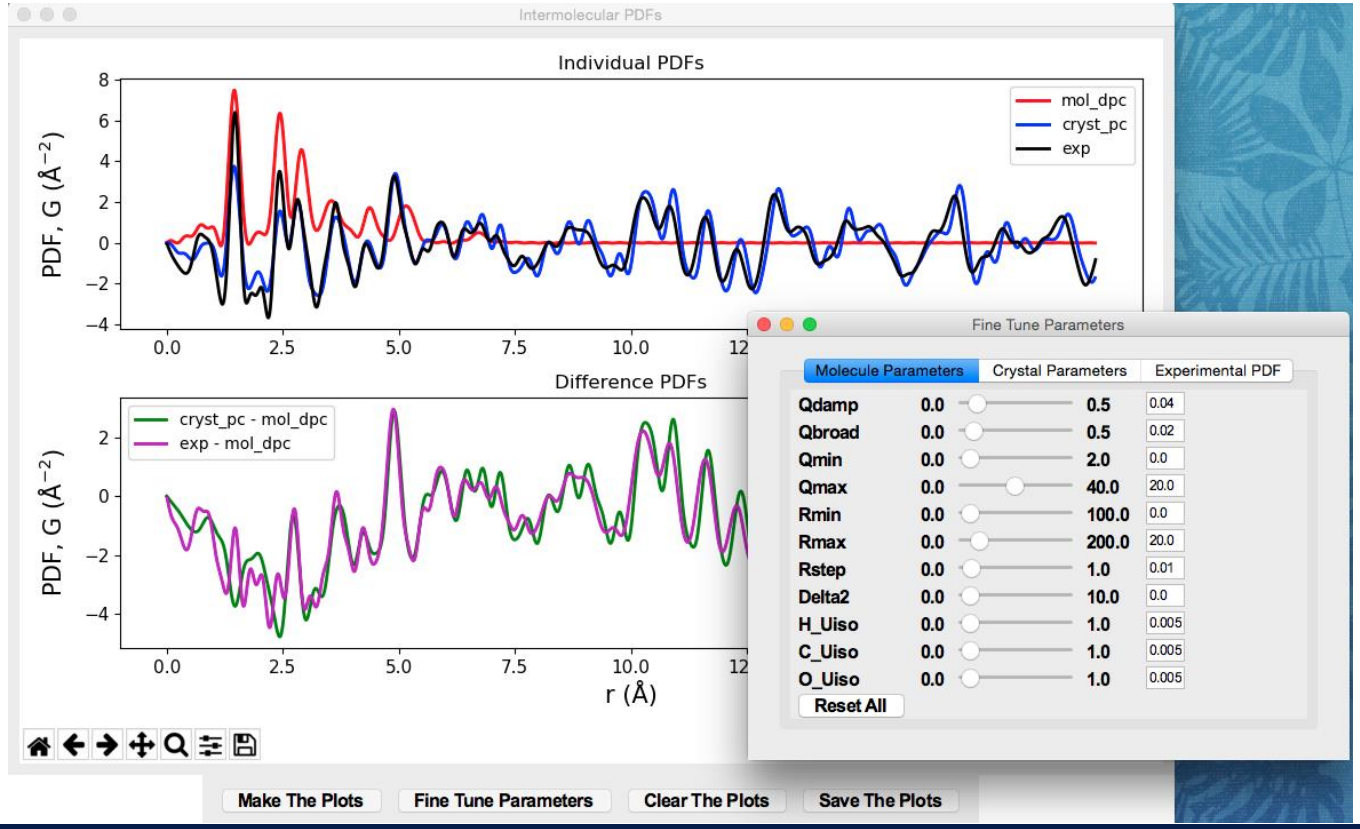
Users may use xINTERPDF to study the molecule-molecule interaction in organics. The program considers three common scenarios—the organic compound of interest is (1) crystalline (2) amorphous and (3) amorphous solid dispersion (two components). Shown below is an example when compound is crystalline, other two cases are similar.

To extract intermolecular interaction for crystalline organics, 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 Expand Thermals 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 Visualize to plot results.

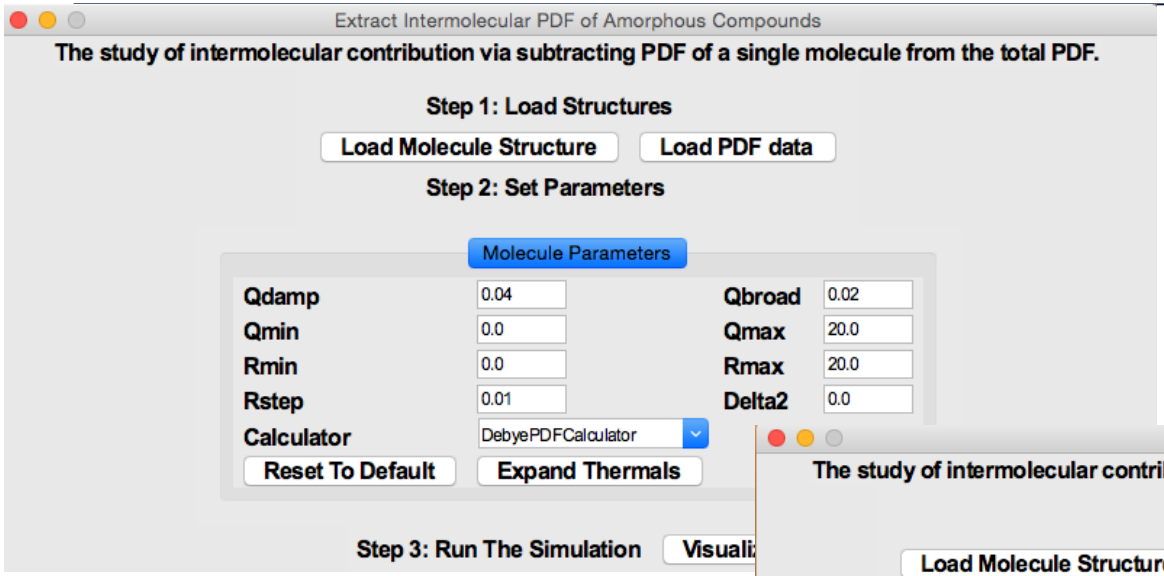


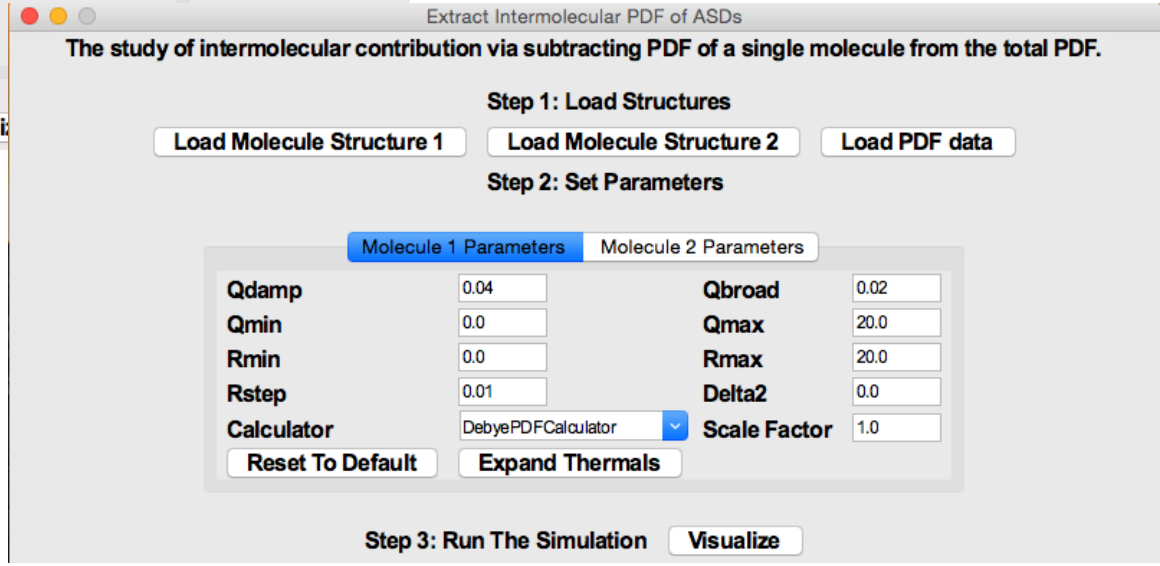
Usage 1: Extracting Intermolecular PDF

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 Fine Tune Parameters to adjust the simulated plots. Hit Save the Plots to save raw data.



Usage 1: Extracting Intermolecular PDF

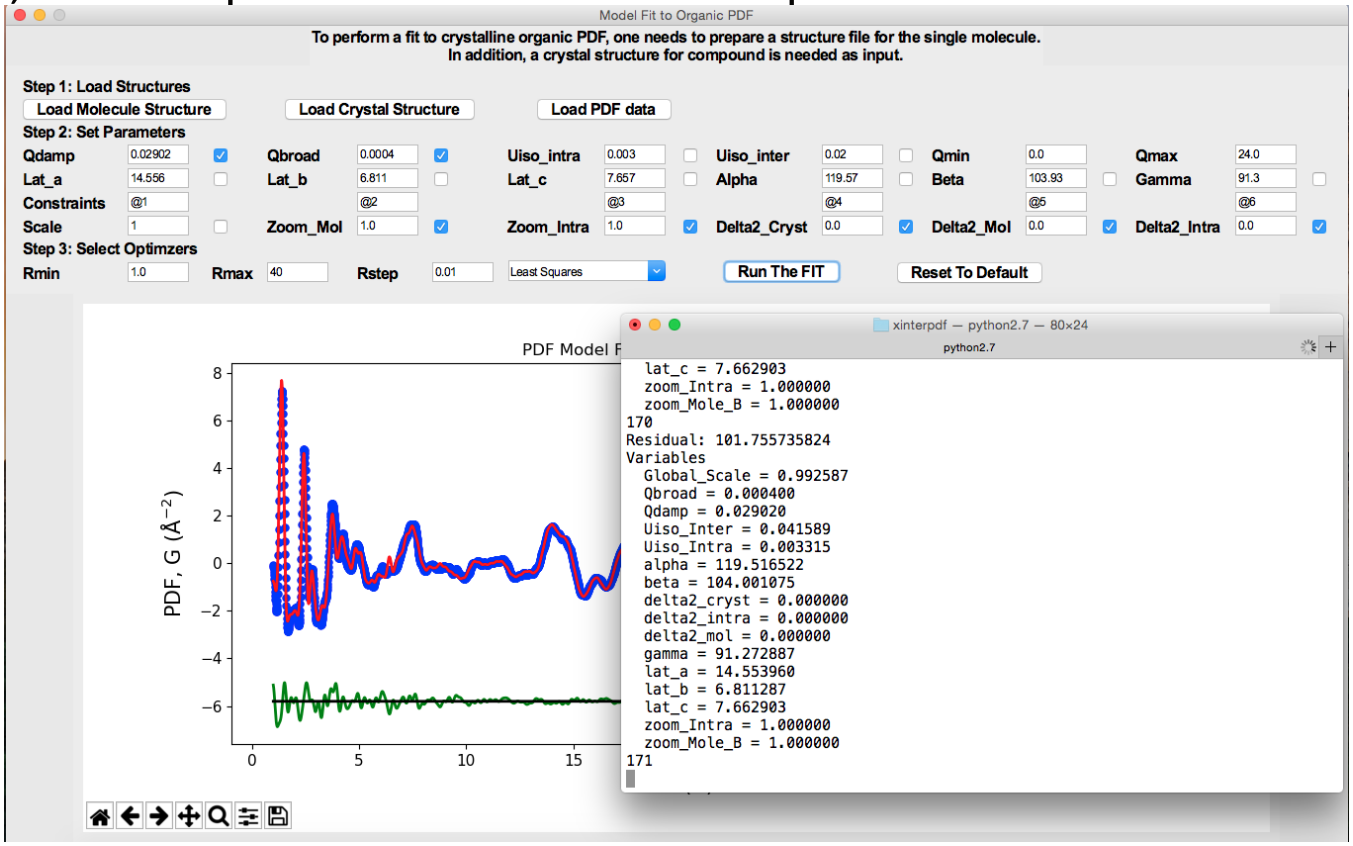




Similar to crystalline case, users can study intermolecular PDFs in amorphous compounds and amorphous solid dispersions. In both cases, only PDF(s) for molecule(s) is simulated and subtracted.

