# SasView Tutorials

## P(r) Inversion Analysis
## in SasView Version 5.x

## www.sasview.org

**Preamble**

SasView was originally developed by the University of Tennessee as part of the Distributed Data Analysis of Neutron Scattering Experiments (DANSE) project funded by the US National Science Foundation (NSF), but is currently being developed as an Open Source project hosted on GitHub and managed by a consortium of scattering facilities. Participating facilities include (in alphabetical order): the Australian National Science & Technology Centre for Neutron Scattering, the Diamond Light Source, the European Spallation Source, the Institut Laue Langevin, the ISIS Pulsed Neutron & Muon Source, the National Institute of Standards & Technology Center for Neutron Research, the Oak Ridge National Laboratory Neutron Sciences Directorate, and the Technical University Delft Reactor Institute.

SasView is distributed under a 'Three-clause' BSD licence which you may read here: https://github.com/SasView/sasview/blob/master/LICENSE.TXT

SasView is free to download and use, including for commercial purposes.

© 2009-2019 UMD, UTK, NIST, ORNL, ISIS, ESS, ANSTO, ILL, TUD, DLS

**If you make use of SasView**

If you use SasView to do productive scientific research that leads to a publication, we ask that you acknowledge use of the program with the following text:

> *This work benefited from the use of the SasView application, originally developed under NSF Award DMR-0520547. SasView also contains code developed with funding from the EU Horizon 2020 programme under the SINE2020 project Grant No 654000.*

**Contributors to this Tutorial**

James Doutch (james.doutch@stfc.ac.uk)

## Learning Objective

This tutorial will demonstrate how to perform P(r) inversion on 1D ('intensity' versus Q) datasets in SasView. The P(r) function is the real space distance distribution function. It is most commonly used in the analysis of small-angle scattering data from monodisperse protein solutions but can be applied to data from more generic systems. This form of analysis should, in general, be used on systems in which the Guinier region shows no evidence of contributions from aggregation or the structure factor S(q). This tutorial does not aim to cover all scenarios that solution scatterers, particularly biologists, may encounter in analysing their data – a full workflow for biological solution scatters should include accurate determination of the radius of gyration, estimation of molecular weight and inspection of basic SAS plots such as the Kratky plot!

It is assumed that the reader has some familiarity with pair distance distribution functions and the principles of P(r) inversion analysis. If not, this reference provides an overview:

- Moore, PB. *J. Appl. Cryst.*, (1980), 13, 168-175

A brief overview of the principles is presented in the SasView User Documentation.

> *The program interface shown in this tutorial is SasView Version 5.0.1 running on a Windows platform but, apart from a few small differences in look and functionality, this tutorial is generally applicable to SasView 5.x running on any platform.*

## Glossary

| | |
|---|---|
| Distance distribution function | A real-space function that describes spatial relations in the system whose datasets are being analysed. The form of the function can be used to infer the overall shape/internal structure of the scattering objects. |
| | |
| Fourier Transform | A mathematical 'tool' that decomposes a measured signal into a sum of sine or cosine functions. |
| | |
| Inverse-space | See reciprocal-space |
| | |
| Real-space | Real world coordinates. The opposite of inverse-space or reciprocal space. |
| Reciprocal-space | The Fourier Transform of real world coordinates. The opposite of real-space. In scattering measurements, *Q* is the reciprocal-space equivalent of a real world length scale. |
| Regularization | A mathematical procedure for adding information to a problem to prevent over-fitting of the data. |

| | |
|---|---|
| Transformation | A mathematical process for converting between real-space and inverse-space. |
| Uncertainties | Every experimental measurement, including the measurement of $I(Q)$, is subject to some degree of error (which will, ideally, be included in the dataset). Similarly, the parameters returned by any analysis will have some associated range of uncertainty.<br><br>Parameters with uncertainties that are more than 95% of the parameter value should be viewed with deep suspicion. |

# Running SasView

## Windows

Either select SasView from '**Start**'> '**All Programs**' or, if you asked the installer to create one, double-click on the SasView desktop icon.



