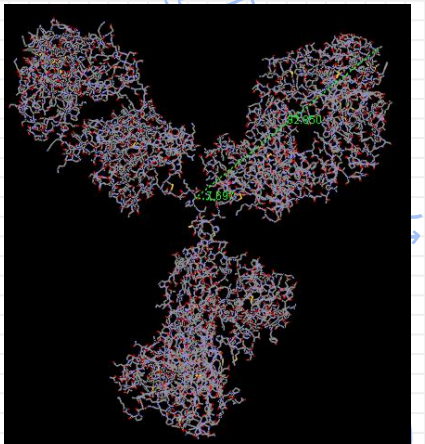
Implementing SANS Analysis Models for Concentrated Protein Solutions

Author: Alex Zheng – Urbana High School, Frederick //
UCLA
Mentor: Dr. Yun Liu
Collaborators: Paul Kienzle, Paul Butler, Jeffrey Krzywon

Background:

- My project develops software to supplement and improve Sasview, particularly in investigating the properties of proteins
  - There are many different focuses and parts to the project
  - Main goal is to develop fit models for Sasview
- By comparing SANS data to these models, size and interactions between proteins are revealed
- Important in cases such as pharmaceuticals
  - Protein-protein structure and interactions are essential to their applications.

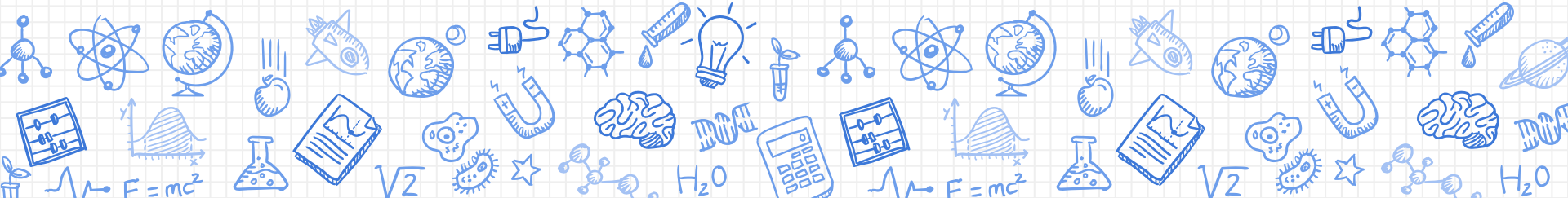


NIST MAB Protein

Special thanks to Dr. Christina Bergonzo for providing the MD Simulation files

Brief Overview:

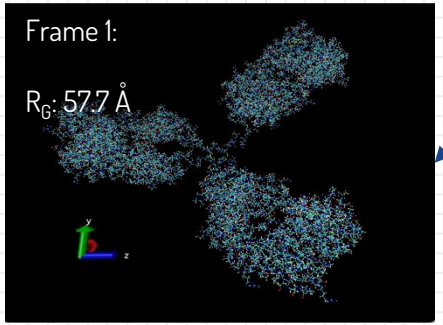
Radius of Gyration (R_G) // Scattering Pattern ($I(Q)$) // Beta Q ($\beta(Q)$)

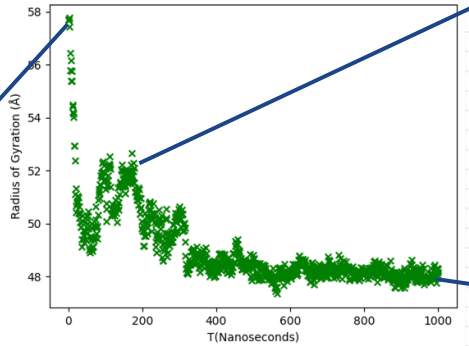


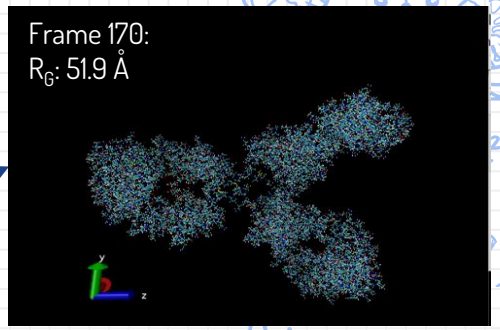
Radius of Gyration: R_G

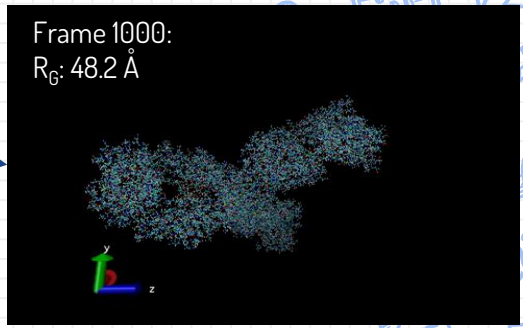
- Size of the molecule (unit: Angstrom (Å), 0.1 nanometer or 10^{-10} meters)
- Used for calculating other properties as well
- R_G is the average distance of all the atoms in the molecule from the center of mass.
  - A smaller R_G means a more compact protein
- We can calculate it with Mass or Scattering Length Density (SLD)

$$R_g^2 = \frac{\sum_i m_i (\vec{r}_i - \vec{r}_c)^2}{\sum_i m_i} = \frac{\sum_i m_i \vec{r}_{i,c}^{\,2}}{\sum_i m_i}$$





