Feb 12, 2021

Fitting of SAXS data using SCATTER

Chris Berndsen[1]
[1]James Madison University

1 Works for me This protocol is published without a DOI.

Chris Berndsen
James Madison University

SUBMIT TO PLOS ONE

ABSTRACT

Scatter is a Java-based Small Angle X-ray Scattering software developed by Rob Rambo. It can be used for processing and analyzing data to calculate Rg, Dmax, and other parameters. If the ATSAS package is installed, Scatter can also access DAMMIN/DAMMIF to produce dummy atom models fitted to the data.

This protocol was written for students at James Madison University to fit SAXS data in instructional or research lab settings.

PROTOCOL CITATION

Chris Berndsen 2021. Fitting of SAXS data using SCATTER . **protocols.io** https://protocols.io/view/fitting-of-saxs-data-using-scatter-bd3bi8in

CREATED

Mar 21, 2020

LAST MODIFIED

Feb 12, 2021

PROTOCOL INTEGER ID

34627

MATERIALS TEXT

Computer
SAXS data
internet access

Scatter 4.0
by Rob Rambo

ATSAS software package (optional)

BEFORE STARTING

Download SCATTER 4: https://bl1231.als.lbl.gov/scatter/
- Scatter on a Mac requires Java and X11 tools
- Scatter on a Windows system requires updated Java

Recommended but not required
Download ATSAS: https://www.embl-hamburg.de/biosaxs/download.html

Sample Information

1 Record your sample information in the table below. Add rows as needed.

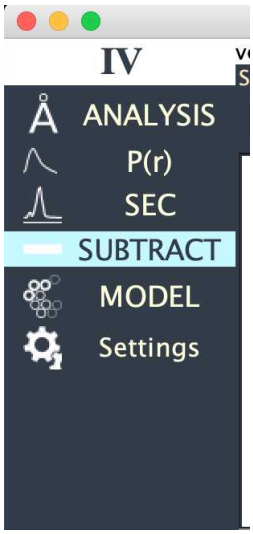
	Sample parameters
Protein/molecule name	
Concentration (mg/ml)	
Uniprot ID	
Number of amino acids/residues	
Crystal structures	
Well number (if SiBYLS HT-SAXS)	
Molecular weight (kDa)	
Sequence	

Buffer subtraction

2 Start

Scatter 4.0
by Rob Rambo

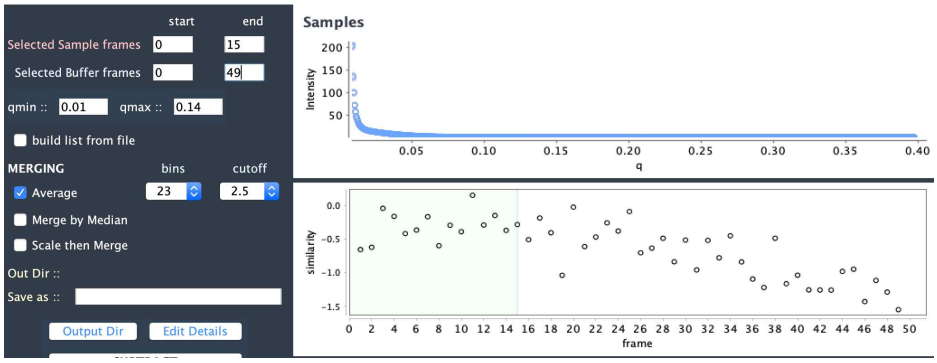
and select Subtract from the left side menu.



3 Drag your .dat files collected on your sample into the top pane.

4 Drag your .dat files collected on your buffer into the bottom pane.

5 Look at the similarity vs. frame window and determine if there are signs of radiation damage. This would shown by a deviation from the average in either the sample or buffer window.



Control panel and data plots for sample data. The lower right plot shows that around frame 24 the sample starts to deviate and may be sensitive to radiation.

