

---

# xINTERPDF User Guide v.0.1.0

Chenyang Shi, Senior Scientist at AbbVie Inc.

Updated June 14<sup>th</sup>, 2018



---

Overview

Installation

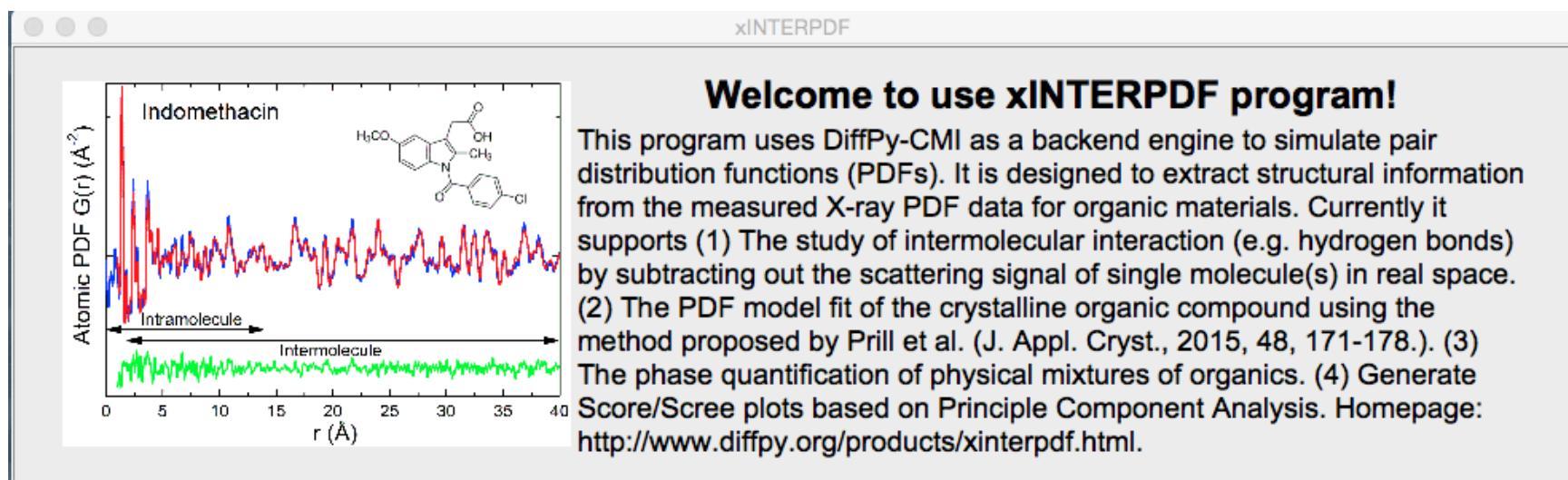
Technical Terms

Examples

Disclosure/Acknowledgement

# Overview

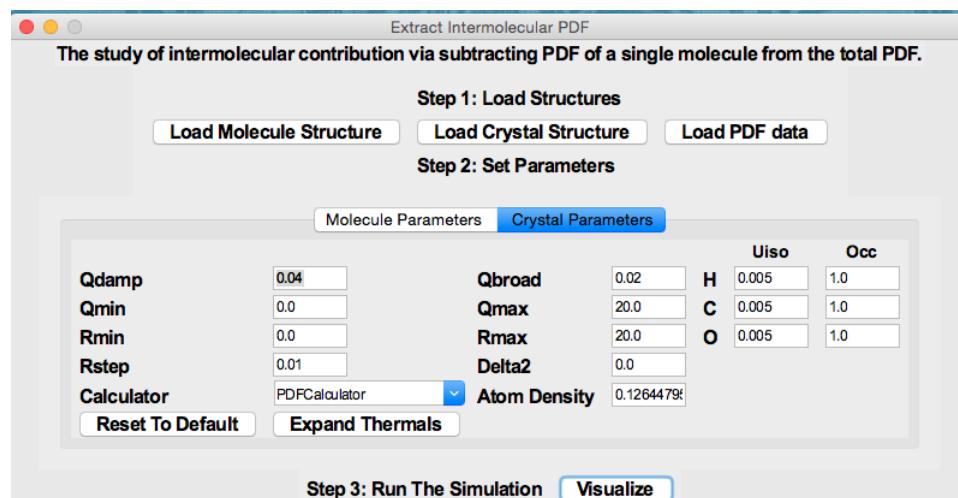
xINTERPDF is written in Python 2.7. It utilizes Tkinter and matplotlib modules for creating a Graphical User Interface (GUI) and plot visualization, and NumPy, SciPy and Scikit-learn (i.e. PCA feature) for scientific calculations. It builds on DiffPy-CMI package (<http://www.diffpy.org/products/diffpycmi/index.html>) 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

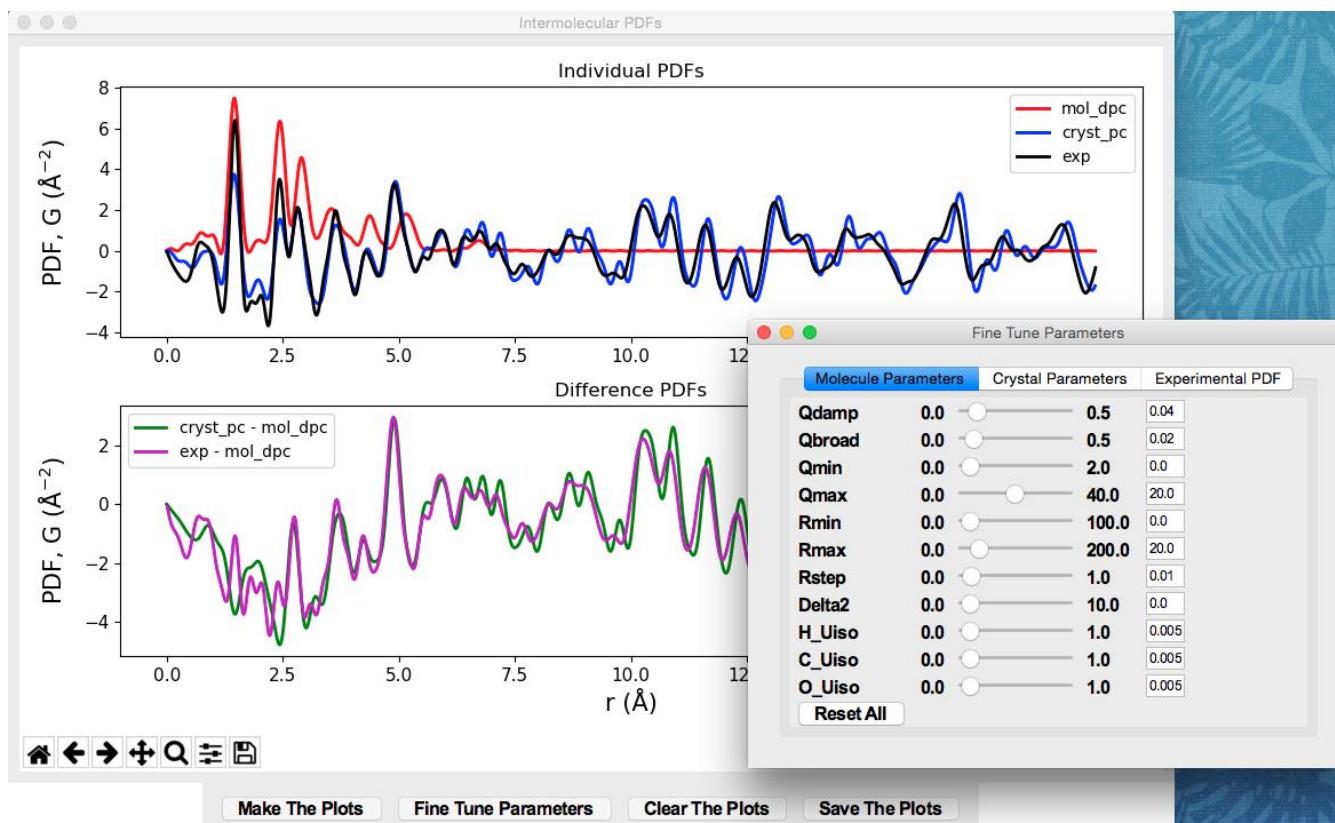
Users may use xINTERPDF to study the molecule-molecule interaction in organics. The program considers three common scenarios—the organic compound of interest is (1) crystalline (2) amorphous and (3) amorphous solid dispersion (two components). Shown below is an example when compound is crystalline, other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



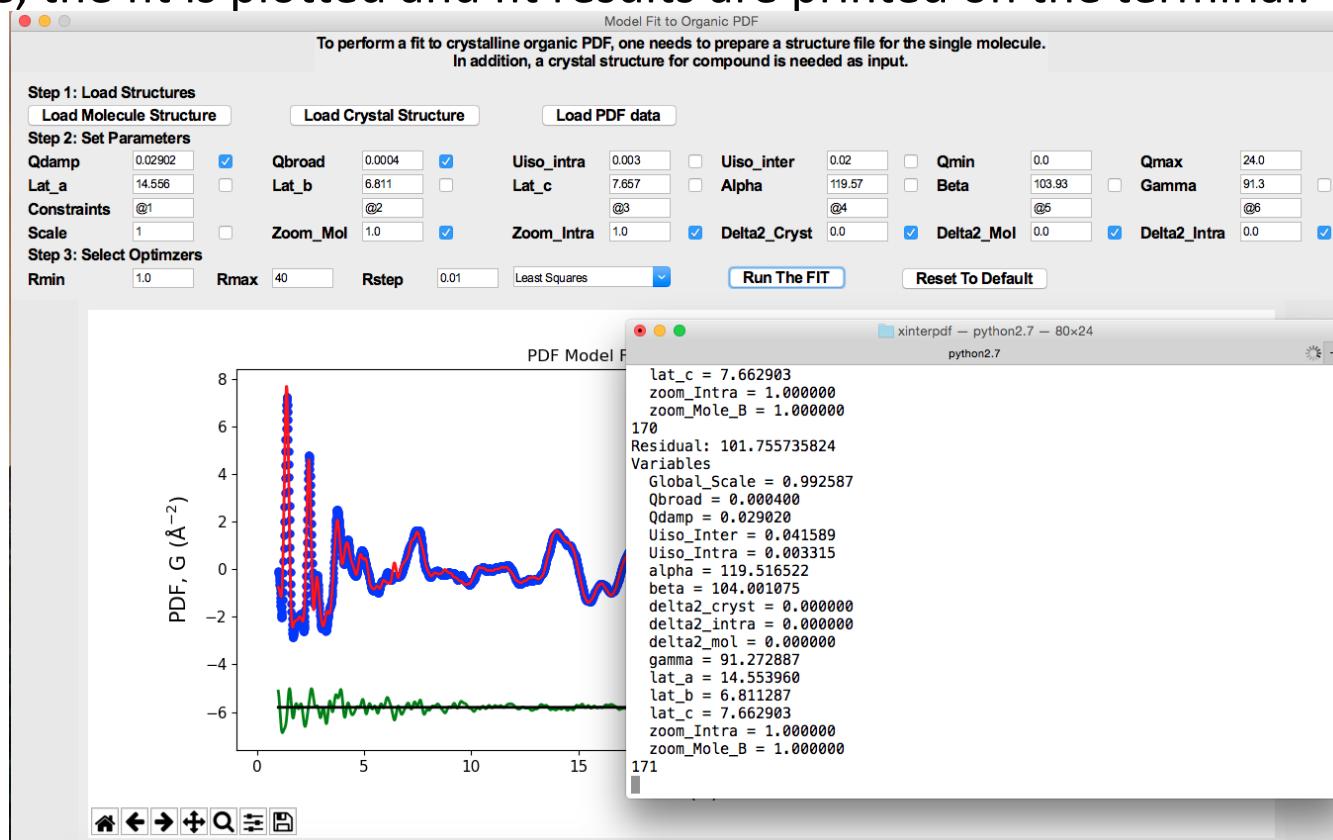
# Usage 1: Extracting Intermolecular PDF

The screenshot shows two windows of the software. The left window is titled "Extract Intermolecular PDF of Amorphous Compounds" and has tabs for "Step 1: Load Structures" (with "Load Molecule Structure" and "Load PDF data" buttons), "Step 2: Set Parameters" (with a "Molecule Parameters" group containing fields for Qdamp, Qmin, Rmin, Rstep, Calculator, Qbroad, Qmax, Rmax, and Delta2, and buttons for "Reset To Default" and "Expand Thermals"), and "Step 3: Run The Simulation" (with "Visualize" and other buttons). The right window is titled "Extract Intermolecular PDF of ASDs" and has tabs for "Step 1: Load Structures" (with "Load Molecule Structure 1", "Load Molecule Structure 2", and "Load PDF data" buttons), "Step 2: Set Parameters" (with "Molecule 1 Parameters" and "Molecule 2 Parameters" groups containing fields for Qdamp, Qmin, Rmin, Rstep, Calculator, Qbroad, Qmax, Rmax, and Delta2, and a "Scale Factor" field set to 1.0, along with "Reset To Default" and "Expand Thermals" buttons), and "Step 3: Run The Simulation" (with "Visualize" and other buttons). Both windows have a status bar at the bottom.