SasView

## Mac OS

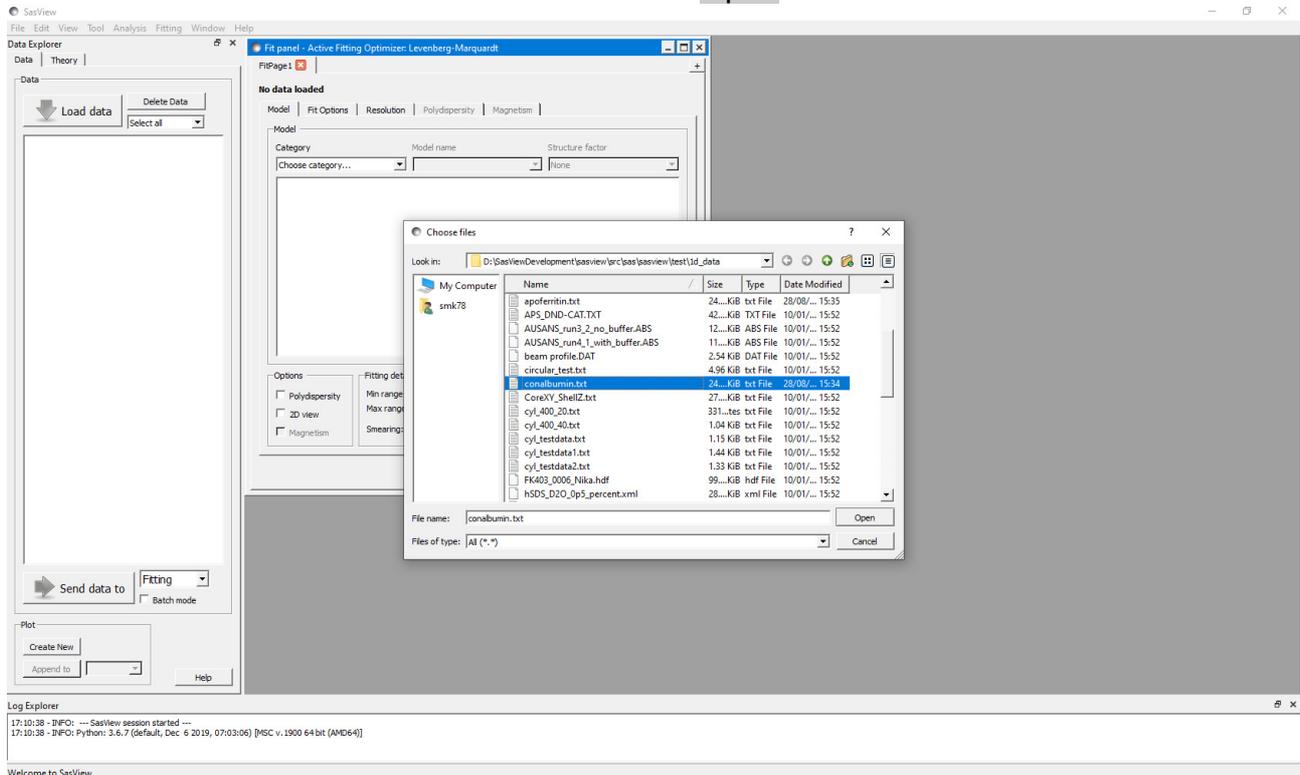Go in to your '**Applications**' folder and select SasView.
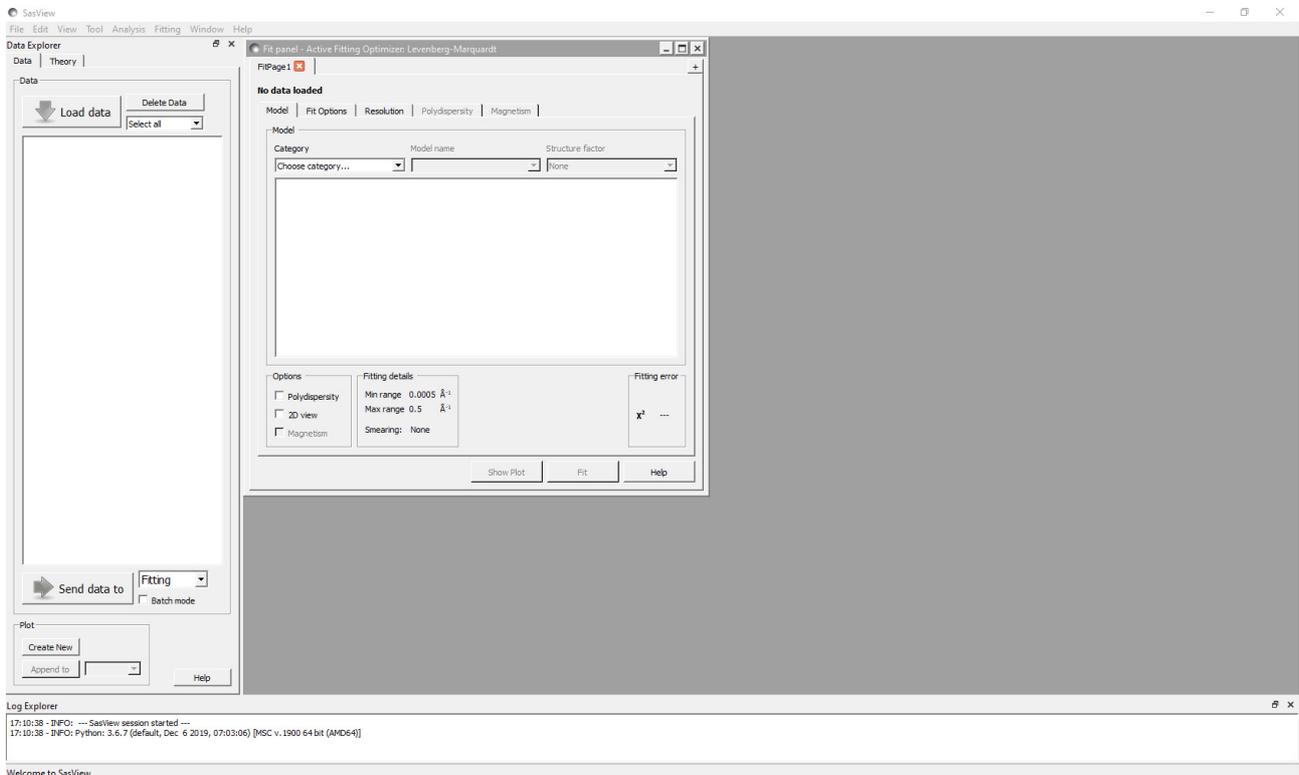
# Example 1