5.1 Set the range in the box for which frames to start and end with in the final file.

6 Set the options for the output directory and files

6.1 Set the output directory as a sub-folder of the data directory named:

```
[Well_number]_[proteinname]_analysis
```

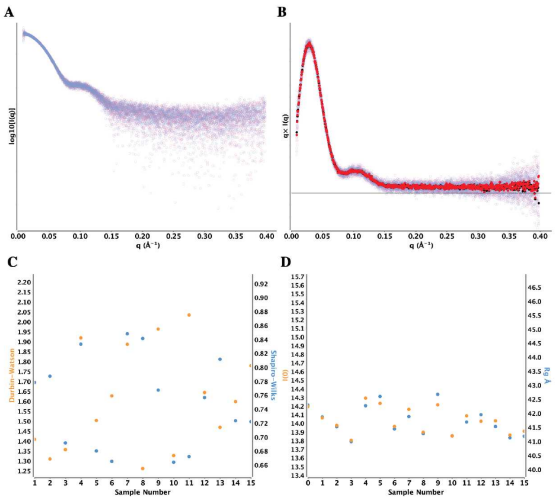
Record the output directory as a note on this step.

6.2 Set the output file name:

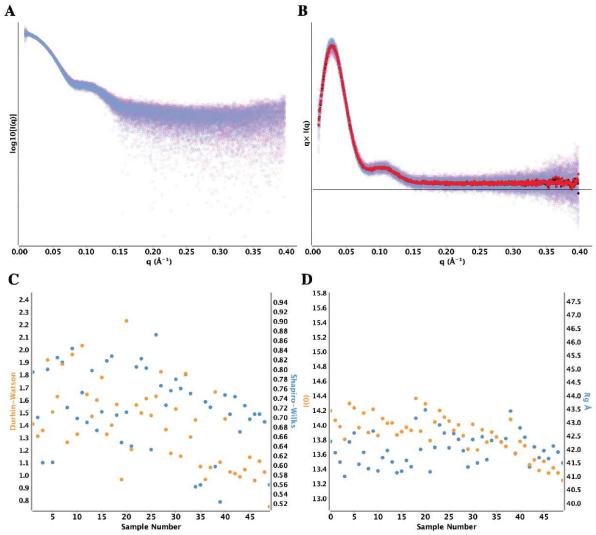
```
[Well_number]_[proteinname]_[#_dat_files]_subtracted
```

6.3 Click the subtract button and wait.

6.4 Look at the PDF report before doing further analysis.



Report plots showing basic characteristics of data set which is seemingly reasonable. Plot A shows log(I) vs. q, Plot B is a re-plot where the data should not go below the horizontal line, Plot C and D show similarity of the frames included in the data set.



Report plots showing basic characteristics of data set with some issues. Plots are as above. Plot C and D show deviations in frames >25 suggesting that they should not be included. This data set should be reprocessed to exclude these later data sets.

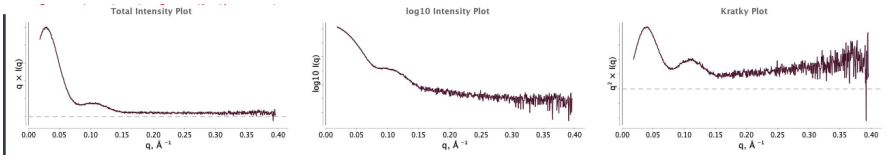
6.5 If the report suggests too many dat files were included, reduce the number of dat files incorporated in step 4.1 and repeat the subtraction steps making sure to change the file name.

6.6 Record the ranges for the sample and buffer in a note below.

Analysis of data

7 Switch to the analysis pane on the left side.

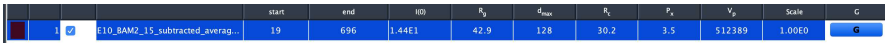
7.1 In the analysis pane are three main areas:



The plot bar

log10 Plot	Volume	Kratky	Errors	Complex?	Scale	Average	Load
Normalized Kratky	Vc Plot	Powerlaw	R_x Analysis	Ratio	Scale to I(0)	Median	Save File
Guinier Peak Analysis	Flexibility	q•I(q) Plot	I(0):Rg Plot	SVD	ScaleMerge	Clear	PDF

The analysis menu



The data set window.