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.

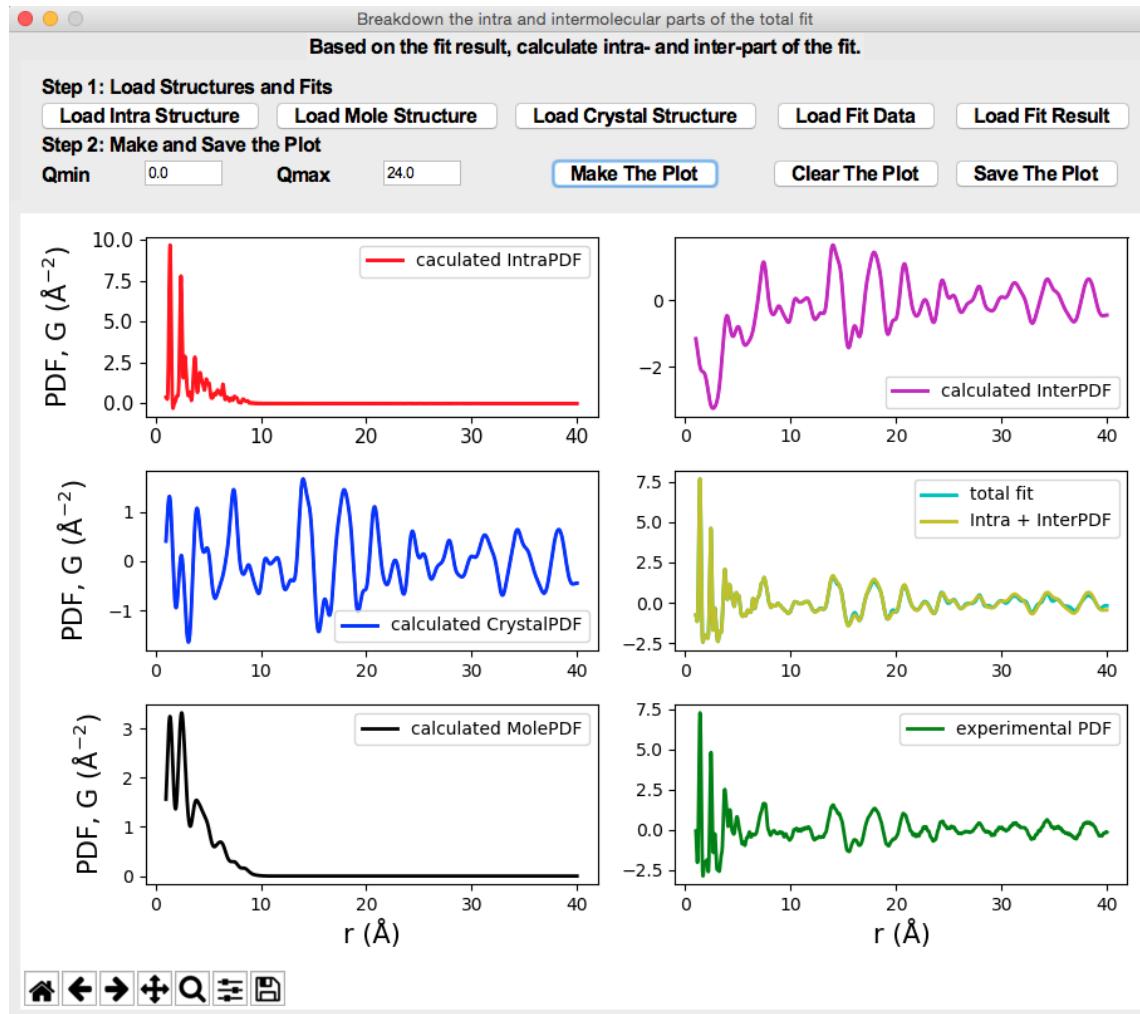
# Usage 2: Model fit to organic crystalline PDF

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 results are plotted and fit results are printed on the terminal.



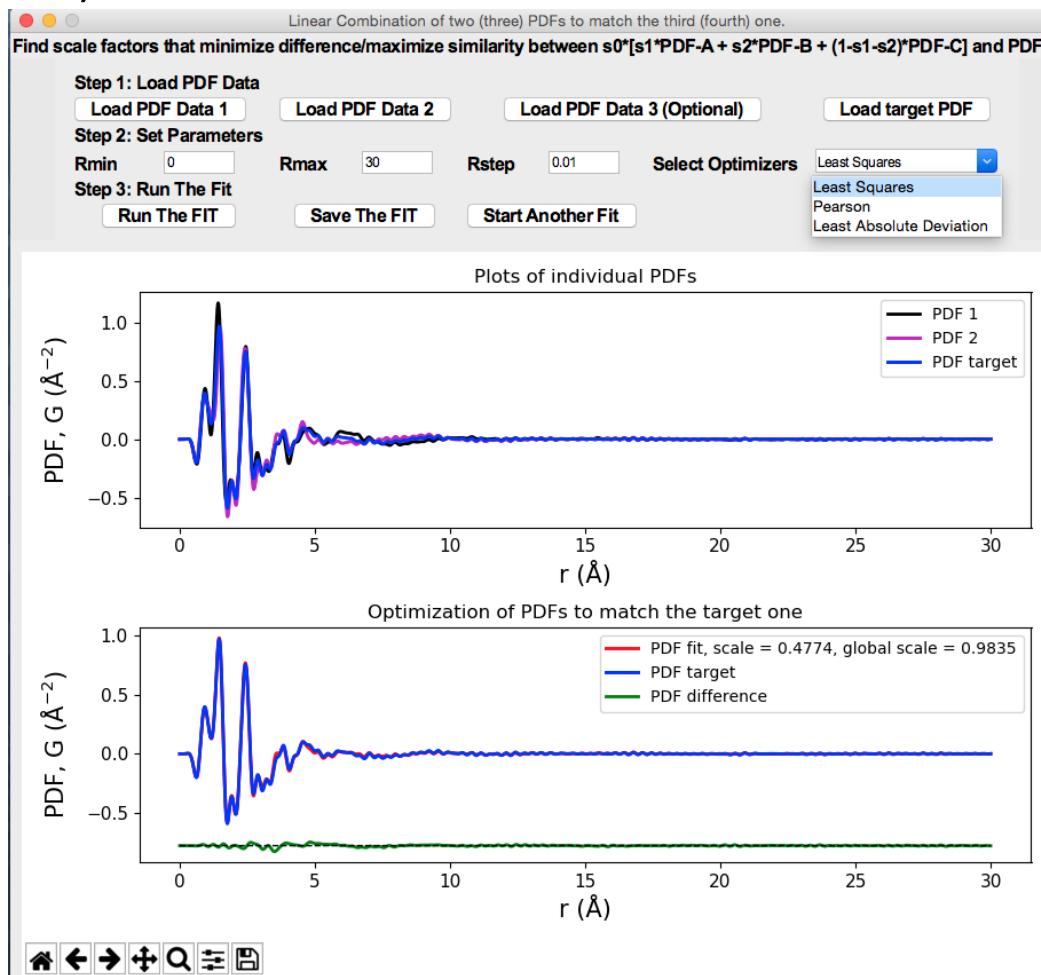
## Usage 3: a breakdown of the model fit to organic crystalline PDF

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



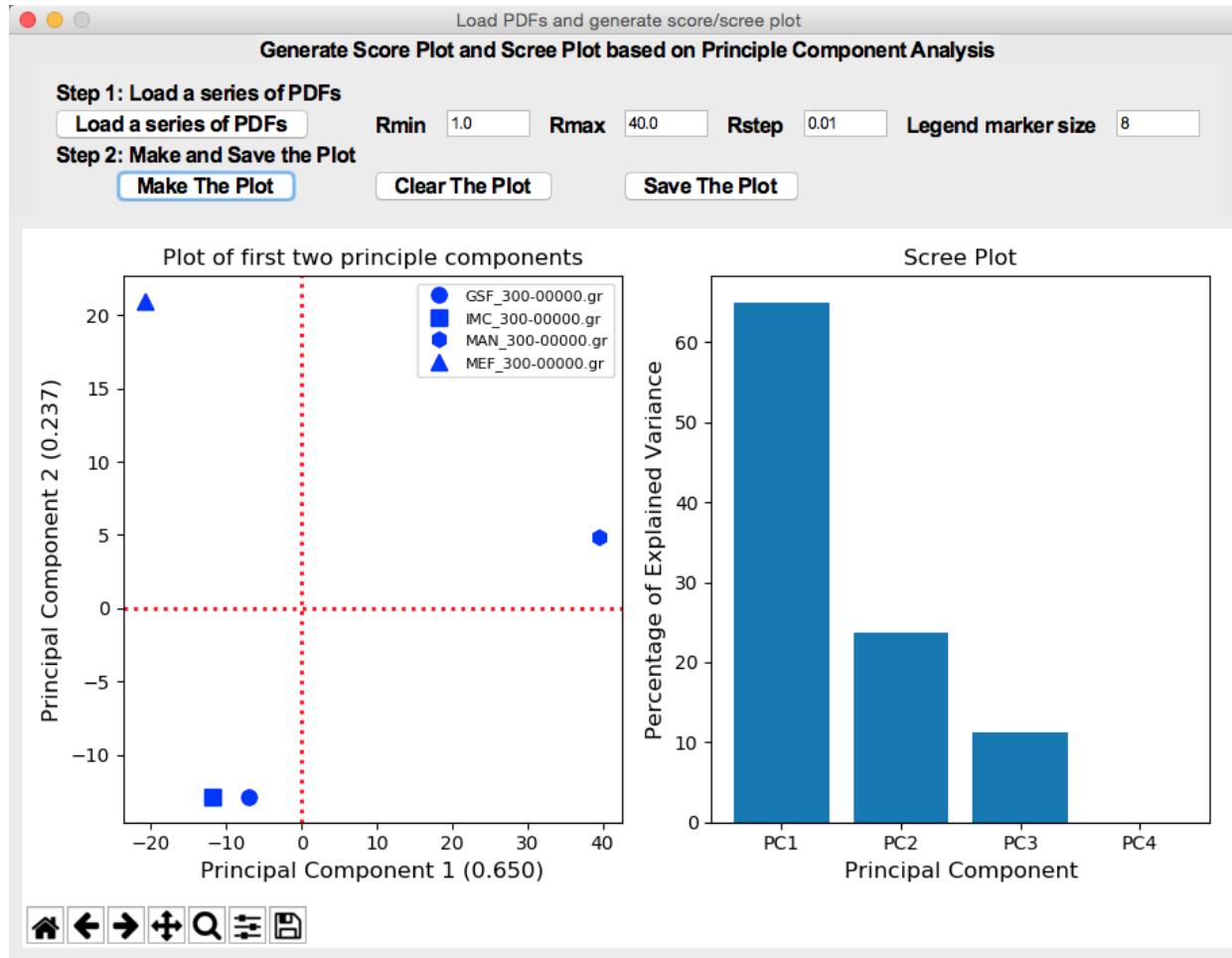
# Usage 4: phase quantification

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

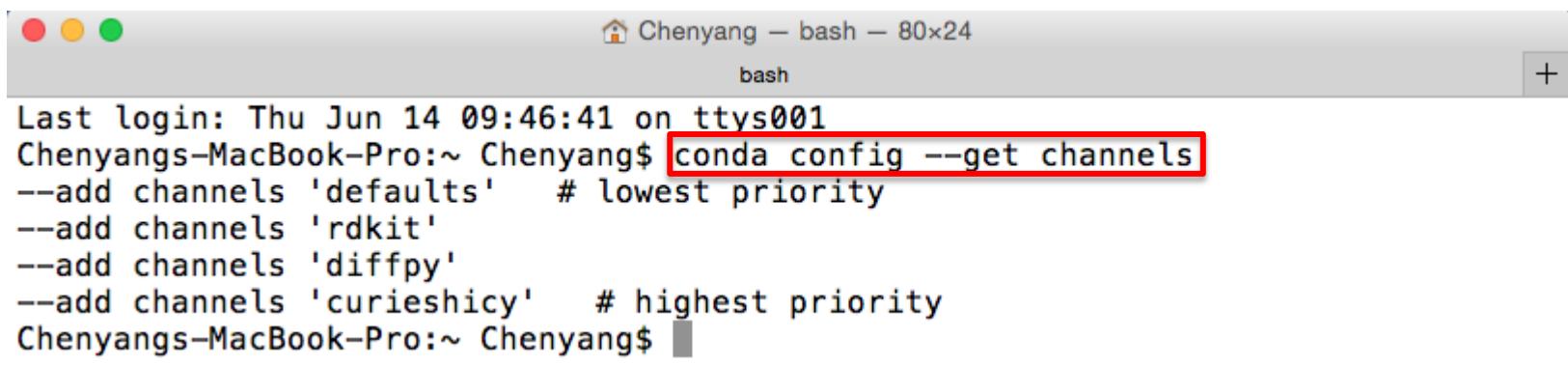
The program also provides an easy interface for generating Scree/Score plots of PDFs based on PCA analysis.



