

Tutorial 3: Model fitting with SAS data

Introduction:

This tutorial covers the following advanced processing of SAXS data:

- Comparing scattering data and crystal structures (q space) - FoXS and CRY SOL

Requirements:

1. BioXTAS RAW, version 2.1.0 (newest).
 - Install instructions are available from: <https://bioxtas-raw.readthedocs.io/en/latest/install.html>
 - This tutorial assumes you are familiar with RAW.
2. ATSAS programs, version $\geq 3.0.1$.
 - Download and install instructions are available from: <http://www.embl-hamburg.de/biosaxs/download.html>
 - Requires a free registration for academic users. Industrial users must pay to use.
3. Internet connection.

Other useful materials:

1. There are RAW tutorial videos, which can be viewed here: <https://bioxtas-raw.readthedocs.io/en/latest/videos.html>
2. ATSAS
 - Manuals: <http://www.embl-hamburg.de/biosaxs/manuals/>
 - User forum: <http://www.saxier.org/forum/>

Part 1. Compare crystal structure and SAXS data (q space) – FoXS and CRY SOL

Envelopes make for pretty pictures, but usually the best way to test models against SAXS data is to generate theoretical scattering profiles and compare them to your measured scattering profile. In this part of the tutorial you will do this using two different software packages to do this: FoXS and CRY SOL from the ATSAS package.

1. In your system file browser, make a folder called **polymerase_theory** in the **Example_Data/reconstruction_data** folder. Copy the **2POL.pdb** file from the **Example_Data/reconstruction_data/polymerase_complete** folder into this **polymerase_theory** folder.
2. Navigate to the FoXS website: <https://modbase.compbio.ucsf.edu/foxs/>
3. For Input Molecule line select 'Choose File'. Select the **2POL.pdb** file.
4. Click 'Submit Form' to calculate the scattering profile.

foXS

Fast SAXS Profile Computation with Debye Formula

• About FOXS • Web Server • Help • FAQ • Download • Sali Lab • IMP • Links

Type PDB code of input molecule or upload files in PDB format (zip file with several PDBs can be uploaded):

Input molecule: (PDB:chainId e.g. 6lyz:A) or upload file: No file chosen

Experimental profile: No file chosen (optional) [sample input](#)


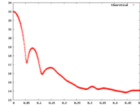
e-mail address: (optional, the results are sent to this address)

Advanced Options

NEW! MultiFoXS Now with conformational sampling and multi-state modeling, try [here](#)

If you use FoXS, please cite:
 Schneidman-Duhovny D, Hammel M, Tainer JA, and Sali A. Accurate SAXS profile computation and its assessment by contrast variation experiments. Biophysical Journal 2013. 105 (4), 962-974
 Schneidman-Duhovny D, Hammel M, Tainer JA, and Sali A. FoXS, FoXSDock and MultiFoXS: Single-state and multi-state structural modeling of proteins and their complexes based on SAXS profiles NAR 2016 [[FREE Full Text](#)]
 Contact: dina@salilab.org

5. On the results page that appears, click on the Profile file name (**2POL.pdb.dat**) to download. Move that downloaded file to your **polymerase_theory** folder and rename it **2POL_foxs.dat**.

Fast SAXS Profile Computation with Debye Formula

[About FOXS](#) • [Web Server](#) • [Help](#) • [FAQ](#) • [Download](#) • [Sali Lab](#) • [IMP](#) • [Links](#)

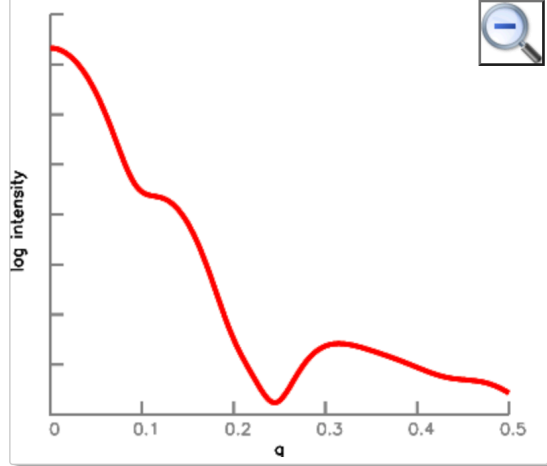
PDB files


[2POL.pdb](#)

Profile file

-

Can't see interactive display? [Use old interface](#)