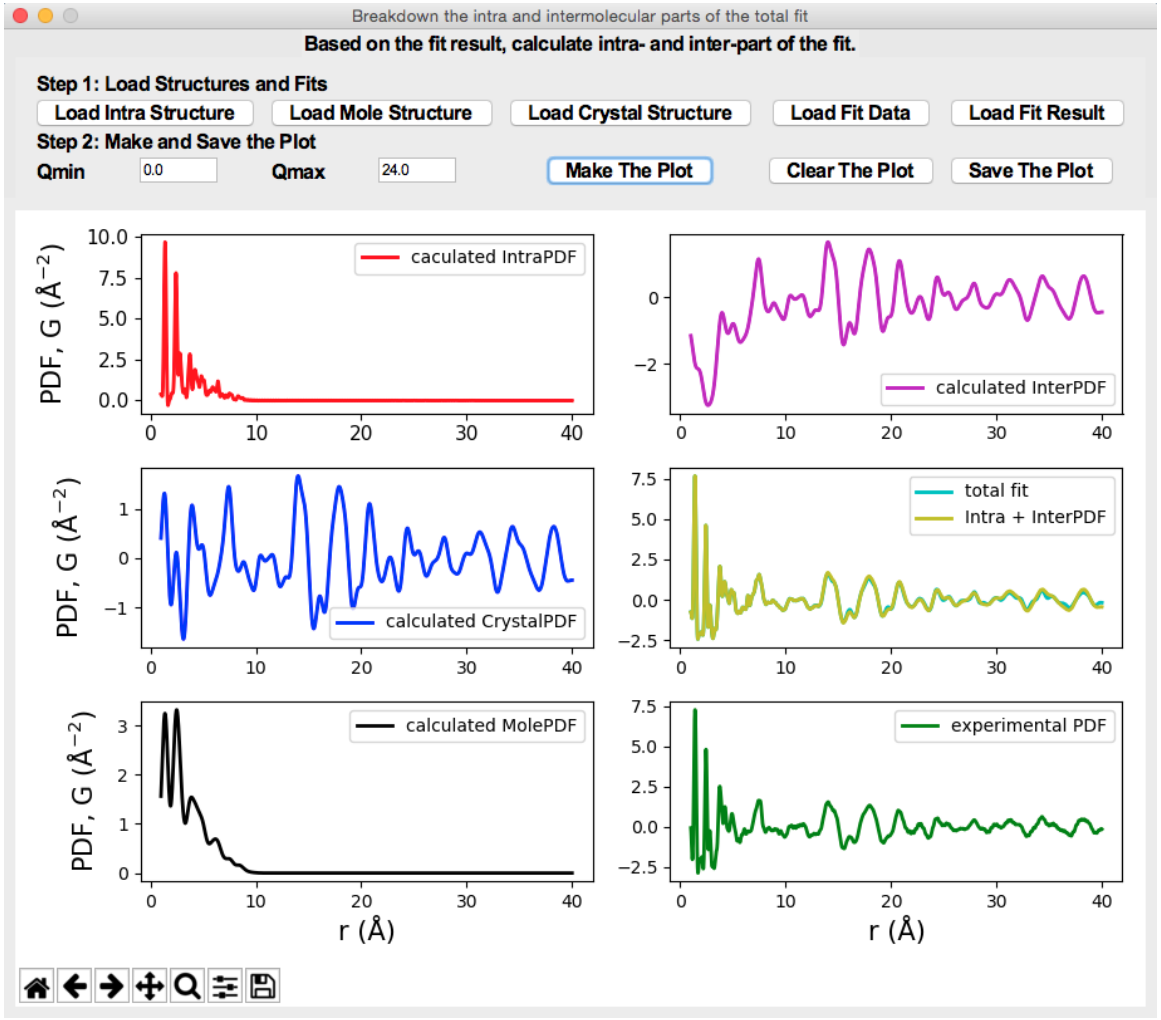
Usage 2: Model fit to organic crystalline PDF

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 Run the FIT to start the refinement. After the fit is complete, the fit is plotted and fit results are printed on the terminal.



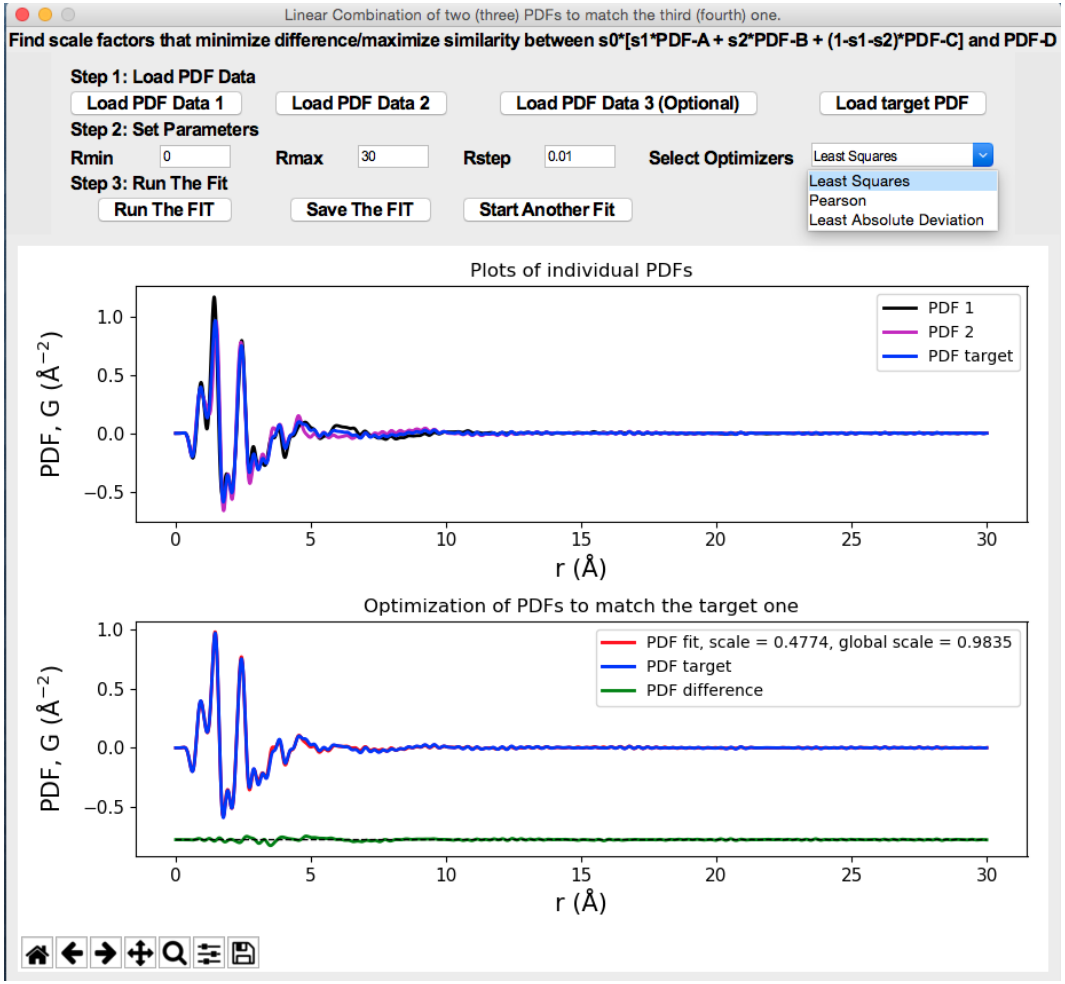
Usage 3: a breakdown of the model fit to organic crystalline PDF

After the PDF fit to crystalline X-ray PDF is done, the users can break down the total fit to intra- and intermolecular PDFs.



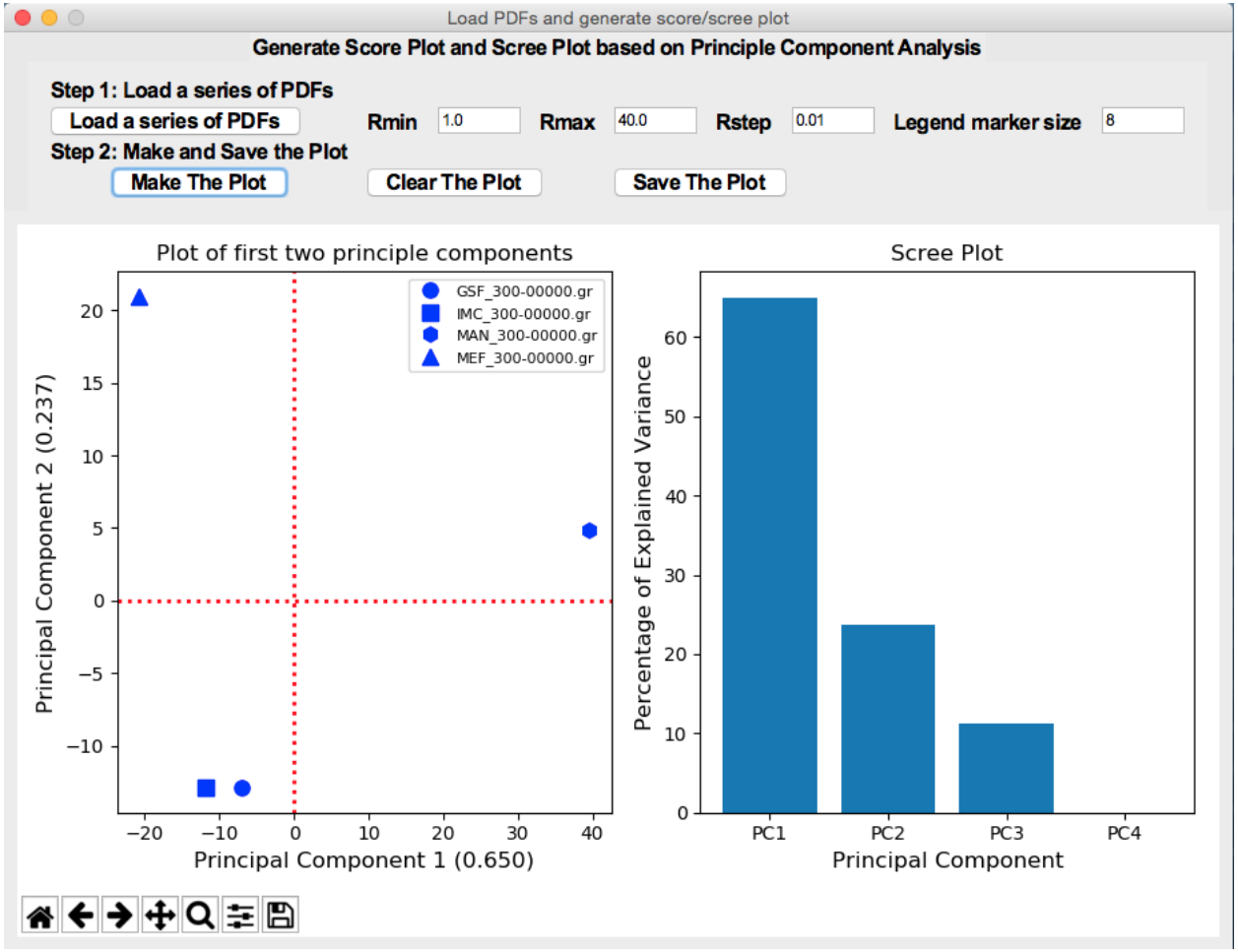
Usage 4: phase quantification

The program offers an easy functionality that takes in the PDFs of 2 or 3 pure materials and compares against a target PDF (e.g. physical mixture or amorphous solid dispersion of them). The scale factors are returned which encodes the info about phase fraction.



Usage 5: Principle Component Analysis

The program also provides an easy interface for generating Scree/Score plots of PDFs based on PCA analysis.

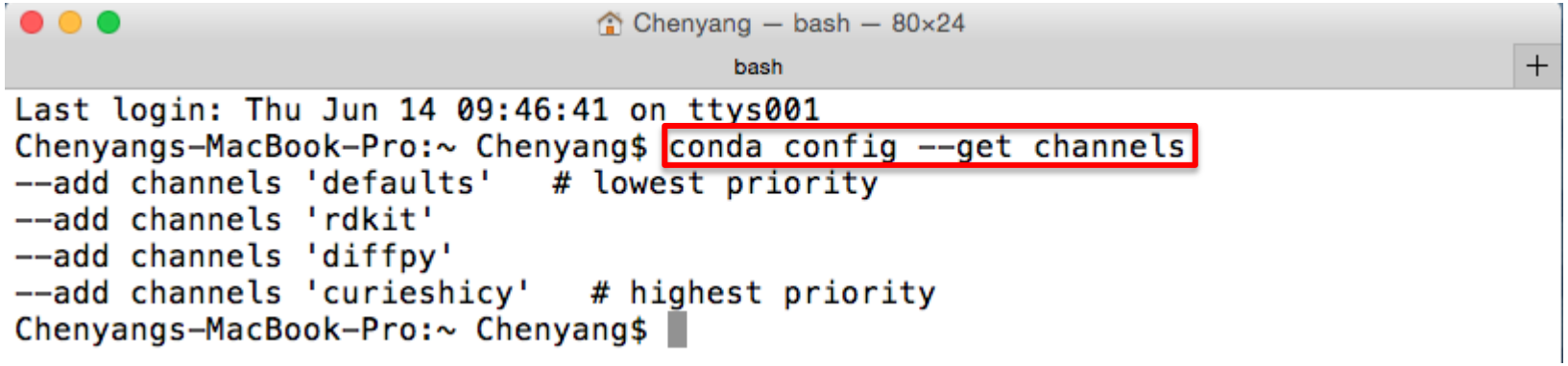


Installation

xINTERPDF can be installed on Linux and macOS 64 bit computers. The easiest way to install it is through **conda**. Here is an example of installing it on macOS 10.10.3.

