SasView Tutorials

Basic 1D Data Fitting
in SasView Versions 3.x/4.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.

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

Steve King (stephen.king@stfc.ac.uk)

Learning Objective

This tutorial will demonstrate how to fit <u>individual</u> 1D ('intensity' versus Q) datasets in SasView and showcase some of the associated functionality of the program. Simultaneous or batch fitting of multiple 1D datasets, and the fitting of 2D datasets, are considered in separate tutorials.

It is assumed that the reader has some familiarity with the purpose and principles of data fitting. If not, these Wikipedia articles provide an overview:

- https://en.wikipedia.org/wiki/Curve_fitting
- https://en.wikipedia.org/wiki/Mathematical_optimization

The program interface shown in this tutorial is SasView Version 4.2 running on a Windows platform but, apart from a few small differences in look and functionality, this tutorial is generally applicable to any earlier version of SasView on any platform. However, there is a separate tutorial for using the new program interface released with SasView 5.0.

Glossary

a priori information	Known facts about the system whose datasets are being modelled that can guide the selection of model or model parameters.
Chi-square (X^2, 'Chi2')	A statistical test of how well a chosen model fits the data with a given set of model parameters. In SasView this means $$X^2 = \sum \left(\left(I(Q)_{meas} - I(Q)_{calc} \right)^2 / E(Q)^2 \right)$$ where I is the scattering intensity and E is the error on the intensity value. Clearly, as $X^2 \to 0$, the better the model fit is. NB: The SasView interface actually reports a variation of chi-square called the <u>reduced chi-square</u>, sometimes referred to as the '<u>goodness-of-fit</u>' $$X^2_{Reduced} = \sum \left(\left(I(Q)_{meas} - I(Q)_{calc} \right)^2 / E(Q)^2 \right) / \left(N_{pts} - N_{params} \right)$$ where N_{pts} is the number of data points in the dataset and N_{params} is the number of model parameters being optimised (which may be less than the total number of parameters in the model!). As $X_R^2 \to 1$, the better the model fit is.
Compute [button]	Perform a direct calculation of the model with the current parameters but <u>without any optimisation</u>

Correlated parameters	To obtain the best solutions the optimiser needs to explore the widest possible parameter space. That is best achieved if every parameter can be considered independent of every other parameter. However, there will be instances where this is not the case and one or more parameters may be correlated. In such cases the best approach is to fix the values of some of correlated parameters using *a priori* information and then optimise the remaining values. A particularly common instance of correlated parameters encountered when model-fitting SAS data is when the components of the forward scattering intensity $$I(0) = \phi V (\Delta \rho)^2$$ where Φ is the volume fraction of scatterers, V is the volume of one scatterer (and so dependent on size parameters), and Δρ is the contrast (difference in SLDs), are separated out as individual parameters in a model. Correlated parameters often manifest themselves in the SasView interface as having very large uncertainties.
Fit [button]	Perform a calculation of the model <u>optimising the selected parameters</u>
Model-fitting	The process of finding a good yet <u>physically-realistic</u> mathematical description ('solution') for a dataset or collection of datasets. The procedure employed to achieve this is called optimisation.
OpenCL	A low-level software framework that allows calculations to be distributed between any compatible processors (eg, GPUs as well as CPUs) on the host computer. OpenCL can speed up demanding model-fits if suitable hardware is available.
OpenMP	A software framework that permits shared-memory multi-processing (ie, parallelisation) of calculations. OpenMP can speed up demanding model-fits if suitable hardware is available.
Optimiser	The mathematical algorithm used to perform the model-fitting.
Polydispersity	Where one or more model parameters have a distribution of values. SasView allows for 2 types of polydispersity: *Size* where, for example, the radii and/or lengths of the scatterers have a distribution of values *This will apply in most instances*

	Angular	where the scatterers are anisometric in shape (eg, cylinders) and exhibit preferred orientations with respect to Q, for example, in under shear or in a magnetic field
Reduced Chi2	See Chi-square	
Residual	The difference between the measured and calculated values of a function at a given point. In SasView this means: $$R(Q) = I(Q)_{meas} - I(Q)_{calc}$$ where *I* is the scattering intensity. NB: The SasView interface actually reports a variation of the residual called the <u>normalised residual</u> $$R(Q)_{Normalised} = (I(Q)_{meas} - I(Q)_{calc})/E(Q)$$ where *E* is the error on the intensity value. A normalised residual can be thought of as the number of standard deviations between the measured value and the calculated value. Thus, for a good fit, 68% of the values will lie within -1 < $R(Q)_{Normalised}$ < +1, and 95% within +/-2. Residuals larger than +/-3 indicate significant problems with either the input data or the choice of model or model parameters.	
SLD	Abbreviation for Scattering Length Density, a measure of the ability of a molecule to scatter. Strictly speaking, SLD is a SANS quantity, so if fitting SAXS data use electron density values in their place. SLD values (neutron and X-ray) can be calculated with the SLD Calculator Tool in SasView.	
Smearing	Sometimes the instrumental geometry used to acquire the experimental data has an impact on the clarity of features in the scattering curve. For example, peaks or fringes might be slightly broadened. This is known as Q-resolution smearing. To compensate for this effect SasView can add a resolution contribution into a model calculation (which by definition will be exact) to make it more representative of what has been measured experimentally. SasView provides 4 smearing options:	

|		
	None	no smearing correction is applied
	Use dQ Data	the measured Q-resolution for each data point is used to apply a smearing correction *This is the default if dQ data is present*
	Custom Pinhole	SasView will apply a smearing correction suitable for data measured on 'pinhole geometry' instruments (eg, most SAXS/SANS instruments)
	Custom Slit	SasView will apply a smearing correction suitable for data measured on 'slit geometry' instruments (eg, most USAXS/USANS instruments)
Theory	The name SasView gives to a model calculation.	
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 optimisation will have some associated range of uncertainty. Parameters with uncertainties that are more than 95% of the parameter value should be viewed with deep suspicion. NB: Unhelpfully, but for reasons of space, the SasView interface actually labels parameter uncertainties as errors.	
Weighting	An optimiser can be instructed to pay less or more attention to data points in a dataset by changing the weighting of those data points. SasView provides 4 weighting options:	
	No Weighting	all data points will be weighted equally
	Use dI Data	the data points will be inversely weighted according to their measured intensity errors (ie, less prominence will be given to data points with large errors) *This is the default if dI data is present*
	Use sqrt |(I Data)|	the data points will be inversely weighted according to the square root of their intensity values
	Use |(I Data)|	the data points will be inversely weighted according to their intensity values

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.

Mac OS

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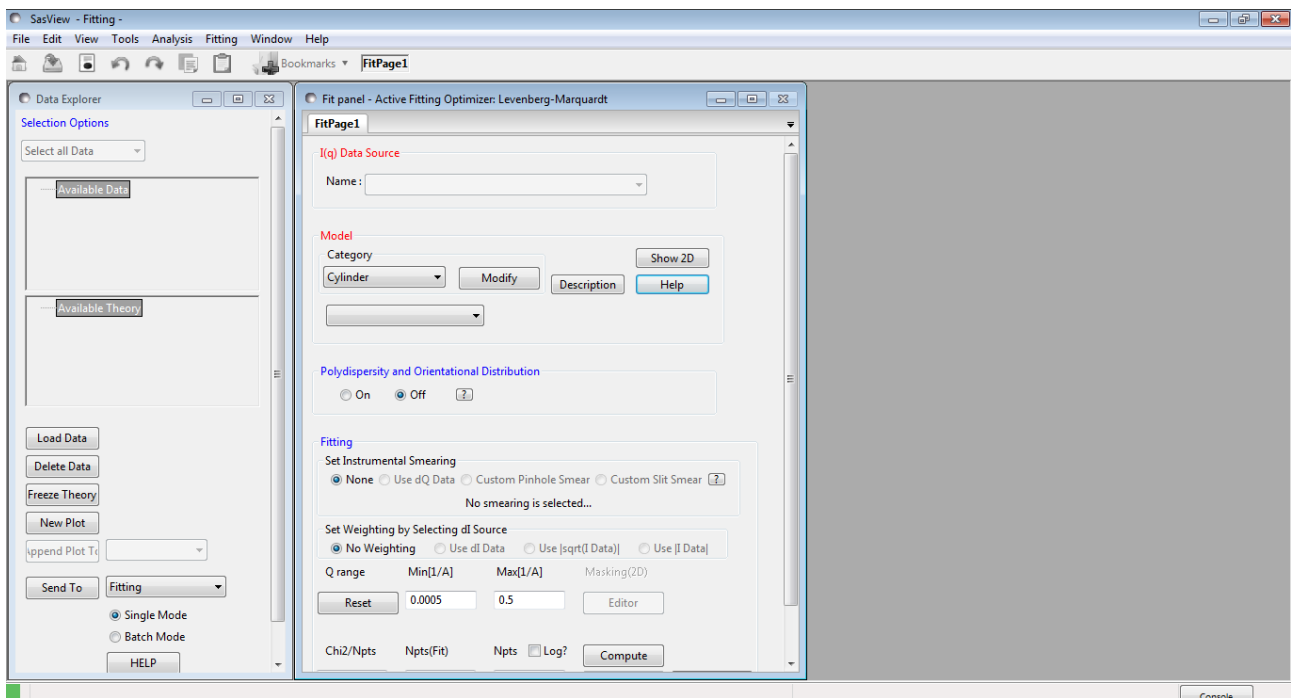
Example 1

This demonstrates a simple model fit, including the impact of correlated parameters and resolution smearing.

