Overview of Current Methods in the Study of Biomolecular Structure by SANS in the Dilute Solution Limit

Susan Krueger
(and many others who I will try to acknowledge along the way)

Low - q Seminar
October 1, 2014
Biology SANS Applications
Dilute Limit

Biological Macromolecules in Bulk Solution

- Proteins
- Nucleic acids
- Protein-nucleic acid complexes
- Multi-subunit protein complexes
- Protein-lipid complexes
- Membranes and membrane components
The Scattering Process: A Review

The Scattering Process

\[ e^{i \mathbf{k}_i \cdot \mathbf{r}} \rightarrow e^{i \mathbf{k}_i \cdot \mathbf{r}} + f(\Omega) \frac{e^{i \mathbf{k}_f \cdot \mathbf{r}}}{r} \]
The Scattering Process

\[
\frac{dW}{d\Omega} = \text{neutrons/sec scattered into solid angle } d\Omega
\]

\[
\text{neutrons/sec/cm}^2 \text{ on target}
\]

\[
\Phi \text{ (beam flux)}
\]

Differential scattering cross section

\[\sigma = \int \frac{d\sigma}{d\Omega} d\Omega = 4\pi |f(\Omega)|^2 = 4\pi |b|^2 \text{ in units of area (cm}^2)\]

1 barn = $10^{-24}$ cm$^2$

Total scattering cross section

Scattering length (per atom)

\[\sigma_{coh} = 4\pi(\bar{b})^2\]

\[\sigma_{inc} = 4\pi[(\bar{b}^2) - (\bar{b})^2]\]

\[\sigma = \sigma_{coh} + \sigma_{inc} + \sigma_{abs}\]
Neutron Cross Sections

<table>
<thead>
<tr>
<th>Atomic nucleus</th>
<th>$b_{coh}$ (fm)</th>
<th>$\sigma_{coh}$ (barns)</th>
<th>$\sigma_{inc}$ (barns)</th>
<th>$\sigma_{abs}$ (barns)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$^1$H</td>
<td>- 3.741</td>
<td>1.758</td>
<td>80.26</td>
<td>0.333</td>
</tr>
<tr>
<td>$^2$D</td>
<td>+ 6.671</td>
<td>5.592</td>
<td>2.05</td>
<td>0.000</td>
</tr>
<tr>
<td>$^{12}$C</td>
<td>+ 6.651</td>
<td>5.559</td>
<td>0.0</td>
<td>0.004</td>
</tr>
<tr>
<td>$^{14}$N</td>
<td>+ 9.37</td>
<td>11.03</td>
<td>0.5</td>
<td>1.91</td>
</tr>
<tr>
<td>$^{16}$O</td>
<td>+ 5.803</td>
<td>4.232</td>
<td>0.0</td>
<td>0.000</td>
</tr>
<tr>
<td>$^{31}$P</td>
<td>+ 5.13</td>
<td>3.307</td>
<td>0.005</td>
<td>0.172</td>
</tr>
<tr>
<td>$^{32}$S</td>
<td>+ 2.804</td>
<td>0.988</td>
<td>0.0</td>
<td>0.54</td>
</tr>
</tbody>
</table>

Small Angle Scattering

Constructive interference from structures in the direction of $\vec{q}$

Diffraction length scale

$$d \approx \frac{2\pi}{q} \quad \text{where} \quad q = \frac{4\pi}{\lambda} \sin \theta$$

$$\begin{align*}
2 & \quad \frac{6 \text{ Å}}{d} \\
10 \text{ to } 1000 \text{ Å} & \quad 2 \quad 0.3^\circ \text{ to } 5^\circ
\end{align*}$$
Differential Scattering Cross Section

\[
\frac{d}{d\Omega} = \left| f(\cdot) \right|^2 \sum_{ij} e^{-i\vec{q} \cdot (\vec{r}_i - \vec{r}_j)} \quad \text{where} \quad \vec{q} = \vec{k}_f - \vec{k}_i
\]

Coherent:

\[
\frac{d}{d\Omega} = \sum_{ij} b_i b_j e^{-i\vec{q} \cdot (\vec{r}_i - \vec{r}_j)}
\]