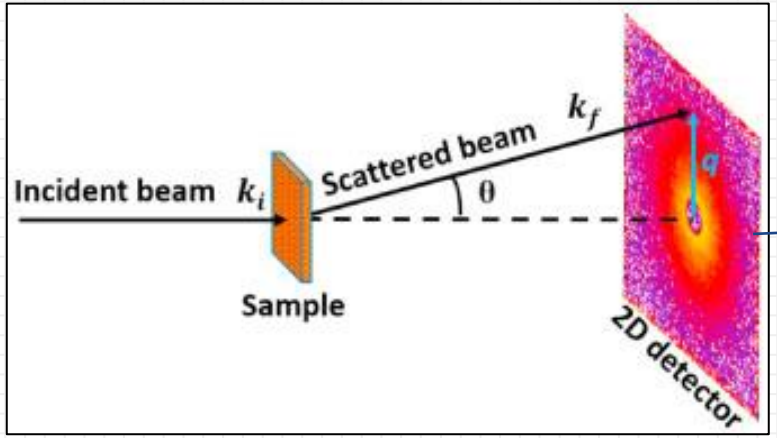


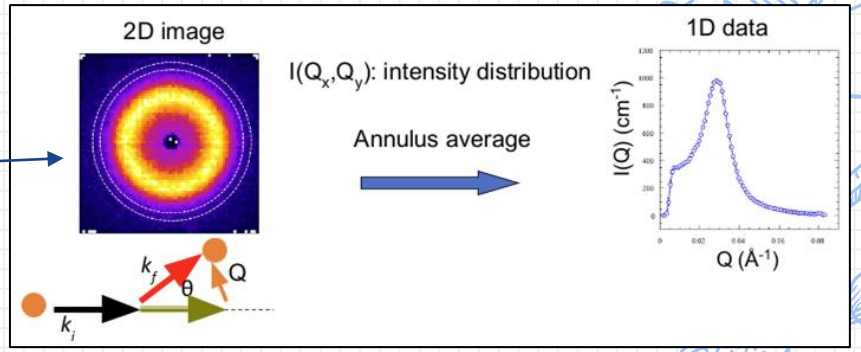


Trajectory A: Protein change over time

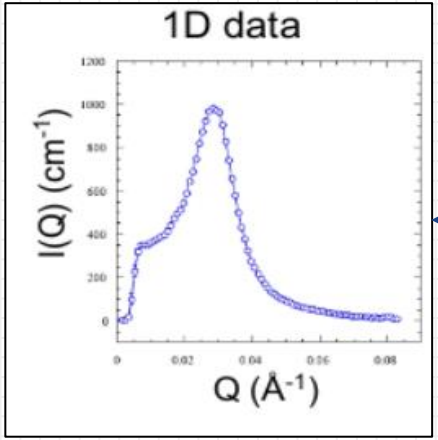
Scattering Pattern: I(Q)

- I(Q) is the 1D scattering pattern obtained from averaging 2D scattering data
  - Essential to understanding interactions between proteins through SANS
  - Dependent on Q (Angle)
  - We can find I(Q) either experimentally through physical scattering in real space (see below) or theoretically through calculations (next slide)

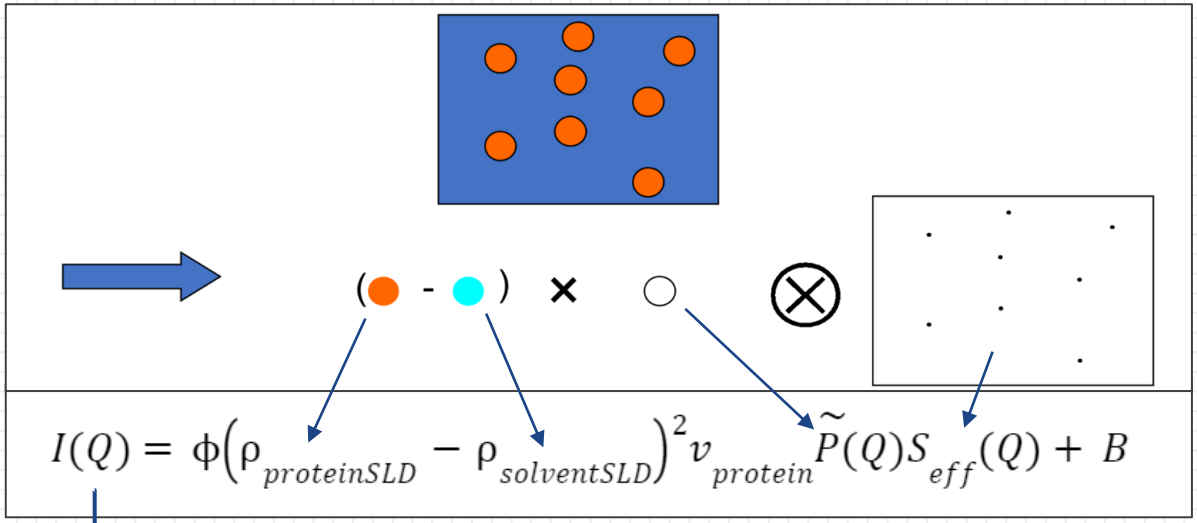




Scattering Pattern: I(Q)



- I(Q) can be also be calculated (see equation)
- This allows us to compare the experimental and calculated I(Q) to ascertain various properties of the sample.
  - More details later