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 **ISIS_Polymer_Blend_RT2.xml** dataset and click the Open button.
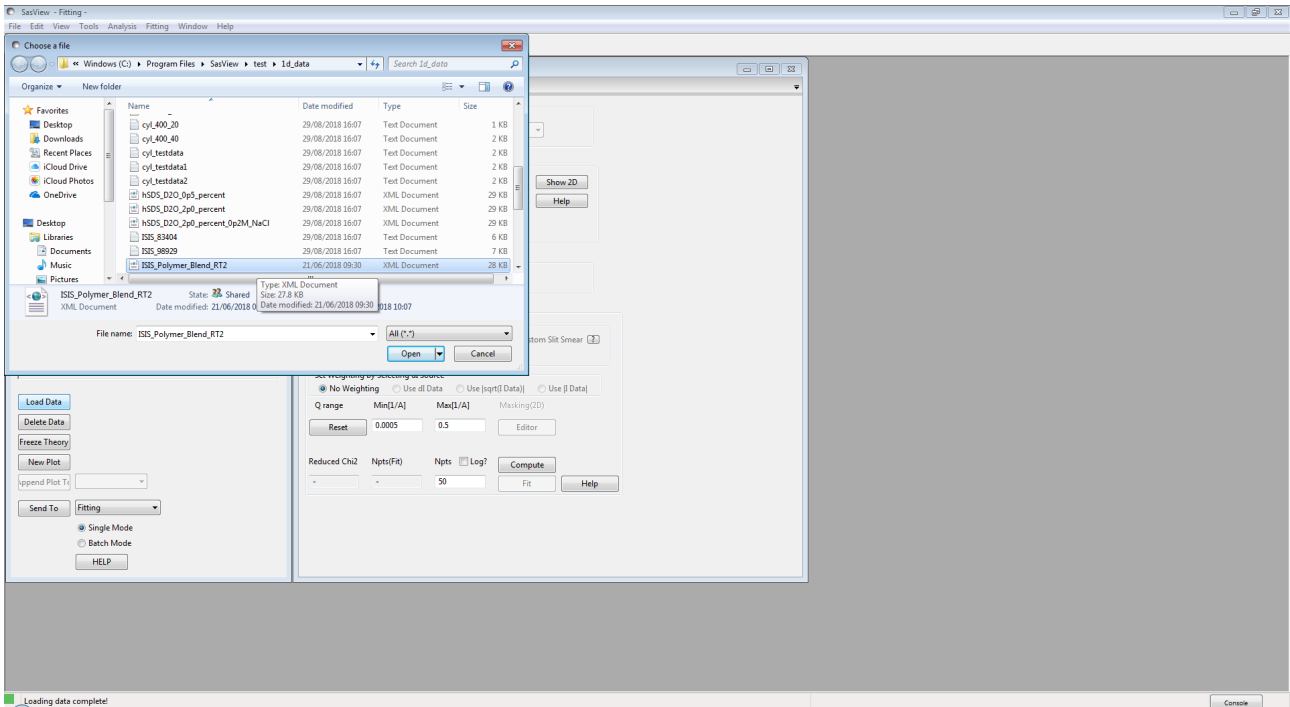
At the bottom of the Data Explorer panel, click the Send To button.
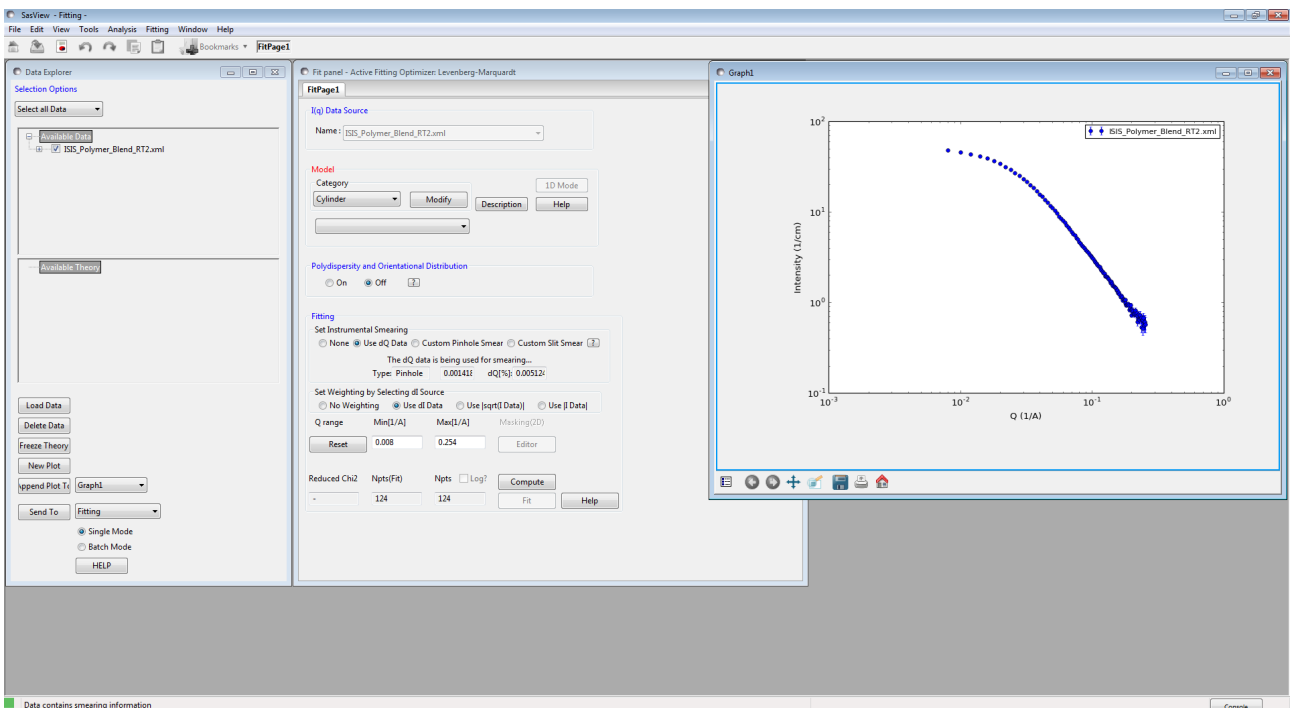
We now have to choose a model to fit to the data. To assist us in this task we have the *a priori* knowledge that this sample is a (solid) polymer blend (as it happens, of h8-polystyrene and d8-polystyrene). In the absence of solvent molecules the conformational characteristics of the polymer chains in the sample are described by a 3D random walk mediated by excluded volume constraints. Polymer physics tells us that this results in a Gaussian distribution of chain density about the centre-of-mass. The scattering from such

a structure was formulated by Debye (1947) and is implemented in SasView as the **mono_gauss_coil** model.

> **Tip:** The names of several models changed between SasView 3.x and SasView 4.x. For example, the mono_gauss_coil model was previously called the Debye model.

