Length-scales and tools in structural biology

SAS bridges the gap between atomic resolution (NMR and crystallography) and the light microscope.
Objects that can be studied by SAS

SAS cannot determine *de novo* the positions of individual atoms/residues in biomacromolecules on the Angstrom scale

Atomic resolution: NMR and/or Crystallography

Scattering basics: Huygens-Fresnel principle

\[ |\vec{k}| = \frac{|\vec{k}+\vec{k}'|}{\lambda} = \frac{2\pi}{\lambda} \]

Incoming X-ray/neutron wave

Many scattering centers, FOURIER transform:

\[ I(Q) = \left\langle \sum_j b_j e^{-iQ \cdot r_j} \right\rangle^2 \]

\[ Q = \frac{|\vec{k}' - \vec{k}|}{\lambda} = \frac{4\pi}{\lambda} \sin \theta \]
Reciprocal relationship between real space and the diffraction pattern

Many scattering centers, Fourier transform:

\[ I(Q) = \left( \sum_{ij} b_i b_j e^{-iQ(r_i - r_j)} \right)^2 \]

Orientational averaging

\[ I(Q) = \sum_{ij} b_i b_j \frac{\sin Q(r_i - r_j)}{Q(r_i - r_j)} \]

Debye equation (1915)

Why “small” angle scattering?

SAS sample conditions and information obtained

Global properties:
- Radius of gyration,
- Molecular weight

Structural details
- Neutrons, X-rays
- Solutions ~ mg/ml
- Volume ~ 10-200 μL
- Mass > 0.1 mg

SAS sensitivity:
- Macromolecules 1 kDa ... ~ MDa
- Linear dimension 10 Å ... ~ 1000 Å

Information obtained:
1) Oligomeric state of macromolecules
2) Shape or conformation (globular, stick etc...)
3) Interaction of different macromolecules
4) Variation of points (1)-(3) as a function of pH, salt, ligands, T, p, ...
5) Contrast variation: visualisation of individual sub-units in situ

Concept of scattering density and contrast

In vacuo:
\[ I(Q) = \left| \sum J_j e^{-iQr_j} \right|^2 \\
I(Q) = \int \frac{J_j}{V_j} e^{-iQr} dV_j \approx \int \rho_{\text{protein}} e^{-iQr} dV_j \]
\[ I(Q) = N \left( \int \rho_{\text{protein}} e^{-iQr} dV \right)^2 \]

In solution:
\[ I(Q) = \left( \int \rho_{\text{protein}} \rho_{\text{solvent}} e^{-iQr} dV \right)^2 \]

Continuum approximation:
\[ \frac{J_j}{V_j} \approx \rho_{\text{protein}} = \text{const} \]

- How are scattering densities calculated?
- Under which conditions is the approximation valid?

Ideal solutions: no inter-particle effects, only form-factors
**Model-free parameters**

**Guinier approximation and radius of gyration**

\[ I(Q) \approx I(0) \exp \left( -\frac{1}{3} R_g^2 Q^2 \right) \]

\[ \ln[I(Q)] \approx \ln[I(0)] - \frac{1}{3} R_g^2 Q^2 \]

\( R_g \leq 1...1.3 \)

(radius from expansion of Debye equation)

\[ R_g^2 = \frac{1}{M} \sum_i m_i r_i^2 \]

**Radius of gyration: For a given molecular weight, a sphere has the smallest \( R_g \), i.e. it is the most compact object**
Relative measurement of molecular weight ($M_r$) by SAS

\[ I(Q) \propto \left| \sum_j b_j e^{-iQr_j} \right|^2 \quad I(0) \propto \left| \sum_j b_j \right|^2 \quad I(0) \propto N \left| \sum_j b_j \right|^2 \]

\[ N = CN d V / M_r \]

\[ \left| \sum_j b_j \right|^2 \propto M_r^2 \]