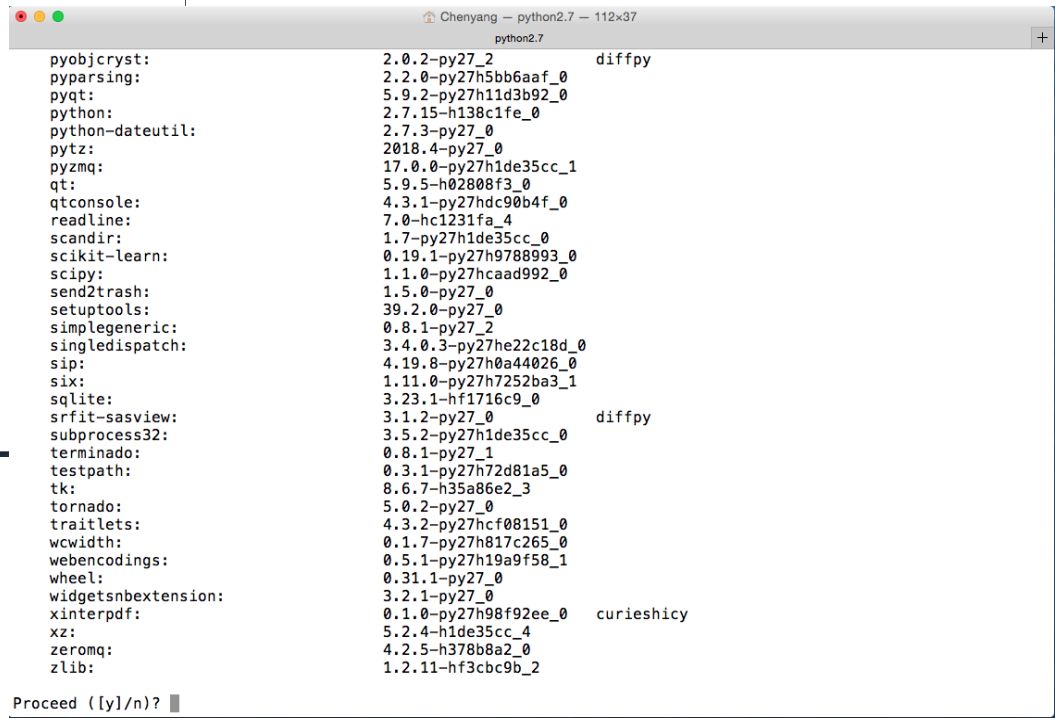
(1) Download Anaconda Distribution for macOS at https://www.anaconda.com/download/?lang=en-us#macos. Select Python 2.7 version to install.