In the FitPage, select the model Category called **shape-independent** and then from the drop-down box below select the **mono_gauss_coil** model.
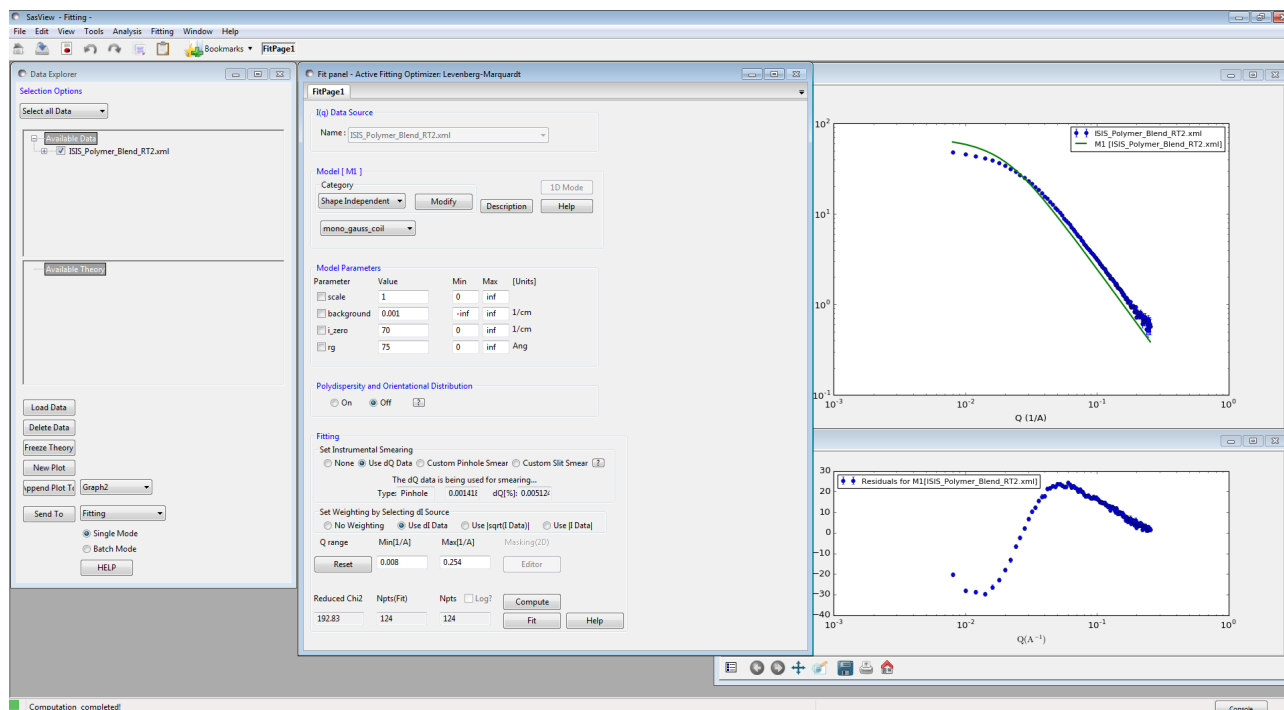

The model loads with some default parameters, the calculated (not optimised at this stage) scattering from which is shown as the green line on the upper graph. The lower graph shows the normalised residuals for the current model parameters.

By default, SasView displays all data on Log_{10} Q-axes, and scattering data on Log_{10} I-axes. To change this, right-click on a graph and select Change Scale. For this example, linear Q-axes are better.

In the lower half of the FitPage notice that SasView has chosen to apply resolution smearing using the measured dQ data, and will also weight the datapoints with the measured dI data. It has done this because those data are present in the dataset.

> **Tip:** If present in the dataset, use dI and/or dQ data to achieve a more robust solution.

If you want to examine the contents of the dataset, right-click on the upper graph, select ISIS_Polymer_Blend_RT2.xml and then DataInfo. A separate Data Info window will appear. The amount of information presented by DataInfo will depend on the file format of the dataset (CanSAS-standard xml files contain the most information) and the compliance of the data reduction program that wrote them.
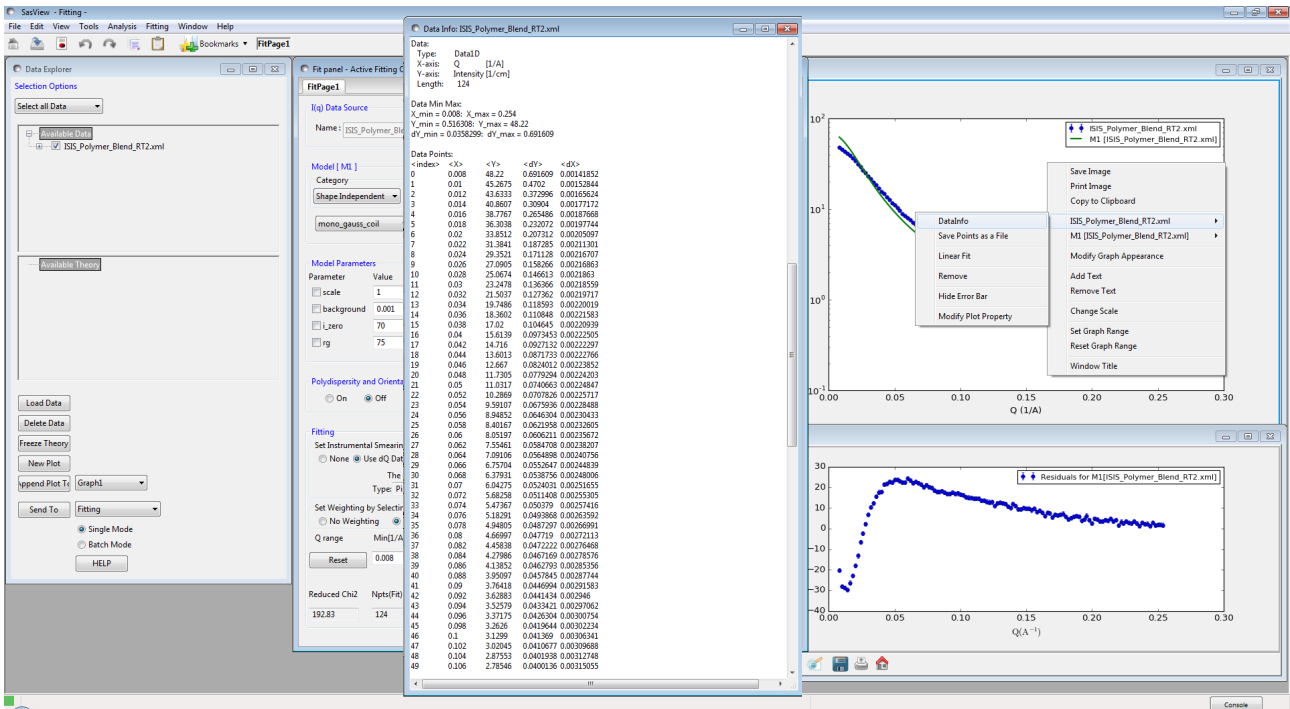
Looking at the upper graph, it is clear that the statistical quality of the measured data degrades at higher-Q values. We can exclude these points from the fitting using the Q range boxes at the bottom of the FitPage. Clicking in one or other of these boxes brings up two vertical bars on the upper graph; a red one for the minimum-Q limit, and a purple one for the maximum-Q limit.

Change the Q-limits either by typing values in the boxes, or by grabbing a line with the mouse pointer, left-clicking, and then dragging the line left or right. For this example, set the maximum-Q value to 0.19 Å⁻¹.
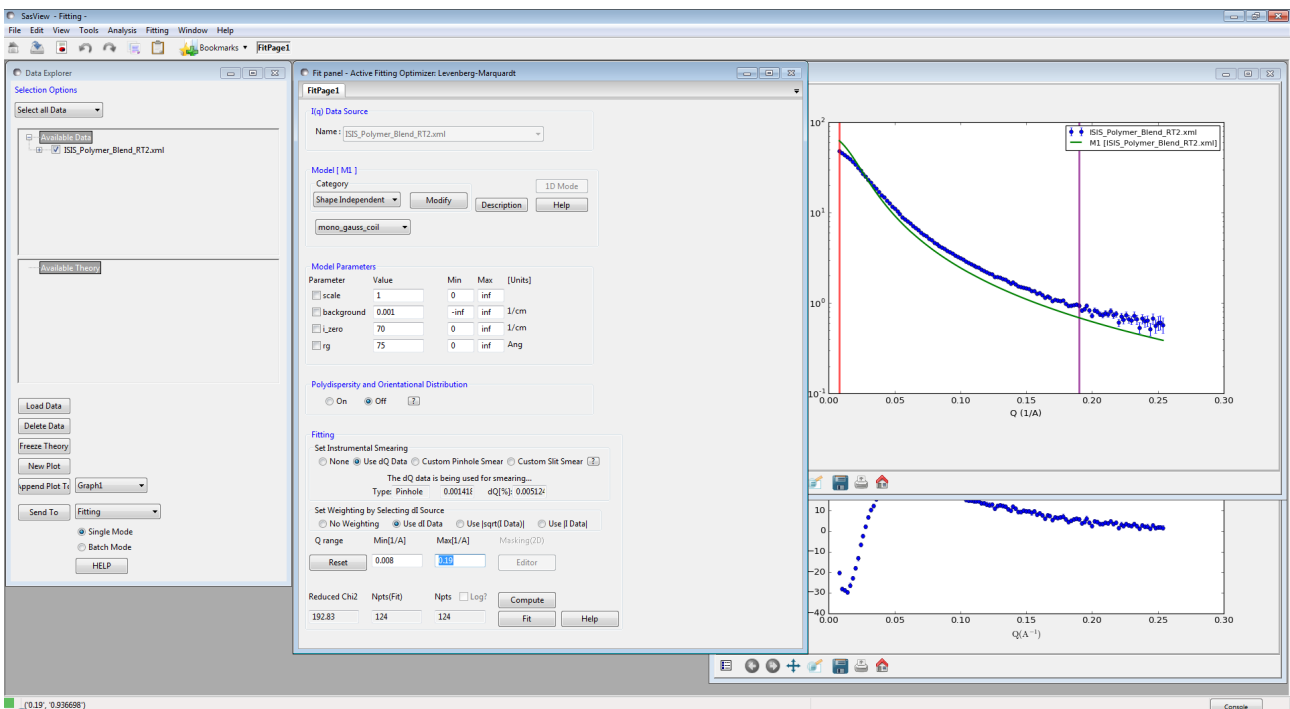
Each of the 4 parameters in this model has a checkbox next to it. When the box is checked and a fit is run, the optimiser will adjust the value of the parameter.