φ - Scale

ρ - Scattering Length Density; effective size of interaction field, related to probability of scattering

v - Volume

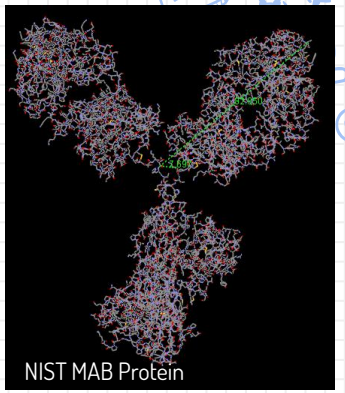
$\tilde{P}(Q)$ - Normalized Form Factor; Particle Shape

$S_{eff}(Q)$ - Effective Structure Factor; Protein Location & Interaction

B - Background

Beta Q: β(Q)

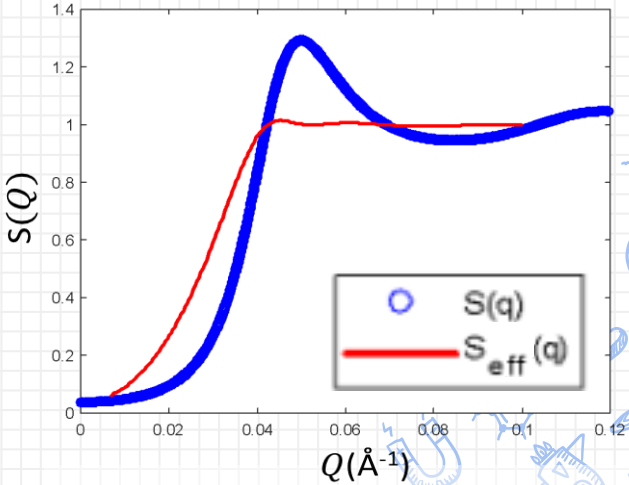
- Beta Q is a correction factor that accounts for the **orientation and interaction between proteins** in solution
- I(Q) (SANS Scattering pattern) is calculated with $S_{eff}(Q)$
  - Results from the experiments is $S_{eff}(Q)$ rather than S(Q)
  - S(Q) assumes a spherical shape (β(Q) of 1)
  - β(Q) is used to transform S(Q) to $S_{eff}(Q)$
- It is important to calculate the Effective Structure factor to accurately obtain the Scattering patterns



$$I(Q) = \Phi\left(\rho_{proteinSLD} - \rho_{solventSLD}\right)^2 v_{protein} \widetilde{P}(Q) S_{eff}(Q) + B$$

$$S_{eff}(Q) = 1 + \beta(Q)(S(Q) - 1)$$

$$\beta(Q) = \frac{\langle F(Q^-)\rangle^2}{\langle F^2(Q^-)\rangle} = \frac{\langle F(Q^-)\rangle^2}{P_A(Q)}$$



The effect of the static beta approximation

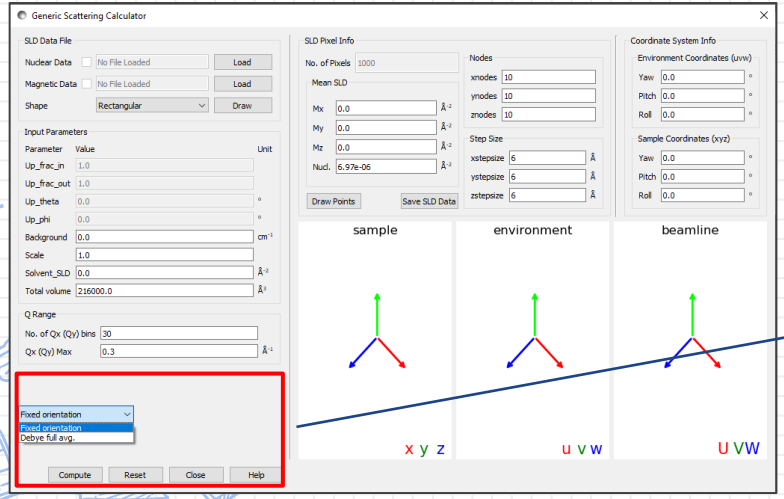
7

Project 1: R_G and β(Q)

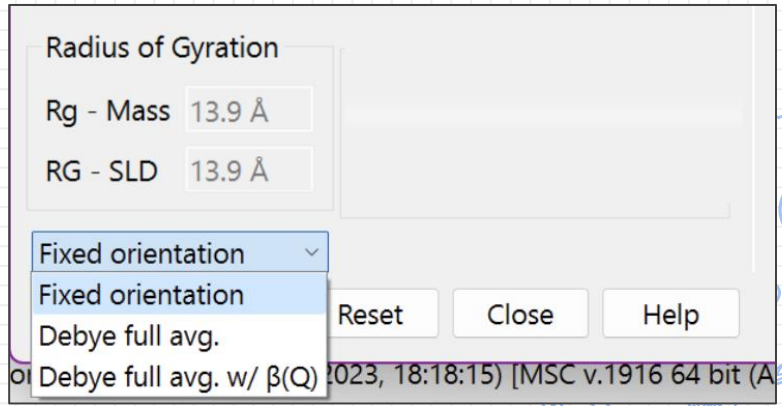
2	ATOM	1	N	ASP	X	1	29.350	18.970	134.130	0.00	0.00
3	ATOM	2	H1	ASP	X	1	29.580	19.450	134.990	0.00	0.00
4	ATOM	3	H2	ASP	X	1	30.210	18.550	133.810	0.00	0.00
5	ATOM	4	H3	ASP	X	1	28.670	18.240	134.290	0.00	0.00
6	ATOM	5	CA	ASP	X	1	28.810	19.950	133.150	0.00	0.00
7	ATOM	6	HA	ASP	X	1	27.860	20.210	133.610	0.00	0.00
8	ATOM	7	CB	ASP	X	1	29.590	21.270	133.050	0.00	0.00
9	ATOM	8	HB2	ASP	X	1	29.020	22.000	132.480	0.00	0.00

PDB file example

- Calculate R_G and β(Q) from Nuclear data (PDB Files)
- Integration into the Generic Scattering Calculator of Sasview to make it more accessible
- Involves revamping the GUI as well as working in the backend to properly integrate features.