# Installation

xINTERPDI 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.



```
Last login: Thu Jun 14 09:46:41 on ttys001
Chenyangs-MacBook-Pro:~ Chenyang$ conda config --get channels
--add channels 'defaults'    # lowest priority
--add channels 'rdkit'
--add channels 'diffpy'
--add channels 'curieshicy'   # highest priority
Chenyangs-MacBook-Pro:~ Chenyang$
```

# Installation

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

The terminal window shows the command `conda create -c curieshicy -n xinterpdf xinterpdf` highlighted with a red box. The output shows the package plan and the list of packages to be installed, including `xinterpdf` and its dependencies.

```
Chenyangs-MacBook-Pro:~ Chenyang$ conda create -c curieshicy -n xinterpdf xinterpdf
Solving environment: done
## Package Plan ##
environment location: /Users/Chenyang/anaconda2/envs/xinterpdf

added / updated specs:
  - xinterpdf

The following NEW packages will be INSTALLED:

appnope:           0.1.0-py27hb466136_0
backports:         1.0-py27hb4f9756_1
backports.shutil_get_terminal_size: 1.0.0-py27hc9115de_2
backports._abc:    0.5-py27h6972548_0
blas:              1.0-mkl
bleach:            2.1.3-py27_0
boost:              1.61.0-py27_0
ca-certificates:  2018.03.07-0
certifi:            2018.4.16-py27_0
configparser:      3.5.0-py27hc7edf1b_0
cycler:             0.10.0-py27hfc73c78_0
dbus:               1.13.2-h760590f_1
decorator:          4.3.0-py27_0
diffpy-cmi:        2.1-hfb420b8_0      diffpy
diffpy.srfit:      1.3-py27h491f910_0    diffpy
diffpy.srreal:     1.2-py27_0      diffpy
diffpy.structure:  1.3.5-py27_0      diffpy
diffpy.utils:       1.2.2-py27_0      diffpy
entrypoints:        0.2.3-py27hd680fb1_2
enum34:              1.1.6-py27hf475452_1
expat:                2.2.5-hb8e80ba_0
freetype:             2.8-h12048fb_1
functools32:        3.2.3.2-py27h8ceab06_1
futures:             3.2.0-py27h1b80678_0

pyobjcryst:        2.0.2-py27_2      diffpy
pyparsing:          2.2.0-py27h5bb6aaf_0
pyqt:                5.9.2-py27h11d3b92_0
python:              2.7.15-h138c1fe_0
python-dateutil:   2.7.3-py27_0
pytz:                2018.4-py27_0
pyzmq:               17.0.0-py27h1de35cc_1
qt:                  5.9.5-h02808f3_0
qtconsole:          4.3.1-py27hd90b4f_0
readline:             7.0-hc1231fa_4
scandir:             1.7-py27h1de35cc_0
scikit-learn:       0.19.1-py27h9788993_0
scipy:                1.1.0-py27hcaad992_0
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
srfr-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)?
```