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



Installation

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

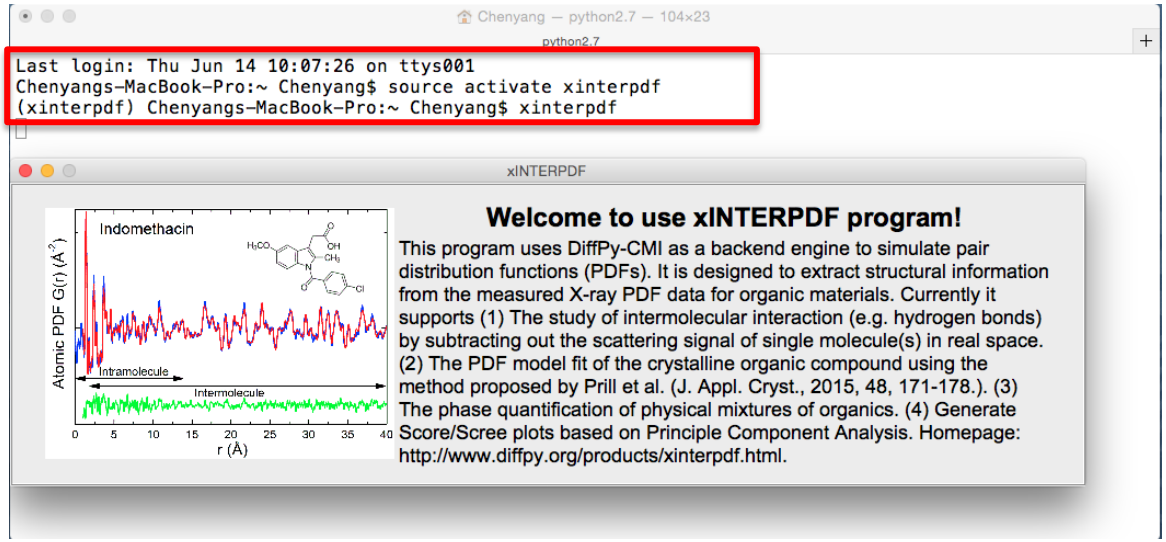


Installation

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

```
Chenyang — bash — 112×37
bash
    send2trash:                   1.5.0-py27_0
    setuptools:                   39.2.0-py27_0
    simplegeneric:                0.8.1-py27_2
    singledispatch:               3.4.0.3-py27he22c18d_0
    sip:                          4.19.8-py27h0a44026_0
    six:                          1.11.0-py27h7252ba3_1
    sqlite:                       3.23.1-hf1716c9_0
    srfit-sasview:                3.1.2-py27_0          diffpy
    subprocess32:                 3.5.2-py27h1de35cc_0
    terminado:                    0.8.1-py27_1
    testpath:                     0.3.1-py27h72d81a5_0
    tk:                           8.6.7-h35a86e2_3
    tornado:                      5.0.2-py27_0
    traitlets:                    4.3.2-py27hcf08151_0
    wcwidth:                      0.1.7-py27h817c265_0
    webencodings:                 0.5.1-py27h19a9f58_1
    wheel:                        0.31.1-py27_0
    widgetsnbextension:           3.2.1-py27_0
    xinterpdf:                    0.1.0-py27h98f92ee_0  curieshicy
    xz:                           5.2.4-h1de35cc_4
    zeromq:                       4.2.5-h378b8a2_0
    zlib:                         1.2.11-hf3cbc9b_2

Proceed ([y]/n)? y

Preparing transaction: done
Verifying transaction: done
Executing transaction: done
#
# To activate this environment, use:
# > source activate xinterpdf
#
# To deactivate an active environment, use:
# > source deactivate
#

Chenyangs-MacBook-Pro:~ Chenyang$
```

```
Chenyang — python2.7 — 104×23
python2.7
Last login: Thu Jun 14 10:07:26 on ttys001
Chenyangs-MacBook-Pro:~ Chenyang$ source activate xinterpdf
(xinterpdf) Chenyangs-MacBook-Pro:~ Chenyang$ 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_Recipe_macOS_Linux/Conda_Build_Recipe_macOS/xinterpdf (macOS) or https://github.com/curieshicy/xINTERPDF/tree/master/Conda_Recipe_macOS_Linux/Conda_Build_Recipe_Linux/xinterpdf (Linux). To start the program, in a terminal, navigate to the folder where you put both files, and type **python cli.py** to invoke the main window. Make sure you have installed Diffpy-CMI, matplotlib (2.0.2) and Scikit-Learn (0.19.1). 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**; Scikit-Learn can be installed by **conda install scikit-learn=0.19.1**

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). Specifically PDF Calculator (PC) and Debye PDF Calculator (DPC) are used to simulate PDF in real and reciprocal spaces, respectively.

PDF Calculator (PC)

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

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

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)

When using DPC to calculate PDF, first the reduced structure factor, F(Q), defined as Q[S(Q)-1], is calculated by Debye sum* and then it is Fourier transformed to obtain PDF, G(r). The relevant equations are listed below.

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

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

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

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

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

*Debye, P. (1915). *Ann. Phys.* **351**, 809-823.

More details see Chapter 3 of *Underneath the Bragg peaks: Structural Analysis of Complex Materials*, 2nd ed.; Elsevier: Amsterdam, The Netherlands, 2013.

Explanation of parameters

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

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

Q_{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.

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.

Q_{max}:

-PC: Upper bound of the experimental Q-range used. Affects the termination ripples. Not used when is set to zero.

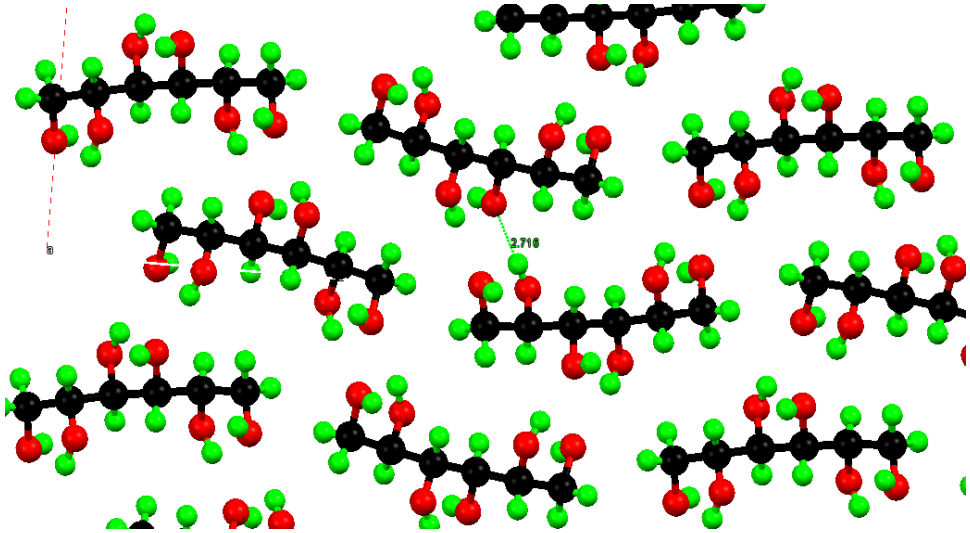
-DPC: Upper bound of the Q-grid for the calculated F(Q). Affects the termination ripples.

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

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

R_{step}: Spacing in the calculated r-grid. r-values are at the multiples of rstep.

Example 1: Hydrogen bonds in D-mannitol



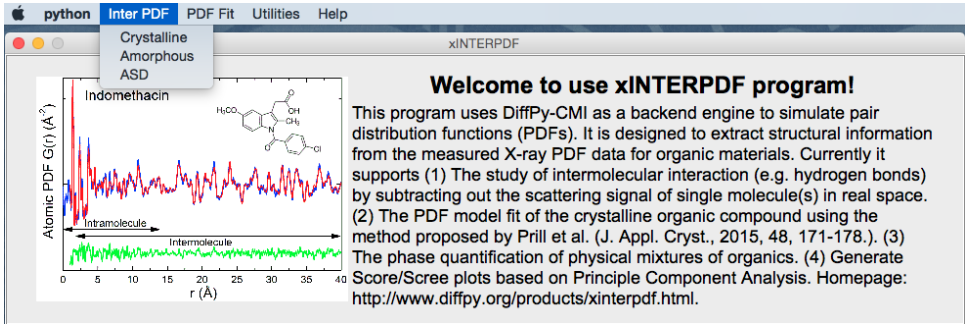
D-mannitol is a polyalcohol with rich hydrogen bonds. Based on its room temperature (283-303 K) structure as reported by Kim et al, the nearest O…O bond distance is ~2.72 Å.

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 Å^{-1}, Q_{broad}=0.0004 Å^{-1}. A Q_{max} of 24 Å^{-1} was used for Fourier transform. The software program PDFgetX2 was used.

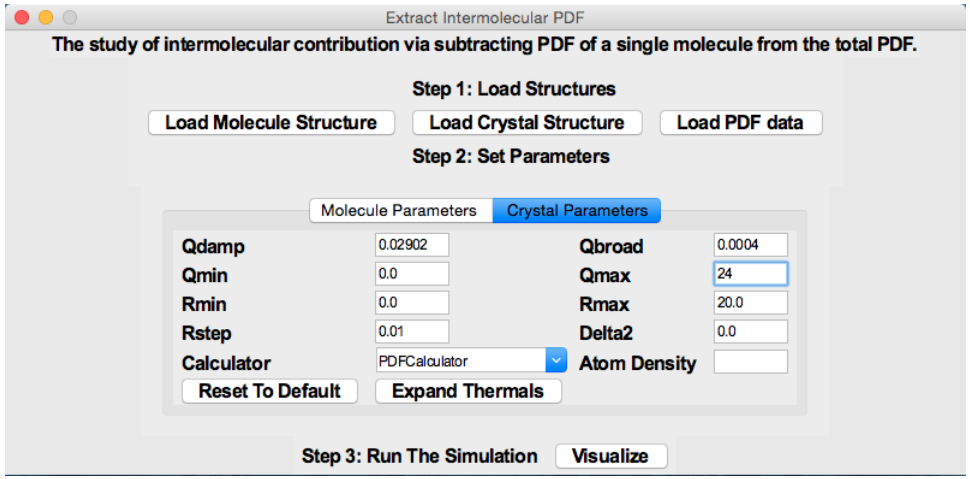
Examples files are available at https://github.com/curieshicy/xINTERPDF/tree/master/Examples.

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.

Example 1: Hydrogen bonds in D-mannitol



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

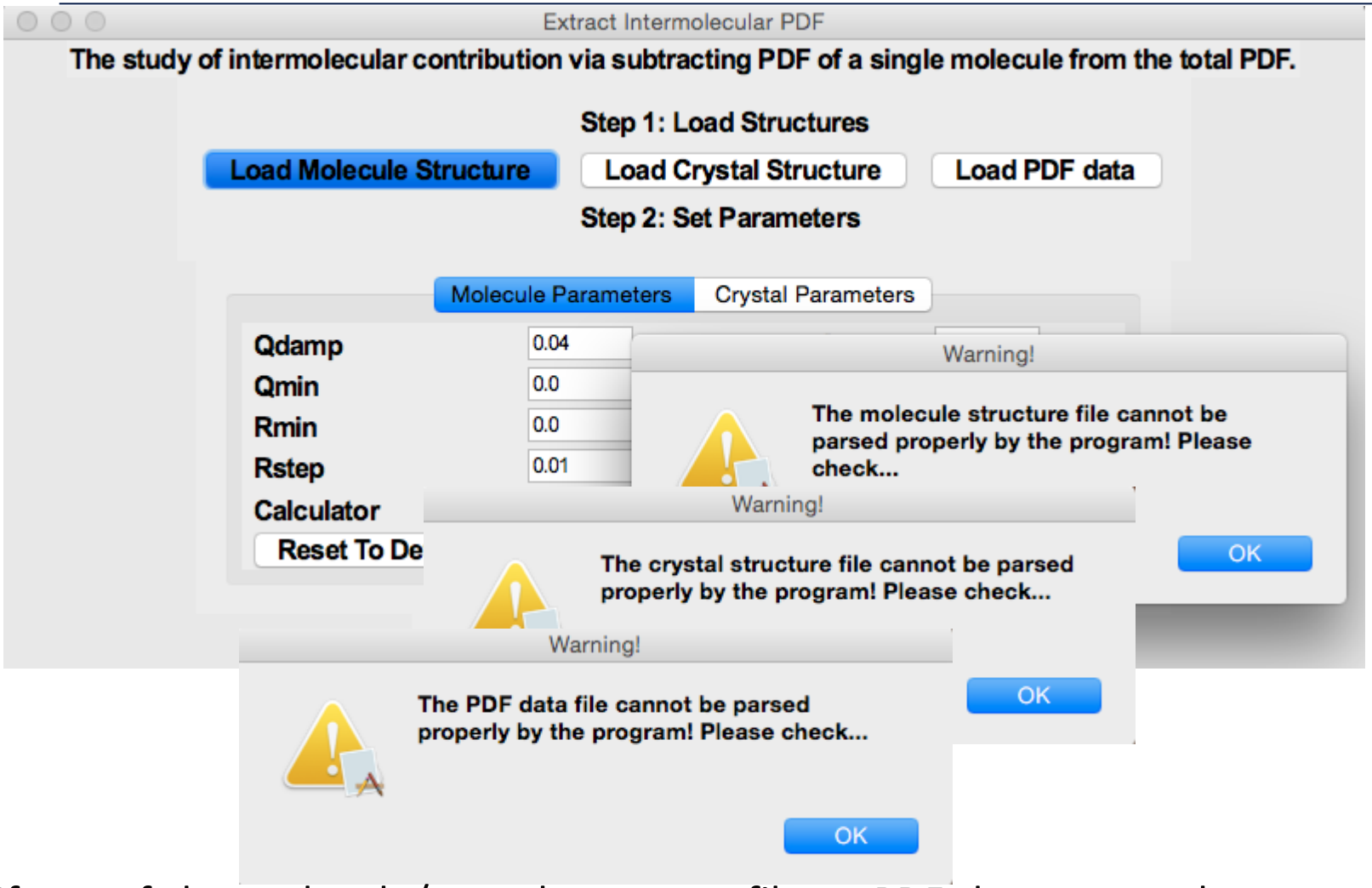


Following the Steps: first load in files for molecule (MAN.xyz), crystal (MAN.cif) and PDF data (MAN_300-00000.gr). In step 2, set Q_{damp} and Q_{broad} values to those determined from cerium oxide calibrant; set Q_{max} to 24 $Å^{-1}$.

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

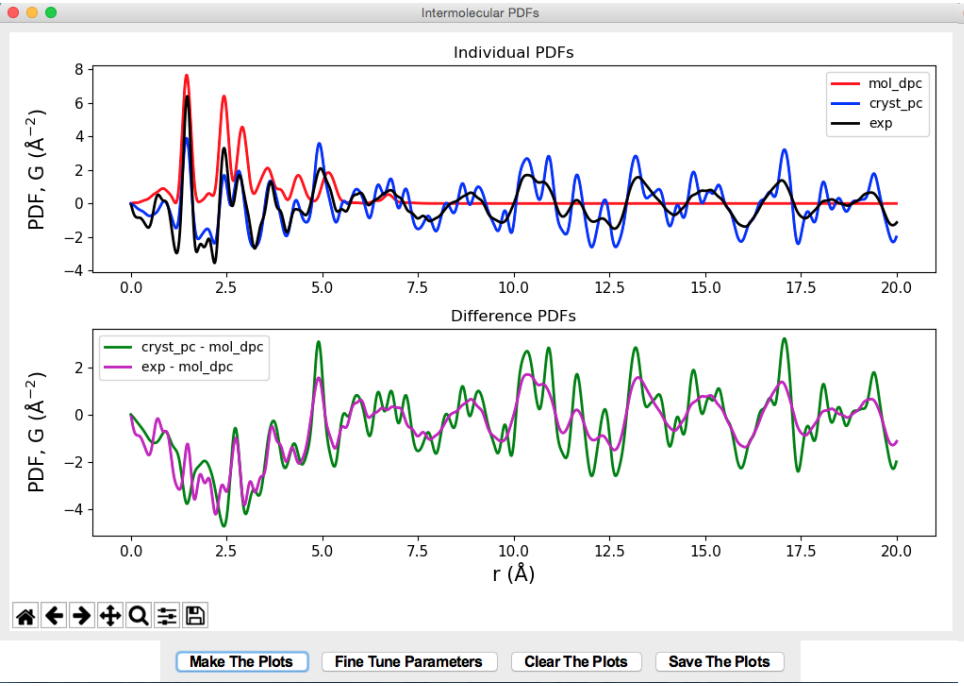
In step 3, click **Visualize** to see the plots.

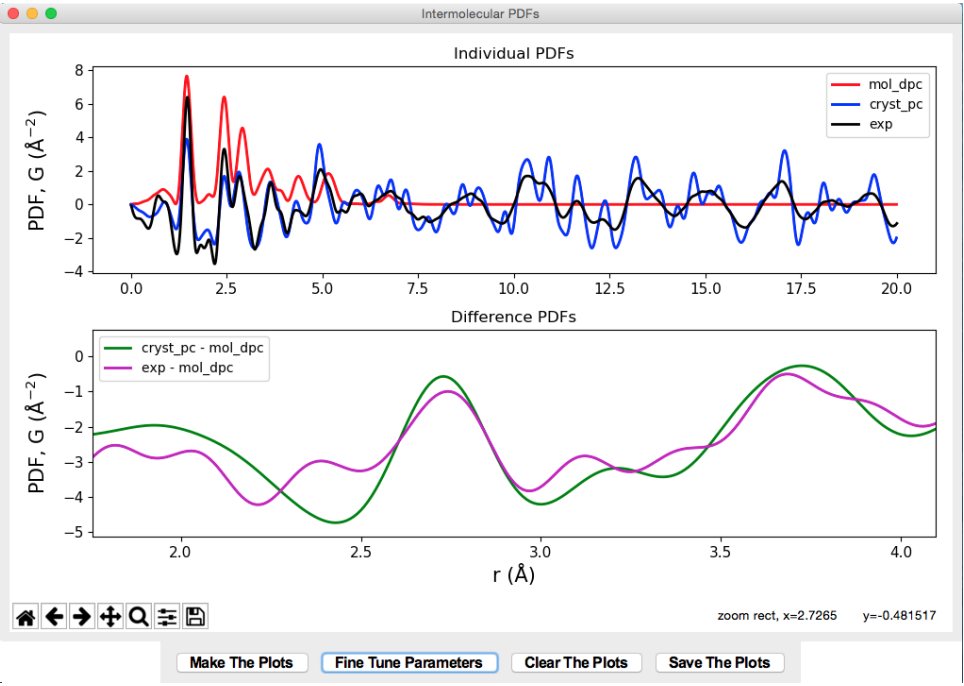
Example 1: Hydrogen bonds in D-mannitol



If any of the molecule/crystal structure file or PDF data cannot be recognized by the program, an error message will be displayed to alert the users.

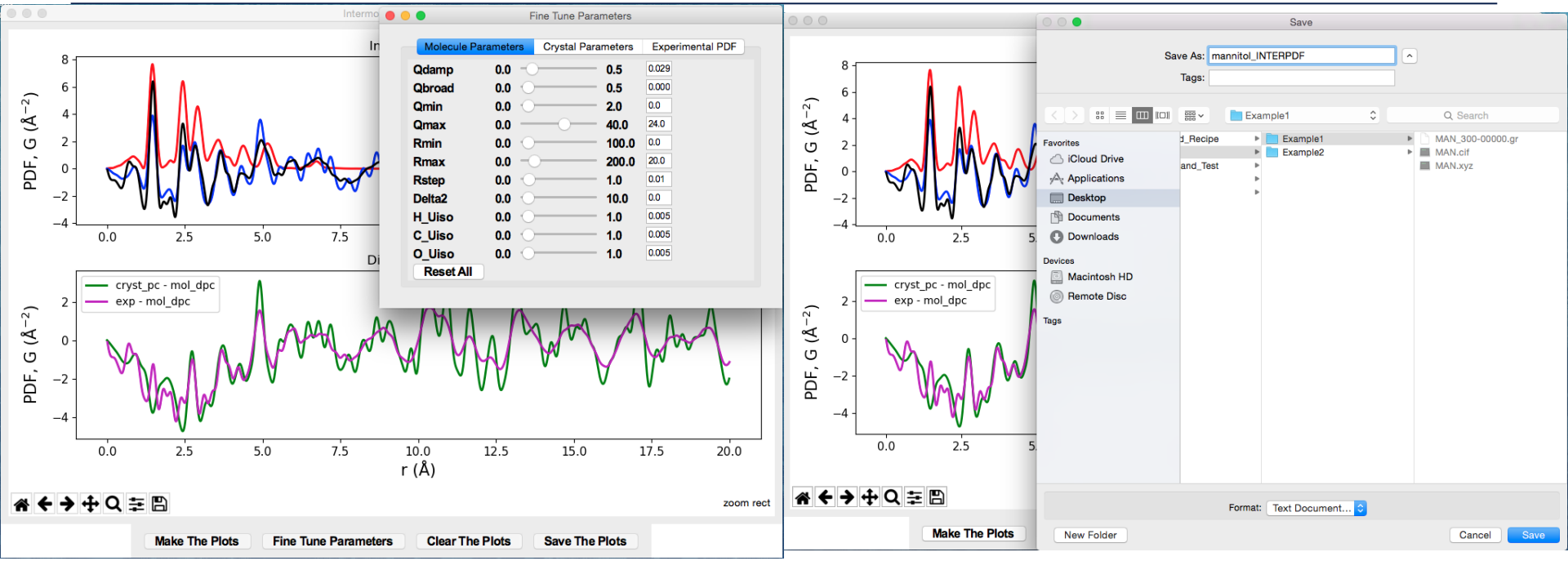
Example 1: Hydrogen bonds in D-mannitol





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.

Example 1: Hydrogen bonds in D-mannitol



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.

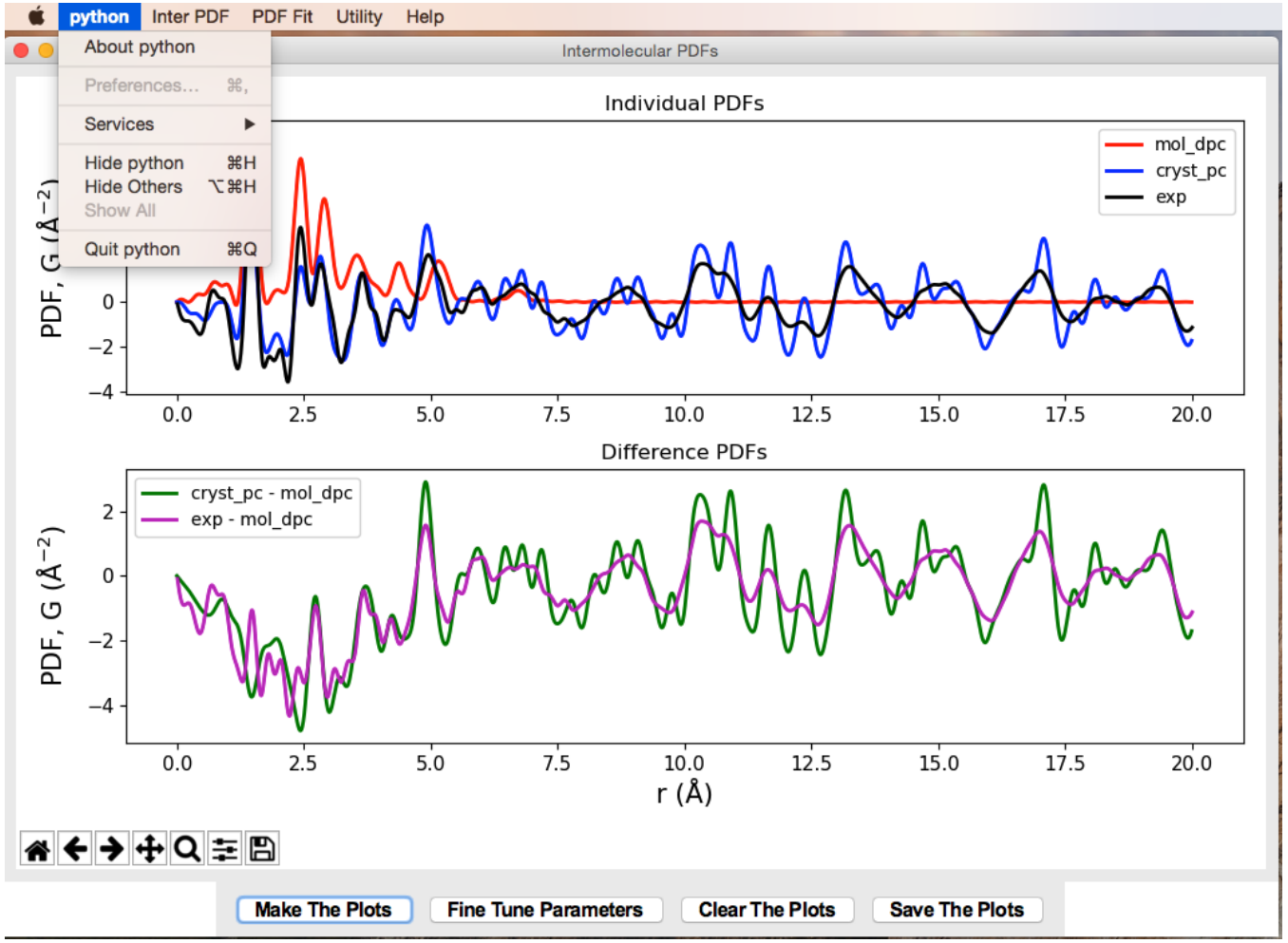
Example 1: Hydrogen bonds in D-mannitol

mannitol_INTERPDF.txt