Replace:

\[
\sum_i b_i \rightarrow \int_V (\vec{r}) \, d\vec{r} \quad \text{where} \quad \rho(\vec{r}) = b_i \delta(\vec{r} - \vec{r}_i)
\]

\[
\frac{d}{d\Omega}(\vec{q}) = \frac{1}{N} \left| \int_V (\vec{r}) e^{i\vec{q} \cdot \vec{x}} \, d\vec{r} \right|^2
\]

Scattering length density
Differential Scattering Cross Section

Normalizing by sample volume:

\[
\frac{d\Sigma}{d\Omega}(\vec{q}) = \frac{N}{V} \frac{d\sigma}{d\Omega}(\vec{q}) = \frac{1}{V} \left| \int_V \rho(\vec{r}) e^{i \vec{q} \cdot \vec{r}} d\vec{r} \right|^2
\]

Rayleigh-Gans Equation

\[
(\vec{r}) = f(\vec{r}) + b_i \frac{N}{V}
\]

- Average scattering length density
- Fluctuations about the average
- Uniform scattering length density
Neutron Scattering Length Densities

H₂O: $\rho = -0.562 \times 10^{10}$ cm⁻²
D₂O: $\rho = 6.404 \times 10^{10}$ cm⁻²
H₂O:D₂O mixture: $\rho = (-0.562 + 6.966y) \times 10^{10}$ cm⁻², where $y$ = fraction of D₂O

<table>
<thead>
<tr>
<th></th>
<th>$\rho$ in H₂O (10¹⁰ cm⁻²)</th>
<th>$\rho$ in D₂O (10¹⁰ cm⁻²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>protein</td>
<td>1.8</td>
<td>3.2</td>
</tr>
<tr>
<td>DNA</td>
<td>3.4</td>
<td>4.1</td>
</tr>
<tr>
<td>RNA</td>
<td>3.6</td>
<td>4.3</td>
</tr>
<tr>
<td>PC Lipid Head Group</td>
<td>1.1</td>
<td>1.1</td>
</tr>
<tr>
<td>CH₂</td>
<td>-0.31</td>
<td>-0.31</td>
</tr>
<tr>
<td>CH₃</td>
<td>-0.85</td>
<td>-0.85</td>
</tr>
</tbody>
</table>

SLDs in H₂O and D₂O depend on H:D exchange.
The Scattered Intensity

Rayleigh-Gans Equation:

\[ I(q) \propto \frac{d\Sigma}{d\Omega}(\vec{q}) = \frac{1}{V} \left| \int_V \rho(\vec{r}) e^{i\vec{q} \cdot \vec{r}} \, d\vec{r} \right|^2 \]

Assume there are \( N \) randomly-oriented, homogeneous particles:

\[ \frac{d}{d\Omega}(\vec{q}) = \frac{N}{V} 2V^2 \left\langle \left| \frac{1}{V} e^{i\vec{q} \cdot \vec{\tau}} \, d\vec{\tau} \right|^2 \right\rangle \]

Particle volume

Average over all orientations

\( F(\vec{q}) \) (depends on particle shape)
Macromolecules in Solution

Reciprocal Space

\[ l(q) = \int_{V} e^{i \hat{q} \cdot \hat{r}} d\hat{r} \]

Macromolecule in Solvent

Uniform Scattering Length Density, \( \rho \), in \( V \)

Solvent of Infinite Extent
(Not Observed!)

Solvent Scattering Length Density, \( \rho_s \), in \( V \)

\[ I(q) = \frac{N}{V} (\ )^2 V^2 \left\langle |F(\hat{r})|^2 \right\rangle, \text{ where} \]

Real Space

\[ \mu \]

\[ \rho_s \delta(0) \]

\[ s \int_{V} e^{i \hat{q} \cdot \hat{r}} d\hat{r} \]
The Guinier Approximation

\[ I(q) = I(0) \exp\left( -\frac{q^2 R_g^2}{3} \right) \]

valid when \( q R_g \leq 1 \)

\[ I(0) = \frac{d \sum(0)}{d} = \frac{N}{V} ( \ldots )^2 V^2 \]

\( R_g: \) radius of gyration (about the particle C.M.)

\[ R_g^2 = \frac{\int \rho(\vec{r}) r^2 \, d\vec{r}}{\int \rho(\vec{r}) \, d\vec{r}} \]
Guinier Plot

Get $R_g$ and $I(0)$ from Guinier fit

$$\ln[I(q)] = \ln[I(0)] - \frac{q^2 R_g^2}{3}$$

$q_{\text{max}} R_g \leq 1$
The Forward Scattering on an Absolute Scale

\[ I(0) = \frac{d \sum(0)}{d} = \left( \frac{N}{V} \right)^2 V^2 \]