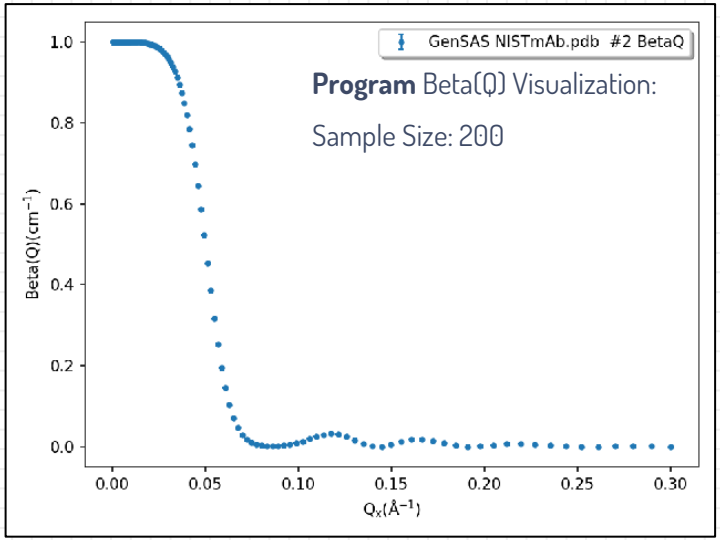
Before



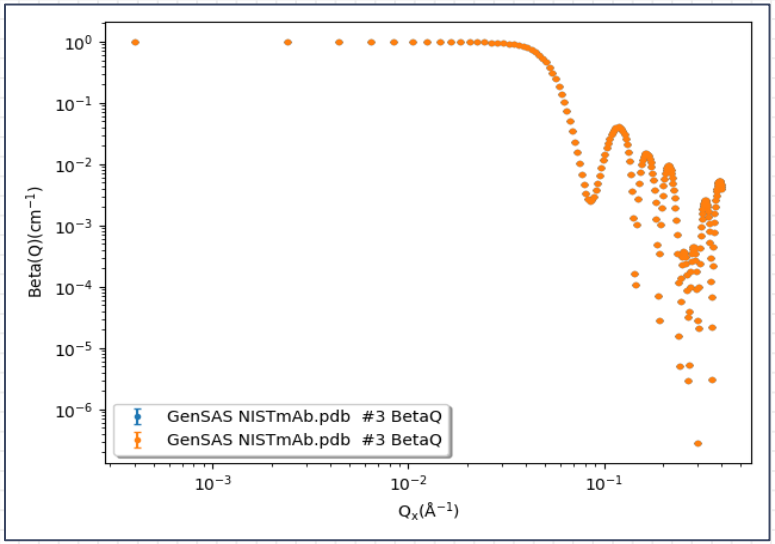
After

8

Generated β(Q)

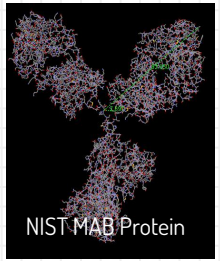


Separate Program Visualization



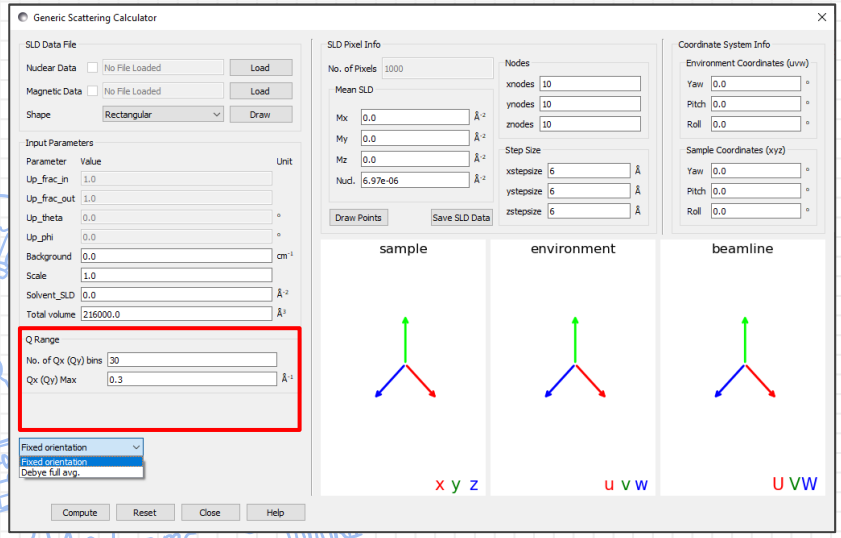
Post-Sasview Integration

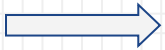
Notice how the separate program visualization has a linear spacing between the points and is plotted on a linear scale. In contrast, the Sasview program has linear spacing, but is plotted on a log scale.