```
# ****************************************************************************************************
# ****************************************************************************************************
# ******Here are the raw data for the plotting the curves.**********************************************
# ******You have chosen to use DebyePDFCalculator 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 ***************
# ****************************************************************************************************
# ****************************************************************************************************
#
0.0000000000000000e+00 0.0000000000000000e+00 0.0000000000000000e+00 -9.2800110000000509e-02 0.0000000000000000e+00 -9.2800110000000509e-02
1.0000000000000021e-02 6.2124431919152300 20e-03 -1.0593895557895623 7e-02 -1.8414150000000013 5e-01 -1.6806332747704794 13e-01 -1.9035394319191523 14e-01
2.0000000000000042e-02 1.2225909323276090 01e-02 -2.1499218870443124 32e-02 -2.7261720000000004 0e-01 -3.3725128193719214 33e-02 -2.8484310932327611 3e-01
2.9999999999999889e-02 1.7862823253750714 78e-02 -3.3009277023760955 07e-02 -3.5692170000000008 4e-01 -5.0872100277511669 85e-02 -3.7478452325375072 32e-01
4.0000000000000083e-02 2.2968729869229598 25e-02 -4.5382099117205566 83e-02 -4.3589549999999991 3e-01 -6.8350828986435158 14e-02 -4.5886422986922958 96e-01
5.0000000000000278e-02 2.7421566633332141 59e-02 -5.8825397912140718 61e-02 -5.0856389999999995 77e-01 -8.6246964545472856 73e-02 -5.3598546663333213 05e-01
5.9999999999999778e-02 3.1139166657049969 36e-02 -7.3484401702927748 53e-02 -5.7416659999999997 14e-01 -1.0462356835997771 44e-01 -6.0530576665704993 73e-01
7.0000000000000666e-02 3.4081368565815763 07e-02 -8.9433303601190533 57e-02 -6.3217920000000005 22e-01 -1.2351467216700628 97e-01 -6.6626056856581583 61e-01
8.0000000000000167e-02 3.6236904467153288 61e-02 -1.0667082082780514 34e-01 -6.8232300000000001 28e-01 -1.4290772529495843 20e-01 -7.1855990446715334 31e-01
8.9999999999999667e-02 3.7701883693465033 93e-02 -1.2512012726906007 74e-01 -7.2456640000000005 49e-01 -1.6282201096252510 44e-01 -7.6226828369346510 96e-01
1.0000000000000056e-01 3.8587631462883964 51e-02 -1.4463317178995951 93e-01 -7.5911439999999996 70e-01 -1.8322080325284348 38e-01 -7.9770203146288398 70e-01
1.1000000000000006e-01 3.9044584724493944 24e-02 -1.6499914283681532 68e-01 -7.8639090000000000 38e-01 -2.0404372756130928 49e-01 -8.2543548472449390 64e-01
1.1999999999999956e-01 3.9254480262952945 46e-02 -1.8595660101698793 80e-01 -8.0701120000000003 96e-01 -2.2521108127994088 34e-01 -8.4626568026295301 28e-01
1.3000000000000044e-01 3.9420806701530315 69e-02 -2.0720858921156826 15e-01 -8.2175030000000004 46e-01 -2.4662939591309857 02e-01 -8.6117110670153040 89e-01
1.4000000000000133e-01 3.9761399485552545 44e-02 -2.2843985646661002 51e-01 -8.3150480000000004 39e-01 -2.6820125595216259 83e-01 -8.7126619948555261 71e-01
1.4999999999999944e-01 4.0511429795101840 11e-02 -2.4933520735928008 11e-01 -8.3725400000000005 37e-01 -2.8984663715438191 42e-01 -8.7776542979510185 92e-01
1.6000000000000033e-01 4.1863838300732210 41e-02 -2.6959791995064597 00e-01 -8.4001979999999998 32e-01 -3.1146175825137817 35e-01 -8.8188363830073224 21e-01