*This computes the distance distribution function for the glycoprotein conalbumin.*

In the Data Explorer panel, click the Load Data button, and navigate to the **\test\1d_data** folder in the SasView installation directory.

Select the **conalbumin.txt** dataset and click the Open button.



This dataset is SANS data collected on the SANS2d instrument (ISIS Neutron and Muon Source, Harwell) from conalbumin (or ovotransferrin) - a suggested PDB entry would be 1N04.  This structure forms a dimer in solution.
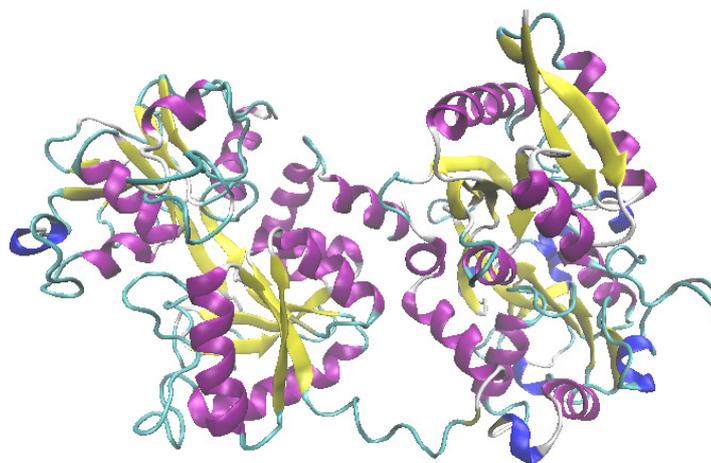


*Structure of conalbumin derived from protein crystallography (PDB entry 1N04)*

The small angle scattering patterns from dilute protein molecules in solution typically exhibit a number of distinctive regions, with a Guinier region at low q from which the radius of gyration is obtained - this region, typically <0.02Å$^{-1}$ should be carefully checked for any signs of aggregation or structure factor, which have a deleterious effect on the P(r) function.  The mid q region would indicate the overall shape, and any high portions above ~0.15Å$^{-1}$ may show some conformational detail, depending on the quality of the data.

Once the data has been loaded, at the bottom of the Data Explorer select 'Inversion' in the drop down menu, then click on 'Send data to'.  The FitPage transforms into a P(r) Inversion analysis page. At the bottom of this, click 'Calculate'.  The measured data is displayed, along with a window showing the derived function.