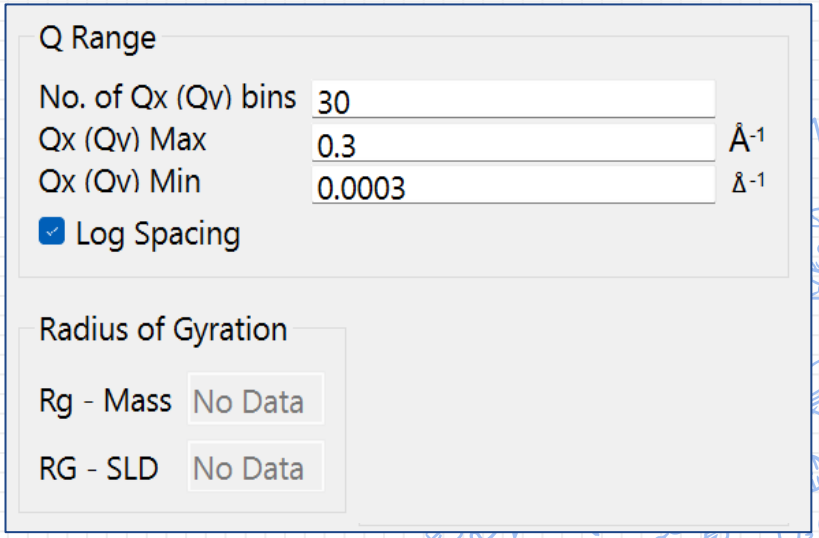


Project 2: Adjusting Q Spacing

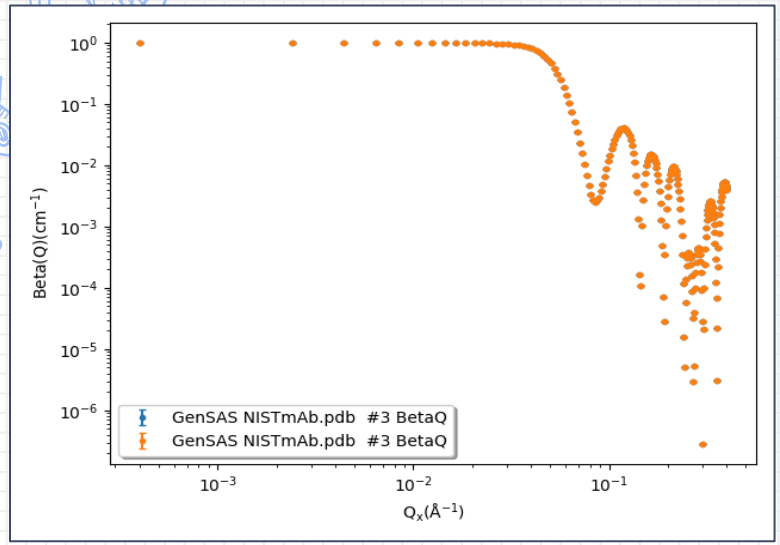
- Sasview utilizes linear spacing between points, but is plotted on a log scale.
- At times, log spacing is more useful than linear spacing
- Adjust the program to allow the user to choose between linear and log spacing.
- Allow User to set a QMinimum as well
  - Default set to .001 * QMax in the past