# 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 — 112x37
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
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 — 104x23
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
```

Welcome to use xINTERPDF program!

This program uses DiffPy-CMI as a backend engine to simulate pair distribution functions (PDFs). It is designed to extract structural information from the measured X-ray PDF data for organic materials. Currently it supports (1) The study of intermolecular interaction (e.g. hydrogen bonds) by subtracting out the scattering signal of single molecule(s) in real space. (2) The PDF model fit of the crystalline organic compound using the method proposed by Prill et al. (J. Appl. Cryst., 2015, 48, 171-178.). (3) The phase quantification of physical mixtures of organics. (4) Generate Score/Scree plots based on Principle Component Analysis. Homepage: <http://www.diffpy.org/products/xinterpdf.html>.

# 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/xinterpdf](https://github.com/curieshicy/xINTERPDF/tree/master/Conda_Build_Recipe/xinterpdf). 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

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

Where  $\sigma_{ij}$  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  $\langle f \rangle$  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.  $\rho_0$  is the atomic density.  $\sigma'_{ij}$  is the root mean squared displacement coming from the atomic displacement parameters (ADP) tensors of the atom-pair.  $\delta_1$  and  $\delta_2$  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.

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

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

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

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

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<sub>damp</sub>**: 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<sub>broad</sub>**: 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<sub>min</sub>**:

-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<sub>max</sub>**:

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

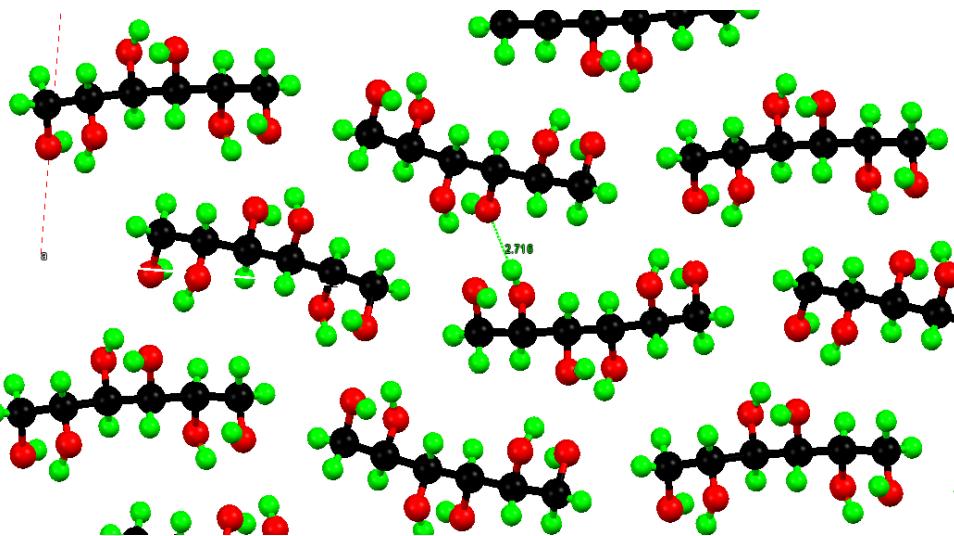
-DPC: Upper bound of the Q-grid for the calculated F(Q). Affects the termination ripples.

**R<sub>min</sub>**: Lower bound of the r-grid for PDF calculation

**R<sub>max</sub>**: Upper bound of the r-grid for PDF calculation.

**R<sub>step</sub>**: 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  $\sim 2.72 \text{ \AA}$ .

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:  $Q_{\text{damp}} = 0.02902 \text{ \AA}^{-1}$ ,  $Q_{\text{broad}} = 0.0004 \text{ \AA}^{-1}$ . A  $Q_{\text{max}}$  of  $24 \text{ \AA}^{-1}$  was used for Fourier transform. The software program PDFgetX2 was used.

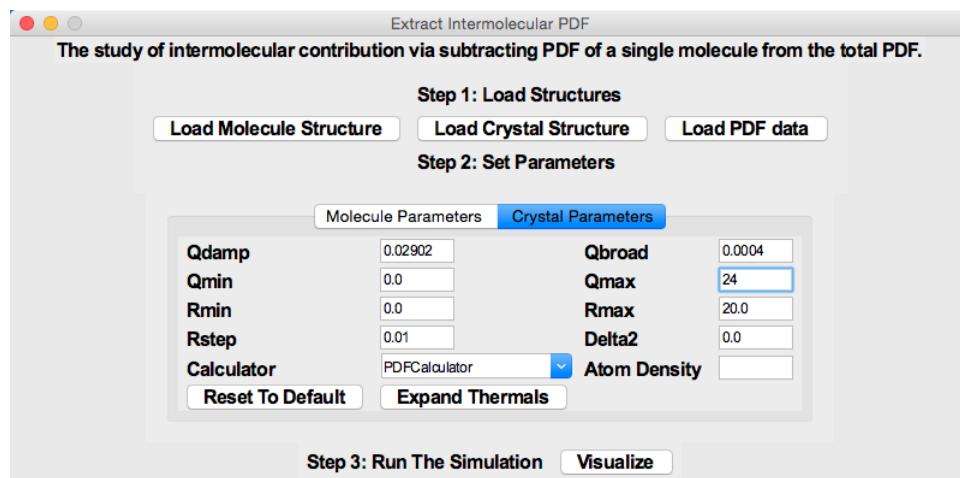
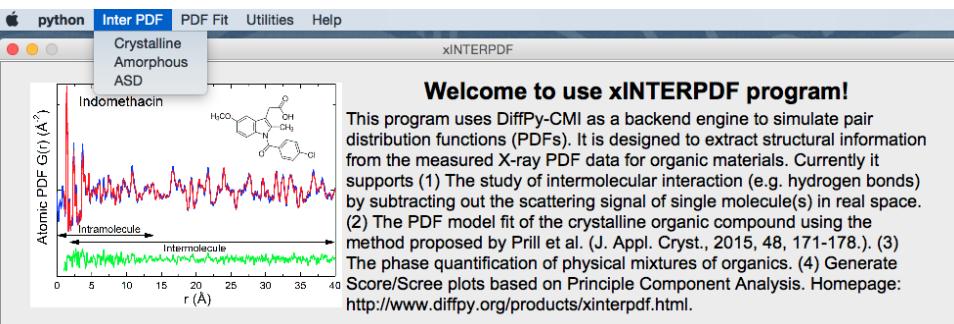
Examples files are available at

<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



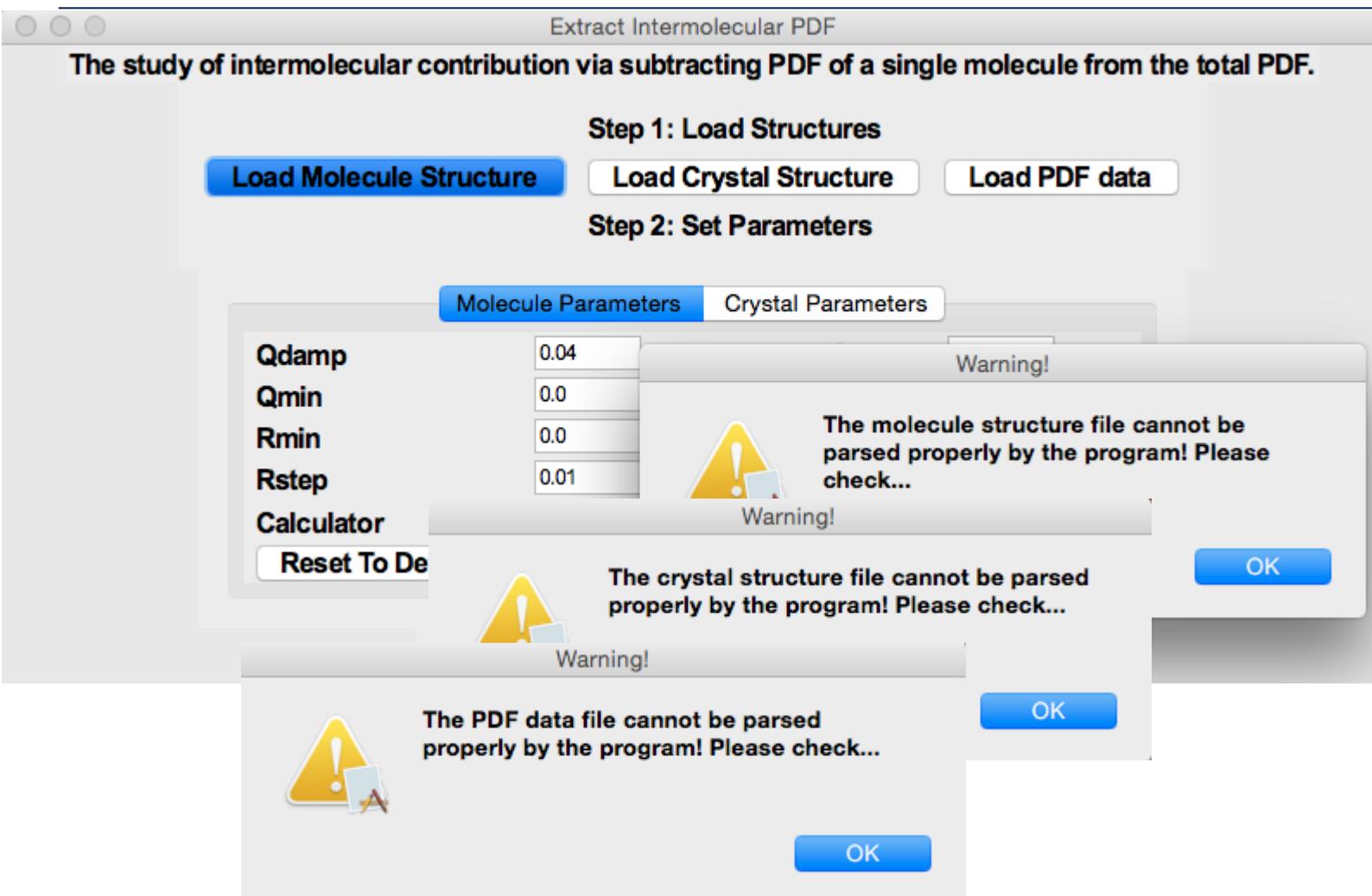
After invoke the main window in a terminal, select from drop-down menu, **Inter PDF /Crystalline**, which brings up a new window titled **Extract Intermolecular PDF**.

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  $Q_{\text{damp}}$  and  $Q_{\text{broad}}$  values to those determined from cerium oxide calibrant; set  $Q_{\text{max}}$  to  $24 \text{ \AA}^{-1}$ .

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

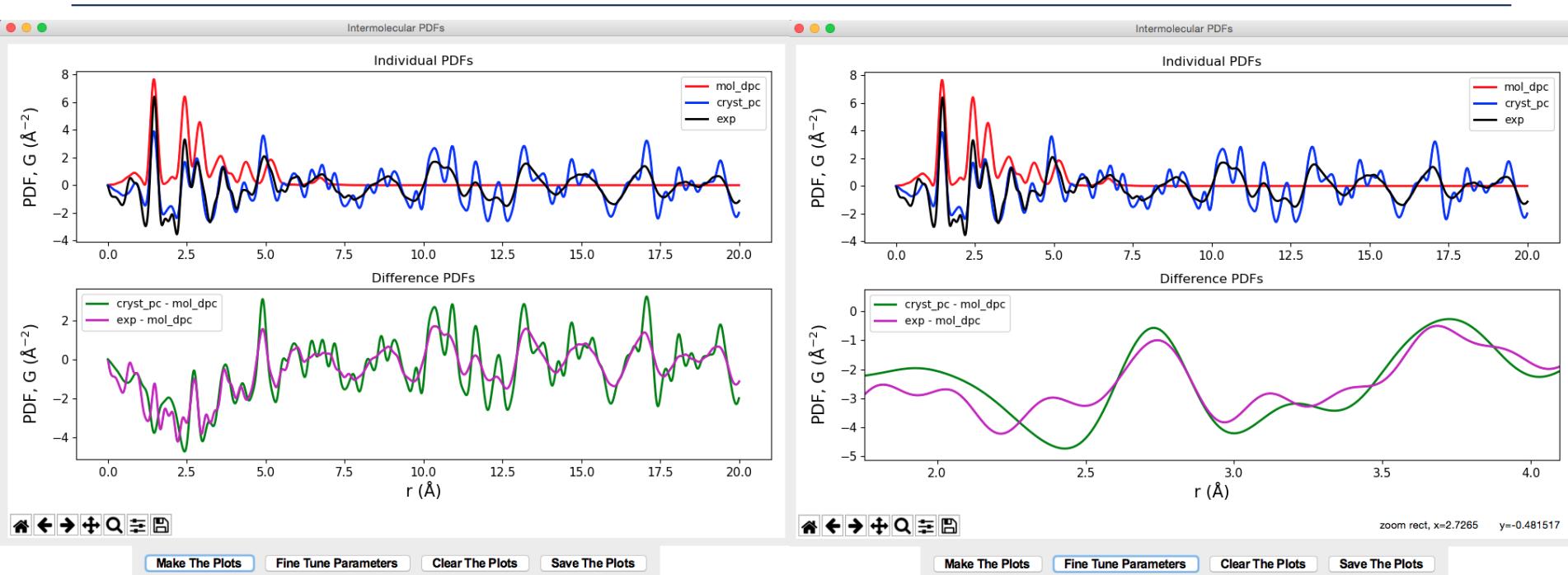
In step 3, click **Visualize** to see the plots.

# Example 1: Hydrogen bonds in D-mannitol



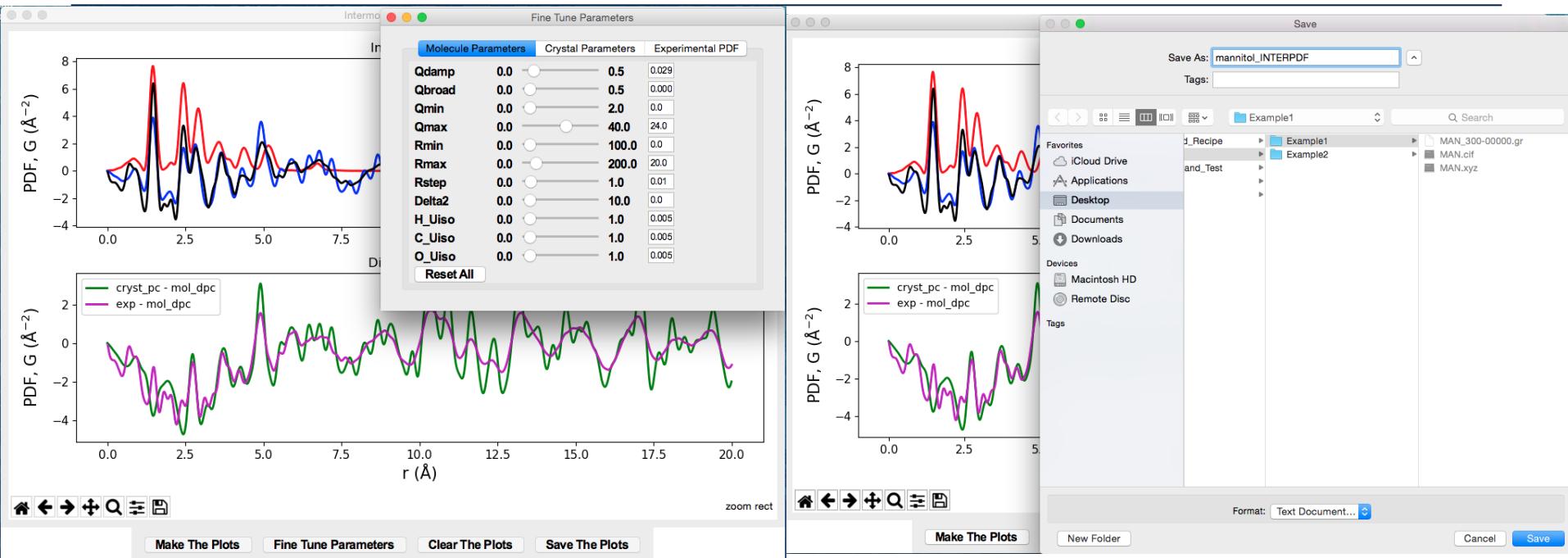
If any of the molecule/crystal structure file or PDF data cannot be recognized by the program, an error message will be displayed to alert the users.

# 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.

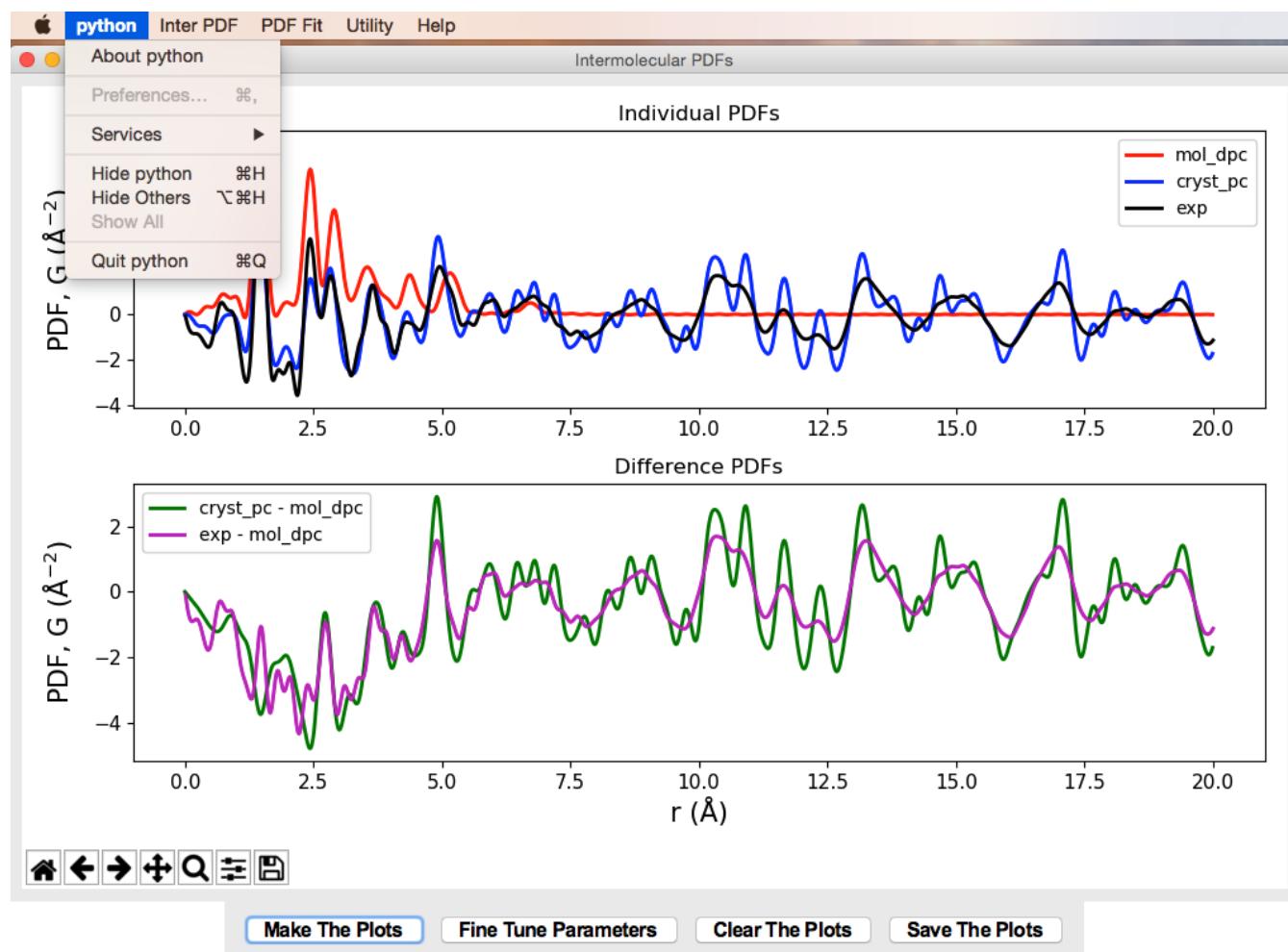
Click **Clear the Plots** to erase the plots. Click **Save The Plots** to save all raw data.

# Example 1: Hydrogen bonds in D-mannitol

