Basics of X-ray scattering by solutions

D.Svergun, EMBL-Hamburg

Small-angle scattering in structural biology

Radiation sources:
- X-ray tube ($\lambda = 0.1 - 0.2 \text{ nm}$)
- Synchrotron ($\lambda = 0.05 - 0.5 \text{ nm}$)
- Thermal neutrons ($\lambda = 0.1 - 1 \text{ nm}$)

Sample

Incident beam

Wave vector $k$, $k = 2\pi/\lambda$

Solvent

Scattered beam, $k_f$

Detector

Data analysis

Shape determination

Rigid body modelling

Missing fragments

Oligomeric mixtures

Hierarchical systems

Flexible systems

Resolution, nm:
- 3.1
- 2.0
- 1.6
- 1.0
- 0.8

Additional information:
- Homology models
- Atomic models
- Distances
- Orientations
- Interfaces

Complementary techniques:
- MS
- EM
- Crystalllography
- NMR
- Bioinformatics
- Biochemistry
- AUC
- FRET
- EPR

Scattering curve $I(s)$

Sample

Solvent

Incident beam

Wave vector $k$, $k = 2\pi/\lambda$

Scattered beam, $k_f$

Detector

Resolution, nm:
- 3.1
- 2.0
- 1.6
- 1.0
- 0.8

Additional information:
- Homology models
- Atomic models
- Distances
- Orientations
- Interfaces

Complementary techniques:
- MS
- EM
- Crystalllography
- NMR
- Bioinformatics
- Biochemistry
- AUC
- FRET
- EPR

Scattering curve $I(s)$

Sample

Solvent

Incident beam

Wave vector $k$, $k = 2\pi/\lambda$

Scattered beam, $k_f$

Detector

Resolution, nm:
- 3.1
- 2.0
- 1.6
- 1.0
- 0.8

Additional information:
- Homology models
- Atomic models
- Distances
- Orientations
- Interfaces

Complementary techniques:
- MS
- EM
- Crystalllography
- NMR
- Bioinformatics
- Biochemistry
- AUC
- FRET
- EPR

Scattering curve $I(s)$

Sample

Solvent

Incident beam

Wave vector $k$, $k = 2\pi/\lambda$

Scattered beam, $k_f$

Detector

Resolution, nm:
- 3.1
- 2.0
- 1.6
- 1.0
- 0.8

Additional information:
- Homology models
- Atomic models
- Distances
- Orientations
- Interfaces

Complementary techniques:
- MS
- EM
- Crystalllography
- NMR
- Bioinformatics
- Biochemistry
- AUC
- FRET
- EPR
General principles of solution SAXS

Radiation sources:
- X-ray generator ($\lambda = 0.1 - 0.2$ nm)
- Synchrotron ($\lambda = 0.03 - 0.35$ nm)
- Thermal neutrons ($\lambda = 0.2 - 1$ nm)

Small-angle scattering: experiment

Monochromatic beam
Wave vector $k$, $k = 2\pi/\lambda$

Scattering vector $s = k_1 - k$
Scattering by matter

- **X-rays** are scattered mostly by electrons
- **Thermal neutrons** are scattered mostly by nuclei
- Scattering amplitude from an ensemble of atoms $A(s)$ is the Fourier transform of the scattering length density distribution in the sample $\rho(r)$
- Experimentally, scattering intensity $I(s) = [A(s)]^2$ is measured.

Notations

The momentum transfer (i.e. the modulus of the scattering vector) is denoted here as $s = 4\pi \sin(\theta)/\lambda$.

There are also different letters used, like

$$Q = q = s = h = k = 4\pi \sin(\theta)/\lambda$$

Sometimes, the variable $S = 2\sin(\theta)/\lambda = 2\pi s$ is used, and to add to the confusion, also denoted as “s”, or $\mu$ or yet another letter. Always check the definition for the momentum transfer in a paper.
Small-angle scattering: contrast

To obtain scattering from the particles, matrix scattering must be subtracted, which also permits to significantly reduce contribution from parasitic background (slits, sample holder etc).

Contrast $\Delta \rho = \langle \rho(r) - \rho_s \rangle$, where $\rho_s$ is the scattering density of the matrix, may be very small for biological samples.