8 If coming straight from subtraction, the average and median data sets will already be loaded. If not, drag the average or median dataset into the lower panel to the data set window.

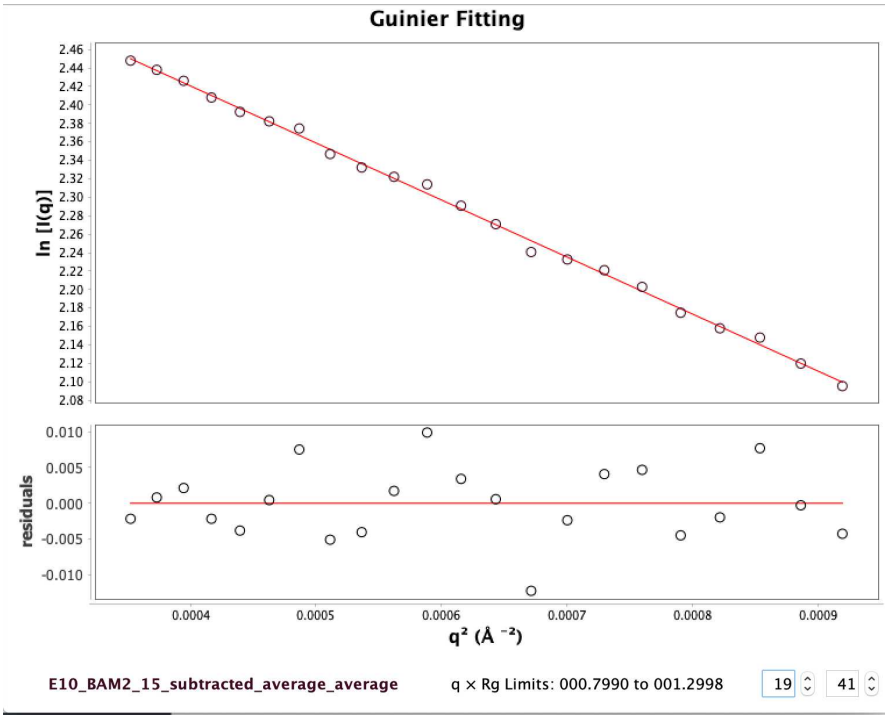
You can load multiple data sets to analyze and compare. To analyze a data set, check the box next to the data set name and uncheck all the others.

9 Press the G button at the right side of the data set window to begin Guinier analysis which will give an estimate of the radius of gyration.

The radius of gyration is average electron density weighted squared distance of the scatters from the center of the object. This can be thought of the radius of the molecule if the molecule was a perfect sphere.

9.1 The G button brings up two panels, one a re-plot of the data to linearize the beginning portions of the data and a residuals plot of the fit line in the top panel to the data. Adjust the number of points using the arrows at the bottom right until the top panel red ine fits the data and the residuals. Residuals should be randomly scattered around the horizontal line.

Important: the q x Rg limits should be between 0 and 1.3.