<div style="background-color: #e6e68a">
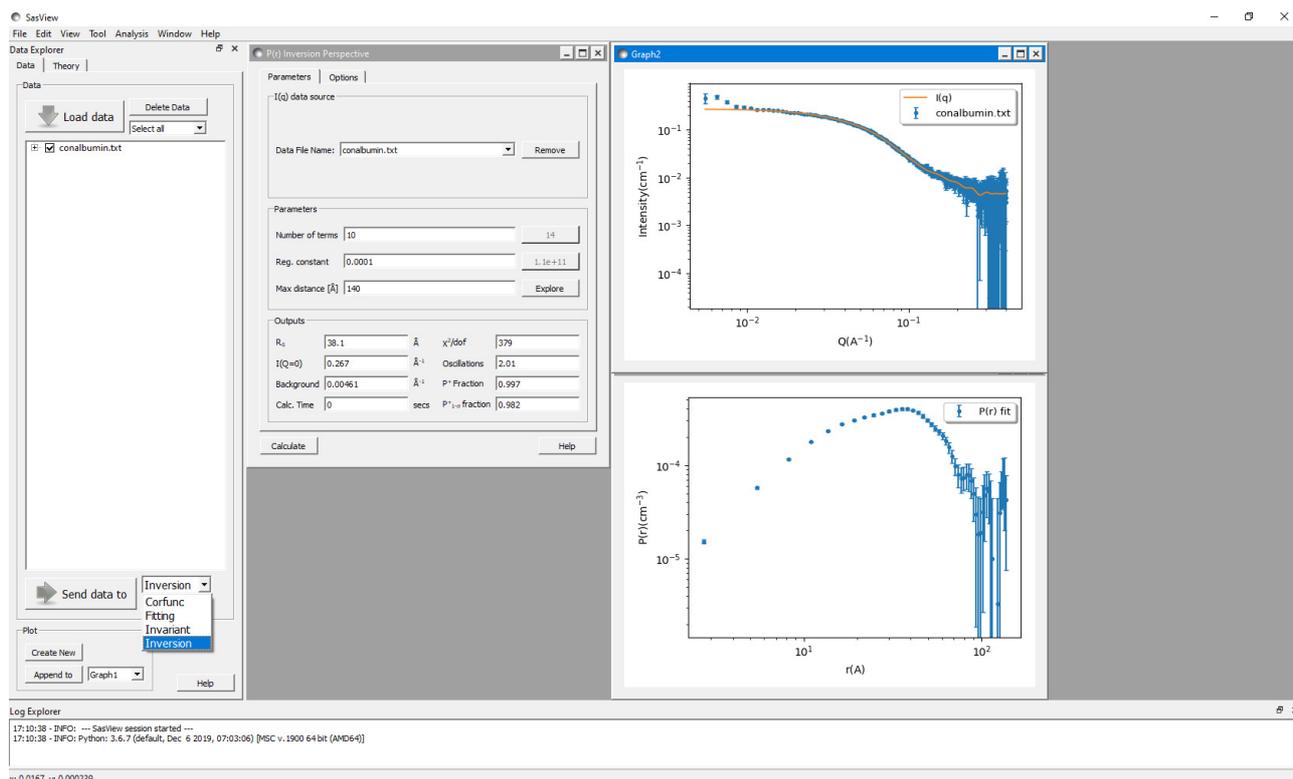
**Tip:** If you want to change the axis scales to alter the appearance of the plot, right-click on the graph and select Change Scale.

</div>

Under the 'Options' tab there is a box to enter any slit smearing data, if appropriate for your instrument.  Otherwise the height and width boxes may be left blank.

The q range used for the regularisation can be adjusted using the 'Q range' boxes. Generally, one would want to use as much of the Guinier region as possible, but do remove any 'noisy' points near the beamstop (i.e. at $q_{min}$) which your reduction software has not masked off!!  Here, let us restrict the $q_{min}$ to 0.015Å$^{-1}$.  At high q, for most protein systems it may be better to place $q_{max}$ near 0.2Å$^{-1}$ in the first instance – let's do this here! Using higher values of q will require the background subtraction to be highly accurate.  You may be able to infer how well your buffer has been subtracted by inspection of the Kratky plot – a kick up at high q may suggest problems and should cause you to lower $q_{max}$.  In addition, SANS data in higher fractions of $H_2O$ may present problems in this regard. Adjust $q_{max}$ down accordingly!  For small errors in background subtraction, SasView can

account for this using the 'Estimate background level' radio button. Here we will leave this feature turned on – it is possible to make a manual entry should you wish.
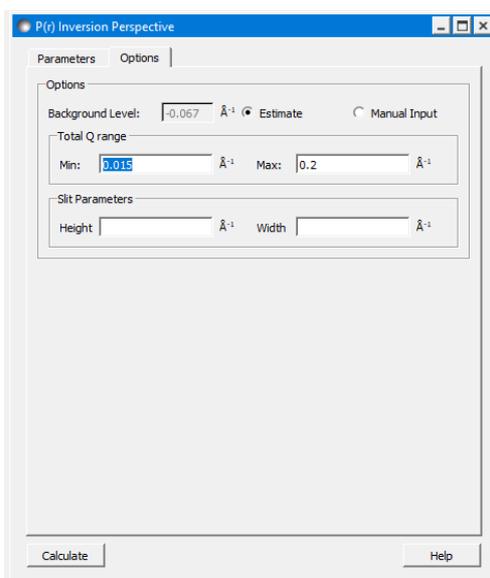


We can see in the plot above that a regularisation has already been computed using the default value of 140Å. Notice how the fit line to the experimental data (in green) looks quite good but the actual computed P(r) function oscillates rapidly out to high values of D. This is a simple indication that the solution is incorrect!!