Relative calibration to a known standard (BSA, lysozyme) in same buffer conditions:

\[ I_{prot}(0) = I_{standard}(0) \frac{C_{prot}M_{prot}}{C_{standard}M_{standard}} \]

Characteristic and distance distribution function
Characteristic and distance distribution function $\rho(r)$

Debye equation in integral form

$$ I(Q) = \int \rho(r_1) \rho(r_2) \frac{\sin Q(r_1 - r_2)}{Q(r_1 - r_2)} dV_1 dV_2 $$

$$ \gamma_o(r) = \frac{\gamma(r)}{\gamma(0)} = \frac{\langle V_c(r) \rangle}{V} $$

$\gamma_o(r)$ as a function of $r$

$\rho(r) = r^2 \gamma(r)$

Figure 2.2. Characteristic function $\gamma_0(r)$ and distance distribution function $\rho(r)$ for a solid sphere of radius $R$.

Figure 2.1. The concept of "shifted volume" $V_c(r)$.

SAS provides information on distances on different length-scales

Examples of $I(Q)$ and $p(r)$

$$I(Q) = \left[ \frac{3 \sin(QR) - QR \cos(QR)}{(QR)^3} \right]^2$$

Some words about resolution…

Nominal definition:

\[ R = \frac{2\pi}{Q_{\text{max}}} \]

Inter-subunit distance with a precision of about 1-2 Å!

Inter-subunit position/orientations less well-defined!
Modelling

Sophisticated data analysis and modelling

- Ab initio structure analysis
  ![Ab initio structure analysis](image)

- Rigid body modelling
  ![Rigid body modelling](image)

- Validation of structural models
  ![Validation of structural models](image)

Match with experimental scattering curve

Compatible with experimental scattering curve

Tutorials/lectures Maxim, Daniel, Alexey

F. Gabel (May 18th 2015) EMBO Practical Course
Different possibilities of modelling

Monodispersity is paramount (AUC, gels, SEC-MALLS)!

Flexible systems

Important check: molecular weight!
Oligomeric equilibria

Again: important check is molecular weight!

Neutrons vs. X-rays
SAXS vs. SANS: scattering processes

- X-ray scattering length is proportional to number of electrons
- Neutron scattering length depends irregularly on atom and isotope

Atoms have a form-factor for X-rays but nuclei don’t for neutrons...

Practical calculation of scattering densities

Example glycine in H₂O:

\[ \rho = \frac{2.067 + 0.94 + 0.58 + 3\times(-0.37)}{66.4} \times 10^{-12} \text{cm/Å}^3 = 2.68 \times 10^{10} \text{cm}^{-2} \]
SAXS vs. SANS: some practical aspects

<table>
<thead>
<tr>
<th></th>
<th>sample amount</th>
<th>flux</th>
<th>contrast</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAXS</td>
<td>~15 µl @ 1-10 mg/ml</td>
<td>high</td>
<td>weak</td>
</tr>
<tr>
<td>SANS</td>
<td>~150 µl @ 1-10 mg/ml</td>
<td>low</td>
<td>high (D$_2$O)</td>
</tr>
</tbody>
</table>

parasitic scattering at low Q

flat background

SAXS, ID02, 10s exposure time

SANS, D22, 20min exposure time

H$_2$O

D$_2$O

SANS vs. SAXS instruments

D22 (ILL): SANS

Quartz cuvette (SANS)

Capillary (SAXS)

BM29 (ESRF): SAXS

Exposure times:

~ 10-20 minutes (D22)
~ 1-10 seconds (ID02)
~ 10-100 seconds (BM29)

Including sample change:

~ 10-20 minutes (D22)
~ 5 minutes (ID02)
~ 2-3 minutes (BM29)
Understanding contrast: destructive interference in SANS

Idea of contrast variation

Vary scattering behaviour of (parts of) solutes with respect to solvent
Natural Contrast in SANS

\[ I(Q) = \left( \int (\rho_{\text{protein}} - \rho_{\text{solute}}) \cdot e^{-\lambda d} \cdot dV \right) \]