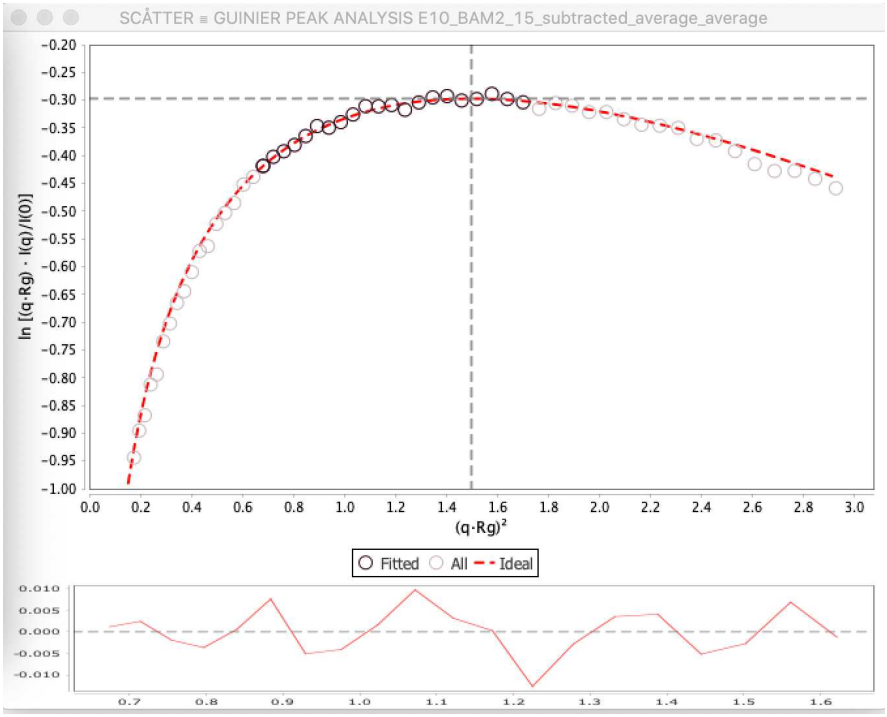
9.2 Record the start and end q^2 values below

	Start	End
Rg (Guinier) data range		

9.3 Close the window and the new Rg will appear in the data set window under Rg.

9.4 Click to the Guinier peak analysis to bring up a window to check your fitting. Ideally the data peaks at the crossing of the dashed lines and the dashed red line goes through the data points.

The Gunier peak window will also appear and you can adjust the fit if needed. If you do adjust the fit, record the new start and end number in a new row in step 9.2 above and note that you changed the range later



Guinier peak analysis window

10 Press the R_x analysis button in the analysis menu to estimate the radius of the cross section.

This radius will be identical to the Rg for a perfect sphere. Other shapes will have a value less than Rg and this can be useful for suggesting deviations from globular. For example for B-form DNA, the cross-section radius is ~10 A.

10.1 The window that appears will have a re-plot of the data and a residual plot of the data below. Fit the data so that the q x Rg is between 0 and 1.3 and the residuals are randomly scattered around the red line.

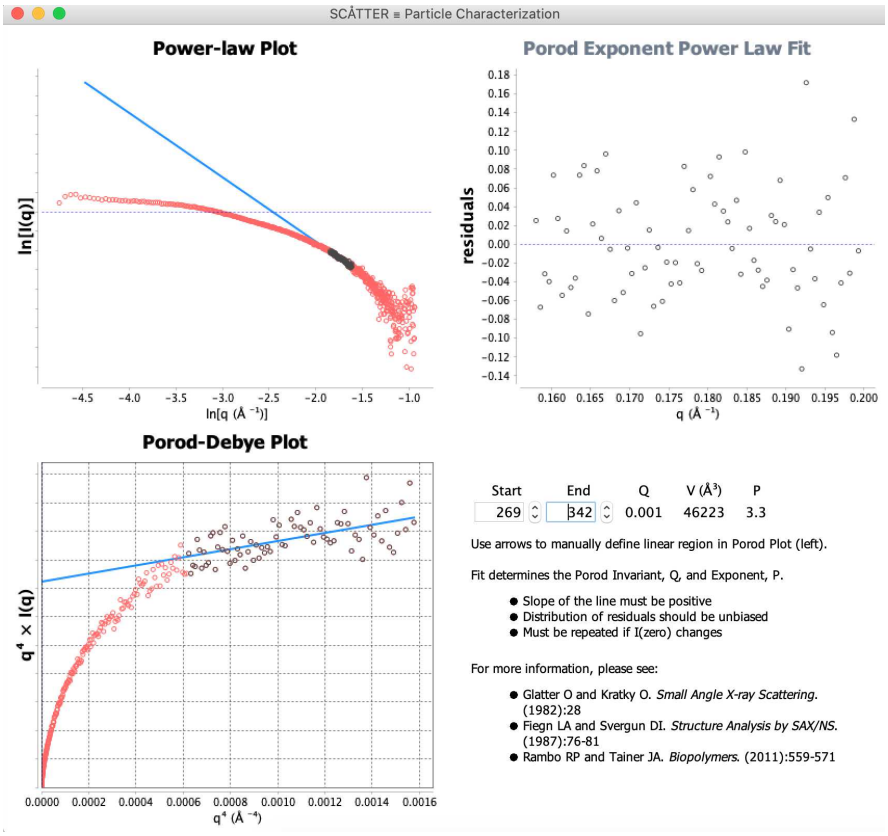
The cross section radius is determined by fitting the data at q values immediately adjacent to the Guinier region but slight larger.