```
# ****
# *****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 ****
# ****
# ****
# 0.000000000000000e+00 0.000000000000000e+00 0.000000000000000e+00 -9.280011000000000509e-02 0.000000000000000e+00 -9.280011000000000509e-02
1.00000000000000021e-02 6.212443191915230020e-03 -1.059388955578956237e-02 -1.841415000000000135e-01 -1.680633274770479413e-02 -1.903539431919152314e-01
2.00000000000000042e-02 1.222590932327609001e-02 -2.149921887044312432e-02 -2.72617200000000040e-01 -3.372512819371921433e-02 -2.848431093232761113e-01
2.99999999999999889e-02 1.786282325375071478e-02 -3.300927702376095507e-02 -3.56921700000000084e-01 -5.087210027751166985e-02 -3.747845232537507232e-01
4.000000000000000083e-02 2.296872986922959825e-02 -4.538209911720556683e-02 -4.358954999999999913e-01 -6.835082898643515814e-02 -4.588642298692295896e-01
5.000000000000000278e-02 2.742156663333214159e-02 -5.882539791214071861e-02 -5.085638999999999577e-01 -8.624696454547285673e-02 -5.35985466633321305e-01
5.99999999999999778e-02 3.113916665704996936e-02 -7.348440170292774853e-02 -5.7416659999999997144e-01 -1.04623568359977144e-01 -6.053057666570499373e-01
7.000000000000000666e-02 3.408136856581576307e-02 -8.94330360119053357e-02 -6.321792000000000522e-01 -1.235146721670062897e-01 -6.62605685658158361e-01
8.000000000000000167e-02 3.623690446715328861e-02 -1.066708208278051434e-01 -6.823230000000000128e-01 -1.429077252949584320e-01 -7.185599044671533431e-01
8.999999999999999667e-02 3.770188369346503393e-02 -1.251201272690600774e-01 -7.245664000000000549e-01 -1.628220109625251044e-01 -7.62268236934651096e-01
1.00000000000000056e-01 3.858763146288396451e-02 -1.446331717899595193e-01 -7.59114399999999670e-01 -1.832208032528434838e-01 -7.977020314628839870e-01
1.10000000000000006e-01 3.904458472449394424e-02 -1.649991428368153268e-01 -7.863909000000000038e-01 -2.040437275613092849e-01 -8.254354847244939064e-01
1.19999999999999956e-01 3.925448026295294546e-02 -1.859566010169879380e-01 -8.07011200000000396e-01 -2.252110812799408834e-01 -8.462656802629530128e-01