JSmol

PDB file show/hide	R _g	# atoms	Profile file
2POL <input checked="" type="checkbox"/>	32.09	5688	2POL.pdb.dat

If you use FoXS, please cite:

Schneidman-Duhovny D, Hammel M, Tainer JA, and Sali A. Accurate SAXS profile computation and its assessment by contrast variation experiments. *Biophysical Journal* 2013. 105 (4), 962-974

Schneidman-Duhovny D, Hammel M, Tainer JA, and Sali A. FoXS, FoXSDock and MultiFoXS: Single-state and multi-state structural modeling of proteins and their complexes based on SAXS profiles based on NAR 2016 [[FREE Full Text](#)]

Contact: dina@sallilab.org

6. Open RAW if it is not already open. If it is open, clear all data loaded in RAW (unless your reconstruction is still running. If it is, just remove any items in the manipulation panel).
7. Load the **polymerase.dat** file in the **Example_Data/reconstruction_data** folder.
8. Carry out Guinier, molecular weight, and GNOM analysis on the scattering profile. Save the **polymerase.out** file in the **reconstruction_data** folder.
9. Load the **2POL_foxs.dat** file in the **polymerase_theory** folder.
10. The theoretical scattering profiles extends from $q=0$ to $q=0.5$, a much wider range than the measured profile. To make comparison easy you can trim the q range of the **2POL_foxs.dat** profile to match the **polymerase.dat** profile. Use the triangle to show more options for the scattering profile and adjust the q_{\min} until it is 0.01. Adjust q_{\max} until it is 0.24.
11. Star the **polymerase.dat** file, right click on the **2POL_foxs.dat** file and select the Other Options->Superimpose. In the dialog window that pops up, select 'Scale'.

- *Question:* How well do the theoretical scattering profile and the measured profile agree? Can you adjust the scale factor to get them in better agreement?
 - *Note:* Typically, experimental scattering profiles do not agree perfectly with profiles generated from crystal structures 'blindly'. Most programs will fit the crystal structure profile to the experimental data by altering the hydration layer in the modeled scattering, as well as fitting an overall scale factor and (sometimes) a constant offset.
- 12.** To get a better fit from FoXS we can upload a scattering profile to fit the data to. FoXS will adjust parameters like the hydration layer and excluded volume to improve fitting (you can turn this off in the advanced options), as well as the overall scale. Return to the initial FoXS page and add the **polymerase.dat** file as the Experimental profile. Submit the job again to rerun the calculation.
- 13.** You'll see in the results that this time the residual between the data and fit are shown, and the χ^2 of the fit is calculated as well. Download the **2POL_polymerase.fit** file. Copy it into your **polymerase_theory** folder and rename it **2POL_foxs_fit.fit**.
- 14.** Open the **2POL_foxs_fit.fit** file in RAW. You'll notice that it loads in 2 curves. The 2POL_foxs_fit.fit is the experimental data. The 2POL_foxs_fit_FIT is the generated theoretical scattering profile fit to the data.
- *Question:* What differences do you observe between the blind curve and the fit curve?
 - *Note:* On the fit results page, FOXS reports values of c_1 and c_2 . These are an overall scale factor for the excluded volume and an estimate of the average number of water molecules for an exposed solvent atom. These are fit to adjust the scattering profile. Significant changes in these values between different fits can minimize differences between the fits, for example if you're fitting several different datasets to a structure. This may lead to very different conclusions. For example. The χ^2 of the fit data is 1.55, whereas without fitting, using the default values of c_1 and c_2 , it is 4.87. This is the difference between a good fit and a poor fit. So pay attention to these parameters. There are some methods that attempt to treat the hydration layer more accurately, such as the WAXSiS tool (<http://waxsis.uni-goettingen.de/>) which runs a short explicit-solvent all-atom MD simulation then calculates the scattering from that. You might try this for comparison.
- 15.** Another common tool for generating theoretical scattering profiles is CRY SOL from the ATSAS package. We will use CRY SOL to generate a scattering profile from the crystal structure that is fit to the experimental data. Copy the **polymerase.dat** file into the **polymerase_theory** folder.
- 16.** Open a terminal/command prompt and navigate to the **polymerase_theory** folder (Appendix A).
- 17.** Type "crysol 2POL.pdb polymerase.dat" (without quotes) at the command prompt and hit enter.