10.2 Record the start and end q^2 values below

	Start	End
Rc		

11 Press the Volume button in the analysis menu to estimate the volume of the sample

11.1 Adjust the start and end values until the blue line in the Porod-Debye plot fits in the latter portion of the plot with a positive slope (example shown below). The residuals plot should show random scatter around the line.



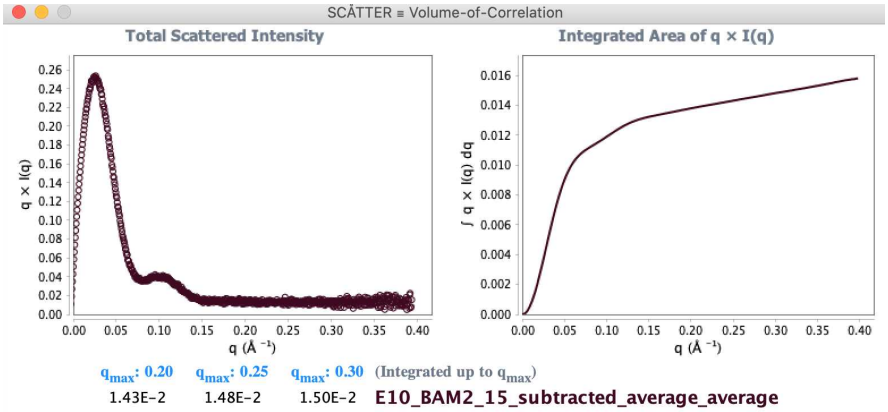
For multi-domain proteins, this analysis can be tricky to do.

11.2 Record the values in the table.

A	B
Volume (A^3)	
Q	
P	
start	
end	

Values of P less than 4 can suggest some disorder in the molecule. The Volume should approximate the volume of the crystal structure. Deviations can suggest aggregation, quaternary structure, or flexibility. Analysis of the Kratky plot and/or the data plot can help distinguish between these possibilities

12 Press the V_c plot button to display the re-plotted data.



Guinier extrapolated dataset (left) plotted as q · I(q). The integral approaches a constant value at high q (right).

- Integral of q · I(q) is proportional to the particle's correlation length, l_c.
- The ratio of I(0) to l_c defines the Volume-per-Correlation length, V_c.
- Strong divergence at high q suggests poor buffer subtraction.
- Ratio of V_c^2/R_g defines an invariant proportional to mass.

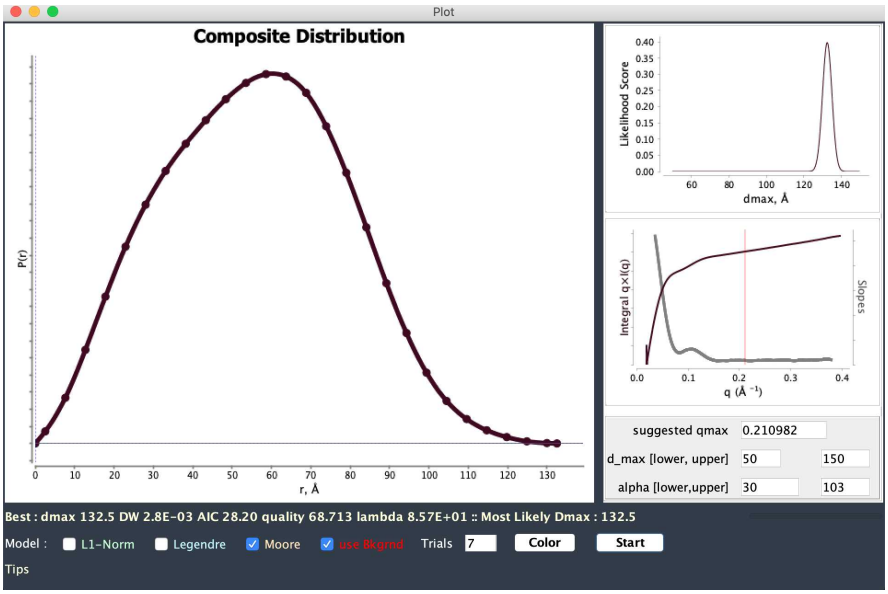
The left side plot should be continous and not show any gaps. The right side plot should level out at high q values. Bad buffer subtraction can show up as discontinuities in the total scattered intensity plot or a non-level integrated area plot.