- Number density of particles
- Concentration of particles
- Molecular weight
- Partial Specific Volume

\[ I(0) = \frac{c N_A}{M_w} \left( \frac{M_w \bar{V}}{N_A} \right)^2 = \frac{c M_w \bar{V}^2}{N_A} \left( \frac{N}{V} \right)^2 \]
Modified Guinier Analyses

Rod-shaped Particles where \( l \gg r \)

\[
I(q) \approx \frac{1}{q} I_c(0) \exp\left(-q^2 R_c^2 / 2\right)
\]

Plot \( \ln[q \cdot I(q)] \) vs \( q^2 \); \( R_c \) related to cross-sectional radius

Disk-Shaped particles where \( r \gg t \)

\[
I(q) \approx \frac{1}{q^2} I_t(0) \exp\left(-q^2 R_t^2 / 12\right)
\]

Plot \( \ln[q^2 \cdot I(q)] \) vs \( q^2 \); \( R_t \) related to thickness
Low Resolution Shape Models

![Graph showing the relationship between wave vector modulus (q) and intensity (I(q)). The graph compares monomer and dimer models. The y-axis represents the intensity in cm⁻¹, and the x-axis represents q in Å⁻¹. The graph illustrates the difference in intensity between the monomer and dimer models as q increases.]
Guinier Fits

- **Monomer**
  \[ R_g = 14.7 \text{Å}; \quad I(0) = 1.0 \text{ cm}^{-1} \]
- **Dimer**
  \[ R_g = 21.0 \text{Å}; \quad I(0) = 2.0 \text{ cm}^{-1} \]
Simple Dimer Models

\[ R_g = 21.0 \text{Å} \]

\[ R_g = 29.0 \text{Å} \]
Distance Distribution Function

\[ P(r) = r^2 \gamma(r) \]
\[ \gamma(r) = \frac{1}{2\pi^2 r} \int qI(q) \sin(qr) dq \]

Debye-Porod Correlation Function

\[ 4\pi P(r) \equiv \text{number of distances within the molecule} \]
\[ I(q) = 4\pi V \int_{0}^{D_{\text{max}}} P(r) \frac{\sin(qr)}{qr} dr \]
\[ D_{\text{max}} \equiv \text{maximum distance within the molecule} \]
\[ P(0) = 0 \]
\[ P(2r \geq D_{\text{max}}) = 0 \]
Distance Distribution Function

\[ R_g = 21.0 \text{Å} \]

\[ R_g = 29.0 \text{Å} \]
$P(r)$ for All-atom Models

Dimer 1, $(B/2 = 755 \, \text{Å}^2)$
$R_G = 26.08 \, \text{Å} \quad D_{\text{max}} = 80 \, \text{Å}$

Dimer 3, $(B/2 = 406 \, \text{Å}^2)$
$R_G = 28.3 \, \text{Å} \quad D_{\text{max}} = 90 \, \text{Å}$

Dimer 2, $(B/2 = 923 \, \text{Å}^2)$
$R_G = 34.04 \, \text{Å} \quad D_{\text{max}} = 120 \, \text{Å}$

Dimer 4, $(B/2 = 255 \, \text{Å}^2)$
$R_G = 30.4 \, \text{Å} \quad D_{\text{max}} = 100 \, \text{Å}$
Advanced Modeling Techniques

Modeling efforts are underway at most neutron and x-ray facilities.
All the modeling in the world doesn’t help if your sample isn’t good…

- Concentration: 1-5 mg/mL → **monodisperse, non-interacting**
- Volume: 350-700 μL per sample
- Data collection time: 0.5-6 hrs per sample
- Typical biology experiment: 1-3 days
- Deuterated solvent is highly desirable.
- Multiple concentrations are usually necessary.
- Specific deuteration may be necessary.
- Multiple solvents of different deuteration → **contrast variation**
Practical Issues

Sample Preparation Challenges

- **Mondispersity**: suitable solvent conditions
- **Complex formation**: under multiple contrast conditions (D$_2$O effects)
- **Large quantities**: needed for multiple contrasts
- **Deuterium labeling**: many neutron facilities now have labs to support users
- **Analytical tools**: SEC-MALS, AUC (SE and SV), DLS

Experiment Planning Tools

- **Contrast Calculator** (SASSIE)
- **Starting All-atom and Hybrid hi-res/low-res Structures** (for *in silico* contrast variation experiments)

- proteins
- nucleic acids
- lipids
- carbohydrates
Intrinsically Disordered Proteins

A Definition:
Proteins or regions of proteins that fail to form specific 3-D structure under physiologic conditions in vitro.
from Le Gall et al., J. Biomolecular Struct. and Dynamics 24, 325 (2007)

Backbone Ramachandran angles vary significantly over time with no specific equilibrium values.