Check all 4 boxes and then click the Fit button.
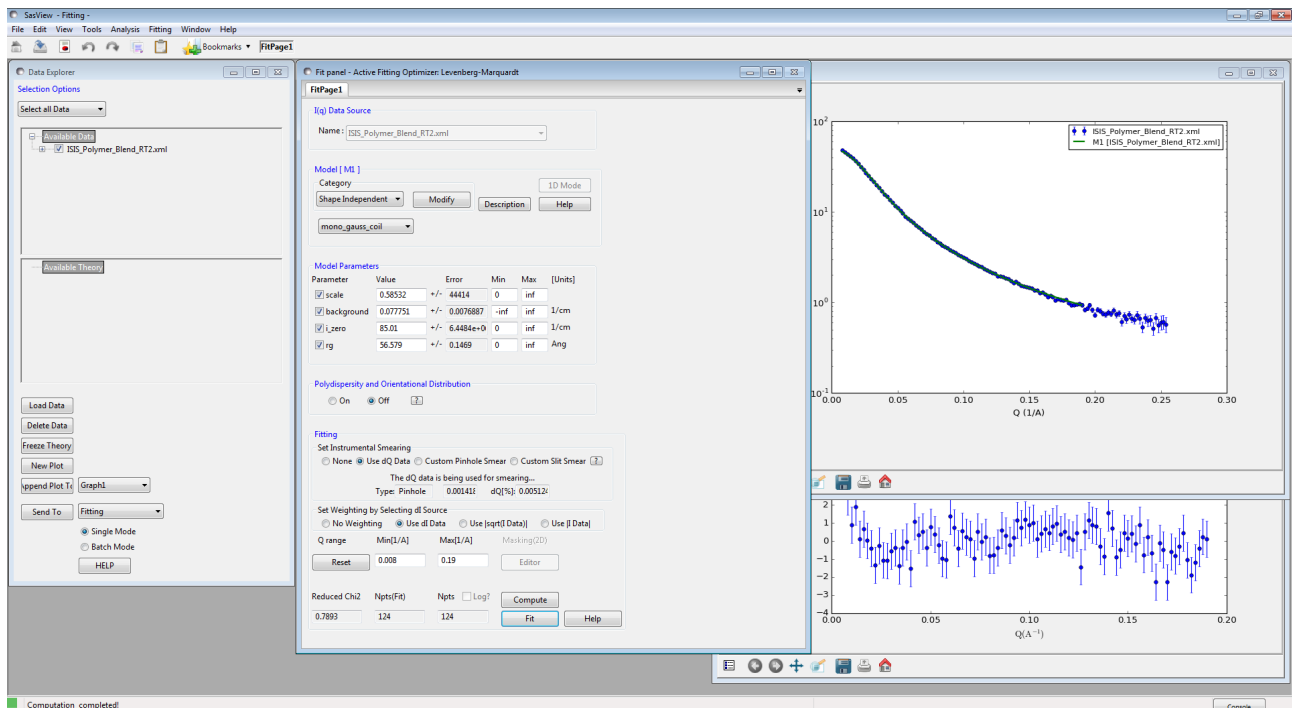
Several things happen very quickly:

- The progression of the fit is displayed in the Status Bar at the bottom of the SasView window
- Parameters and their uncertainties update in the FitPage
- The Reduced Chi2 (chi-squared) value updates in the FitPage
- The green line on the upper graph updates its position and shape
- The residuals plot in the lower graph updates
- Finally, *Computation Completed* appears in the Status Bar

What do you think of the solution? It looks good, right? The model calculation ('theory' curve) is running through the measured datapoints, the normalised residuals are all within +/-2 standard deviations, and the Reduced Chi2 is a very respectable 0.78.

But, look closely at the model parameters and their uncertainties. Two of the parameters, *scale* and *i_zero*, have nonsensical uncertainties. Why?

To understand this it is instructive to examine the help documentation for the model. Click the Help button on the FitPage (not in the Menu Bar) alongside the model name. The documentation tells us that the function that SasView is fitting is

$$I(Q) = scale \cdot I_0 \cdot P(Q) + background$$

So, in this model *scale* and *i_zero* are correlated parameters; changing one of them affects the other and the optimiser is unable to home in on a robust solution. To fit the data with this model we only need one of these parameters.

> **Tip:** SasView adds an overall scale factor (*scale*) and a Q-independent *background* to all its models for consistency, even though some models might not need one or other of these parameters. If in doubt, consult the model help documentation.

Uncheck the *scale* parameter and set its value to 1.0. Then run the fit again (click the Fit button) to get a final solution.
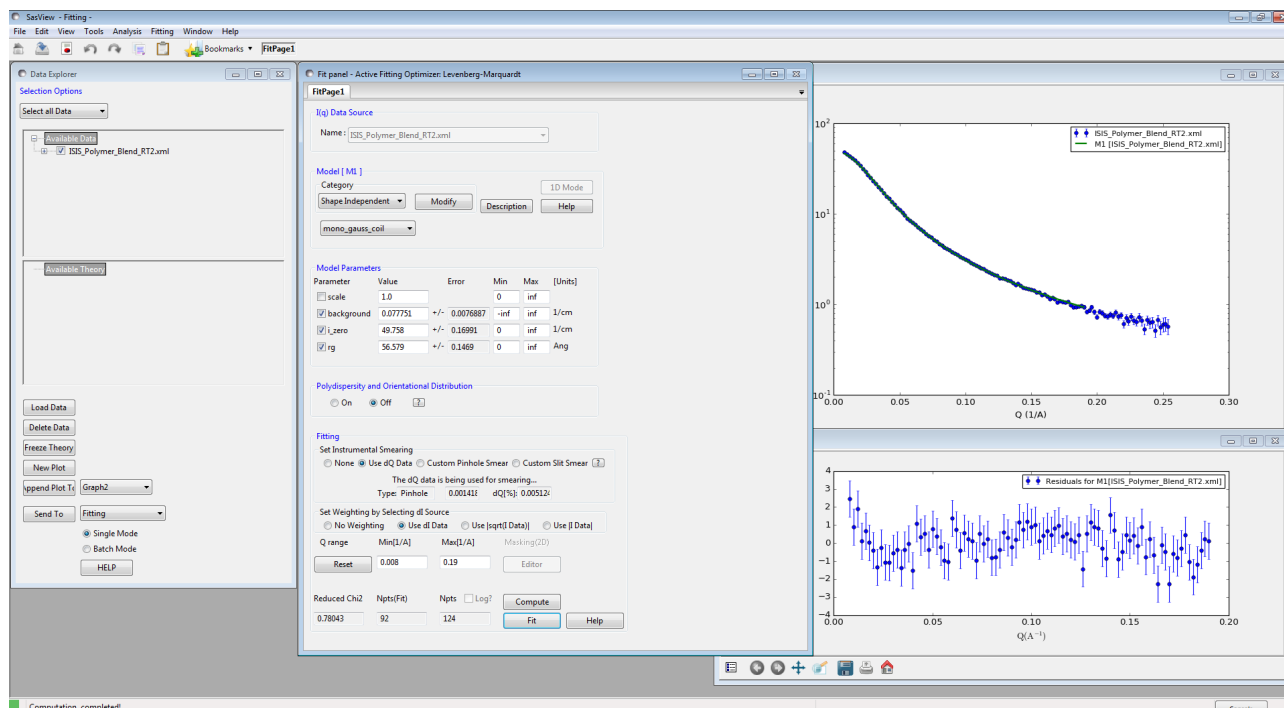
If you want to record the solution you have a number of options:

- If you want the parameter values and their uncertainties, go to the Menu Bar and select Edit and either Copy Params (for output suitable for a text editor) or Copy Params as... (for output suitable for Excel or LaTeX)

- If you want a copy of the graphs, right-click on a graph and select either Save Image (as .ps/eps, .jpg, .pdf, .png, .raw, .svg, or .tif), Print Image, or Copy to Clipboard

- If you want a file of the model function (the green line) and/or residuals, right-click on the appropriate graph and select either M1 [ISIS_Polymer_Blend_RT2.xml] or Residuals for M1 [ISIS_Polymer_Blend_RT2.xml] and then Save Points as a File (as plain ASCII text: .txt, or in CanSAS1D format: .xml)

- If you want a two-page pdf report containing all pertinent information including copies of the graphs, go to the Menu Bar and select Edit and Report Results

and finally
- A SasView session can be saved and reloaded as an 'Analysis' (an individual model fit), or as a 'Project' (everything done since starting your SasView session) by going to the Menu Bar and selecting File and either Save Analysis or Save Project.

We shall now examine some additional aspects of model-fitting this dataset.