P(r) plot

13 Press the P(r) button on the right side to start real-space analysis.

14 Right click on the data set and select the FIND DMAX option in the pop-up menu. This will bring up a window to automatically estimate D_{MAX}, or the maximum distance between atoms in the molecule.

14.1 In the resulting window, Scatter will predict an optimal data range to calculate Dmax. The default settings for Dmax are to search between 50 and 200 Ångstroms, however this can be adjusted for smaller proteins or to a narrower range, which will speed up calculations.



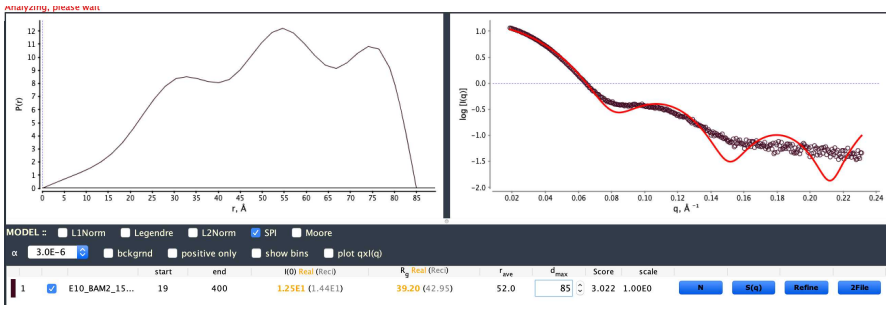
14.2 Press the start button to calculate and wait for the progress bar to finish.

14.3 The top left panel should show a normally distributed peak surrounding the most likely D_{max} value. The number will be shown below the plots. The left side plot will show a rough P(r) plot.

14.4 Record the Dmax value here

Dmax	

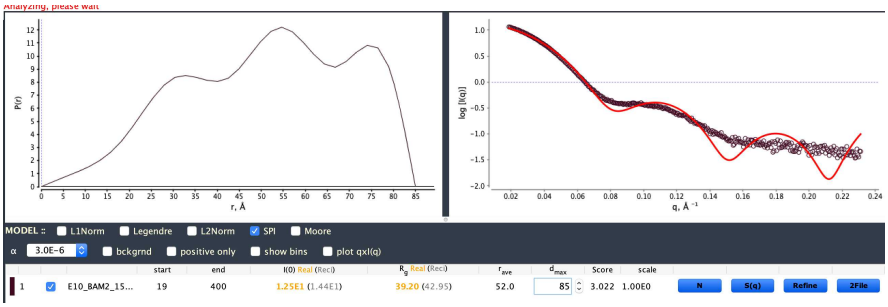
15 In the data set window of the P(r) area, input the D_{max} from 14.4 into the Dmax option.



15.1 Generally data above q = 0.25 don't add much to the fitting, so it is fine to truncate the data used for fitting. Adjust the end value and record the start and end values below.