You will note that the 'Number of terms' and 'Regularisation' boxes have suggested values filled in – these are continuously recomputed and you should use them by clicking on the suggested value unless you have a good reason not to!
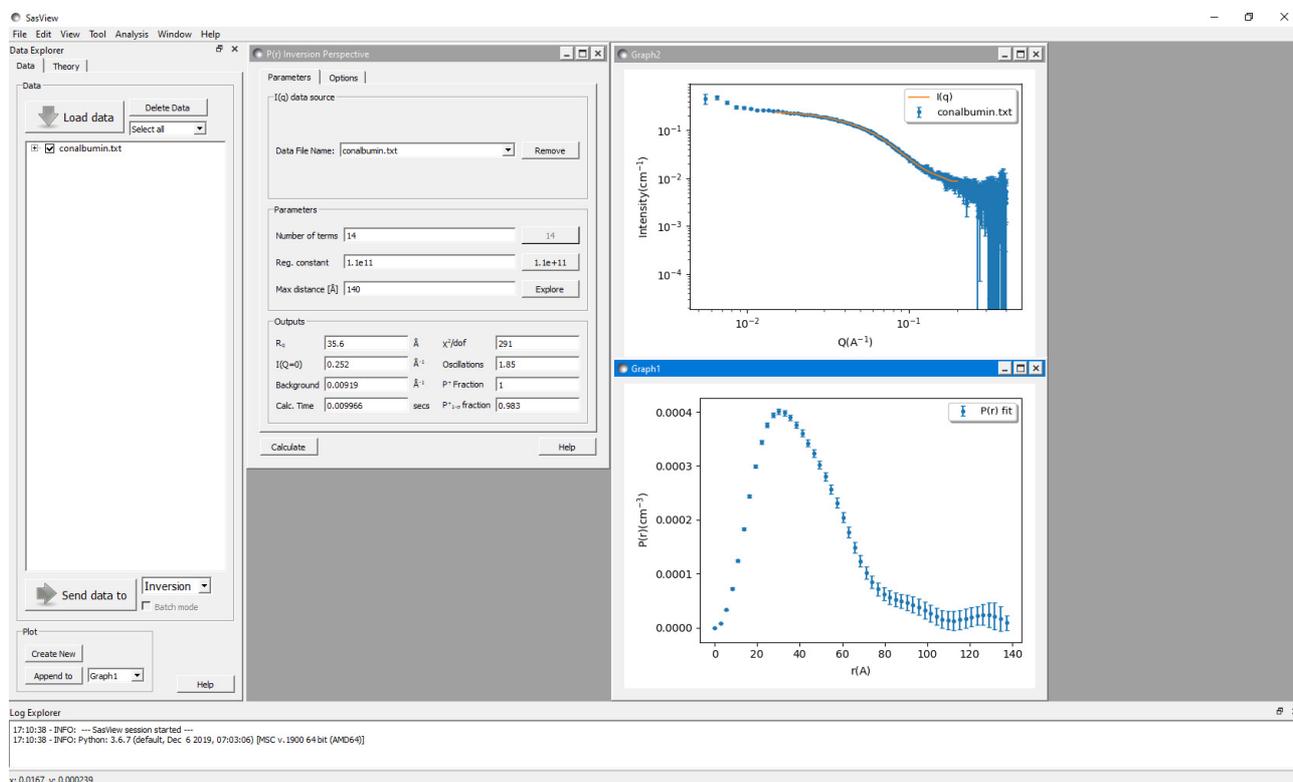
We should explore a range of possible r values to get it right….
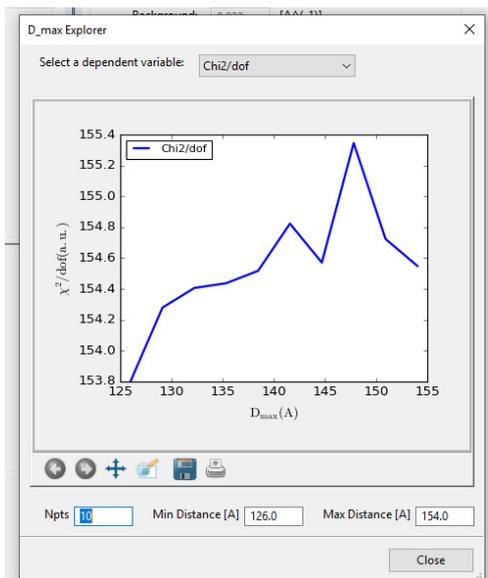


Clicking the 'Explore' button will bring up a separate window..