- *Note:* The first file is the crystal structure from which to generate a theoretical profile. The second file is the experimental data file to fit the theoretical profile to.
- *Note:* Again, this fits parameters to adjust the excluded volume and hydration layer contrast (Ra and Dro respectively).
- *Note:* You can also use CRYSQL without fitting to data.

```


(base) Jesses-iMac:polymerase_theory jesse$ crysol 2POL.pdb polymerase.dat



```
*** ----- ***
*** C R Y S O L Wintel/UNIX/Linux version 2.8.4 ***
*** Please reference: D.Svergun, C.Barberato ***
*** & M.H.J.Koch (1995) J. Appl.Cryst., 28, 768-773 ***
*** Version (LMAX=99) for small and wide angles ***
*** Last revised --- 10/12/18 10:00 ***
*** Copyright (c) ATSAS Team ***
*** EMBL, Hamburg Outstation, 1995 - 2018 ***
*** ----- ***
```



-----



```
Program options :
0 - evaluate scattering amplitudes and envelope
1 - evaluate only envelope and Flms
2 - read CRYSQL information from a .sav file
Type crysol -help for batch mode usage
```



-----



```
Following file names will be used:
2POL01.log -- CRYSQL log-file (ASCII)
2POL01.fit -- fit to experimental data (ASCII)
```



-----



```
----- Reciprocal space grid -----
(in s = 4*pi*sin(theta)/lambda [1/angstrom])
Read atoms and evaluate geometrical center ...
Number of atoms read : 5688
Number of discarded waters : 150
Percent processed 10 20 30 40 50 60 70 80 90 100
Processing atoms :>>>>>>>>>>>>>>>>>>>>>>>>>>>>
For 858 zero directions radius 3.21 assumed
Center of the excess electron density: -0.005 0.150 0.030
Processing envelope:>>>>>>>>>>>>>>>>>>>>>>>>>>>>
Processing envelope:>>>>>>>>>>>>>>>>>>>>>>>>>>>>
--- Structural parameters (sizes in angstroms) ---
Electron rg : 32.06 Envelope Rg : 29.81
Shape Rg : 32.09 Envelope volume : 0.1241E+06
Shell volume : 0.3312E+05 Envelope surface : 0.1011E+05
Shell Rg : 38.76 Envelope radius : 48.24
Shell width : 3.000 Envelope diameter : 94.93
Molecular weight: 0.8115E+05 Dry volume : 0.9835E+05
Displaced volume: 0.1013E+06 Average atomic rad.: 1.620
Number of residuals : 732
Maximum angle in the data file : 0.2401
Number of experimental points : 404
Angular units multiplied by : 1.000
Number of points after regridding : 201
Number of experimental points used : 201
----- Fitting the experimental data ... -----
2POL.pdb Dro:0.052 Ra:1.800 Vol: 93680. Chi^2: 1.468
Rg from the slope of net intensity : 32.92
--WARNING: Recompute the intensity with LM= 32
Average electron density : 0.4631
Data fit saved to file 2POL01.fit
(base) Jesses-iMac:polymerase_theory jesse$
```