Return to the FitPage and change the Instrumental Smearing to None. Then run the Fit again. Notice anything? The parameter values and the Reduced Chi2 value have all subtly changed. This is to be expected; Q-resolution affects the measured data, so it must have a bearing on the model fit.

Now change the model to the one called **poly_gauss_coil**. This model is an extension of Debye's original scattering law to include a Schulz-Zimm molecular weight distribution. This model has one extra parameter: the polydispersity of the polymer (M_w/M_n). Set the *polydispersity* to 1.03 (but leave it unchecked; this is more *a priori* information) and run the Fit again.

The parameter values change again, but there is a marked improvement (decrease) in the Reduced Chi2 value.

Lastly, reset Instrumental Smearing to Use dQ Data and run the Fit one last time.

Here is a summary of the fit results:

Param: Model	scale	bckgrd	bckgrd_err	i_zero	i_zero_err	rg	rg_err	poly disp	reduced chi2
mono use dQ	1	0.0777	0.0076	49.75	0.169	56.57	0.146	n/a	0.780
mono none	1	0.0758	0.0076	49.65	0.169	56.44	0.146	n/a	0.771
poly none	1	0.0676	0.0077	50.07	0.172	57.36	0.150	1.03	0.659
poly use dQ	1	0.0695	0.0077	50.17	0.172	57.50	0.150	1.03	0.669

So for this dataset (sample) allowing for even a tiny amount of molecular weight polydispersity – thereby maximising our use of *a priori* information - gives rise to a significant improvement in the quality of the fit! And though it does appear that including the Q-resolution smearing slightly degrades the quality of the fit in this example, it should be remembered that instrumental smearing is always present (and so should be handled) and this dataset was devoid of any fine structure (peaks, fringes).

Example 2

This demonstrates a fit requiring a structure factor and introduces the use of multiple FitPages.

From the **\test\1d_data** folder in the SasView installation directory, load the **hSDS_D2O_2p0_percent.xml** dataset and send the data for fitting.

This dataset is the SANS from a 2 %^{w}/$_w$ solution of the widely-studied surfactant h25-sodium dodecyl sulphate (SDS) in heavy water (D_2O). So let us consider the available *a priori* information.

The critical micelle concentration for SDS is about 0.2 %^{w}/$_w$, so the scattering must represent a concentrated micellar solution (though not too concentrated else the scattering would show diffraction peaks from a liquid crystalline lattice). SDS is also an anionic surfactant, so the micelles will be charged. Interactions between the charged micelles can be expected to manifest themselves in the scattering as a structure factor. And in the absence of electrolyte SDS micelles are reported to be approximately spherical.

To start with, select the **sphere** model.
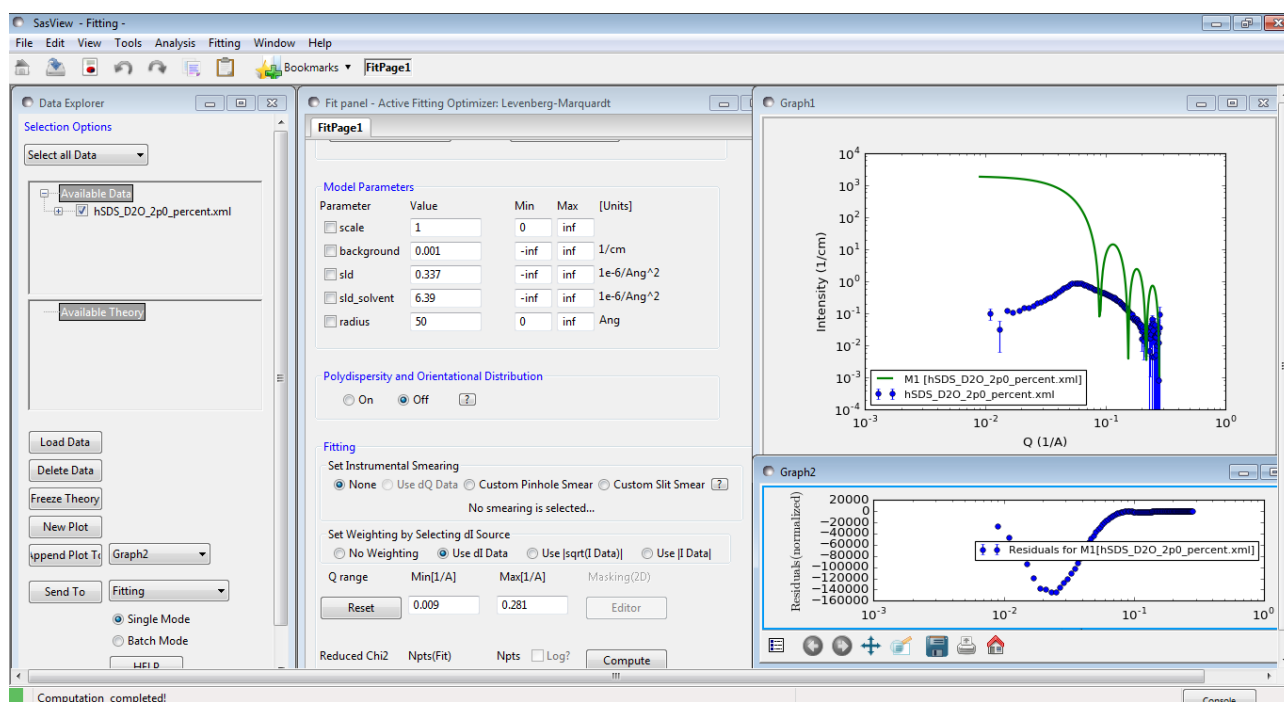
This model has 5 parameters. Two of them, *sld* (for the SDS) and *sld_solvent* (for the D_2O), can be calculated explicitly.

Go to the Menu Bar, select Tools, and then SLD Calculator. Enter the respective empirical formulae and densities in the relevant boxes and click Calculate (leave the neutron wavelength as 6 Å):

- SDS: Na1C12H25S1O4   ; density 1.01 g/cm^3
- D_2O: D2O1   ; density 1.11 g/cm^3

Hopefully you got values of *sld* = 0.337x10^{-6} Å^{-2} and *sld_solvent* = 6.39x10^{-6} Å^{-2}?