This window gives the chi-squared fit to the data versus degrees of freedom (other statistical models are available in the 'dependent variable' drop-down if desired). The correct value should be at a minimum – here we see this is around 125Å (in fact it is probably lower than this). Input this value into the P(r) Control Panel and set the new suggested number of terms and regularisation, and then hit compute!
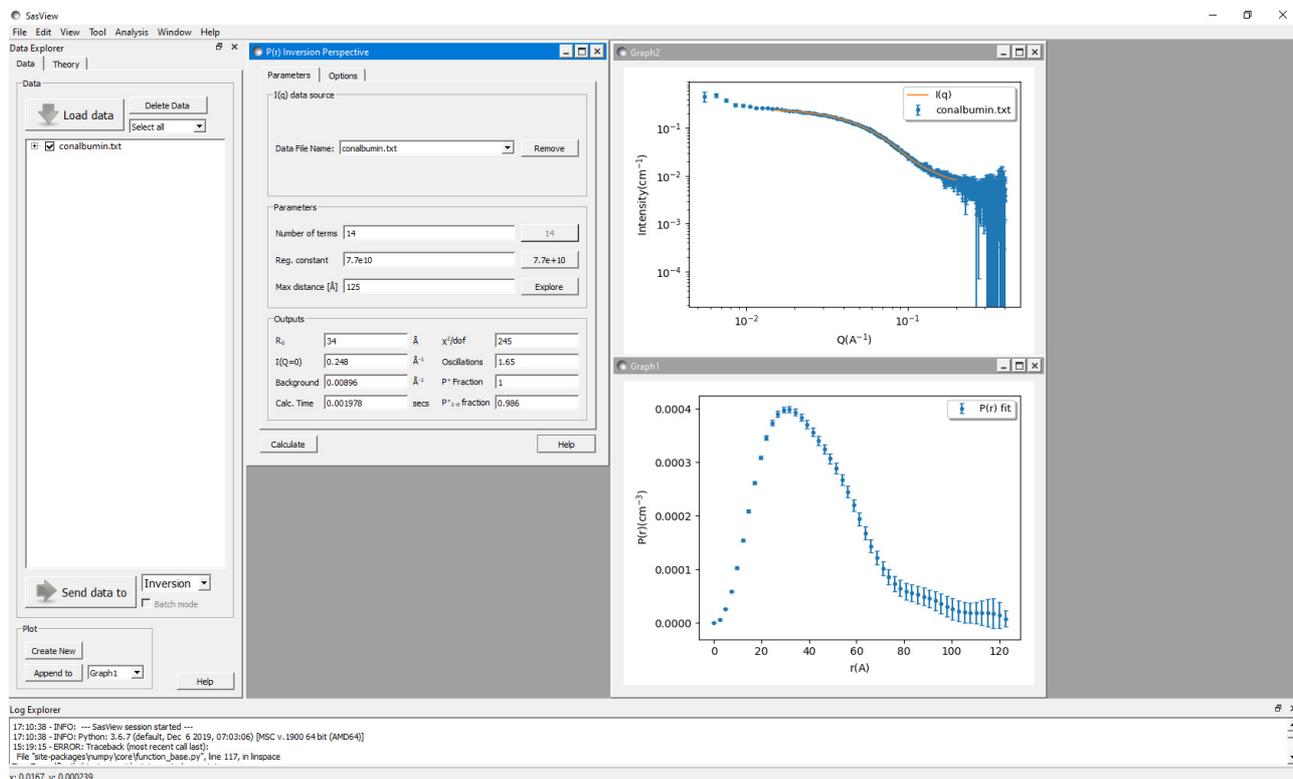