Software Needs:
>10,000 structures in minutes
“Thumbs up” or “Thumbs down” on model structures

http://www.disprot.org/
SASSIE: An Approach to Structure Modeling

- Geometric ensemble sampling based on energetics of dihedral angle motion.

**Build Starting Structure**
- Include “missing” residues

**Dihedral Phase Space Search**
- Pick regions to vary
- Pick $\Delta \Phi$, $\Delta \Psi$, overlap (basis and cutoff)

**Generate Structures**

**Calculate SAS**

**Filter Results**
- Compare to SAS data
| BUILD | CLEAN up and organize coordinate files  
|       | Build topology files (force field)  
|       | All-atom or coarse-grain |
| TOOLS | Structure alignment, centering, translation, rotation  
|       | Coordinate manipulation, data interpolation  
|       | Contrast calculator, experiment planning |
| INTERACT | Graphically “move” structures  
|         | Calculate SANS and reflectivity curves in real time |
| SIMULATE | Protein/RNA dihedral search; grid search  
|          | Structure minimization; torsion angle  
|          | MD; normal modes; free energy solvation |
| CALCULATE | SANS, SAXS, reflectometry  
|           | EM to SANS; HYDROPRO  
|           | Spin echo; backscattering, TOF |
| ANALYZE | Chi-squared filtering to SAS data  
|         | Density plots of conformation space |
All-atom Modeling: SASSIE

Worst Fits

Best Fits

Any $\chi^2$; Any Rg

SANS data
interpolated data
best fit
worst fit
average (n=102,384)

$q (\text{Å}^{-1})$

$I(q)$
PARP-1 Solutions: SASSIE

One structure from an ensemble that fits the SAXS and SANS data.
Antibody Ensemble: SASSIE

Best-Fit Structures

We can use the average fit as a form factor

Single Stranded RNA: SASSIE
Crystal structure is available for the N-terminal domains.
Solution Structure of N-terminal Domain of MCM DNA Helicase

*Rg = 78 Å ± 1 Å

*Rg ~ 47 Å

Rg suggest a larger molecule. Mw suggests double hexamer (or a mixture of single and double hexamers).

>30 missing C-terminal residues
Build a Starting Structure

Possible flexible regions: 88-107 and 243-274
Chi-squared Filtering

Flexible: 243-274

Flexible: 88-107 and 243-274
Best Fits

$\chi^2 = 0.87$

$\chi^2 = 3.5$
Possible flexible regions:
89-108, 244-246 and 580-600
Chi-squared Filtering

Flexible: 89-108 and 580-600

Flexible: 89-108, 244-246 and 580-600
Best Fits

Flexible: 89-108 and 580-600

Flexible: 89-108, 244-246 and 580-600
Structure Comparison

Original Starting Structure

Best Fit Structure Allowing Three Flexible Regions
Compound Particles

Model as an assembly of uniform particle subunits.

\[ I(q) \propto \left( (\Delta \rho)_1 \int_{V_1} e^{i\mathbf{q} \cdot \mathbf{r}} d\mathbf{r}_1 + (\Delta \rho)_2 \int_{V_2} e^{i\mathbf{q} \cdot \mathbf{r}} d\mathbf{r}_2 \right)^2 = \]

\[ (\Delta \rho)_1^2 \left\langle |F_1(q)|^2 \right\rangle + (\Delta \rho)_2^2 \left\langle |F_2(q)|^2 \right\rangle + (\Delta \rho)_1 (\Delta \rho)_2 |F_1| |F_2| \frac{\sin(qr_{12})}{qr_{12}} \]
Compound Particles

\[ I_1(q) = (\Delta \rho)_1^2 F_1^2 \]

\[ I_2(q) = (\Delta \rho)_2^2 F_2^2 \]

\[ I_{12}(q) = 2(\Delta \rho)_1 (\Delta \rho)_2 F_1 F_2 \frac{\sin(qr_{12})}{qr_{12}} \]