X-rays versus neutrons

- **X-rays**: scattering factor increases with atomic number, no difference between H and D
- **Neutrons**: scattering factor is irregular, may be negative, huge difference between H and D

<table>
<thead>
<tr>
<th>Element</th>
<th>H</th>
<th>D</th>
<th>C</th>
<th>N</th>
<th>O</th>
<th>P</th>
<th>S</th>
<th>Au</th>
</tr>
</thead>
<tbody>
<tr>
<td>At. Weight</td>
<td>1</td>
<td>2</td>
<td>12</td>
<td>14</td>
<td>16</td>
<td>30</td>
<td>32</td>
<td>197</td>
</tr>
<tr>
<td>N electrons</td>
<td>1</td>
<td>1</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>15</td>
<td>16</td>
<td>79</td>
</tr>
<tr>
<td>$b_x, \times 10^{-12}$ cm</td>
<td>0.282</td>
<td>0.282</td>
<td>1.69</td>
<td>1.97</td>
<td>2.16</td>
<td>3.23</td>
<td>4.51</td>
<td>22.3</td>
</tr>
<tr>
<td>$b_n, \times 10^{-12}$ cm</td>
<td>-0.374</td>
<td>0.667</td>
<td>0.665</td>
<td>0.940</td>
<td>0.580</td>
<td>0.510</td>
<td>0.280</td>
<td>0.760</td>
</tr>
</tbody>
</table>

Neutron contrast variation
In the equations below we shall always assume that the solvent scattering has already been subtracted.

**Contrast of electron density**

\[
\tilde{\rho} = 0.43 \\
\rho_0 = 0.335 \\
\Delta \rho = \rho - \rho_0
\]

**Solution of particles**

\[
\text{Solution} \quad \Delta \rho(r) \quad F(c,s) = \ast \quad \text{Motif (protein)} \quad \Delta \rho_p(r) \quad F(0,s) = \ast \quad \text{Lattice} \quad d(r) \quad \delta(c,s)
\]
Solution of particles

For spherically symmetrical particles

\[ I(c,s) = I(0,s) \times S(c,s) \]

- form factor of the particle
- structure factor of the solution

Still valid for globular particles though over a restricted s-range

Solution of particles

- 1 – *monodispersity*: identical particles
- 2 – size and shape polydispersity
- 3 – *ideality*: no intermolecular interactions
- 4 – non ideality: existence of interactions between particles

*In the following, we make the double assumption 1 and 3*
**Ideal and monodisperse solution**

\[ A(s) = \Im[\Delta \rho(r)] = \int \Delta \rho(r) \exp(isr) dr \]

Particles in solution \(\Rightarrow\) thermal motion \(\Rightarrow\) particles have random orientations to X-ray beam. The sample is isotropic. Therefore, only the spherical average of the scattered intensity is experimentally accessible.

**Ideality and monodispersity** \[ I(s) = N_i(s) \]

---

**Crystal** \(\text{versus}\) **solution**

\[ I(c,s) = I(0,s) \times S(c,s) \]

For an ideal crystal, \(I(s)\) is the three-dimensional scattering intensity from the unit cell
S(s) is a sum of \(\delta\)-functions along the directions of the reciprocal space lattice \(s=(ha^*+kb^*+lc^*)\)

For an ideal dilute solution, \(I(s)=I(s)\) is the orientationally averaged intensity of the single particle
S(s) is equal to unity
For an ideal crystal, measured signal is amplified into specific directions allowing measurements to high resolution ($d \approx \lambda$).

For an ideal dilute solution, $I(s)$ is isotropic and concentrates around the primary beam (this is where the name "small-angle scattering" comes from): low resolution ($d \gg \lambda$).