1.7000000000000122e-01 4.3991990101231524 81e-02 -2.8896716687368745 32e-01 -8.4082769999999995 57e-01 -3.3295915697491895 72e-01 -8.8481969010123151 52e-01
1.7999999999999933e-01 4.7034817018615587 84e-02 -3.0723342623724941 85e-01 -8.4067239999999998 63e-01 -3.5426824325586503 41e-01 -8.8770721701861554 64e-01
1.9000000000000022e-01 5.1089407344336534 86e-02 -3.2425097909853162 20e-01 -8.4048650000000000 252e-01 -3.7534038644286815 69e-01 -8.9157590734433656 01e-01
2.0000000000000111e-01 5.6205852499886951 87e-02 -3.3994675222911402 42e-01 -8.4111559999999996 33e-01 -3.9615260472900099 00e-01 -8.9732145249988692 90e-01
```

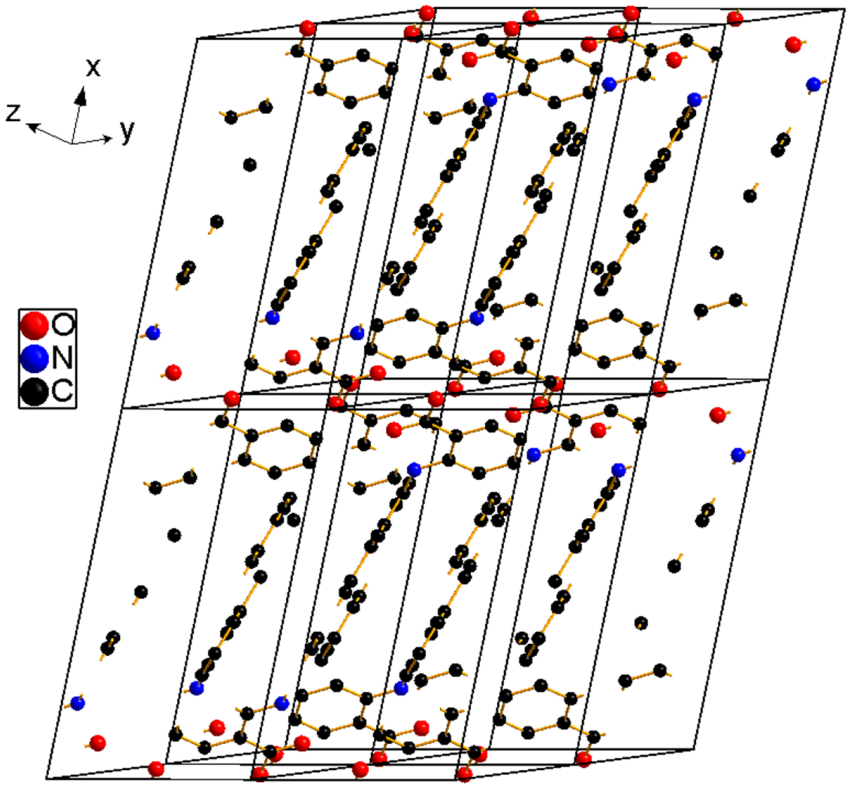
In the file that saved, the header explains the details of each column of data saved. In this example, it describes the Calculators used for simulating PDFs for molecule and crystal, and the scale factor used (in **Fine Tune Parameters** tab) for experimental PDF. Each column, from left to right, corresponds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

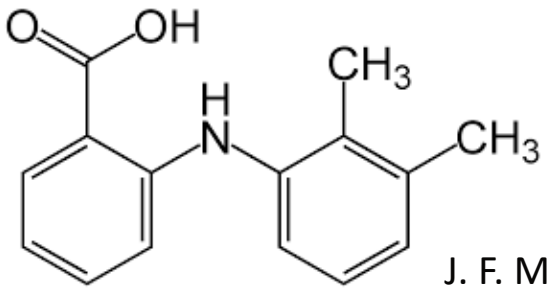
Example 1: Hydrogen bonds in D-mannitol



In the dropdown menu, click **Python**---**Quit Python** to quit. If terminal fails to respond, press Crtl + Z to kill the process.

Example 2: Model fit of MEF crystalline PDF



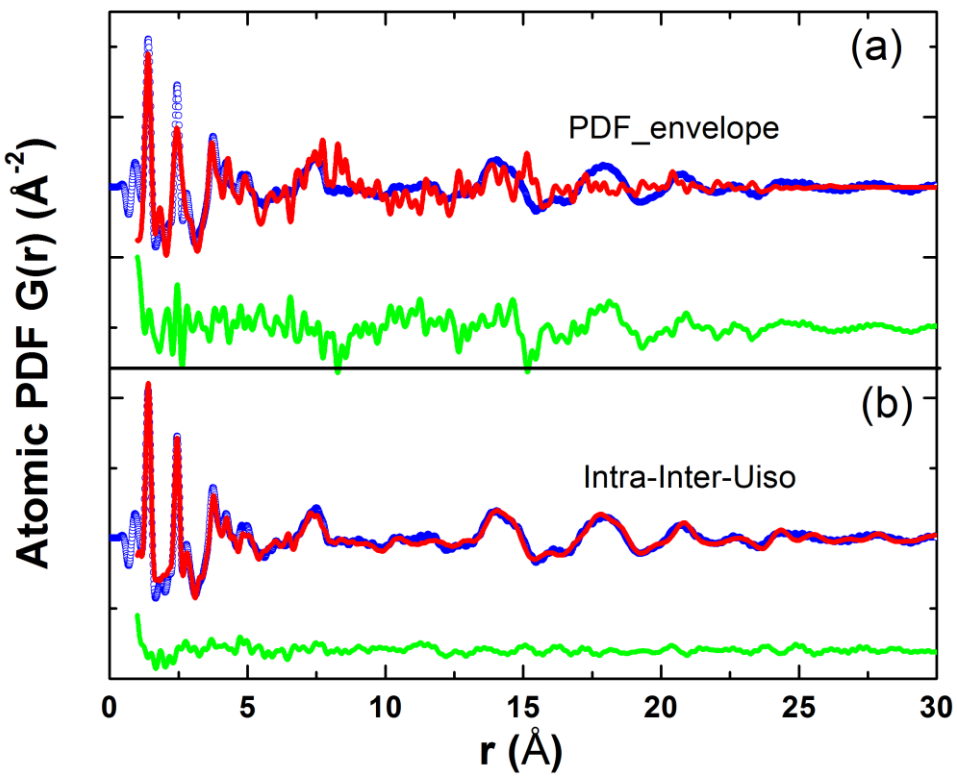


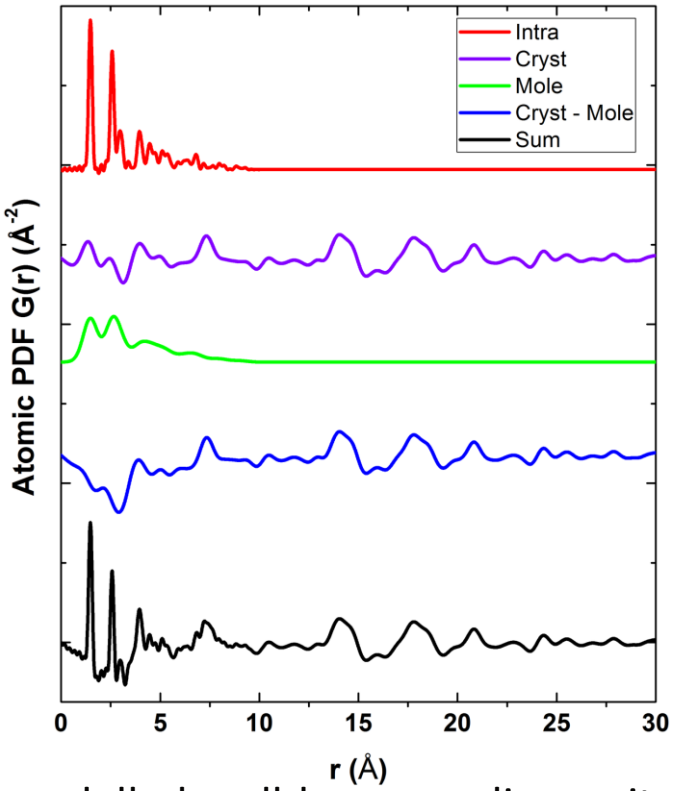
Mefenamic acid (MEF) is a nonsteroidal anti-inflammatory and analgesic drug used to treat mild pain, especially menstrual cramps. The crystal structure of mefenamic acid (Form I) was solved by McConnell and Company in 1976. It has a triclinic structure with a space group $\bar{P}1$.

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

J. F. McConnell and F. Z. Company, *Cryst. Struct. Commun.* 1976, 5, 861-864.

Example 2: Model fit of MEF crystalline PDF

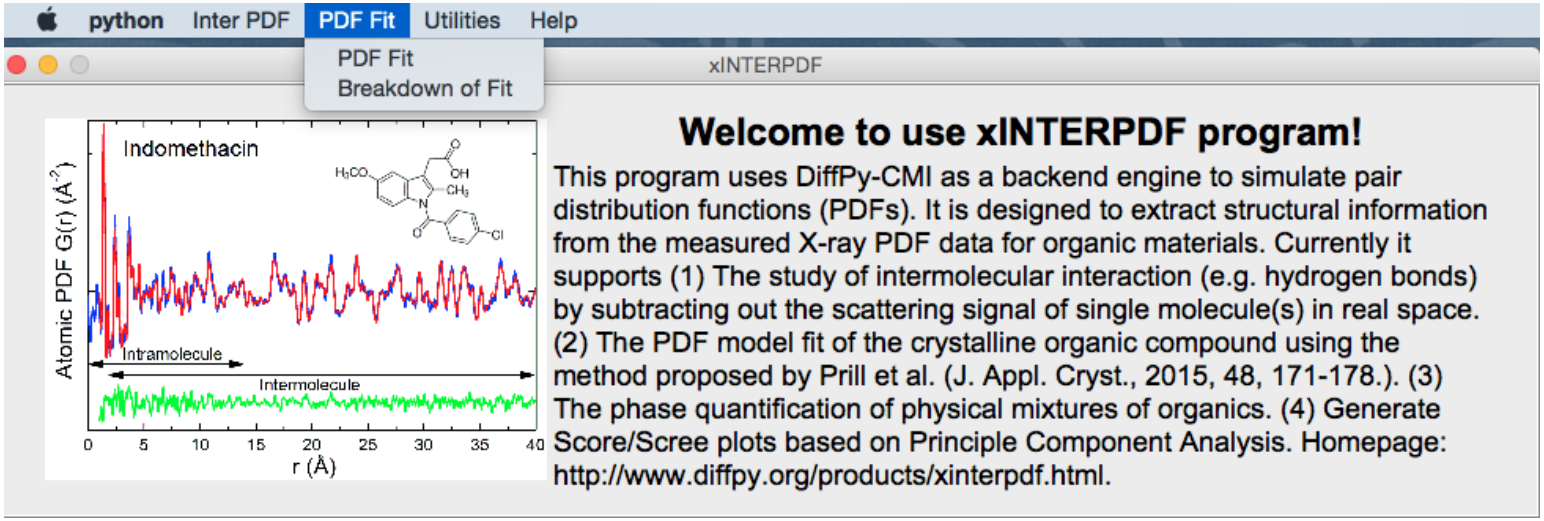




As Prill et al. reported, the PDF of organics cannot be modelled well by expanding unit cell as typically done for inorganic materials (left figure a), because the intermolecular forces are weaker than intramolecular ones. Instead, a model differentiating both intra- and intermolecular contributions 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.

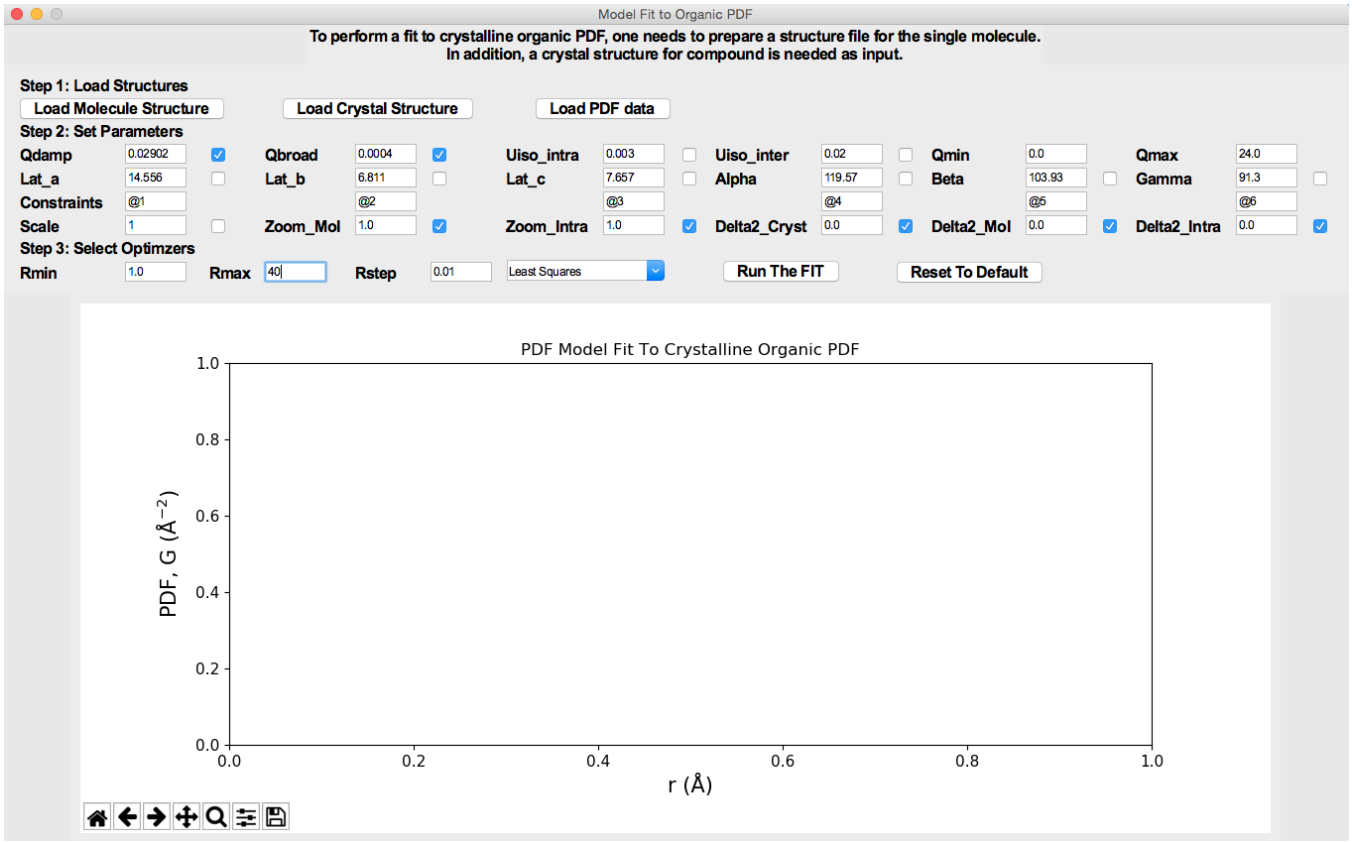
D. Prill, P. Juhás, M. U. Schmidt and S. J. L. Billinge, *J. Appl. Cryst.* 2015, 48, 171-178.

Example 2: Model fit of MEF crystalline PDF



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

Example 2: Model fit of MEF crystalline PDF



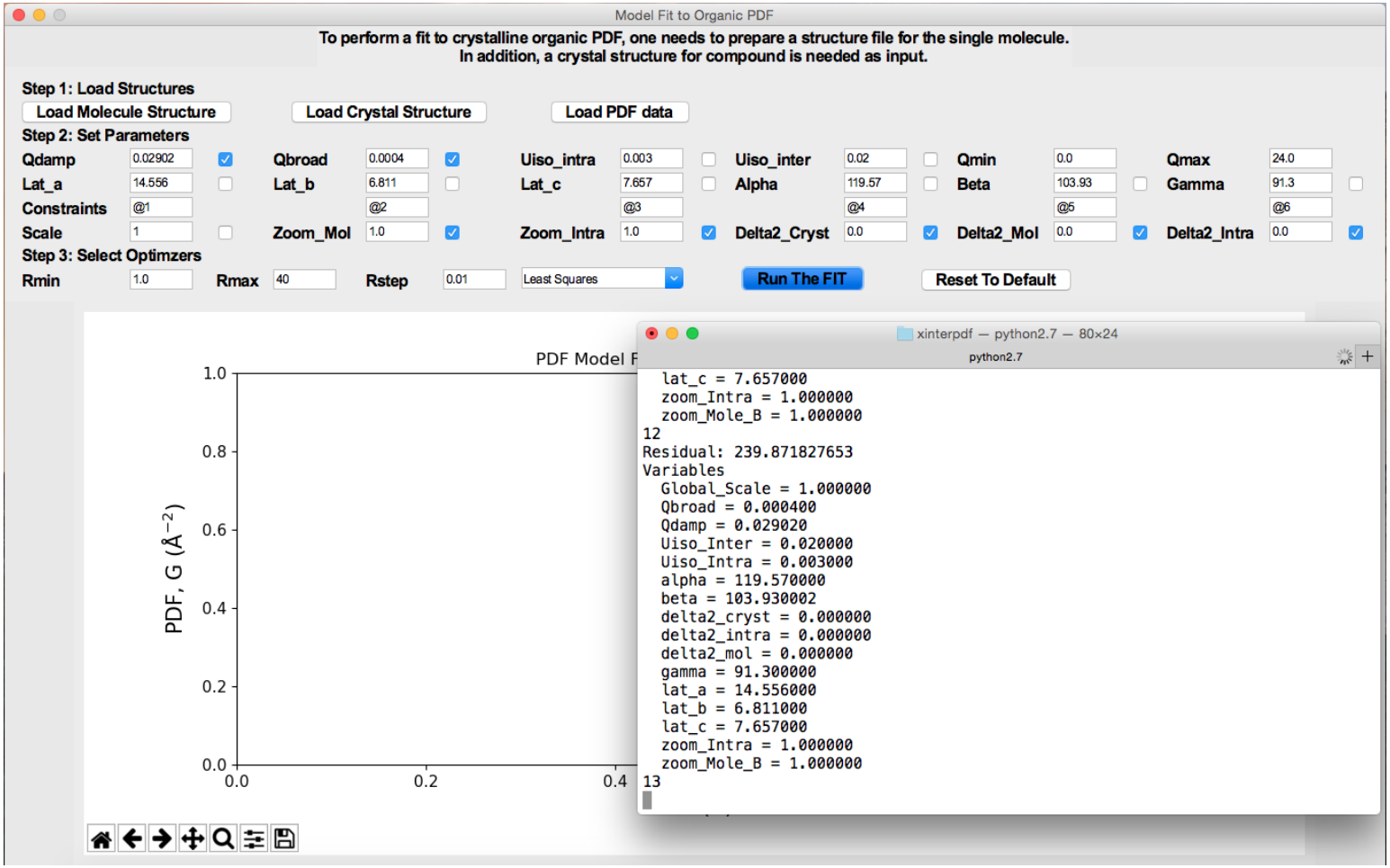
Following the steps, first load in structure files and PDF data (MEF.xyz, MEF.cif and MEF_300-00000.gr). In Step 2, specify the parameters. If a check box is marked, that parameter is fixed during the fit; otherwise is allowed to vary.

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 27, for the meaning of **Mol** and **Intra**.

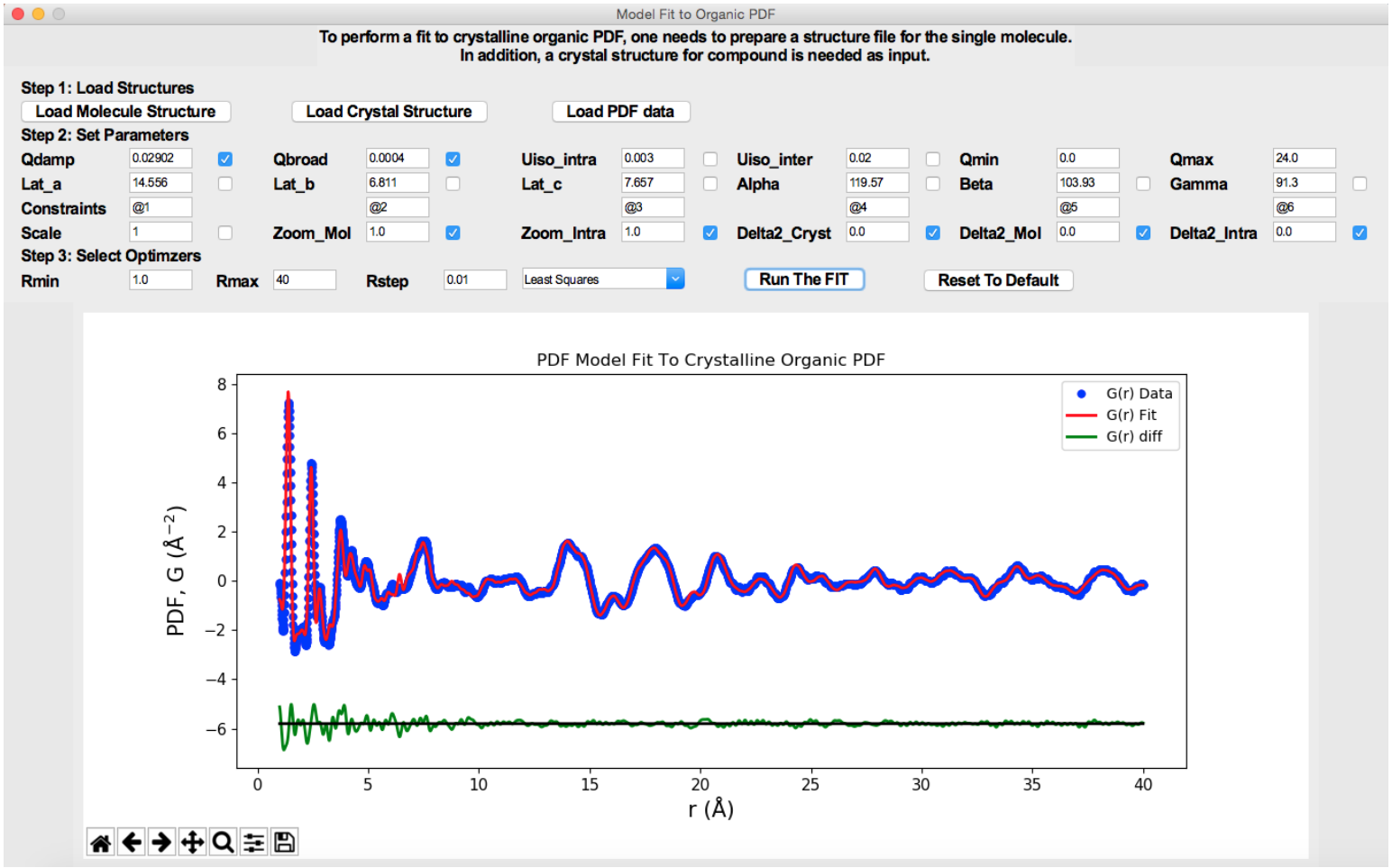
In Step 3, select the fit range and optimizer. Hit **Run The Fit** to start the run.

Example 2: Model fit of MEF crystalline PDF



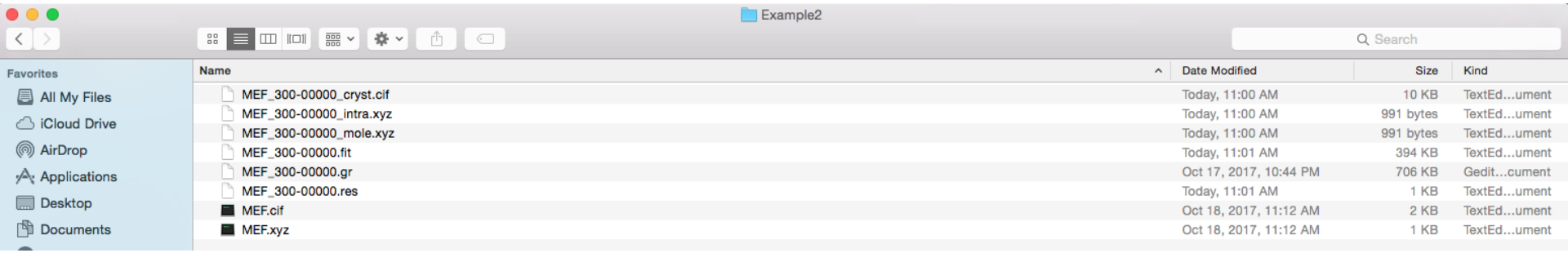
When the fit starts running, the terminal will update fit parameters in real-time.

Example 2: Model fit of MEF crystalline PDF



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

Example 2: Model fit of MEF crystalline PDF

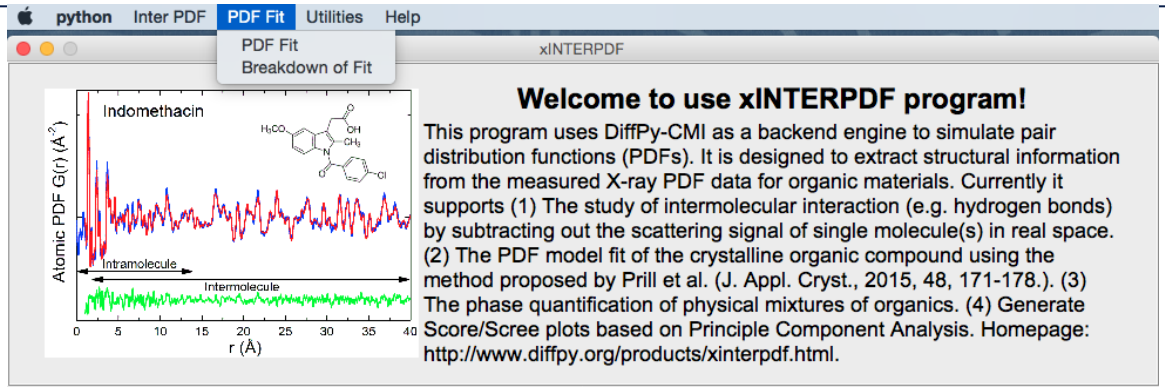


MEF_300-00000.res