1.300000000000000044e-01 3.942080670153031569e-02 -2.072085892115682615e-01 -8.217503000000000446e-01 -2.466293959130985702e-01 -8.611711067015304089e-01
1.400000000000000133e-01 3.976139948555254544e-02 -2.284398564666100251e-01 -8.315048000000000439e-01 -2.682012559521625983e-01 -8.712661994855526171e-01
1.49999999999999944e-01 4.051142979510184011e-02 -2.493352073592800811e-01 -8.372540000000000537e-01 -2.898466371543819142e-01 -8.777654297951018592e-01
1.600000000000000033e-01 4.186383830073221041e-02 -2.695979199506459700e-01 -8.40019799999999832e-01 -3.114617582513781735e-01 -8.818836383007322421e-01
1.700000000000000122e-01 4.399199010123152481e-02 -2.889671668736874532e-01 -8.408276999999999557e-01 -3.329591569749189572e-01 -8.848196901012315152e-01
1.799999999999999933e-01 4.703481701861558784e-02 -3.072334262372494185e-01 -8.40672399999999863e-01 -3.542682432558650341e-01 -8.877072170186155464e-01
1.900000000000000022e-01 5.108940734433653486e-02 -3.242509790985316220e-01 -8.404865000000000252e-01 -3.753403864428681569e-01 -8.915759073443365601e-01
2.000000000000000111e-01 5.620585249988695187e-02 -3.399467522291140242e-01 -8.411155999999999633e-01 -3.961526047290009900e-01 -8.973214524998869290e-01
```

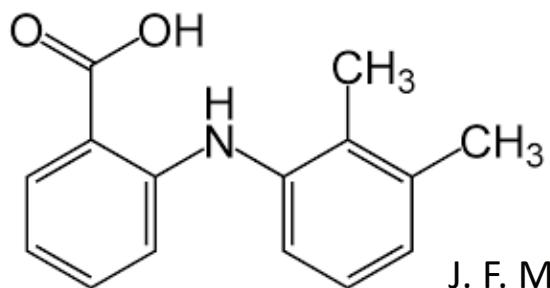
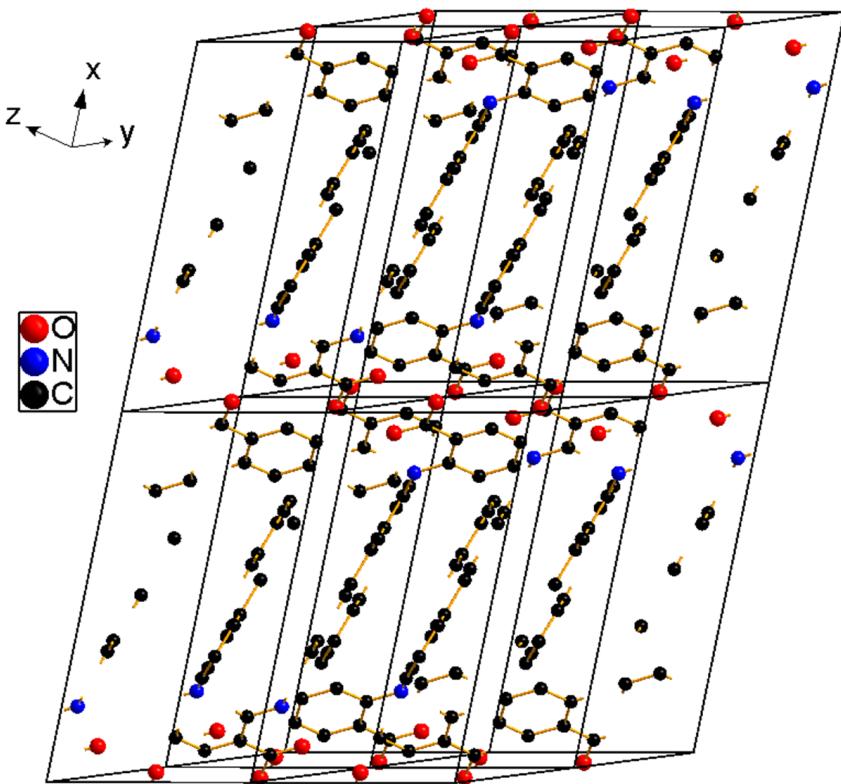
In the file that saved, the header explains the details of each column of data saved. In this example, it describes the Calculators used for simulating PDFs for molecule and crystal, and the scale factor used (in **Fine Tune Parameters** tab) for experimental PDF. Each column, from left to right, corresponds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

# Example 1: Hydrogen bonds in D-mannitol



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

## Example 2: Model fit of MEF crystalline PDF

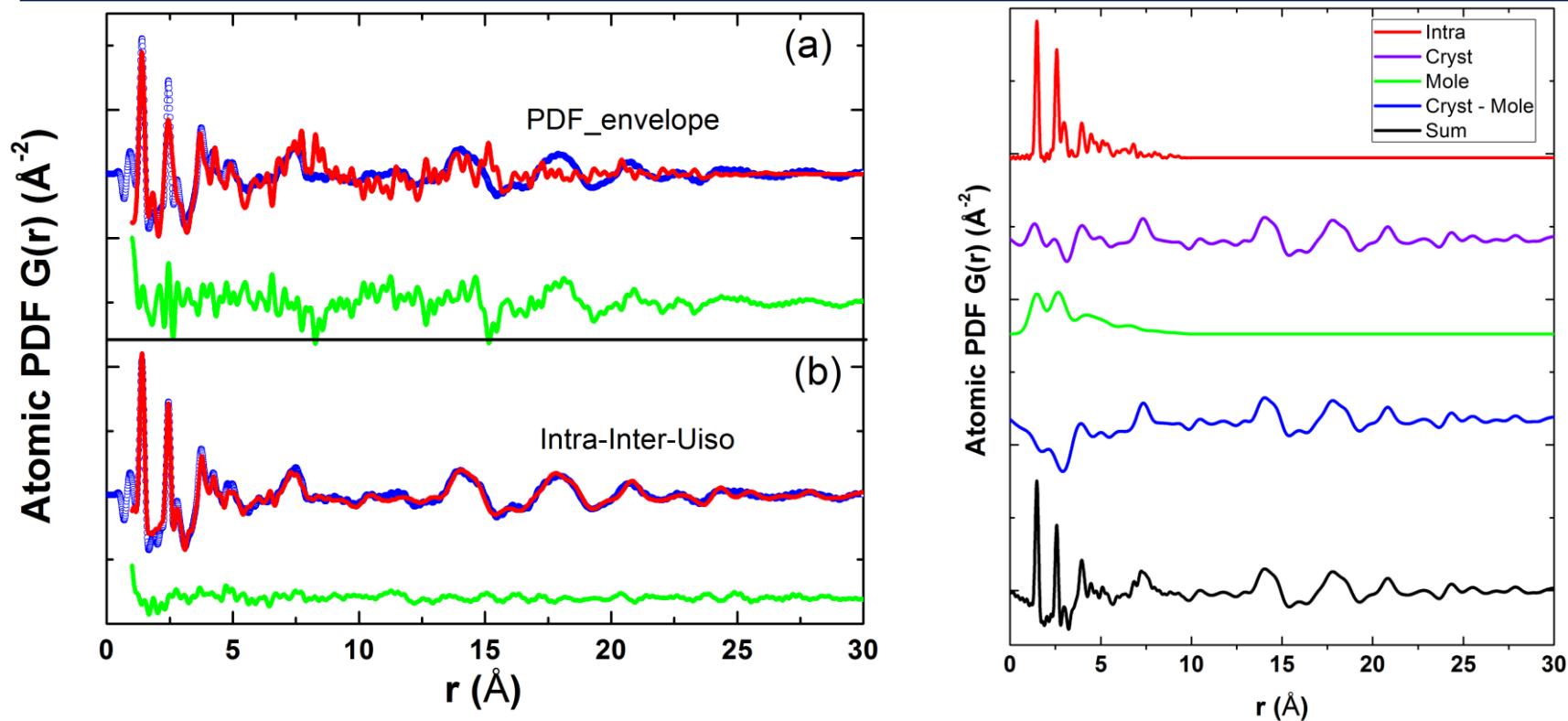


J. F. McConnell and F. Z. Company, *Cryst. Struct. Commun.* 1976, 5, 861-864.