---

**Main equations and overall parameters**

[Image source: ochevidets.ru]
Relation between real and reciprocal space

Using the overall expression for the Fourier transformation one obtains for the spherically averaged single particle intensity

\[
I(s) = \left< A(s) A^*(s) \right> = \int_V \left( \int_V \Delta \rho(r) \Delta \rho(r') \exp\{is(r - r')\} \right) dr dr'
\]

or, taking into account that \(<\exp(isr)\>_{\Omega} = \sin(sr)/sr\) and integrating in spherical coordinates,

\[
I(s) = 4\pi \int_0^{D_{\text{max}}} r^2 \gamma(r) \frac{\sin sr}{sr} dr
\]

where

\[
\gamma(r) = \left\{ \int \Delta \rho(u) \Delta \rho(u + r) du \right\}_{\omega}
\]

Distance distribution function

\[
p(r) = r^2 \gamma(r) = r^2 \gamma_0(r) V \rho^2
\]

\(\gamma_0(r)\) : probability of finding a point at \(r\) from a given point

number of el. vol. \(i \propto V\) - number of el. vol. \(j \propto 4\pi r^2\)

number of pairs \((i,j)\) separated by the distance \(r \propto 4\pi r^2 V \gamma_0(r) = (4\pi/\rho^2)p(r)\)
If the particle is described as a discrete sum of elementary scatterers, (e.g. atoms) with the atomic scattering factors $f_i(s)$ the spherically averaged intensity is

$$I(s) = \sum_{i=1}^{K} \sum_{j=1}^{K} f_i(s)f_j(s) \frac{\sin (sr_{ij})}{sr_{ij}}$$

where $r_{ij} = |r_i - r_j|$

The Debye (1915) formula is widely employed for model calculations.

**Contribution of distances to the scattering pattern**

In isotropic systems, each distance $d = r_{ij}$ contributes a $\sin x/x$ -like term to the intensity.

**Large distances** correspond to high frequencies and only contribute at **low angles** (i.e. at low resolution, where particle shape is seen).

**Short distances** correspond to low frequencies and contribute over a large angular range. Clearly at **high angles** their contribution dominates the scattering pattern.
Small and large proteins: comparison

Guinier law

Near $s=0$ one can insert the McLaurin expansion $\sin(sr)/sr \approx 1-(sr)^2/3!+...$ into the equation for $I(s)$ yielding

$$I(s) = I(0) \left[ 1 - \frac{1}{3} R_g^2 s^2 + O(s^4) \right] \approx I(0) \exp \left( -\frac{1}{3} R_g^2 s^2 \right)$$

This is a classical formula derived by Andre Guinier (1938) in his first SAXS application (to defects in metals). The formula has two parameters, forward scattering and the radius of gyration

$$I(0) = \int_V \int_V \Delta \rho(r) \Delta \rho(r') dr dr' = 4\pi \int_0^{D_{\max}} p(r) dr = (\Delta \rho)^2 V^2$$

$$R_g = \int_0^{D_{\max}} r^2 p(r) dr \left[ 2 \int_0^{D_{\max}} p(r) dr \right]^{-1}$$
Intensity at the origin

\[ i_i(0) = \int_{V_i} \int_{V_i'} \Delta \rho(r) \Delta \rho(r') dV_i dV_i' \]

\[ i_i(0) = \Delta m^2 = (m-m_0)^2 = \left[ \frac{M}{N_A} v_p (\rho - \rho_0) \right]^2 \]

\[ c = \frac{NM}{N_N V} \] is the concentration (w/v), e.g. in mg.ml\(^{-1}\)

\[ I(0) = \frac{cMV}{N_A} \left[ v_p (\rho - \rho_0) \right]^2 \]

---

Intensity at the origin

If: the concentration \( c \) (w/v), ___
the partial specific volume \( v_p \),
the intensity on an absolute scale,
i.e. the number of incident photons
are known,
Then, the **molecular weight** of the particle can be
determined from the value of the intensity at the origin

In practice, MM can be determined from the data on
relative scale by comparison with \( I(0) \) of a reference protein
(e.g. BSA, lysozyme or cytochrom C)
**Radius of gyration**

Radius of gyration: 

\[ R_g^2 = \frac{\int V r^2 \Delta \rho(r) dV}{\int V \Delta \rho(r) dV} \]

\( R_g \) is the quadratic mean of distances to the center of mass weighted by the contrast of electron density.

\( R_g \) is an **index of non sphericity**.