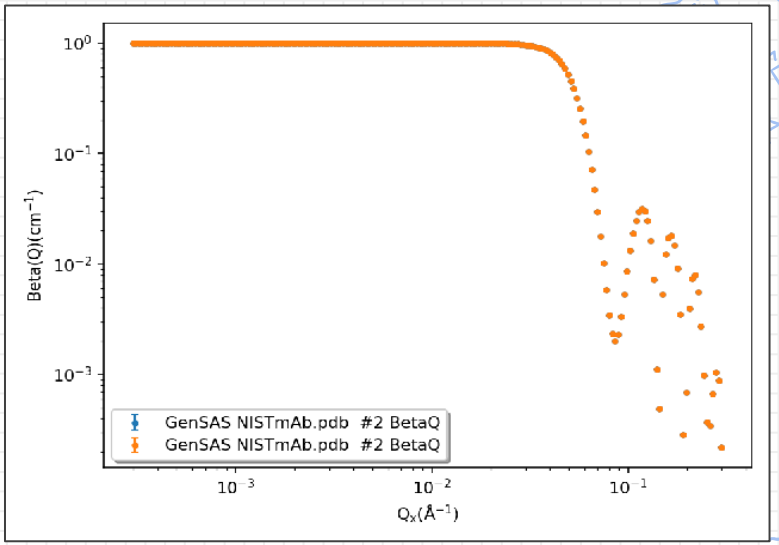




Linear vs Log Spacing



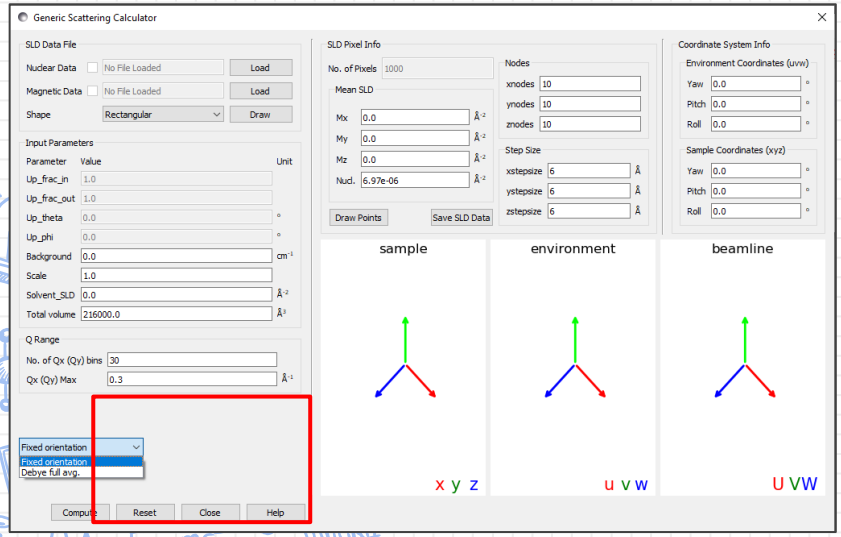
Linear Spacing, Log Scale

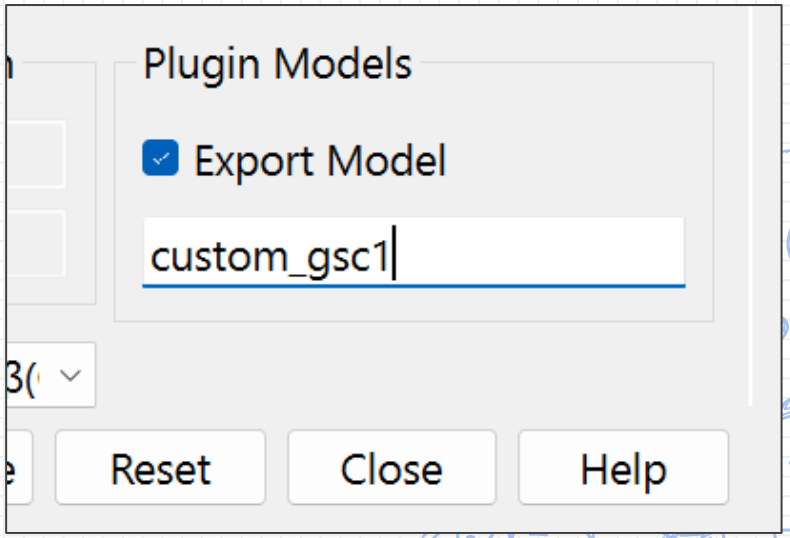


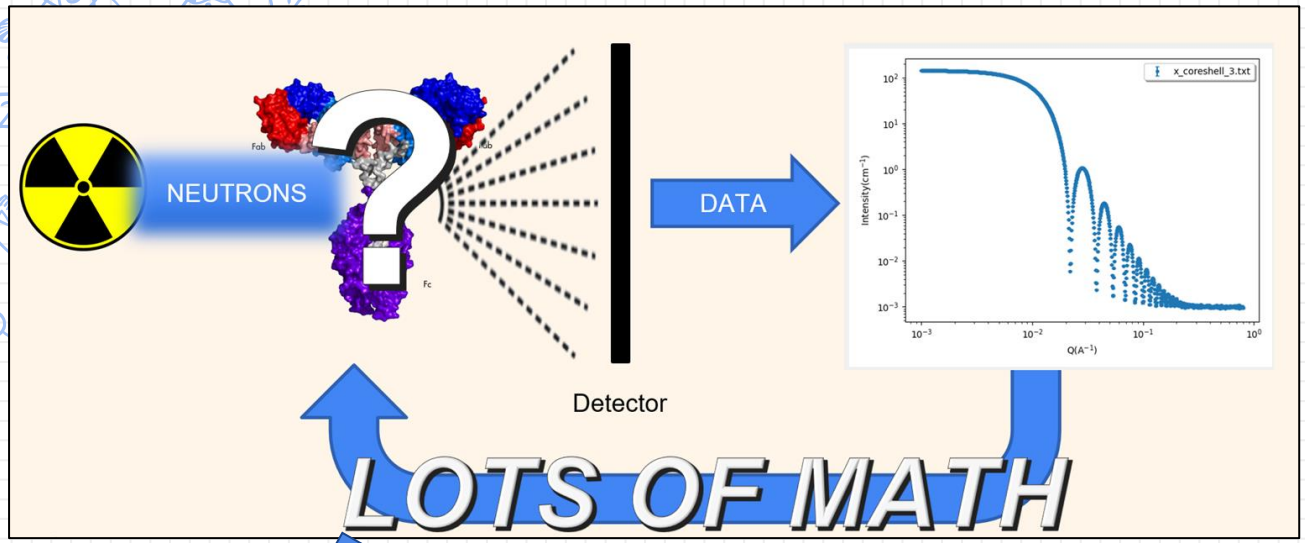
Log Spacing, Log Scale

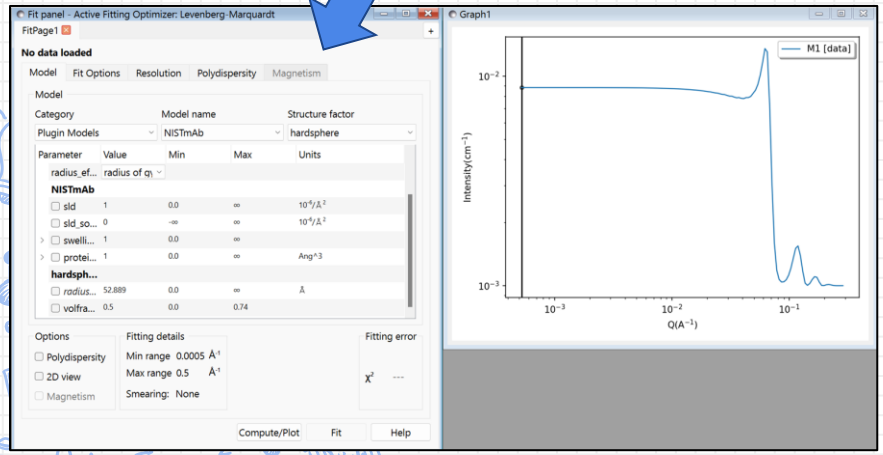
Project 3: Adding Custom Fit Models in the GSC