```

18. This will generate several files. The file with the scattering profile is the **2POL00.fit** file. Load it into RAW.

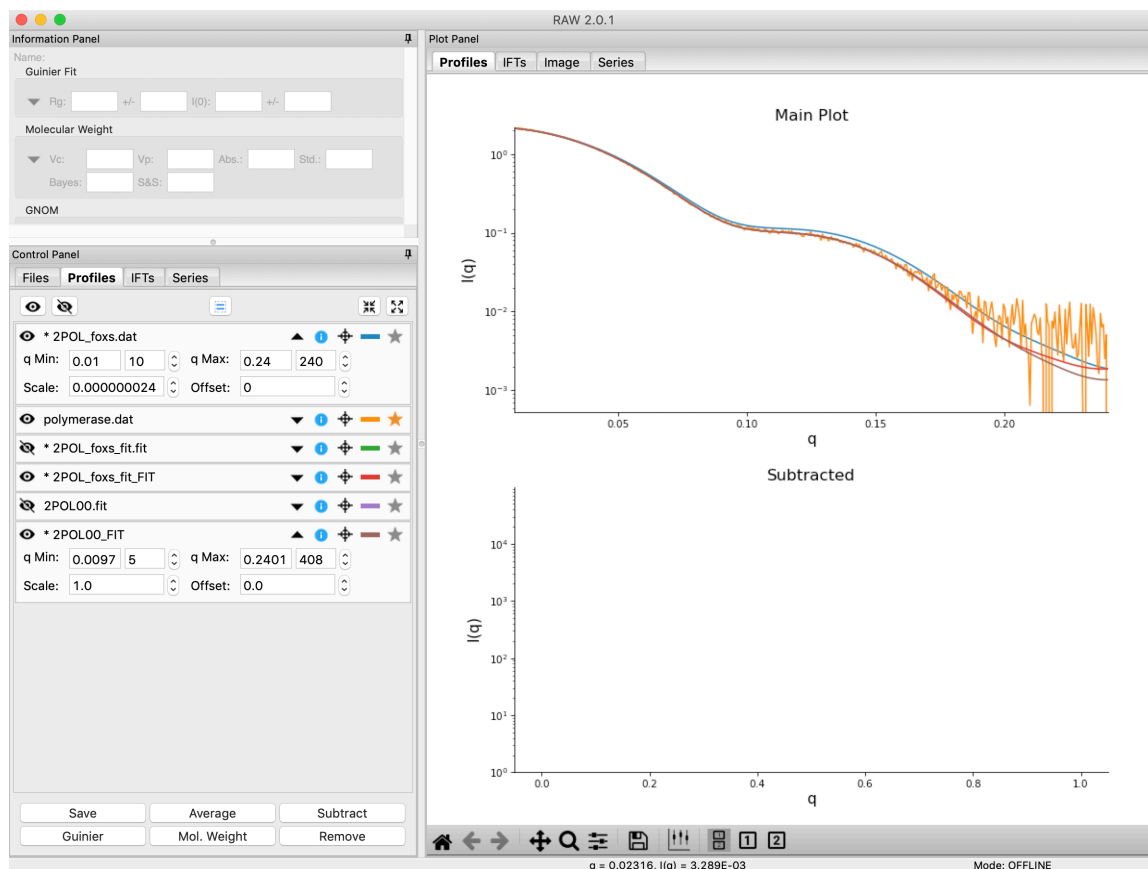
- Note: When it loads, it will load two scattering profiles. The **2POL00.fit** profile is the experimental data (identical to the

polymerase.dat file except that $I(q)$ values of 0 have been added for all q points below the minimum measured q). The **2POL00_FIT** profile is the theoretical profile.

19. Hide the **2POL00.fit** profile.

20. Adjust the q range of the **2POL00_FIT** file until q_{\min} is 0.0097.

- *Question:* How does this theoretical profile compare to those from FoXS?



Appendix A. The command prompt

As part of this tutorial, you will need to be able to open a terminal (mac OS X and Linux) or command prompt (Windows) and navigate to a known directory.

Windows 7:

1. Open a command prompt by clicking on the start menu, searching for "cmd" (no quotes) and running the **cmd** program.
2. Type "cd " (no quotes)
3. Drag the folder you want to move to into the command prompt. It should automatically put the folder path in the command prompt. For example, if you put the **Example_Data** directory on your desktop, and wanted to move to it, you should now see "cd C:\Users\<username>\Desktop\Example_Data" (no quotes) on the command line.
4. Hit enter.
5. The command prompt should show you what directory you are in (listed to the left of the prompt).
6. To check what files are in the current directory, type "dir" (no quotes) and hit enter.

Windows 8:

1. Open a command prompt by clicking on the windows tile and clicking the down arrow to show all apps. In the all apps screen select **Command Prompt** in the Windows System section.
2. Steps 2-6 for Windows 7 also work for Windows 8.

Windows 10:

1. Open a command prompt by clicking on the windows/start menu, selecting All Files, selecting Windows System, and clicking on Command Prompt.
2. Steps 2-6 for Windows 7 also work for Windows 10.

Mac OS X:

1. In the **Applications/Utilities** folder, open the Terminal app.
2. Type "cd " (without quotes).
3. Drag the folder you want to move to into the terminal. It should automatically put the folder path in the command prompt. For example, if you put the **Example_Data** directory on your desktop, and wanted to move to it, you should now see "cd /Users/<username>/Desktop/Example_Data" (no quotes) on the command line.
4. Hit enter.
5. To check what directory you are in, type "pwd" (no quotes) and hit enter.
6. To check what files are in the current directory, type "ls" (no quotes) and hit enter.

Linux:

It depends upon the flavor of Linux you are using. On a many Linux machines:

1. Open the folder in your system file manager.

2. Right click in the folder and select "Open in Terminal"

Otherwise, you can open a terminal and use the "cd" command to change to the proper directory.