This looks better – notice how the fit line through the experimental data looks sensible. The P(r) function may need some extra work – notice how at large r values it does not seem to stabilise. Let's try using a larger range in the 'Explore' window -
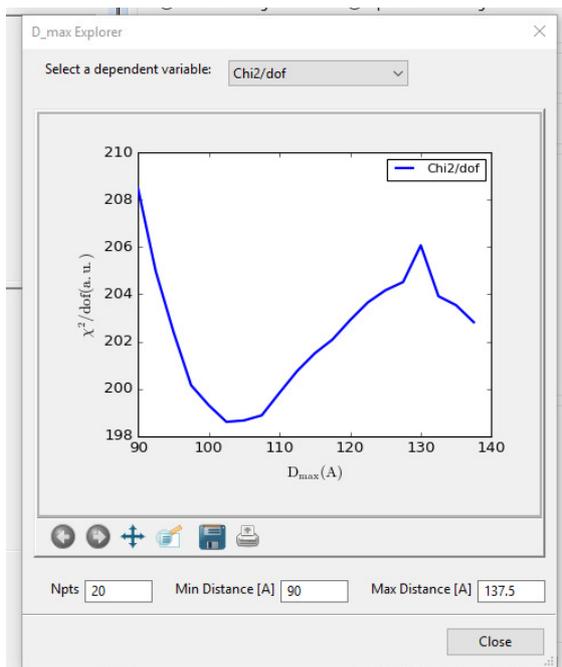
Here we have decreased the minimum distance to 90Å. You may wish to experiment with increasing the number of points – here we have tried 20 rather than 10. This may improve results for certain systems without too much of a computational penalty. Here the minimum is around 102Å.
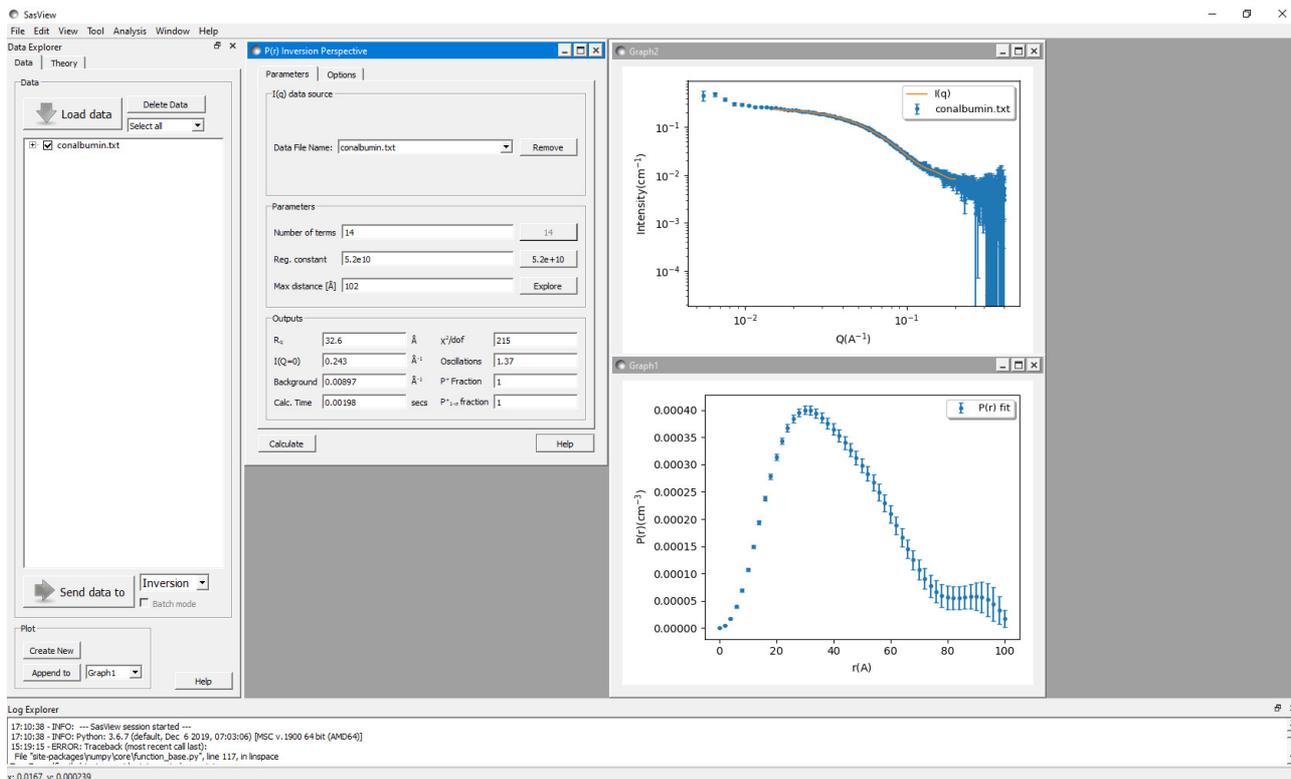


Now, we should inspect the result. In most systems, one would have to make several iterative attempts to find the value at which chi$^2$/dof reaches a minimum. Check the experimental data – see here, that the green fit line is a good fit with the experimental data – it is good here to check for any oscillations or discontinuities in the fit line which is a hint that the transformation has gone wrong. Now, we can check the derived P(r) function. The 'Gaussian' shape suggests we have a globular system, with the maximum distance within the system ~102Å. The function has an extensible part from 80-102Å which should be considered with some care.