- Adding Custom Fit Models into the Generic Scattering Calculator of Sasview
- Takes the protein that has been imported and builds a custom plugin model to calculate and interpolate I(Q).
- Simulates the scattering pattern the protein would produce.
- Now able to fit data to **obtain interaction information** from PDB files in Sasview

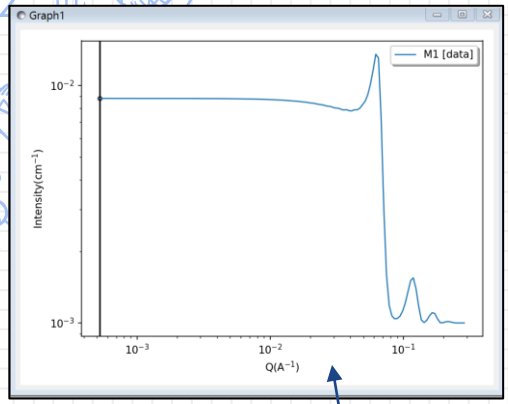


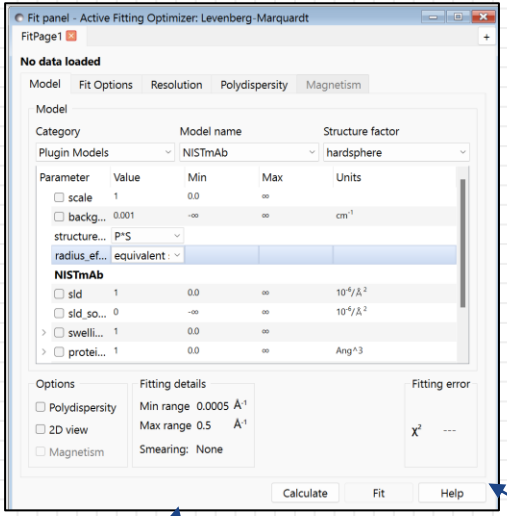


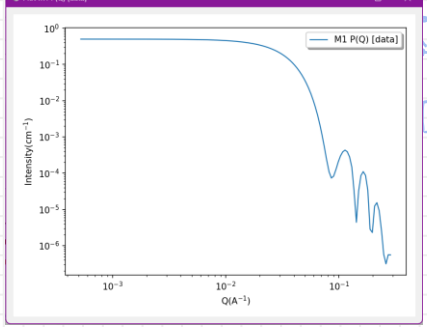


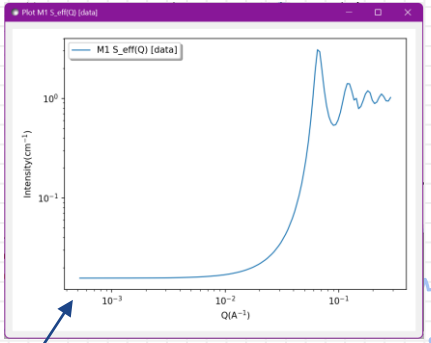


- Simulate the scattering pattern of the protein with our calculated plugin model
- Work backwards by adjusting parameters of the protein minimize the differences (residuals) between the experimental and simulated data.
- Derive meaningful structural information and gain insights into the properties