Enter these values in the FitPage and click Compute. Then check the boxes alongside *scale*, *background* and *radius* and click Fit.
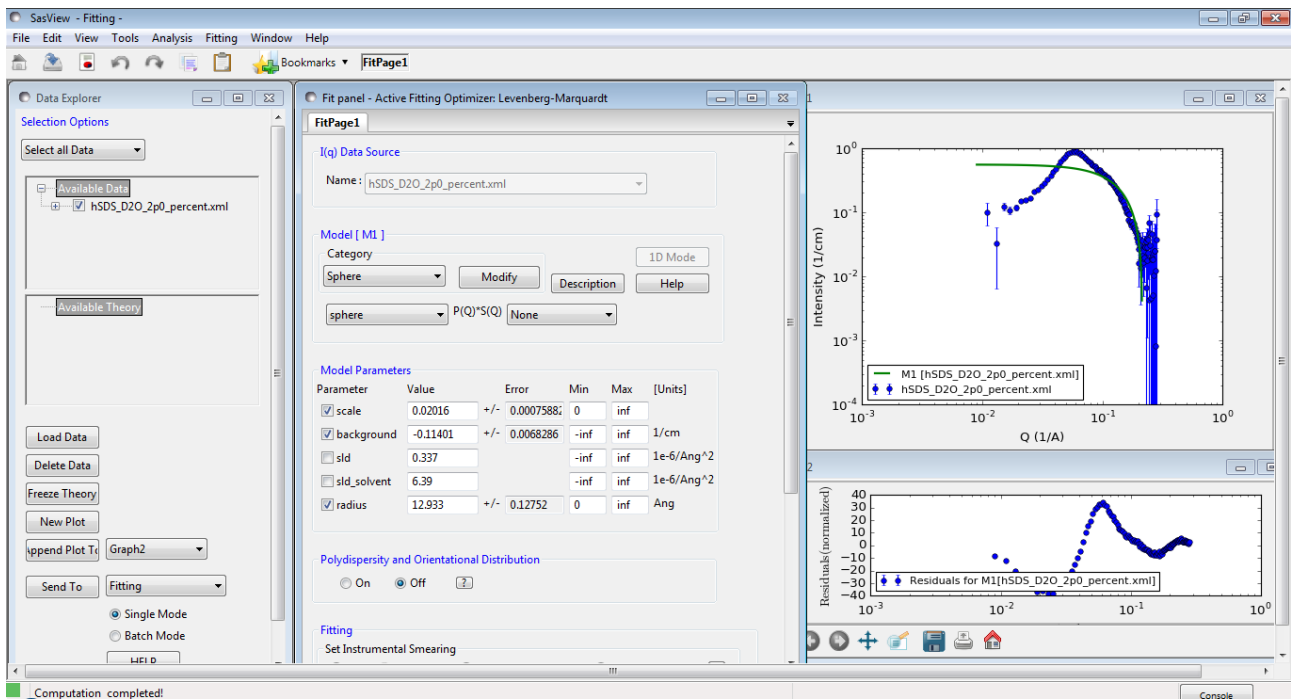
The green theory curve now represents the scattering from <u>non-interacting, monodisperse,</u> spherical micelles ~13 Å in radius. And the *scale* parameter in this case represents the volume fraction of micelles (encouragingly, about 2 %!). But to fit the broad peak – from the interacting micelles – we need to incorporate a structure factor – an S(Q) - into the model.
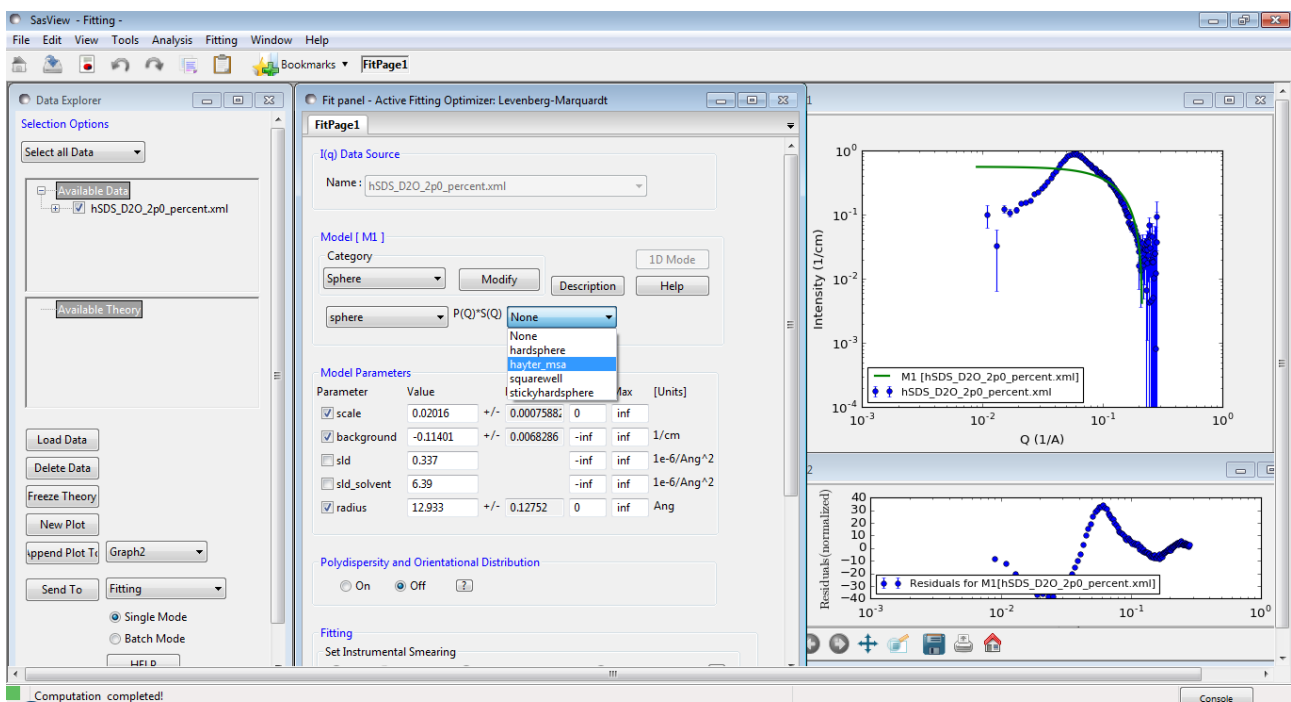
Click on the drop-down box just to the right of the P(Q) model drop-down and select the **hayter_msa** S(Q).

This adds 5 new parameters, but fortunately we know the values of several of them:

- *volfraction*:   2 %$^w/_w$ is approximately (2/1.01) %$^v/_v$ = 0.019
- *charge:*
- *temperature:*   298 K
- *concentration_salt:* 0 M
- *dielectconst:*   78.06 (Vidulich *et al*, *J Phys Chem*, 1967)

However, the *volfraction* parameter now duplicates the *scale* parameter. So set *scale*=1.0 and uncheck it, and check the *volfraction* parameter instead. Also check the *charge* parameter.

Run the fit again.
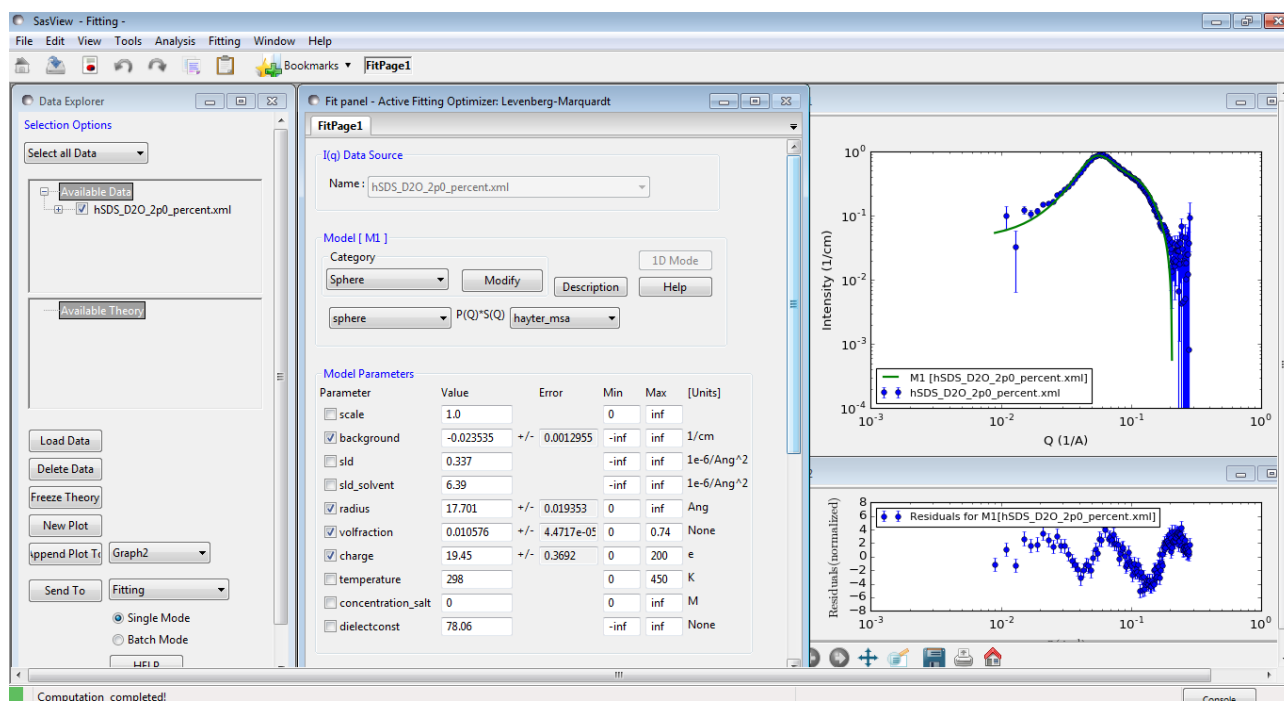
Incorporating the Hayter-Penfold MSA S(Q) has demonstrably improved the fit, but there remains room for improvement: the *Reduced Chi2* is still 6.35 and the residuals range between +/- 4. Also note that the *volfraction* parameter has dropped to ~1 %.

So how might the fit be improved further? One possibility is that water molecules are mixed in with the surfactant sulphate head groups, thereby increasing the effective SLD of the surfactant molecules (and reducing the scattering contrast; ie, *sld – sld_solvent*).