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:  $Q_{\text{damp}} = 0.02902 \text{ \AA}^{-1}$ ,  $Q_{\text{broad}} = 0.0004 \text{ \AA}^{-1}$ . A  $Q_{\text{max}}$  of  $24 \text{ \AA}^{-1}$  was used for Fourier transform. The software program PDFgetX2 was used.

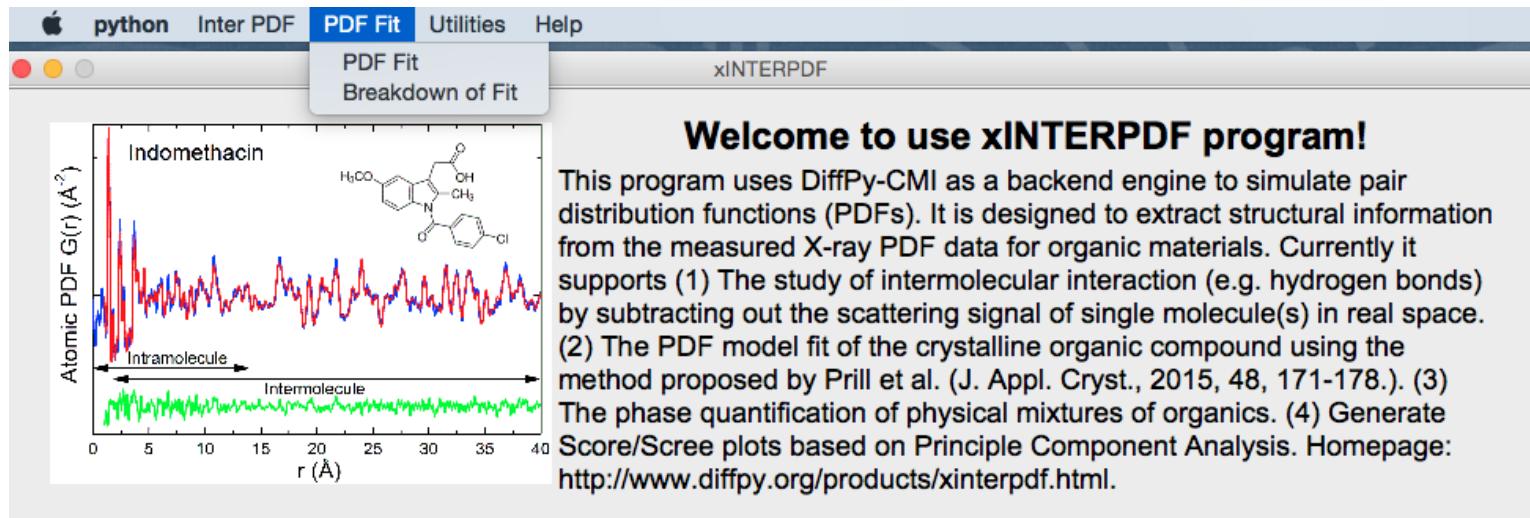
## Example 2: Model fit of MEF crystalline PDF



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  $U_{\text{iso}}$  thermal factor.

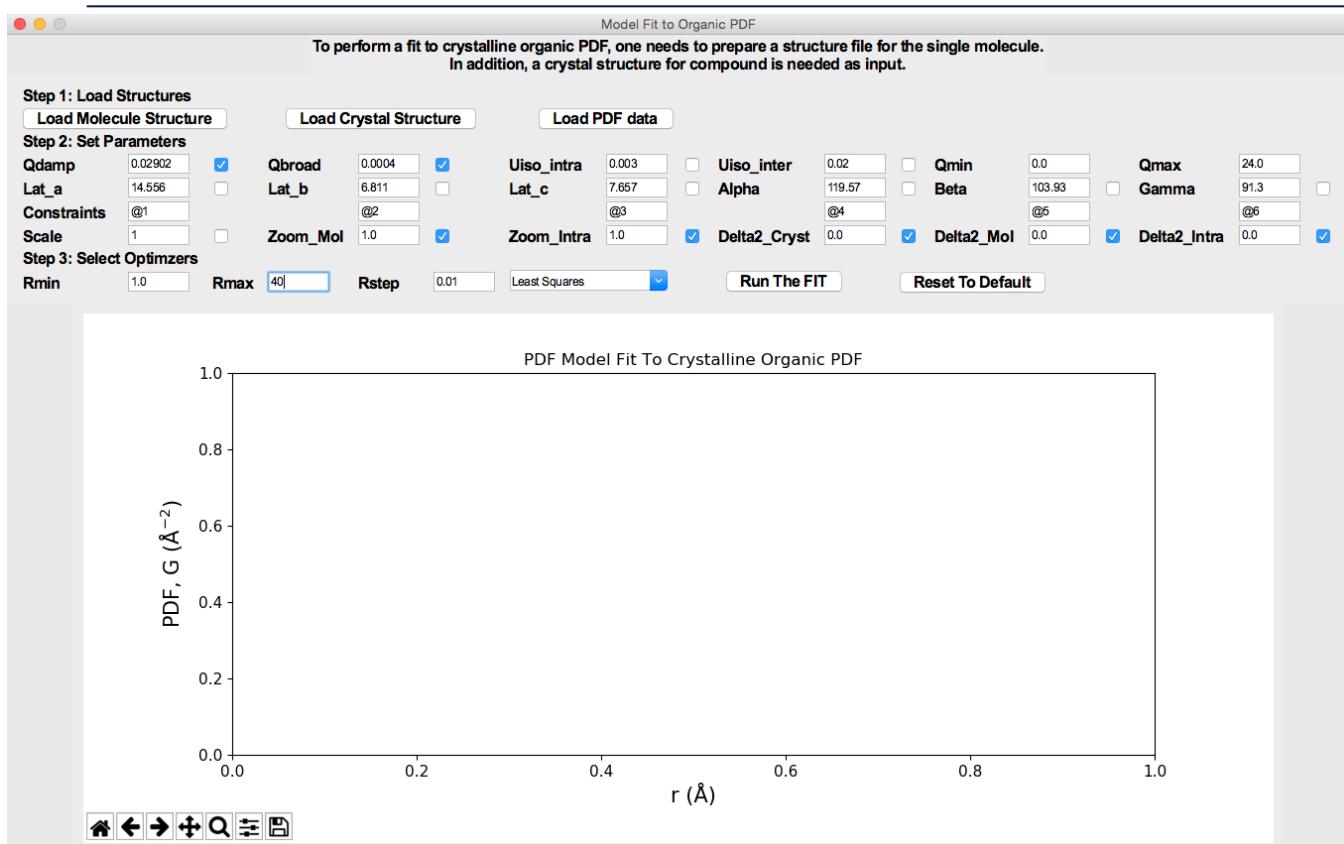
D. Prill, P. Juhás, M. U. Schmidt and S. J. L. Billinge, *J. Appl. Cryst.* 2015, 48, 171-178.

## 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

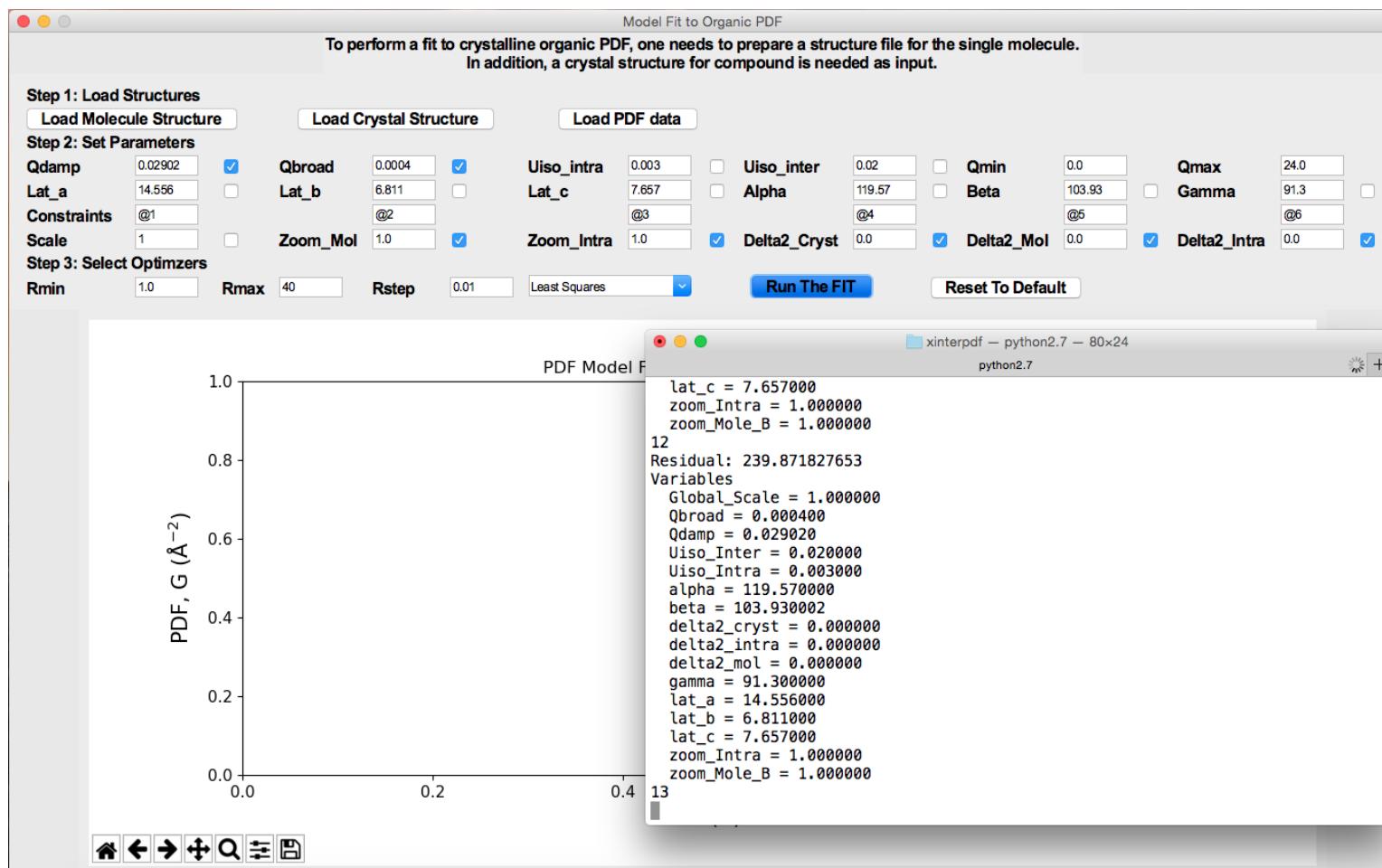


Following the steps, first load in structure files and PDF data (MEF.xyz, MEF.cif and MEF\_300-00000.gr). In Step 2, specify the parameters. If a check box is marked, that parameter is fixed during the fit; otherwise is allowed to vary.

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**.

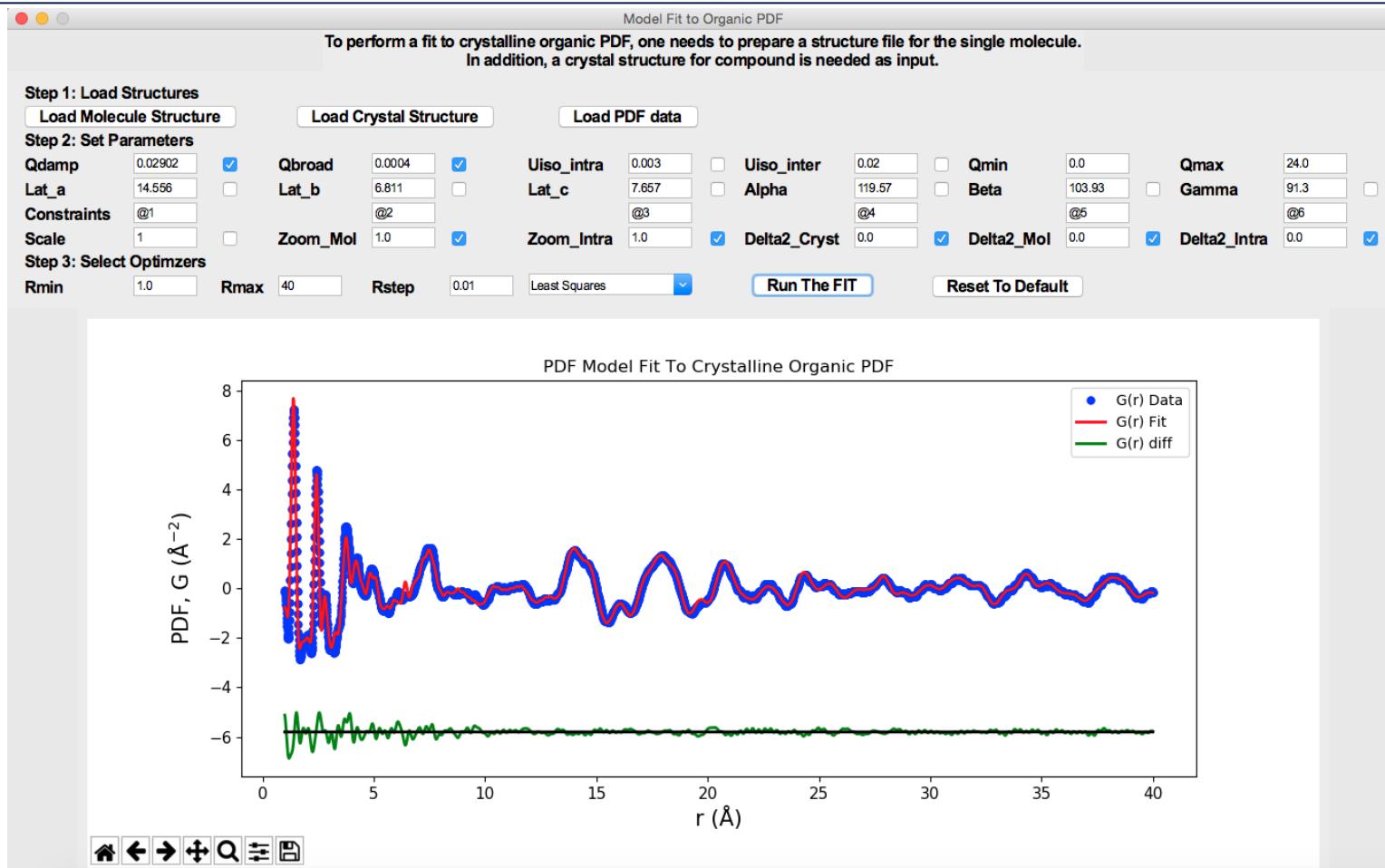
For a tetragonal structure, @1 for **Lat\_a** and **Lat\_b**, @2 for **Lat\_c**. Similar syntax is applied to constrain the angles (**Alpha**, **Beta** and **Gamma**). Since MEF is triclinic, we need six different variables for lattice parameters. **Zoom\_Mol** and **Zoom\_Intra** give the possibilities 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



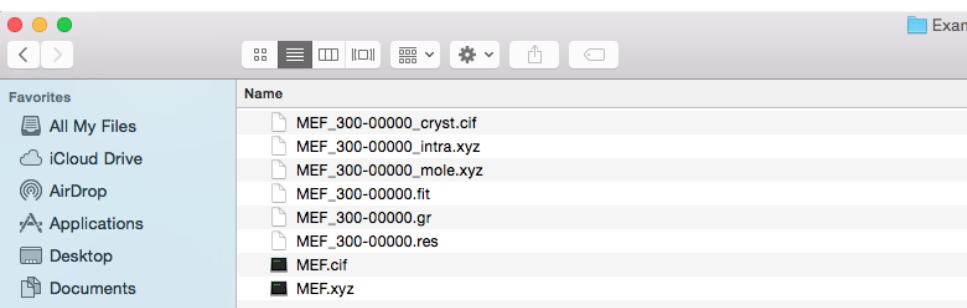
When the fit starts running, the terminal will update fit parameters in real-time.

## 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



The screenshot shows a Mac OS X file browser window titled "Example2". The sidebar on the left lists "Favorites" including "All My Files", "iCloud Drive", "AirDrop", "Applications", "Desktop", and "Documents". The main pane displays a list of files:

Name	Date Modified	Size	Kind
MEF_300-00000_cryst.cif	Today, 11:00 AM	10 KB	TextEd...ument
MEF_300-00000_intra.xyz	Today, 11:00 AM	991 bytes	TextEd...ument
MEF_300-00000_mole.xyz	Today, 11:00 AM	991 bytes	TextEd...ument
