Creating Small Angle Neutron Scattering (SANS) Analysis Models for Concentrated Protein Solutions

Alex Zheng - NIST Center for Neutron Research

1. Background

What is Neutron Scattering?

The NIST Center for Neutron Research (NCNR) utilizes a small 20 MW nuclear reactor to produce neutrons for the numerous instruments and enable neutron scattering experiments, which improves our understanding of various materials and samples. My project focuses on one of the methods of neutron scattering- Small Angle Neutron Scattering, or SANS.

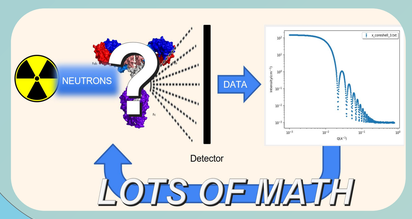
Why Neutron Scattering?

As neutrons only interact with the nucleus of atoms rather than their much larger electron cloud, we can determine various properties of materials. A real world example of this is that the NTSB (National Transportation Safety Board) has come to the NCNR to investigate and analyze material stresses to determine the cause of airliner accidents and crashes.

Background to SANS

In SANS, unknown samples are put into the instrument and bombarded with neutrons to obtain scattering patterns from the detector. From this scattering pattern, we can use a lot of math implemented in analysis software such as Sasview in order to work backwards and discover various properties of the sample.

My work focuses on developing software in Sasview to create fit models to analyze scattering patterns for SANS, particularly for proteins. This reveals properties such as the size and interactions of the proteins. This is especially important in cases such as pharmaceuticals, where these properties define the results of the products, such as toothpaste or shampoos.



NCNR Guide Hall

2. Overview

Sasview

- Sasview is an open source, collaboratively developed software for the analysis of any small angle scattering data
- Compare Small Angle Scattering data to existing shapes such as cylinders or ellipsoids to determine properties such as size
- Various tools:
  - Data Operation
  - SLD calculator
  - Density/Volume calculator
  - Slit Size Calculator
  - Kiessig Thickness Calculator
  - Q Resolution Estimator
  - **Generic Scattering Calculator**
  - Orientation Viewer
  - Python Shell/Editor
  - Image Viewer

My Project:

- Project primarily contributes to the Generic Scattering Calculator
- Generic Scattering Calculator simulates scattering data for different types of sample structures and compare these simulated patterns to their experimental data
  - Understand how parameters, such as size, shape, and distribution, affect the scattering pattern
- Project calculates various properties and builds a custom plugin model with these values

Special Thanks

Mentor: Dr. Yun Liu

CHRNS for sponsoring this program

Collaborators: Paul Kienzle, Paul Butler, Jeff Krzywon

SHIP Directors: Julie Borchers and Leland Harriger

3. Theory and Modelling

Calculated Values-

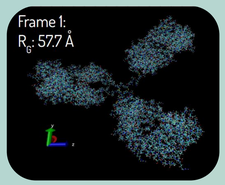
My project requires a deeper mathematical understanding of SANS to calculate three properties- Radius of Gyration (R_G), Scattering Pattern (I(Q)) and Beta Q (β(Q)). This section will dive into these three properties and how we use them.

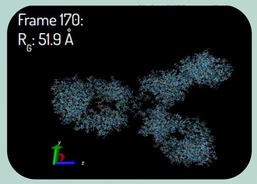
Radius of Gyration (R_G)

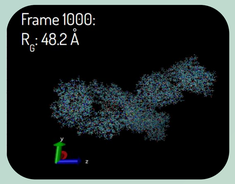
- In simplest terms, Radius of Gyration is the size of the molecule (unit: Angstrom (Å), 0.1 nanometer or 10^{-10} meters)
- Used for calculating other properties as well
- Mathematically, R_G is the average distance of all the atoms in the molecule from the center of mass.
  - A smaller R_G means a more compact protein
- Calculated with Mass (m_i) or Scattering Length Density (b_i); in the latter case, it is called the Guinier Radius

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

$$R_g^2 = \frac{\sum_i b_i (|\vec{r_i} - \vec{r_c}|)^2}{\sum_i b_i}$$

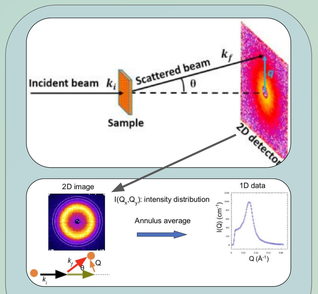




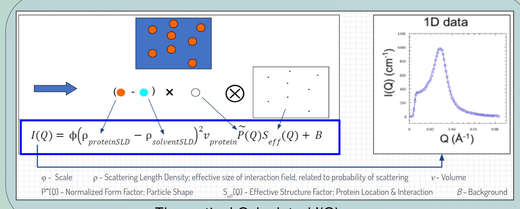


Scattering Pattern I(Q)