For a given volume the smallest \( R_g \) is that of a sphere:

\[ R_g = \sqrt{\frac{3}{5}} R \]

**Ellipsoid of revolution** \((a, b)\)  
**Cylinder** \((D, H)\)  

\[ R_g = \sqrt{\frac{2a^2 + b^2}{5}} \quad R_g = \sqrt{\frac{D^2 + H^2}{12}} \]

---

**Virial coefficient**

In the case of moderate interactions, the intensity at the origin varies with concentration according to:

\[ I(0, c) = \frac{I(0)_{\text{ideal}}}{1 + 2A_2Mc + ...} \]

Where \( A_2 \) is the second virial coefficient which represents pair interactions and \( I(0)_{\text{ideal}} \) is \( \propto \) to \( c \).

\( A_2 \) is evaluated by performing experiments at various concentrations \( c \).

\( A_2 \) is \( \propto \) to the slope of \( c/I(0, c) \) vs \( c \).

To obtain \( I(0, s) \), this extrapolation to infinite dilution is performed for different angles.
Guinier plot example

The law is generally used under its log form:

\[ \ln[I(s)] = \ln[I(0)] - \frac{[sR_g]^2}{3} \]

A linear regression yields two parameters: \( I(0) \) (y-intercept) \( R_g \) from the slope

Validity range: \( 0 < sR_g < 1.3 \)

---

Rods and platelets

In the case of very elongated particles, the radius of gyration of the cross-section can be derived using a similar representation, plotting this time \( sI(s) \) vs \( s^2 \)

\[ sI(s) \approx I_C(0) \exp\left( -\frac{1}{2} R_c^2 s^2 \right) \]

In the case of a platelet, a thickness parameter is derived from a plot of \( s^2I(s) \) vs \( s^2 \):

\[ s^2I(s) \approx I_T(0) \exp\left( - R_t^2 s^2 \right) \]

with \( R_t = T/\sqrt{12} \) \( T \): thickness
Porod invariant and volume

Following the Parseval theorem for Fourier transformations

\[ Q = \int_0^{\infty} s^2 I(s) ds = 2\pi^2 \int_V (\Delta \rho(r))^2 dr \]

Q is called the Porod invariant, which is computed from the intensity but provides the mean square electron density contrast.

For homogeneous particles, \( Q = 2\pi^2 (\Delta \rho)^2 V \), and, taking into account that \( I(0) = (\Delta \rho)^2 V^2 \), the excluded volume of hydrated particle in solution (Porod volume) is

\[ V = 2\pi^2 I(0)/Q. \]

The asymptotic regime: Porod law

Integrating the Fourier transformation for I(s) by parts and using that for particles with a sharp interface \( \gamma'(D_{\text{max}}) = 0 \), one has

\[ I(s) \cong 8\pi s^{-4} \gamma'(0) + O_1 s^{-3} + O_2 s^{-4} + o(s^{-5}) \]

where \( O_1, O_2 \) are oscillating trigonometric terms of the form \( \sin (sD_{\text{max}}) \). The main term responsible for the intensity decay at high angles is therefore proportional to \( s^{-4} \), and this is known as Porod’s law (1949). For homogeneous particles, \( \gamma'(0) \) is equal to \(- (\Delta \rho)^2 S/4\), where \( S \) is the particle surface.
Distance distribution function

$$p(r) = \frac{r^2}{2\pi^2} \int_0^\infty s^2 I(s) \frac{\sin sr}{sr} dr$$

In theory, calculation of $p(r)$ from $I(s)$ is simple.

**Problem**: $I(s)$ is only known over $[s_{\text{min}}, s_{\text{max}}]$ : truncation
- is affected by experimental errors and possible instrumental distortions due to the beam-size and the bandwidth $\Delta \lambda/\lambda$ (neutrons)

$\Rightarrow$ Fourier transform of *incomplete and noisy data* is an *ill-posed problem*.

**Solution**: Indirect Fourier Transform (suggested by O. Glatter, 1977).

$p(r)$ is parameterized on $[0, D_{\text{max}}]$ by a linear combination of orthogonal functions, where $D_{\text{max}}$ is the particle diameter.

Implemented in several programs, including GNOM (part of ATSAS)

---

Distance distribution function

The radius of gyration and the intensity at the origin are derived from $p(r)$ using the following expressions:

$$R_g^2 = \int_0^D r^2 p(r) dr$$

and

$$I(0) = 4\pi \int_0^{D_{\text{max}}} p(r) dr$$

This alternative estimate of $R_g$ makes use of the whole scattering curve, and is much less sensitive to interactions or to the presence of a small fraction of oligomers.