Homogeneous macromolecules can be matched, i.e. made invisible!!!
Not so easy with SAXS…

Contrast in SAXS


**Figure 4.5.** Possibilities of contrast variation in X-ray (a) and neutron (b) scattering. Lines (1)–(4) correspond to low-density lipoproteins, proteins, RNA, and DNA. The scattering density of solutes in H2O/D2O mixtures increases slightly due to H/D exchange. Matching points are denoted by crosses.
An analogon in optics: refractive index

SANS contrast variation: relative arrangement of DNA and protein

Pardon et al. (1975) Nucl. Acids Res. 2(11) 2163-2176

Relative topology of DNA and protein at low resolution before availability of high-resolution models!
Triangulation in the ribosome


Capel et al. (1987) Science 238, 1403-1406

Internal arrangement of proteins before availability of high-resolution models!

Protein-protein complexes


Negative radii of gyration!?

\[
\ln[I(Q)] \approx \ln[I(0)] - \frac{1}{3} R_g^2 Q^2
\]

\[
R_g^2 = \frac{1}{M} \sum_i m_i r_i^2 = \frac{1}{M} \sum_i \Delta \rho_i V_i r_i^2
\]

\[
I(Q) = \left| \sum_j b_j e^{-iQ \cdot r_j} \right|^2
\]

- Scattering in forward direction, \( I(0) \), can be weak (or zero)
- Scattering can get stronger going to higher angles
- Result: “apparent” negative/complex radius of gyration

Can proteins be considered as homogeneous particles?

Membrane proteins and lipids/detergents

Talks by Joe Zaccai, Lise Arleth…

Artificial contrast using deuteration

Protein deuteration not complete but only ~75%!

Careful at high D₂O levels in the solvent: favours oligomerisation/aggregation!
Practical guidelines

DATA COLLECTION
- Measure a concentration range (1-10 mg/ml)
- Check samples for radiation damage

DATA EVALUATION
- Identify potential aggregation using Guinier plot
- Identify potential concentration dependence in the scattering curves
- Extrapolate and merge the data to obtain interference-free scattering curves

DATA ANALYSIS
- Radius of Gyration ($R_g$) from Guinier plot – determine molecule size
- Molecular weight from $R_0$ – determine oligomerization state
- Excluded volume – confirm oligomerization state
- Pair distribution function $P(r)$ – determine distribution of atoms
- Kraty plot – screen for unfolded samples

SOLUTION STRUCTURE MODELING
- Ab-initio shape reconstruction
- Evaluate scattering from atomic structure
- Combine partial atomic structures
- Average independent shapes
- Parse body dock atomic models into SAXS shape
- Compare theoretical and experimental scattering
- Rigid-body and/or flexible structure modeling
- Build assemblies and/or missing linkers
- Build missing domains

Putnam et al. (2007)
Putnam et al. (2007)

**COLLECTION**

- Measure from different concentrations in the range 1-10 mg/ml. X-ray sensitivity can be detected by changes in scattering observed in repeated exposures and typically lead to increases in $R_g$ and $q^2$. X-ray damage can be minimized by adding 5-10% glycerol.

**EVALUATION**

- Calculate $R_g$ and $I(q)$ from the Guinier plot. A nonlinear dependence of $\log I(q)/q^2$ vs. $q^2$ indicates the presence of aggregation. Scattering from aggregated samples strongly influences the entire data set and no further processing should be performed! Aggregation in the sample can be induced or diminished by varying buffer conditions, centrifugation, and filtration.

- Evaluate the effects of interparticle interference $I(q)$ with the scattering of individual particles $S(q)$. Separation of mixed scattering curves from multiple concentrations can reveal concentration-dependent scattering. Decrease in intensity at very small $q$ with increasing protein concentration indicates interparticle repulsion. Repulsion can be dealt with by dilution, increasing ionic strength, and/or extrapolating the data to infinite dilution. Scaling the data to extrapolated scattering is only required in the range where the interference is visible, typically $q \approx 0.1 \text{ Å}^{-1}$.