- Scattering Pattern (I(Q)) is the 1D scattering pattern obtained from averaging 2D scattering data
- Essential to understanding interactions between proteins through SANS
- Dependent on Q (Related to angle)
- We can obtain I(Q) **either experimentally** through physical scattering in real space (see right) **or theoretically** through calculations (see equation below)
- **By comparing the experimental and calculated I(Q) curve, we can ascertain various properties of the sample.**



I(Q) curve generated by real-world SANS experiment



Theoretical Calculated I(Q) curve

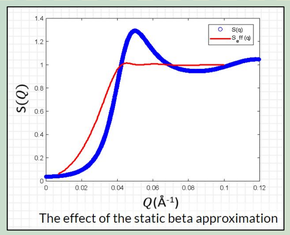
Beta Q (β(Q))

- Beta Q is a correction factor that accounts for the **orientation of anisotropic proteins to extract interactions between proteins** in solution
- I(Q) is calculated with $S_{eff}(Q)$ for anisotropic particles
  - 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

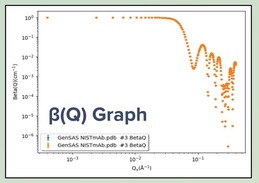
$$I(Q) = \phi(\rho_{proteinSLD} - \rho_{solventSLD})^2 v_{protein} \tilde{P}(Q) S_{eff}(Q) + B$$

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

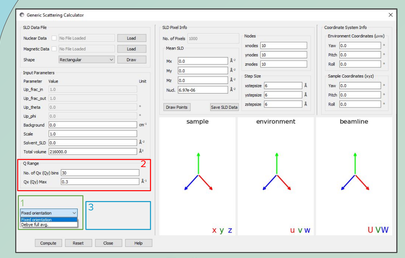
$$\beta(Q) = \frac{\langle F(Q)^- \rangle^2}{\langle F^2(Q^-) \rangle} = \frac{\langle F(Q^-) \rangle^2}{P_n(Q)}$$



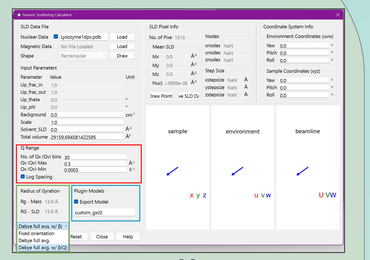
The effect of the static beta approximation



4. GUI Changes



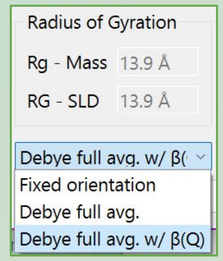
Before



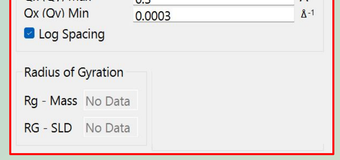
After

Change 1: R_G and β(Q)

- Calculate R_G(Mass & SLD) and β(Q) from atomic coordinates (PDB Files)
  - PDB files are the atomic position of every atom in a protein
- Integration into the Generic Scattering Calculator to make it more accessible
- Involves revamping the GUI as well as working in the backend to properly integrate features.
- Rather than opening a seperate software to calculate β(Q) and R_G, scientists can do it within the Sasview program now.



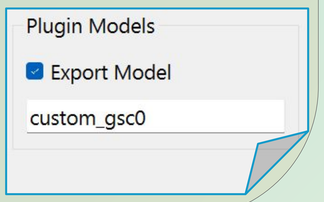
Change 2: Adjusting Q Spacing



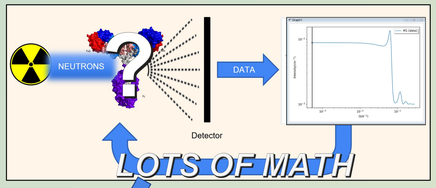
- Sasview utilizes linear spacing between points, but is plotted on a log scale.
- At times, log spacing is more useful than linear spacing
- Adjust the program to allow the user to choose between linear and log spacing.
- Allow User to set a Q Minimum as well
- Default set to .001 * QMax in the past

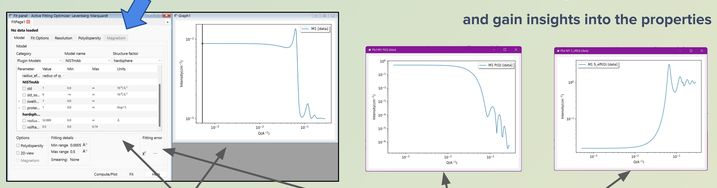
Change 3: Adding Custom Fit Models in the GSC

- 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



5. Summation





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

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