It would also be considered good practice to compare the computed Rg, shown in the output box, with the experimentally derived Rg – if these diverge, this is a danger sign that you are on the wrong track, and would indicate that you have a polydisperse particle mixture, or aggregation/structure factor.  In the example above, a quick derivation of the Rg using the SasView inbuilt Guinier function (found under Fitting and Shape Independent models!) gives an Rg of 33Å, in agreement with the real space function. Careful inspection of the Guinier region is indicated here.  Failing to find a stable solution may be a warning sign that you have an intrinsically disordered or fractal system – such shapes are not amenable to this form of analysis.  If in doubt, there are a number of online tutorials and international training courses specifically for BioSAS that you may find useful!

## Example 2

This is an example of data which should suggest the need to exercise caution in the interpretation of the distance distribution function.
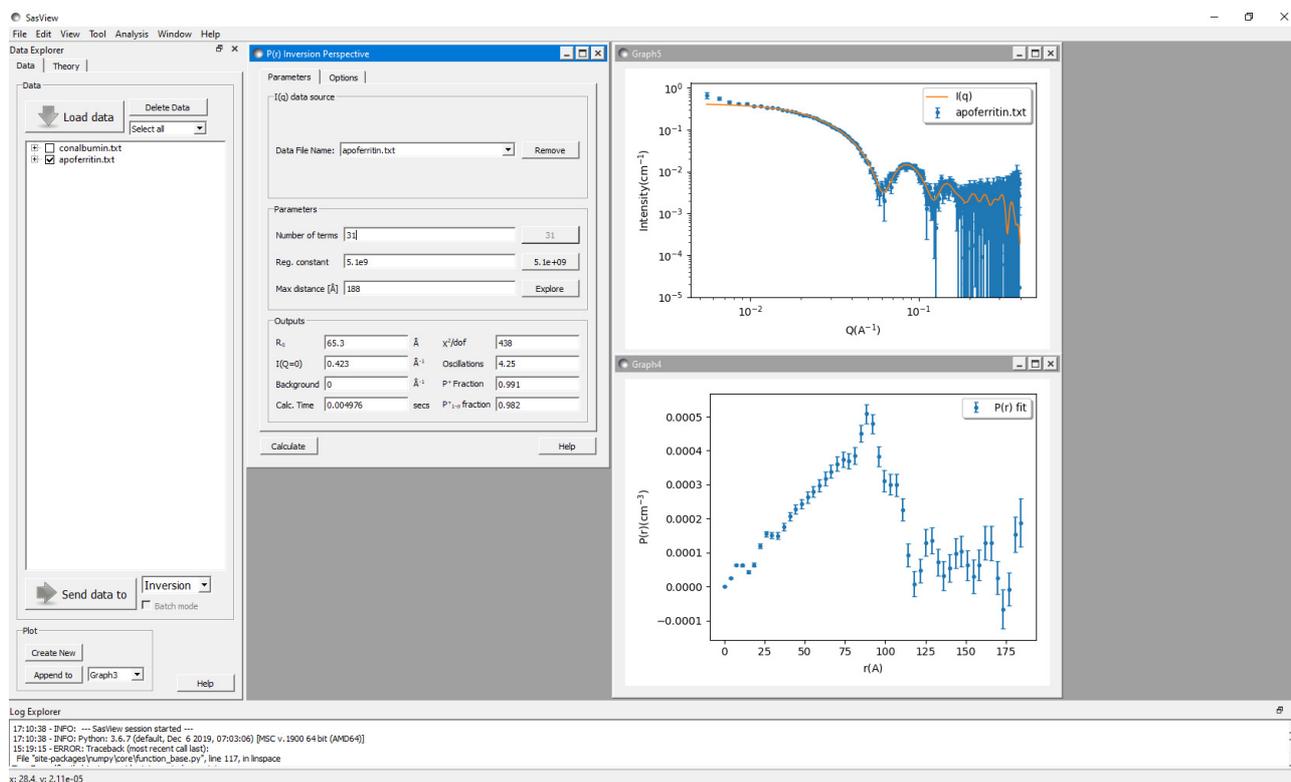


The data here shows a SANS measurement on apoferritin.  Notice that the P(r) function maximum is canted towards the right i.e. towards large values of r.  This could indicate a hollow system (and this is true of apoferritin) but also note the extended section in the P(r) function at large values of r.  This should raise our suspicion that there is something unusual going in this dataset!!!  In this example, it's likely that there is some structure factor, as the shape of the P(r) function is roughly what one might expect from apoferritin (i.e. a core-shell type structure), but the large extension doesn't really make sense – take a look at a typical apoferritin PDB structure e.g. 2W0O!