```
Results written: Tue Feb  6 11:01:55 2018
produced by Chenyang

Cannot compute covariance matrix.
Some quantities invalid due to missing profile uncertainty
Overall (Chi2 and Reduced Chi2 invalid)
--------------------------------------------------------------------------------
Residual        101.75573582
Contributions   101.75573582
Restraints      0.00000000
Chi2            101.75573582
Reduced Chi2    0.02619195
Rw              0.18472270

Variables (Uncertainties invalid)
--------------------------------------------------------------------------------
Global_Scale   9.92587263e-01 +/- 0.00000000e+00
Qbroad         4.00000000e-04 +/- 0.00000000e+00
Qdamp          2.90200000e-02 +/- 0.00000000e+00
Uiso_Inter     4.15886353e-02 +/- 0.00000000e+00
Uiso_Intra     3.31484234e-03 +/- 0.00000000e+00
alpha          1.19516522e+02 +/- 0.00000000e+00
beta           1.04001075e+02 +/- 0.00000000e+00
delta2_cryst   0.00000000e+00 +/- 0.00000000e+00
delta2_intra   0.00000000e+00 +/- 0.00000000e+00
delta2_mol     0.00000000e+00 +/- 0.00000000e+00
gamma          9.12728874e+01 +/- 0.00000000e+00
lat_a          1.45539596e+01 +/- 0.00000000e+00
lat_b          6.81128742e+00 +/- 0.00000000e+00
lat_c          7.66290270e+00 +/- 0.00000000e+00
zoom_Intra     1.00000000e+00 +/- 0.00000000e+00
zoom_Mole_B    1.00000000e+00 +/- 0.00000000e+00