Separate scattering from subunits using contrast variation.
Scattered intensity from the two components can be separated by solving a set of simultaneous equations.

\[ I(q) = \Delta \rho_1^2 l_1(q) + \Delta \rho_1 \Delta \rho_2 l_{12}(q) + \Delta \rho_2^2 l_2(q) \]

\( \Delta \rho_1, \Delta \rho_2: \) contrast for components 1 and 2

\( l_1(q), l_2(q): \) intensity for components 1 and 2

\( l_{12}(q): \) cross-term between components 1 and 2

Recall: \( D_r = r - r_s \)
Neutron Contrast Variation in Biology

Used to separate the scattering contribution from the components in a two-component complex.

- Protein-nucleic acid complexes
- Protein-protein complexes*
- Protein-lipid complexes

*One component must be deuterated.
Contrast Variation

Scattering Length Density ($10^{10}$ cm$^{-2}$) vs. % D$_2$O in Solvent

- CD$_2$
- Deuterated RNA
- Deuterated Protein
- RNA
- DNA
- Water
- Protein
- Lipid Head Group
- CH$_2$

Deuterated Lipid Head Group

CONTRAST

$\Delta \rho$
Contrast Variation Experiment

- **Model data prior to experiment**
  - Predict contrasts values and match points based on chemical composition (amino acid sequence, bases, type of lipid, etc.)
  - Predict SANS curves at various contrasts if a low or high resolution starting model structure is available.

- **Match point determination**
  - Determine the contrast match point for the complex and individual components.

- **Stuhrmann Analysis and Parallel Axis Theorem**
  - Determine $R_g$ for the individual components and the spatial relationship between the two components.

- **Test Model Structures Against ALL Data**
Match Point Determination

For \( q = 0 \):

\[
I(0) = n(\bar{V})^2 = \frac{cM_w}{N_A} (\bar{v})^2
\]

Since \( \Delta \rho \) varies with the %D\textsubscript{2}O in the solvent:

\[
\frac{I(0)}{c} = \frac{M_w}{N_A} (\bar{\nu})^2 \Rightarrow \frac{I(0)}{c} \propto (\bar{\nu})^2 \Rightarrow \sqrt{\frac{I(0)}{c}} \propto %D_2O
\]

Plot \( \sqrt{\frac{I(0)}{c}} \) vs %D\textsubscript{2}O to obtain the match point.
Two Component System: \( R_g \) vs \( \Delta \rho \)

**Stuhrmann Analysis**

\[
R_g^2 = R_o^2 + \frac{a}{D} + \frac{b}{D^2}
\]

\( R_g \) of an equivalent homogeneous complex

\( \beta \neq 0 \Rightarrow \) centers of mass of the two components are not concentric


**Parallel Axis Theorem**

\[
R_g^2 = \frac{1}{V} R_1^2 + \frac{2}{V} R_2^2 + \frac{1}{V} \left( \frac{2 V}{V} \right) D^2
\]

Component 1  Component 2  Cross-term  Distance between centers of mass

Contrast Variation: Biology Example

- Two-component protein in H$_2$O/D$_2$O solvent.
- One subunit is deuterated (d-protein).
- Scattering length densities of the two components are different from one another.

Protein subunit is “matched”.

d-protein subunit is “matched”.
PAI:VN complex with 65% deuterated PAI and non-deuterated VN.
SASSIE and Contrast Variation
SASSIE and Contrast Variation

[Graph showing scattering intensity (I(q)) as a function of q (Å⁻¹) for different D2O concentrations (100%, 85%, 20%, 10%, 0% D2O).]
SASSIE and Contrast Variation

Two structures that fit all of the data (except 100% D$_2$O) well.
Chromatin Solution Structure

Modeling work underway…thanks to new double-stranded DNA Molecular Monte Carlo
Summary: SANS and Biology

- If used correctly, SANS can be a powerful tool to study the structure of biological macromolecules in solution.
- Contrast variation can be easily applied using neutrons.
- Can be combined with SAXS to obtain data at higher q values (q > 0.3 Å⁻¹).
- Excellent facilities are available worldwide. Many groups are actively developing software for structure modeling.
- Sample quality is extremely important. Consistency checks must be performed.
- Use information from other techniques to narrow down the possible model structures.