To test this, check the *sld* parameter and run the fit again.
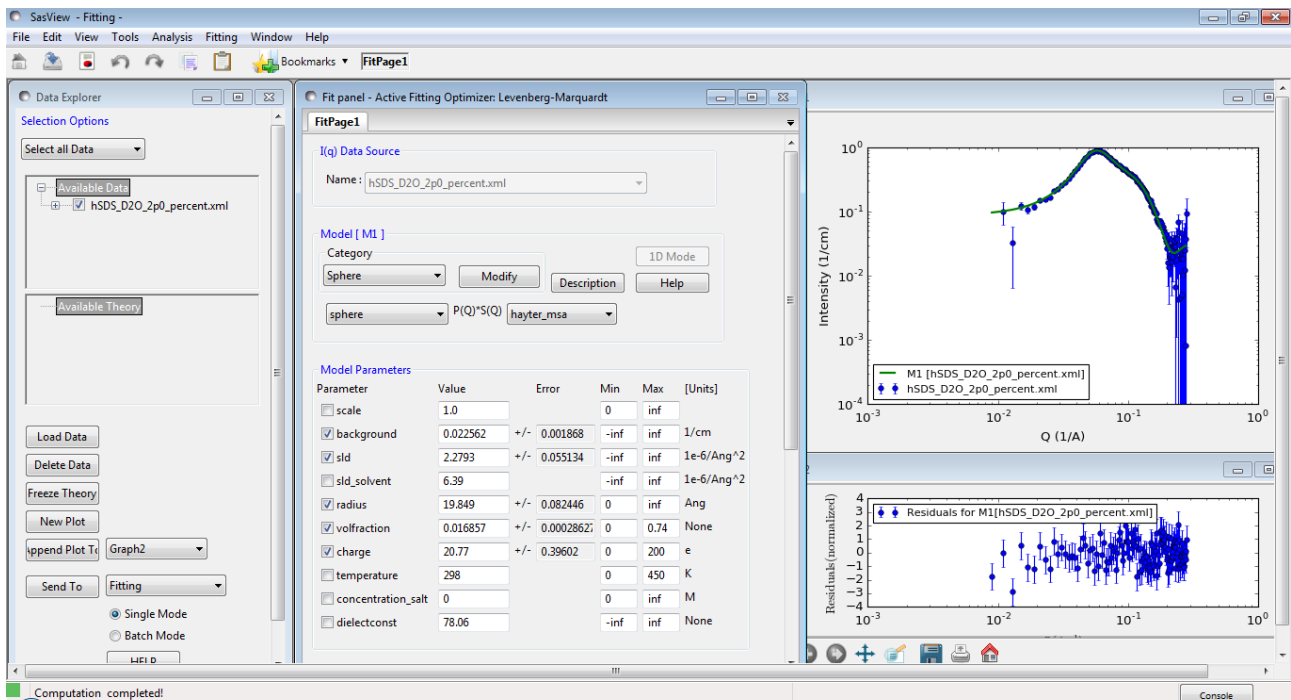
Indeed this does improve the fit further. The *Reduced Chi2* has dropped to 0.855, the residuals are within +/- 2, and the *volfraction* parameter has increased to almost 1.7 %.

But is there scope for an even better fit? Maybe. There is published evidence that suggests SDS micelles are actually slightly ellipsoidal.

To test this, use the Send To button to create a second FitPage and then select the **ellipsoid** model. Do not select an S(Q) for the time being.

Set *sld* and *sld_solvent* to their calculated values.

We know from the fit to the sphere model that the micelles have a radius ~20 Å so set *radius_polar*=20 and *radius_equatorial*=40 (ie, to give the ellipsoids an intial axial ratio of 1:2; a sphere would be 1:1 of course). Check the *scale* and *background* parameters and run the fit.

Now select the **hayter_msa** S(Q) as before and set the known values you used previously. And just like before, set *scale*=1.0 and uncheck it. But this time set the *volfraction* and *charge* parameters to the values determined from the fit to the sphere model; 0.017 and 20, respectively.
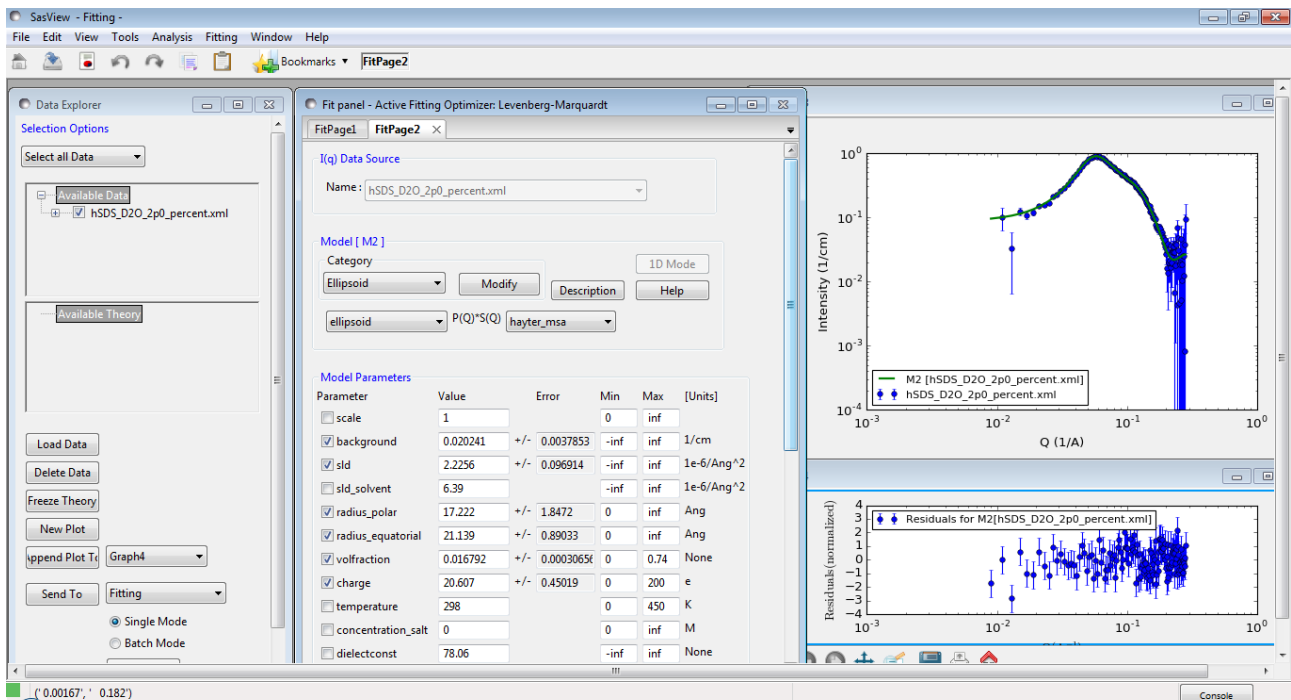
Check *radius_polar* and run the fit.

Uncheck *radius_polar,* check *radius_equatorial* and run the fit.

Now also check *radius_polar* and run the fit.

Now check *volfraction* and *charge* and run the fit.

Finally, check *sld* and run the fit.

A fit at least as good as the one to the sphere model is obtained, suggesting that the micelles are ellipsoids with an axial ratio of 1:1.2. So the SasView fitting is certainly consistent with the literature reports. But unambiguously distinguishing between the two cases would require more precise measurements and higher quality data.

Example 3

This demonstrates a fit requiring a structure factor and polydispersity.

From the **\test\1d_data** folder in the SasView installation directory, load the **Ludox_silica.xml** dataset and send the data for fitting.

This dataset is the SANS from a concentrated dispersion of Ludox® silica nanoparticles in light water (H_2O), so there will once again be an S(Q) contribution from charged spheres.

Select the **sphere** P(Q) model and the **hayter_msa** S(Q) model.

Once again, we have a lot of *a priori* information:

- *sld:*   3.4 $\times 10^{-6}$ Å^{-2} (calculated)
- *sld_solvent:*   -0.56$\times 10^{-6}$ Å^{-2} (calculated)
- *volfraction*:   18 %$^v/_v$ (based on the manufacturers data sheet; 40 %$^w/_w$)
- *temperature:*   298 K
- *concentration_salt:* 0 M
- *dielectconst:*   78.39 (Vidulich *et al*, *J Phys Chem*, 1967)

Enter these values and click Compute.
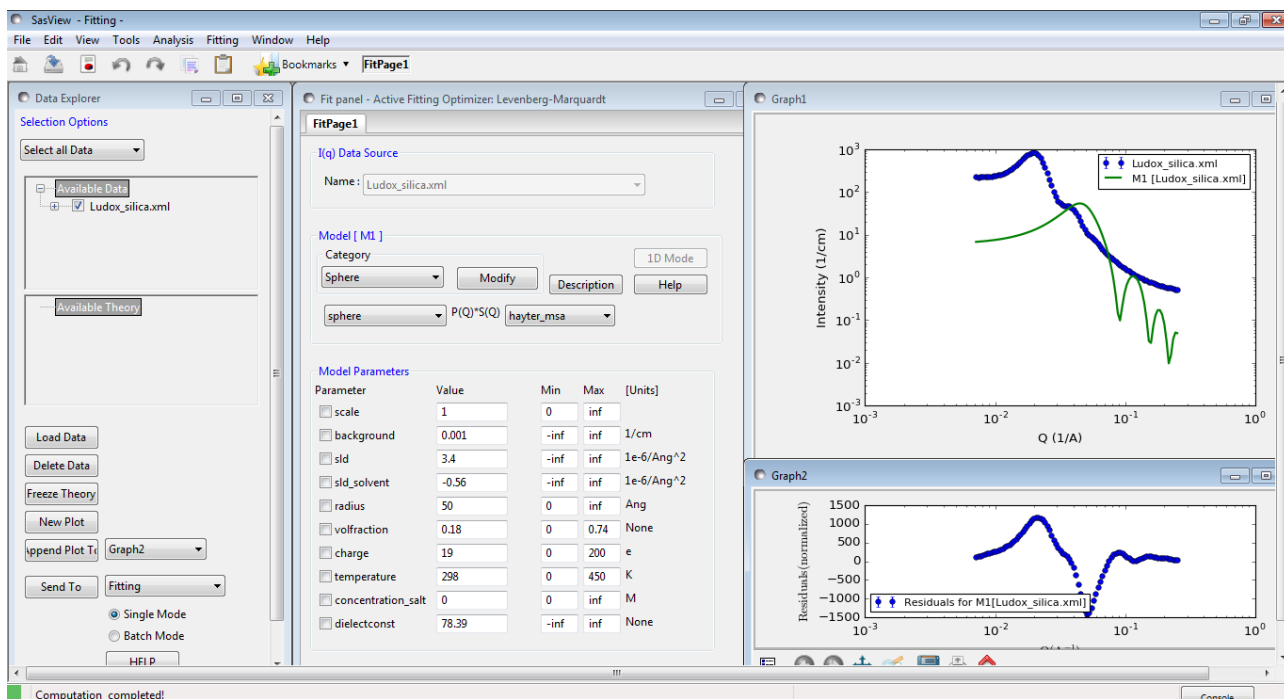
Clearly the particle *radius* is too small at present. Increase it to 150 Å and click Compute again.