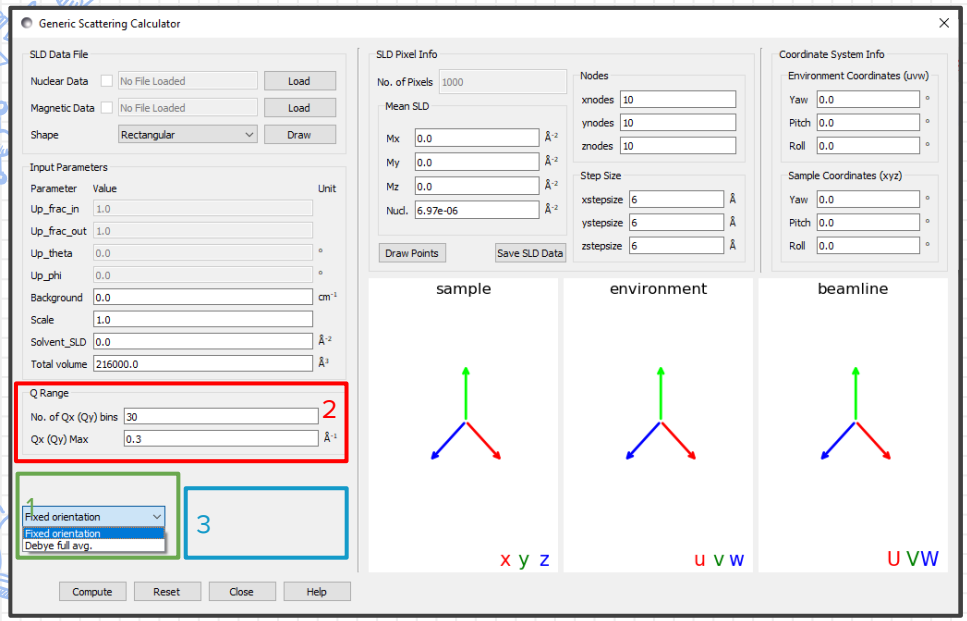


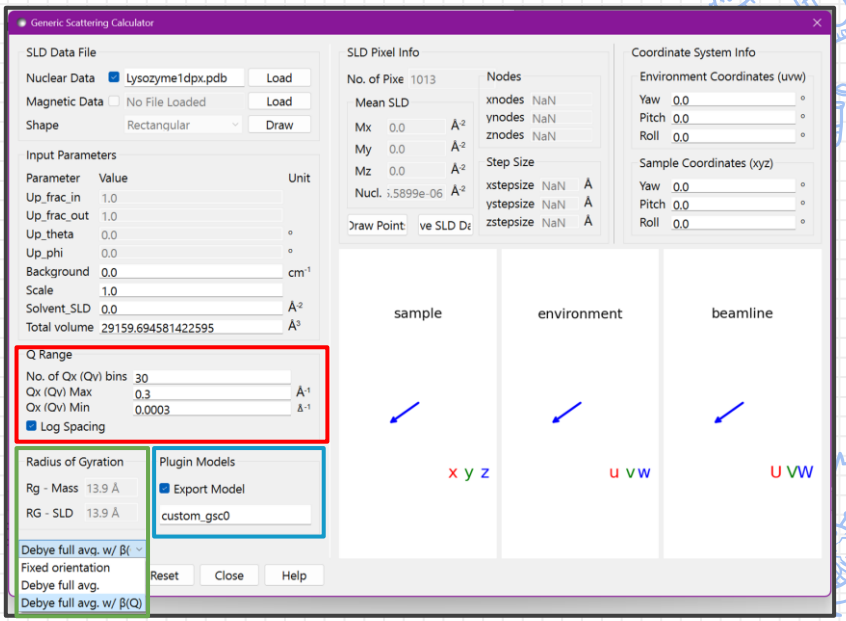


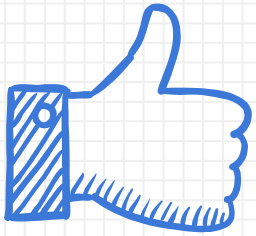


$$I(Q) = \phi\left(\rho_{proteinSLD} - \rho_{solventSLD}\right)^2 v_{protein} \tilde{P}(Q) S_{eff}(Q) + B$$

All Changes:


Generic Scattering Calculator
SLD Data File
Nuclear Data
No File Loaded
Load
Magnetic Data
No File Loaded
Load
Shape
Rectangular
Draw
Input Parameters
Parameter
Value
Unit
Up_frac_in
1.0
Up_frac_out
1.0
Up_theta
0.0
Up_phi
0.0
Background
0.0
Scale
1.0
Solvent_SLD
0.0
Total volume
216000.0
Q Range
No. of Qx (Qy) bins
30
Qx (Qy) Max
0.3
2
1
Fixed orientation
Fixed orientation
Debye full avg.
3
SLD Pixel Info
No. of Pixels
1000
Mean SLD
Mx
0.0
My
0.0
Mz
0.0
Nucl.
6.97e-06
Nodes
xnodes
10
ynodes
10
znodes
10
Step Size
xstepsize
6
ystepsize
6
zstepsize
6
Draw Points
Save SLD Data
Coordinate System Info
Environment Coordinates (uvw)
Yaw
0.0
Pitch
0.0
Roll
0.0
Sample Coordinates (xyz)
Yaw
0.0
Pitch
0.0
Roll
0.0
sample
environment
beamline
x y z
u v w
U VW
Compute
Reset
Close
Help


Generic Scattering Calculator
SLD Data File
Nuclear Data
Lysozyme1dpx.pdb
Load
Magnetic Data
No File Loaded
Load
Shape
Rectangular
Draw
Input Parameters
Parameter
Value
Unit
Up_frac_in
1.0
Up_frac_out
1.0
Up_theta
0.0
Up_phi
0.0
Background
0.0
Scale
1.0
Solvent_SLD
0.0
Total volume
29159.69458142595
Q Range
No. of Qx (Qv) bins
30
Qx (Qv) Max
0.3
Qx (Qv) Min
0.0003
Log Spacing
Radius of Gyration
Rg - Mass
13.9 Å
RG - SLD
13.9 Å
Debye full avg. w/ β(
Fixed orientation
Debye full avg.
Debye full avg. w/ β(Q)
Plugin Models
Export Model
custom_gsc0
SLD Pixel Info
No. of Pixe
1013
Mean SLD
Mx
0.0
My
0.0
Mz
0.0
Nucl.
Draw Point
ve SLD Da
Nodes
xnodes
NaN
ynodes
NaN
znodes
NaN
Step Size
xstepsize
NaN
ystepsize
NaN
zstepsize
NaN
Coordinate System Info
Environment Coordinates (uvw)
Yaw
0.0
Pitch
0.0
Roll
0.0
Sample Coordinates (xyz)
Yaw
0.0
Pitch
0.0
Roll
0.0
sample
environment
beamline
x y z
u v w
U VW
Reset
Close
Help



THANKS!

Any questions?