Variable Correlations greater than 25% (Correlations invalid)
--------------------------------------------------------------------------------
No correlations greater than 25%
```

After the fit is complete, a variety of files will be saved automatically to the folder where structure files exist. These include structure files (.cif and .xyz files). The MEF-300-00000.fit contains raw data for plotting. MEF-300-00000.res contains a detailed summary of fit results. See the screenshot shown to the left.

Example 3: Breakdown of the total PDF fit in Example 2



A functionality **Breakdown of Fit** has been built to calculate intra- and inter-molecular PDF contribution to the total fit. The interface is shown in the next page. It takes in five files generated after a fit to the crystalline organic X-ray PDF (as in example 2).

The five files from example 2 loaded into program are:

MEF-300-00000_intra.xyz

MEF-300-00000_mole.xyz

MEF-300-00000_cryst.cif

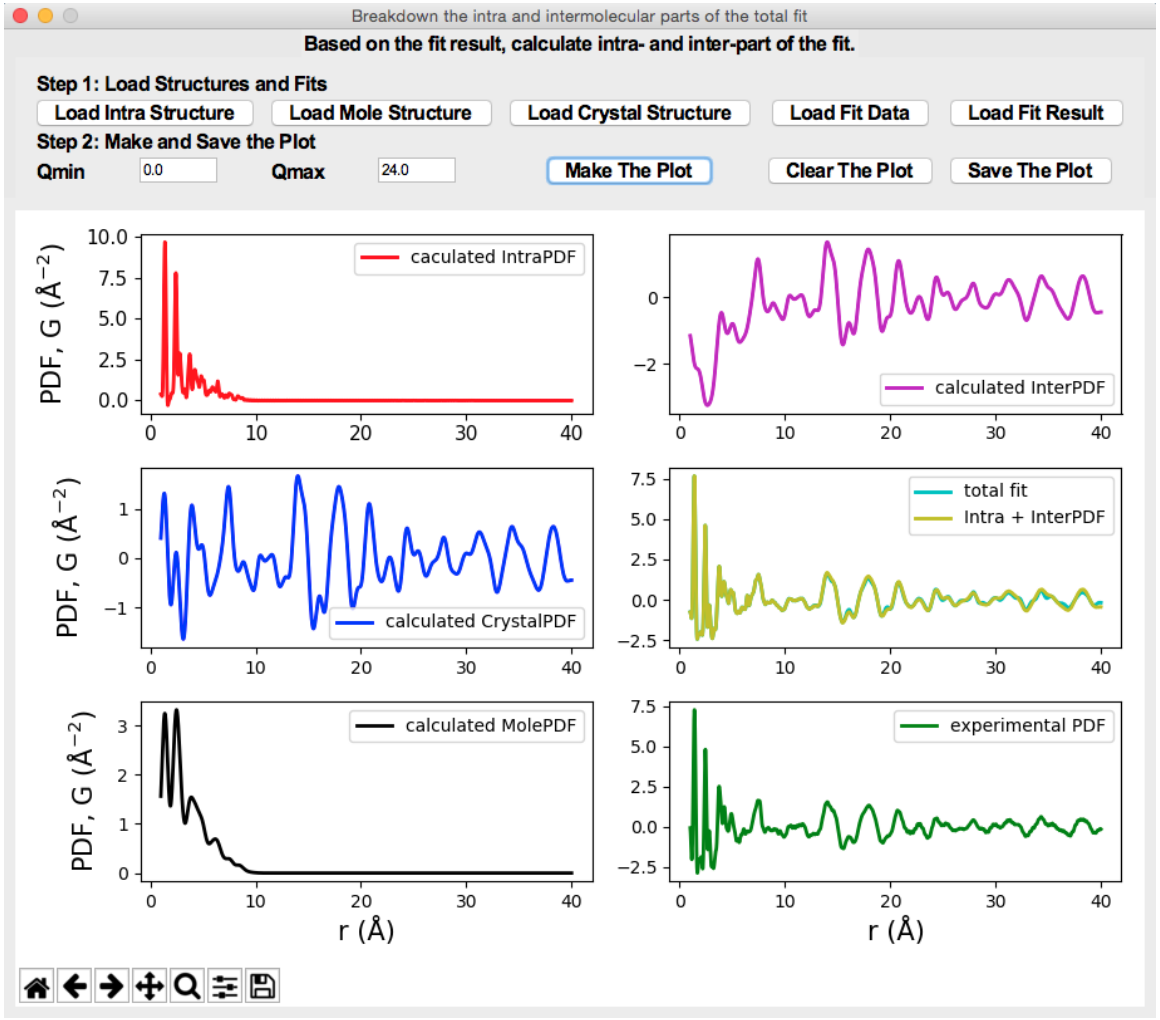
MEF-300-00000.fit

MEF-300-00000.res

Besides, in Step 2, the user needs to specify the **Qmin** and **Qmax** values used for simulation of molecular PDFs.

Example 3: Breakdown of the total PDF fit in Example 2

After loading all five files separately and specifying the Qmin/Qmax, hit **Make The Plot** will generate plots of each contribution to the total fit.



Example 4: Find phase fraction by regression

Phase quantification algorithm:

$$PDF_{simu.} = xA + yB + zC \quad (1)$$

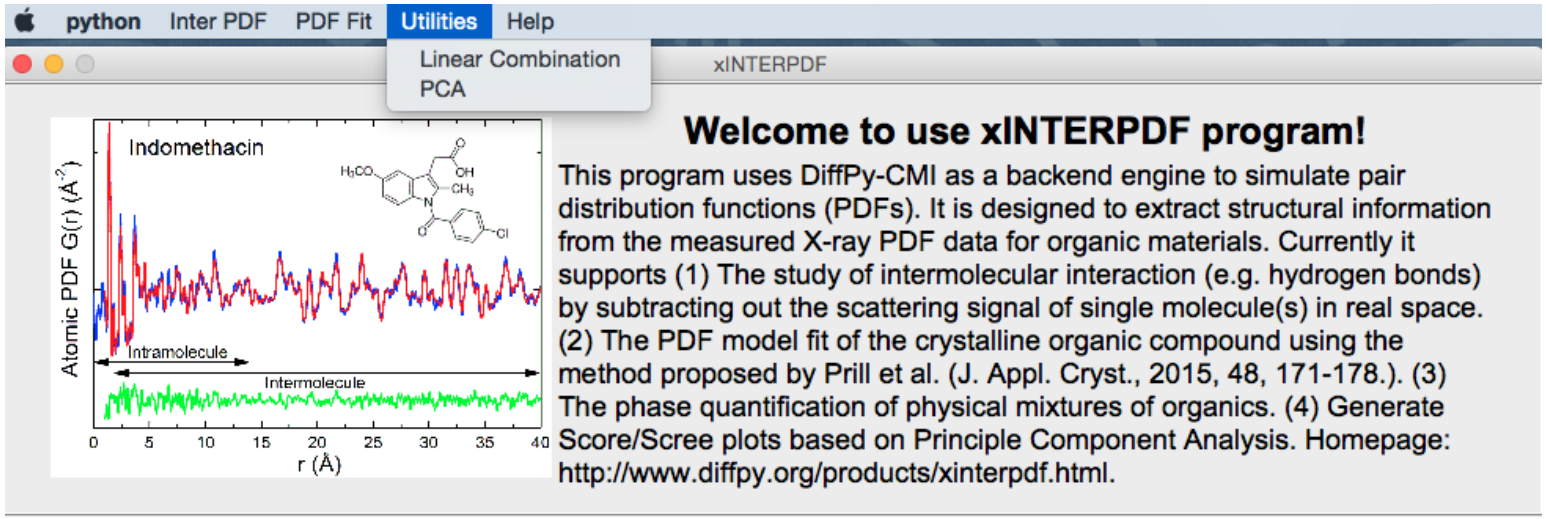
$$x + y + z = 1 \quad (2)$$

$$maximize\ Pearson(PDF_{exp.}, PDF_{simu.}) \quad (3)$$

Or minimize the difference between PDF_{exp} and PDF_{simu}

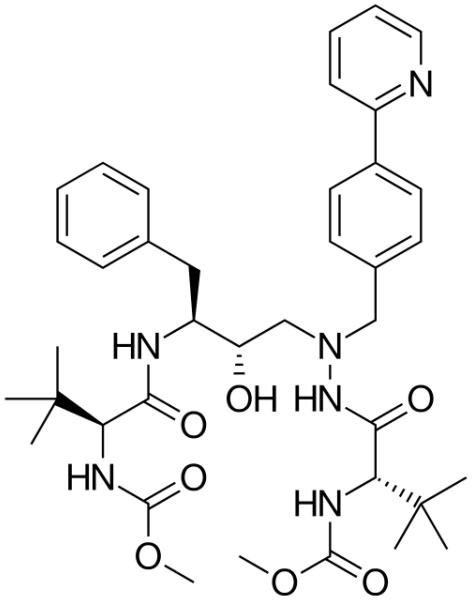
This exercise is also important for *differentiating amorphous solid dispersion (ASDs) and physical mixture* of amorphous drug and polymer, i.e. test the miscibility of drug and polymer. If the final product is a physical mixture, one should be able to scale the PDFs of each component to match the total PDF. However, if it is a pure ASD, the linear combination of PDFs of end members should not yield a desired result. More details see A. Newman et al., *J. Pharm. Sci.* 2008, 97(11):4840-56.

Example 4: Find phase fraction by linear regression

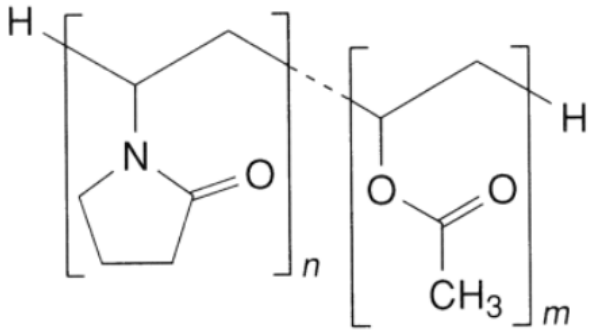


To perform a linear fit, in dropdown menu, click **Utility/Linear Combination** to start the GUI window.

Example 4: Work on amorphous solid dispersion data



Atazanavir



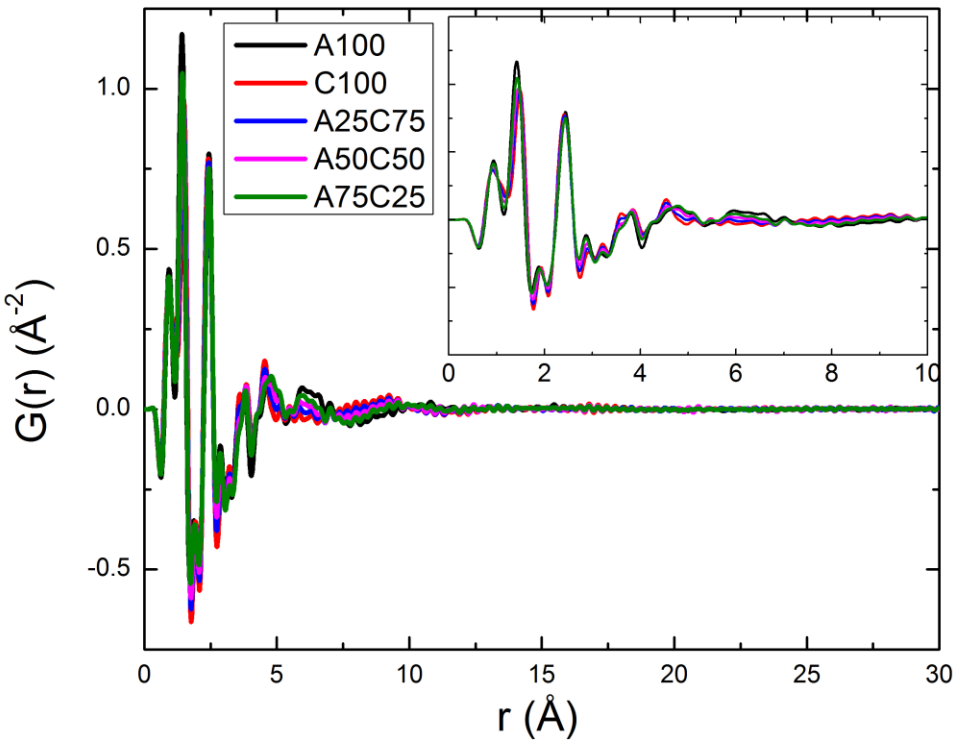
Copovidone

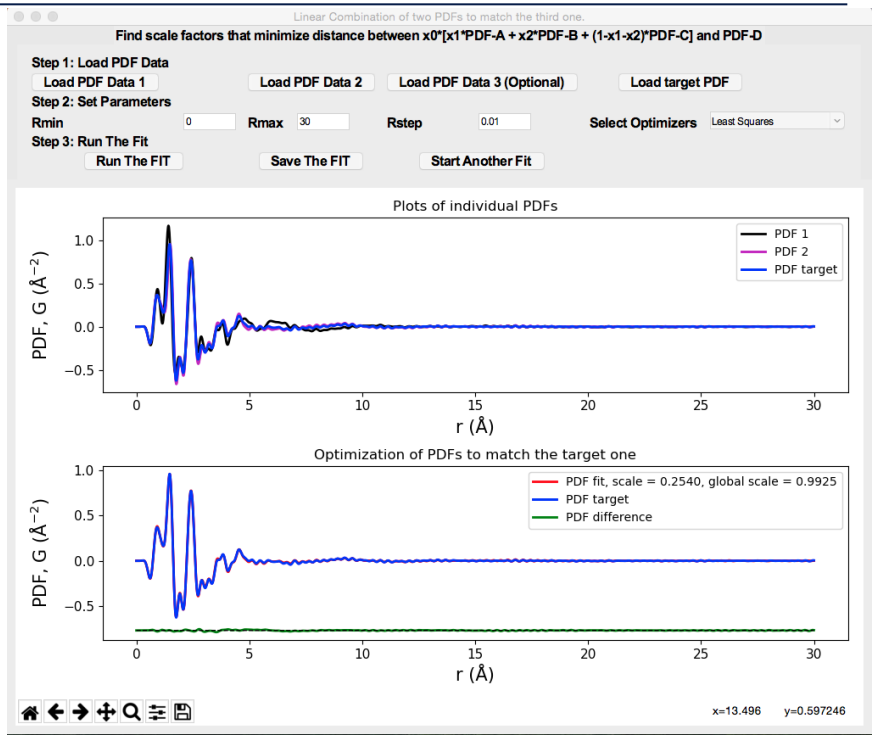
n = 1.16m

Amorphous solid dispersion (ASD) was prepared by rotovap. Polymer weight was corrected for water moisture (6.0 wt%). Based on weight percentage, the samples are named rA25C75, rA50C50, rA75C25, rC100 and mA100 (pure atazanavir amorphous phase was prepared by melt quench method).

Synchrotron X-ray PDF experiment was done at 11-ID-B at 250 K, with a Q_{max} = 20 $Å^{-1}$.

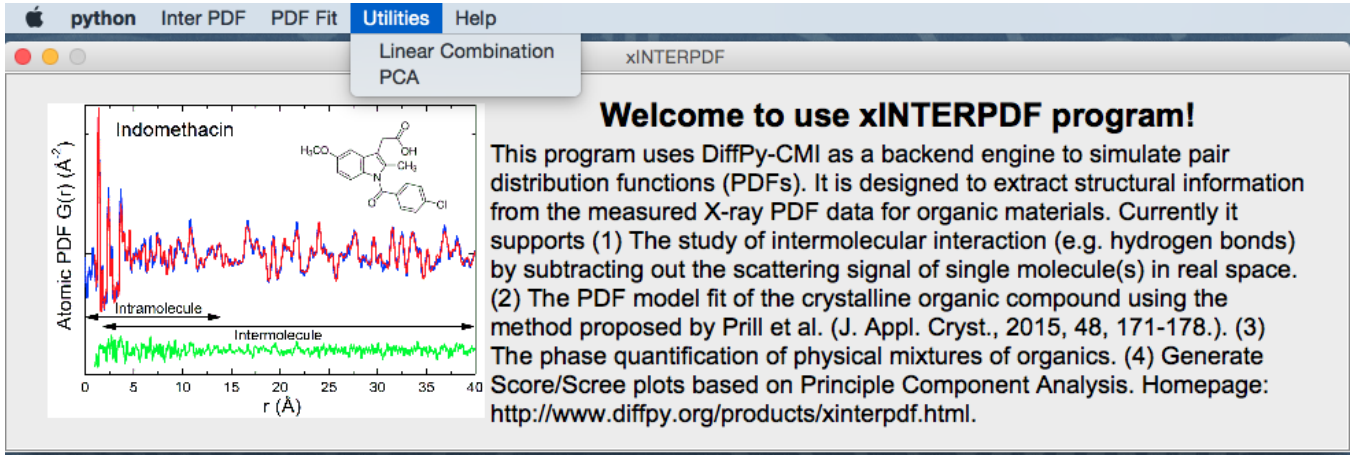
Example 4: Work on amorphous solid dispersion data





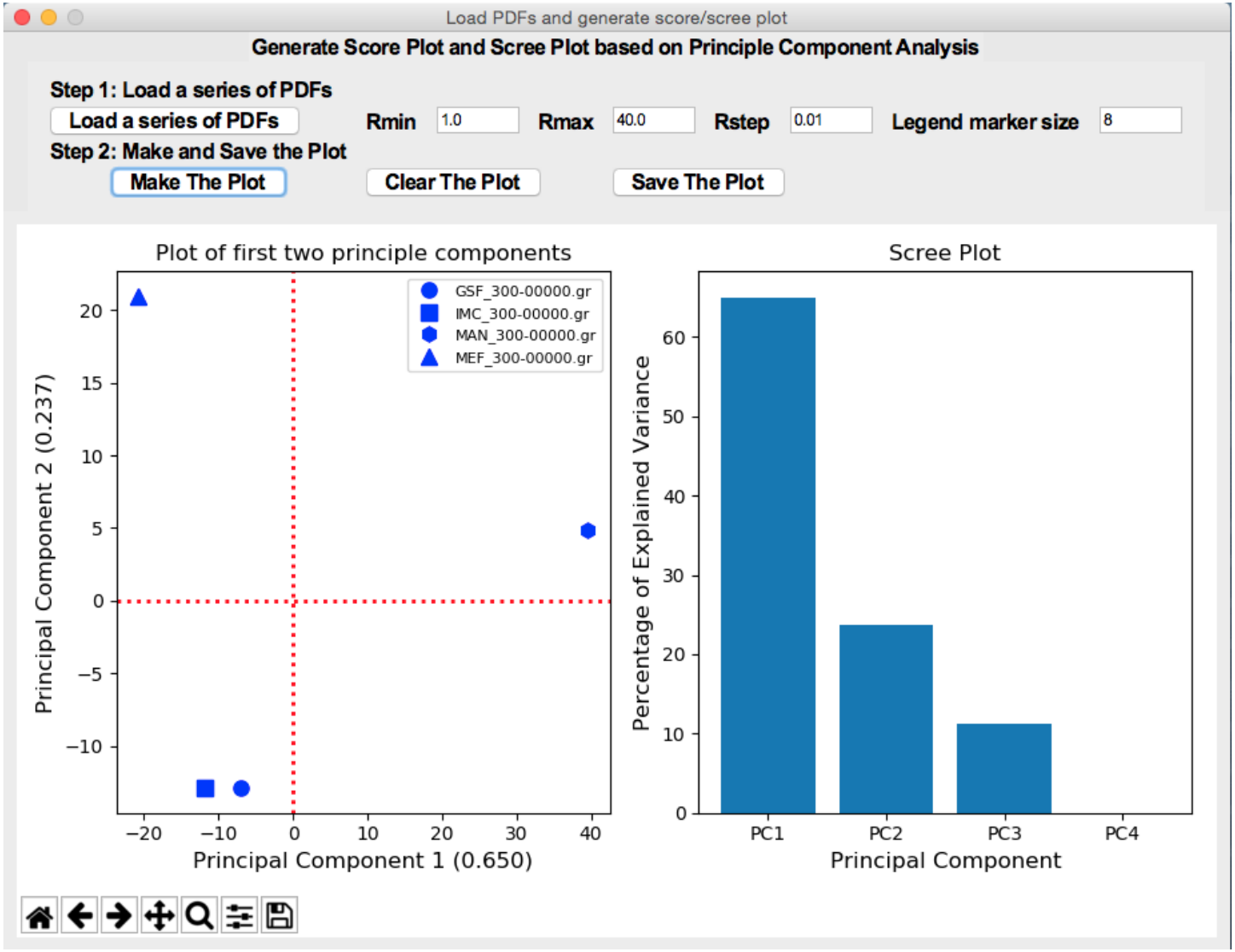
Samples	Nominal	Least Squares	Pearson	Least Absolute Deviation
rA25C75	25.0	25.4	25.4	24.8
rA50C50	50.0	47.7	47.8	46.3
rA75C25	75.0	77.8	77.8	75.8

Example 5: Principle Component Analysis (PCA)



Principle component analysis is an unsupervised learning method, which is used for dimensionality reduction, and clustering. A functionality **PCA** is built in **Utilities**. Given a set of PDF data, the program automatically generates Score plot and Scree plot at the same time. See next page for example usages, where synchrotron X-ray PDFs of indomethacin (IMC), mefenamic acid (MEF), mannitol (MAN) and griseofulvin (GSF) measured at RT are loaded into the program. It can be seen IMC and GSF cluster together in the Score plot, indicating similar structural features. By hitting **Save The Plot**, each principle component of every compound is saved.

Example 5: Principle Component Analysis (PCA)



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 Juhás, Prof. Lian Yu and Dr. Fabia Gozzo.