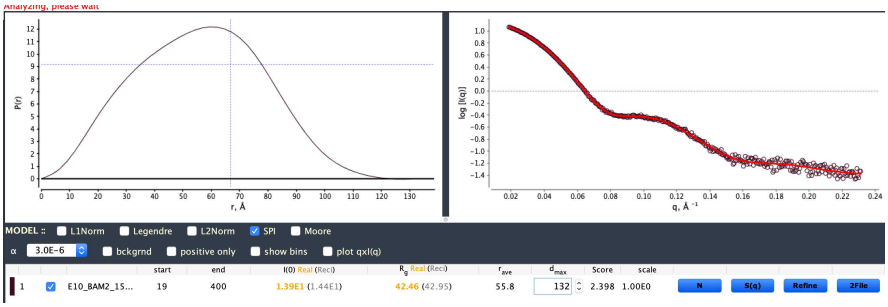
Start	End

15.2 The data below are a bad fit. The P(r) plot should slowly approach zero at high radius values and the fit line should fit well to hte raw data at right.

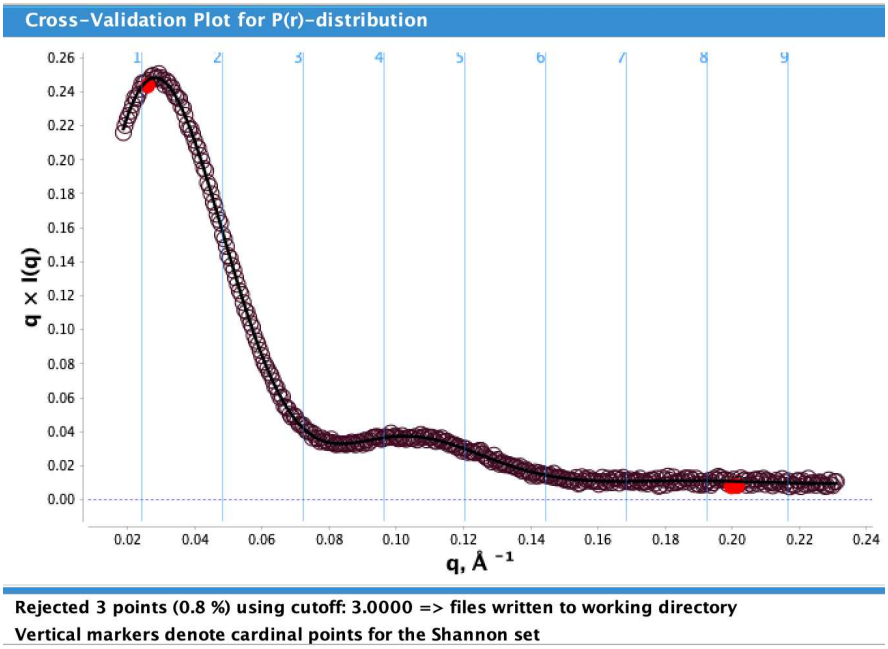


15.3 Adjust D_{max} up or down to improve the fit. The plot in 15.2 indicates a larger D_{max} should be used. If the 0 line in the P(r) plot is significantly removed from the x-axis or the line in the left plot are below the x-axis, then a smaller Dmax should be used.

15.4 Once a reasonable Dmax has been found, the data should appear as below. Press refine to confirm the fit.



15.5 Once refinement has finished, the cross validation plot will appear and outliers are shown in red. Any outliers should be randomly distributed across the data and to a minimum.



15.6 Record the final D_{max} and other parameters below.

A	B
Dmax	
Real Rg	
Real I(0)	
Score	
r (ave)	

15.7 Press the 2File button in the data analysis window to save the P(r) data. The filename should be as

```
[Well_number]_[proteinname]_pr
```

Scatter will save this as a .out file which will be used for model-free analysis. Record where this file is saved in a note in this step.

Exporting the data to file

16 Return to the Analysis window in the left side menu

17 Click the PDF option

17.1 Name the PDF file

```
[Well_number]_[proteinname]_report
```

Indicate the location of this file as a note attached to this step.

17.2 The report contains 4 plots:

1. log(I) v q
2. the normalized Kratky plot, which can indicate disorder and quaternary structure
3. total scattered intensity plot
4. the P(r) distribution

On the second page, are the final numbers (with errors bars).