**DATA ANALYSIS**

- Watch for increasing $R_g$ and $I(q)$ with increasing concentration. These can be due to changes in multimerization state or increasing amounts of aggregation in the sample.

**DIAGNOSING SCATTERING**

- To obtain an ideal scattering curve for the entire $q$-range, the scattering profile must be extrapolated to infinite dilutions at low resolution ($q < 0.1 \text{ Å}^{-1}$) and merged with the scattering profile for larger angles. Accurate large angle data can be obtained by measuring higher concentrations, using longer exposure times, and/or decreasing the sample-to-detector distance.

- If SAXS data are collected from monodisperse samples without interparticle interference, then reconstruction of the solution structure can proceed.

**Molecular weight from $I(q)$ requires a calibration curve and can be used to determine oligomerization state. This can be validated using calculation of the excluded volume.**

**DATA ANALYSIS**

- Molecular weight from $I(q)$ requires a calibration curve and can be used to determine oligomerization state. This can be validated using calculation of the excluded volume.

**Crystalline macromolecules with a $P(r)$ function with a single peak, while elongated macromolecules have a longer tail at large $r$ and have multiple peaks. The maximum length in the particle, $D_{max}$, is the position where the $P(r)$ function returns to zero at large values of $r$. Discrepancies for values of $R_g$ and $I(q)$ calculated from the $P(r)$ function and from the Guinier plot can indicate small amounts of aggregation that primarily affect the low resolution data and the accuracy of the Guinier plot.

**DATA ANALYSIS**

- The Kraty plot identifies unfolded samples. Globular macromolecules follow Poisson's law and have bell-shaped curves. Extended molecules, such as unfolded peptides, lack this peak and have a plateau or are slightly increasing in the larger $q$-range.
Combination of SAXS/SANS with other techniques

Putnam et al. (2007)
SANS allows to go beyond the global shape and study internal structure!

Often problematic to position/orient subunits in a larger complex using SAXS alone...

Internal structure: contrast variation and SANS!
Complex systems: approaches combining SAXS/SANS with other techniques


Structural refinement strategy

System from *Pyrococcus furiosus*

Rigid building blocks from crystal structures + flexible joints

Methyl resonances CSPs of Fib and L7Ae + 452 PRE distance restraints (apo-complex) + SAXS/SANS
Relative positions of FIB proteins within the complex from SANS data

**dFIB 42% D₂O SANS data:**
FIB positions in the complex!

Program MONSA:

**Important restraints for the atomic models!**

Family of refined structures

Network of NMR and SAS restraints


**Talk by Bernd Simon**
Concluding remarks

A few practical comments…

- use SAXS for **homogeneous** systems composed of a **single** body
- SAXS is better suited for **high-throughput**
- SANS good for **complex systems** (protein-DNA/RNA, membrane proteins…)
- SANS has **no radiation damage**
- neutrons only possible at **large facilities** (no “home sources” for the moment!)
- request for beam-time is generally **via an electronic proposal system**
- deadlines are usually twice a year, beamtime is attributed some **months** later
- BAG (“Block allocation group”) systems allow more flexible access
- for continuation proposals, **reports** need to be submitted regularly
- **experiments need to be prepared with great care** (i.e. isotopic effect of D₂O)!!
- “local contacts”, often beamline responsibles, **assist** during experiments
- access (for non-industrial use) is in general **free**
- **no maintenance**, user friendly (software etc….)
Literature

**Basics (scattering, quantum mechanics):**
- The Feynman lectures on Physics, Volume 3: Quantum mechanics (Addison Wesley, 2006)

**Books on small angle (neutron) scattering:**
- Guinier/Fournet: Small angle scattering of X-rays (John Wiley & Sons, 1955)
- Serdyuk, Zaccai, Zaccai: Methods in molecular biophysics (Cambridge University Press, 2007)

**Reviews on SAXS/SANS:**