Now check the *background*, *radius, volfraction*, and *charge* parameters and click Fit.
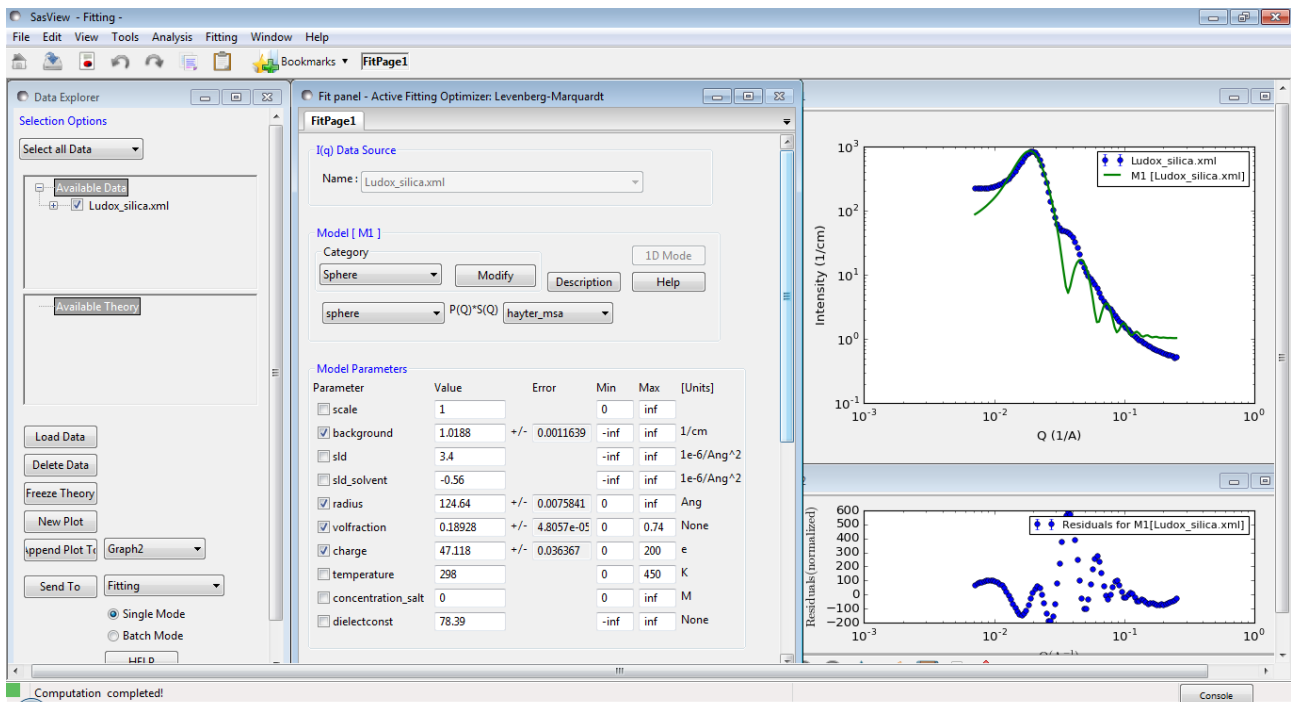
The problem now is that we are once again calculating a theory curve just for monodisperse particles. In Example 2 we managed to obtain a good solution without having to specifically address polydispersity, most likely because the micelles were quite polydisperse in size (ie, had a broad size distribution). In the present dataset, the fringes in the scattering curve suggest there is some polydispersity, but not too much.

Scroll down the FitPage and in the section headed **Polydispersity and Orientational Distribution** click the radio button marked **On**. This provides a further model parameter, the *Distribution of radius.*

Particle size distributions are usually described by a Log-Normal distribution, so use the dropdown to change the *Function* to **lognormal**.

> **Tip:** In SasView 3.x & 4.x the default type of distribution applied is a Gaussian. If you need a different distribution, select it before you set any distribution parameters. The SasView help documentation provides some guidance on appropriate use cases for different distributions.

The width of the chosen polydispersity distribution is determined by the *PD* parameter and the definition of this changes depending on the distribution. For the Log-Normal distribution:

$$PD = \sigma_{normal}$$

where σ_{normal} is the width of the underlying Normal distribution. But whatever the distribution, *PD*=0 represents monodispersity (a very sharp distribution) and *PD*=1 represents an extremely broad distribution.

Enter *PD*=0.15 and click Fit.

Now check *Distribution of radius* and click Fit again.
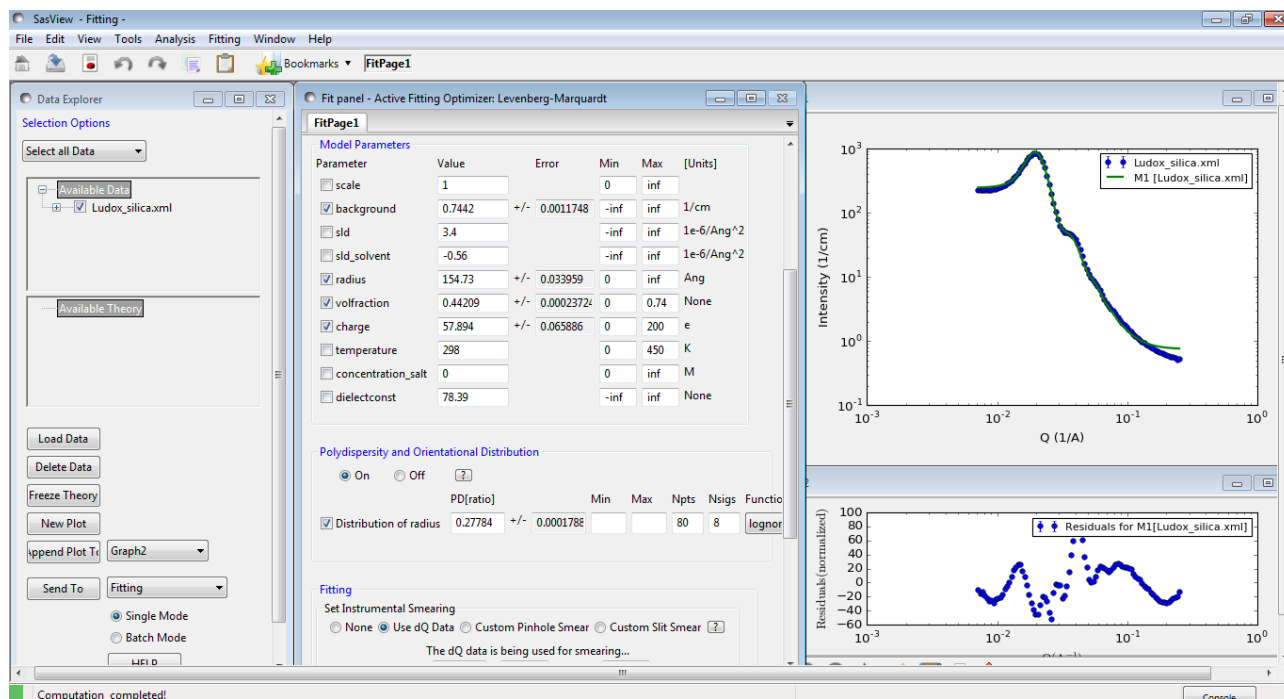
Whilst this looks like a good fit, notice that the *Reduced Chi2* is 780, and the residuals are quite poor. The *volfraction* parameter is also higher than we would like.

Optimising the *sld* value reduces its value slightly from 3.4×10^{-6} to 3.2×10^{-6} but also improves the *Reduced Chi2* to 620 with minimal effect on the other parameters. But this is as good a fit as is likely to be achieved in this instance. The reasons are beyond the scope of this tutorial.

Further Information

For further information, please consult the

> **SasView Tutorial Series**
>
> or
>
> **http://www.sasview.org**
>
> or email
>
> **help@sasview.org**