MEF_300-00000.fit	Today, 11:01 AM	394 KB	TextEd...ument
MEF_300-00000.gr	Oct 17, 2017, 10:44 PM	706 KB	Gedit...cument
MEF_300-00000.res	Today, 11:01 AM	1 KB	TextEd...ument
MEF.cif	Oct 18, 2017, 11:12 AM	2 KB	TextEd...ument
MEF.xyz	Oct 18, 2017, 11:12 AM	1 KB	TextEd...ument

**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)

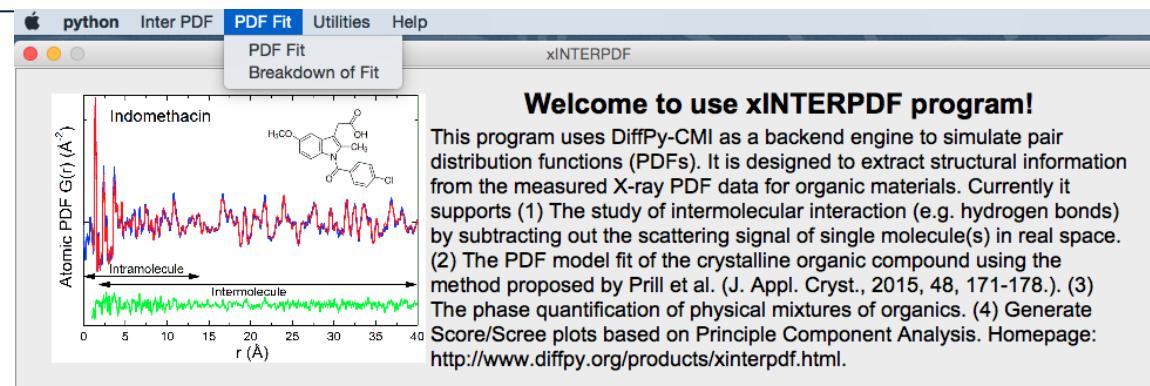
Global_Scale  9.92587263e-01 +/- 0.00000000e+00
Obroad       4.00000000e-04 +/- 0.00000000e+00
Odamp        2.90200000e-02 +/- 0.00000000e+00
Uiiso_Inter  4.15886353e-02 +/- 0.00000000e+00
Uiiso_Intra  3.31484234e-03 +/- 0.00000000e+00
alpha         1.19516522e+02 +/- 0.00000000e+00
beta          1.04001075e+02 +/- 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**

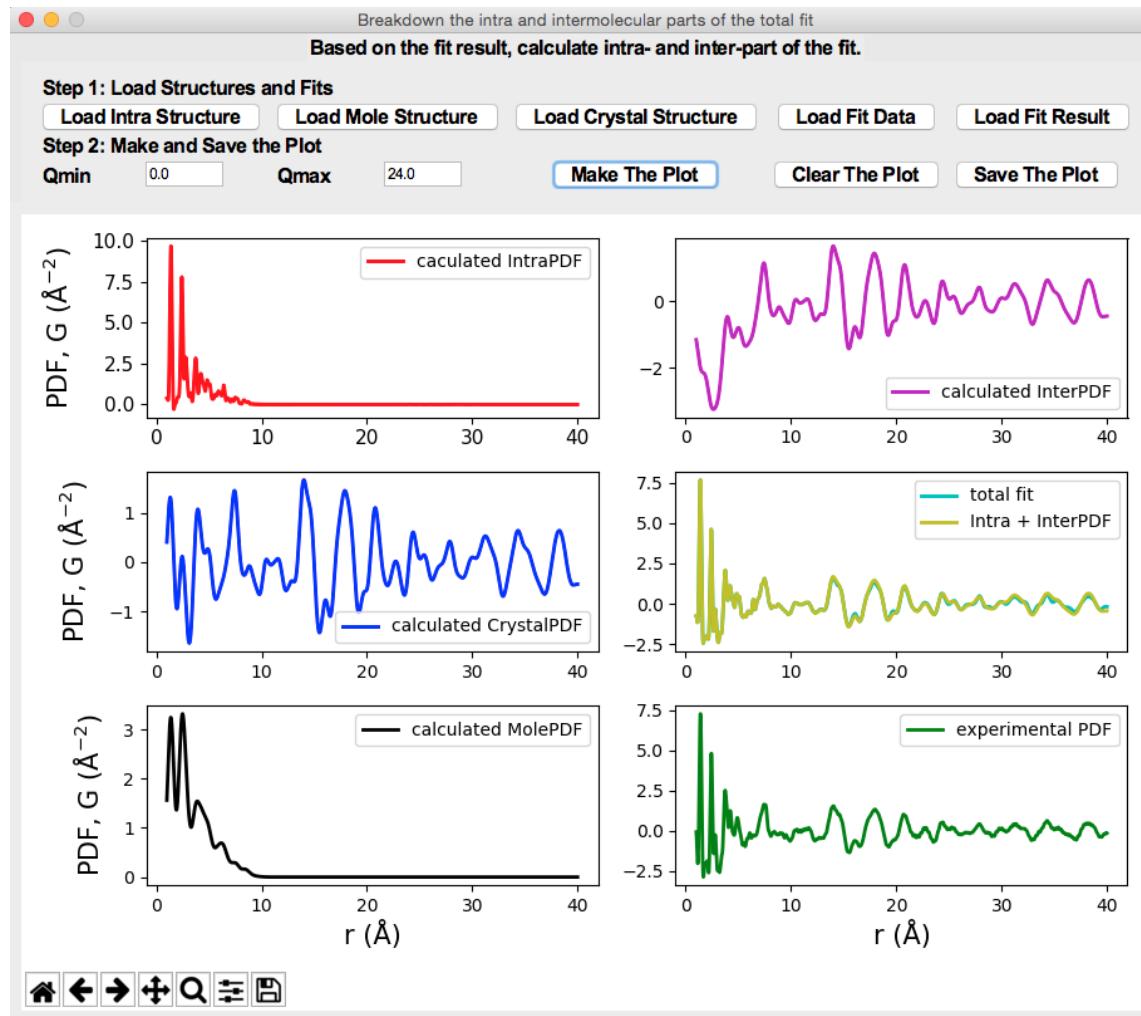
**MEF-300-00000.fit**

**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:

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

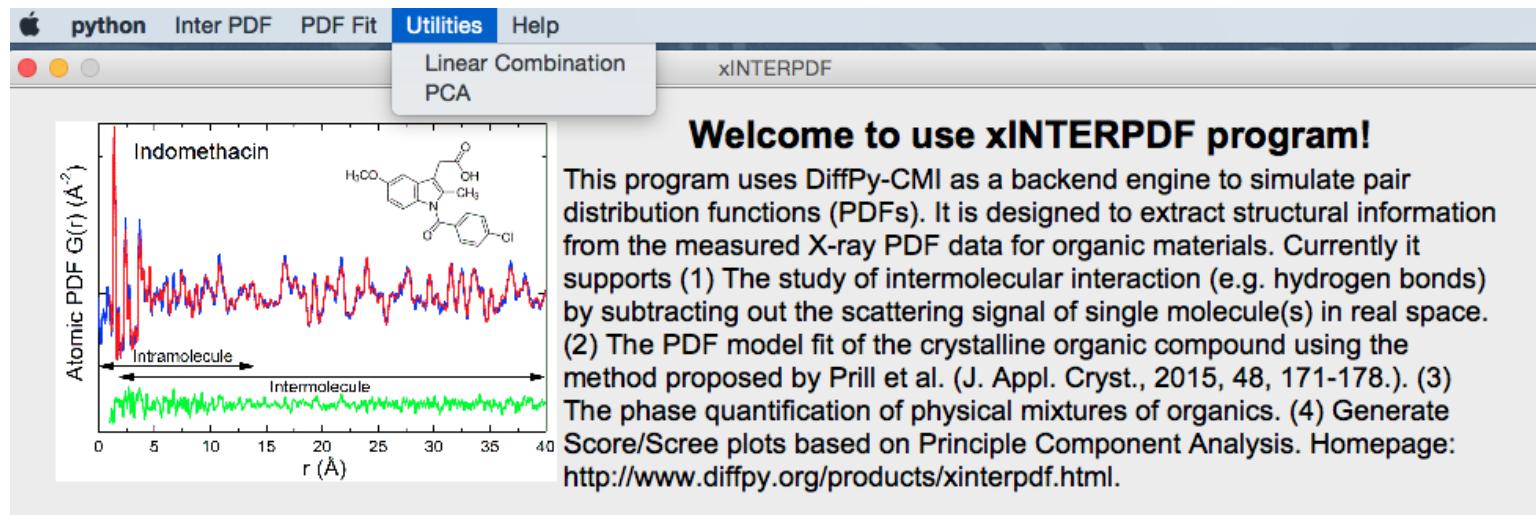
$$x + y + z = 1 \quad (2)$$

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

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.

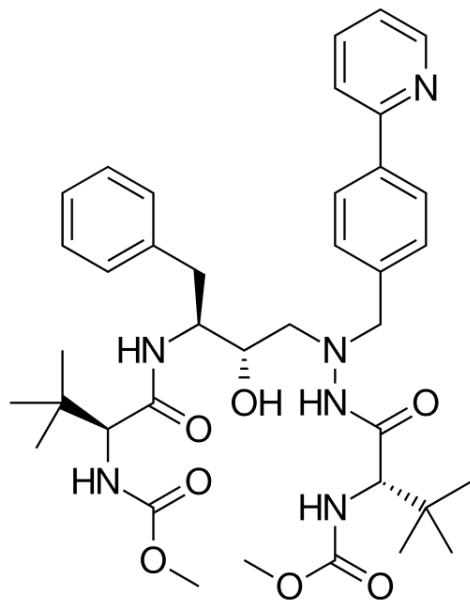
# Example 4: Find phase fraction by linear regression



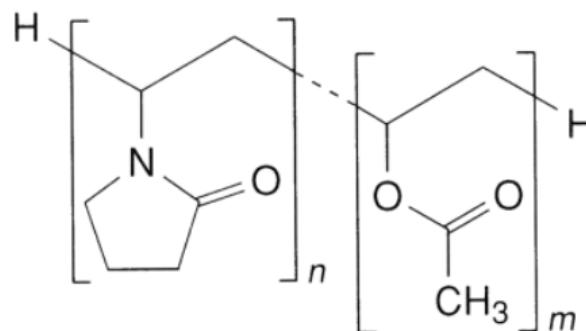
To perform a linear fit, in dropdown menu, click **Utility/Linear Combination** to start the GUI window.

## Example 4: Work on amorphous solid dispersion data

---



Atazanavir



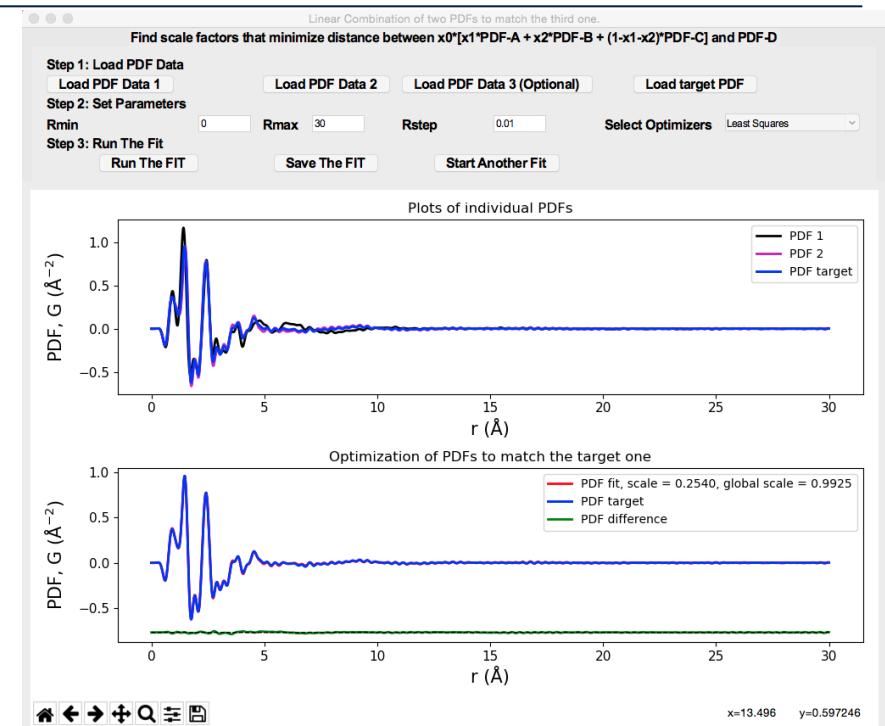
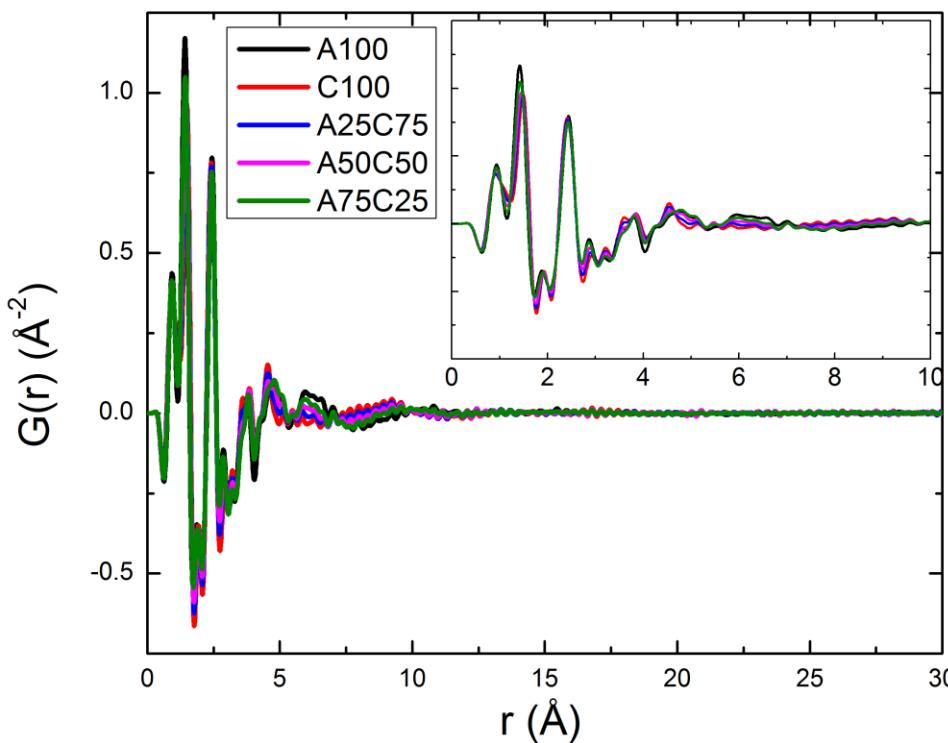
Copovidone

$$n = 1.16m$$

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).

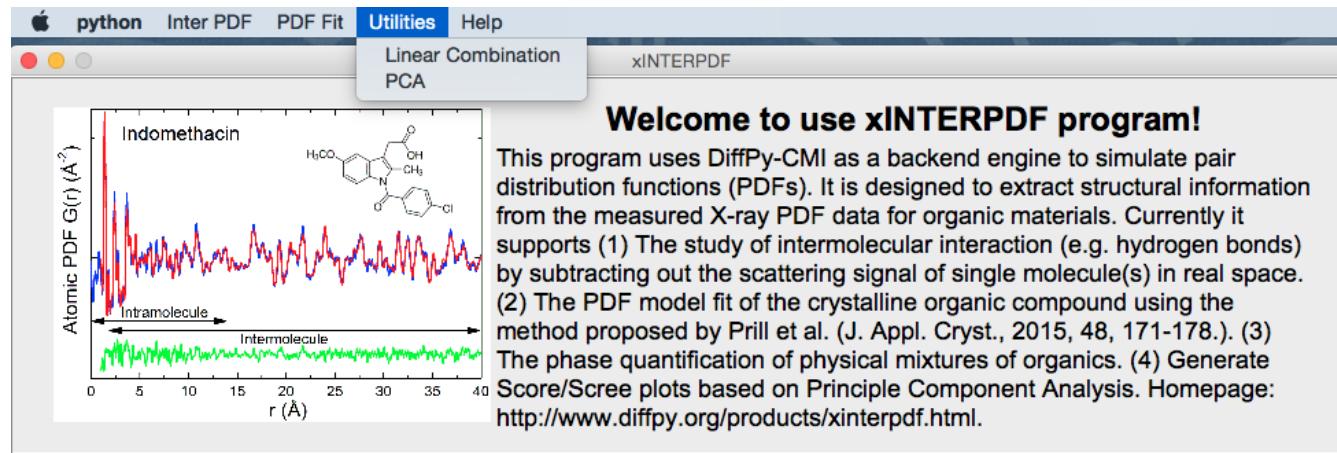
Synchrotron X-ray PDF experiment was done at 11-ID-B at 250 K, with a  $Q_{\max} = 20 \text{ \AA}^{-1}$ .

# Example 4: Work on amorphous solid dispersion data



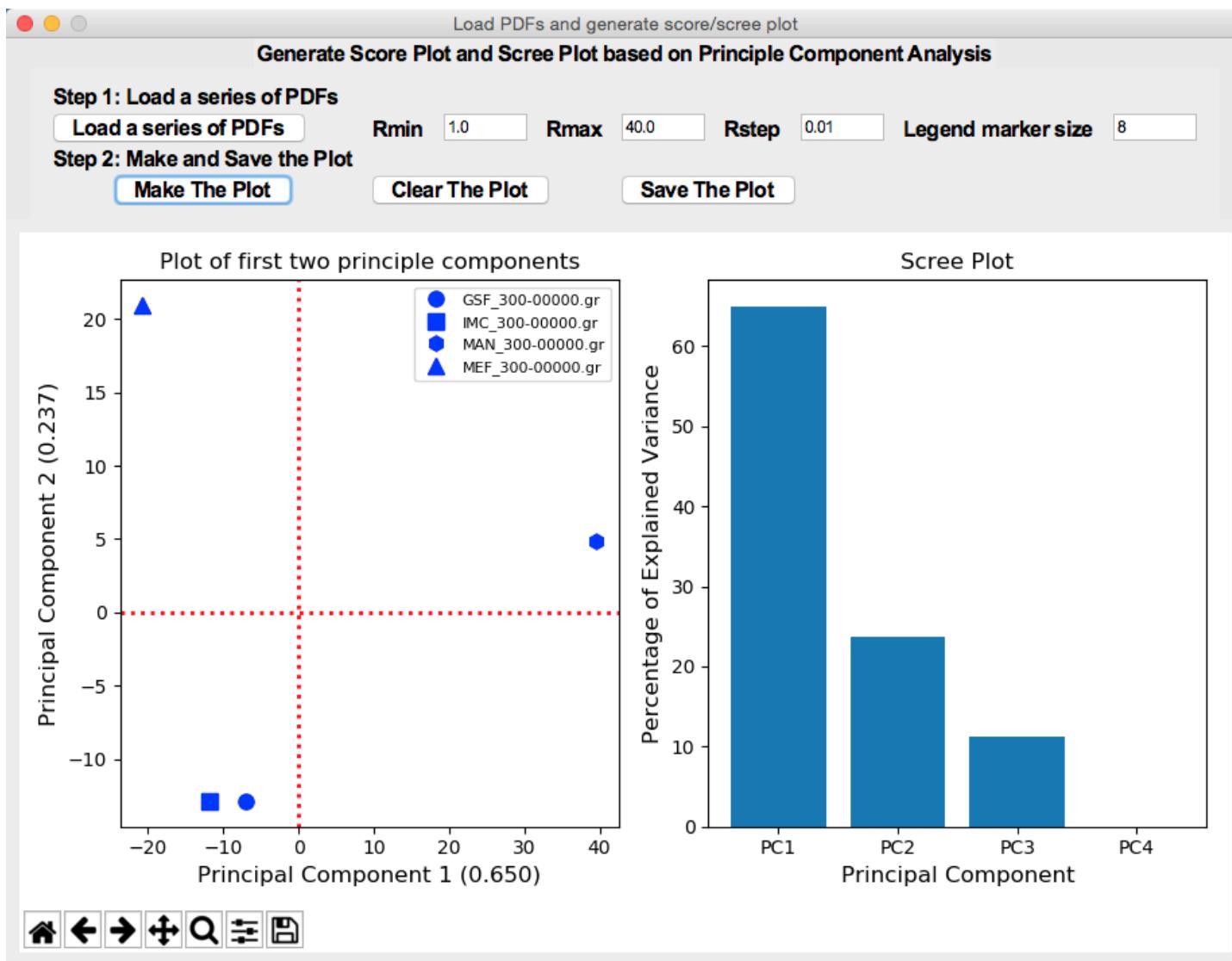
Samples	Nominal	Least Squares	Pearson	Least Absolute Deviation
rA25C75	25.0	25.4	25.4	24.8
rA50C50	50.0	47.7	47.8	46.3
rA75C25	75.0	77.8	77.8	75.8

# 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.