Comparison of both estimates: useful cross-check
Aspartate transcarbamylase from *E. coli* (ATCase)
Heterododecamer
$\left(c_3 \right)_2 \left(r_2 \right)_3$ quasi $D_3$ symmetry
Molecular weight: 306 kDa
Allosteric enzyme

2 conformations:
- $T$: inactive, compact
- $R$: active, expanded

**ATCase: scattering patterns**

$s = 2\sin(\theta)/\lambda \ A^{-1}$
ATCase: \( p(r) \) curves

Note: \( I(s) \) and \( p(r) \) contain the same information, but differently represented.

Kratky plot

A plot of \( s^2I(s) \) vs \( s \) provides a sensitive means of monitoring the degree of compactness of a protein.

Globular particle: bell-shaped curve
Unfolded particle: plateau or increase at large \( s \)-values
Summary of model-independent information

I(0)/c, i.e. molecular mass (from Guinier plot or p(r) function)
Radius of gyration $R_g$ (from Guinier plot or p(r) function)
Radii of gyration of thickness or cross-section (anisometric particles)
Second virial coefficient $A_2$ (extrapolation to infinite dilution)
Maximum particle size $D_{max}$ (from p(r) function)
Particle volume $V$ (from I(0) and Porod invariant)
Specific surface $S/V$ (from I(0), Porod invariant and asymptotics)
Globular or unfoded (From Kratky plot)

Small-angle scattering: experiment

Monochromatic beam
Wave vector $k$, $k=2\pi/\lambda$
Radiation sources:
X-ray generator ($\lambda = 0.1 - 0.2$ nm)
Synchrotron ($\lambda = 0.03 - 0.35$ nm)
Thermal neutrons ($\lambda = 0.2 - 1$ nm)

Scattering vector $s=k_1-k$, $s=4\pi \sin(\theta)/\lambda$

Detector

$2\theta$
Crystal **versus** solution

- Thousands of reflections
- 3D, high resolution
- Data undersampled, $\Delta s = 2\pi/D$

- A few Shannon channels
- 1D, low resolution
- Data oversampled, $\Delta s << \pi/D$
For SAXS solution studies, one does not need to grow crystals.

SAXS is not limited by molecular mass and is applicable under nearly physiological conditions.

Using solution SAXS, one can more easily observe responses to changes in conditions.

SAXS permits for quantitative analysis of complex systems and processes.

In solution, no crystallographic packing forces are present.

Scattering from dilute macromolecular solutions (monodisperse systems)

\[ I(s) = 4\pi \int_0^D p(r) \frac{\sin sr}{sr} dr \]

The scattering is proportional to that of a single particle averaged over all orientations, which allows one to determine size, shape and internal structure of the particle at low (1-10 nm) resolution.
The scattering is related to the shape (or low resolution structure)

Solid sphere

Hollow sphere

Flat disc

Long rod

When biologists go for SAS

This is just trivial case: SAS yields much more
Methods development at EMBL-Hamburg

Data processing and manipulations
- Calibration and normalization
- Raw data processing and radial averaging
- Data manipulations
- Merging and splicing
- Concentration series analysis
- Computation of invariants
- Indirect Fourier transformation
- Simple bodies modelling
- Singular value decomposition
- Analysis of mixtures
- Peak analysis


Used in more than 2000 laboratories worldwide

Ab initio modeling suite
Analysis of mixtures

Data processing and analysis: PRIMUS

Overall parameters

Radius of gyration $R_g$ (Guinier, 1939)

$I(s) = I(0) \exp\left(-\frac{1}{3} R_g^2 s^2\right)$

Maximum size $D_{\text{max}}$: $p(r) = 0$ for $r > D_{\text{max}}$

Excluded particle volume (Porod, 1952)

$V = 2\pi^2 I(0)/Q; \quad Q = \int_0^\infty s^2 I(s) ds$

“Simple” monodisperse systems

Shape and conformational changes of macromolecules and complexes

Validation of high resolution models and oligomeric organization

Rigid body models of complexes using high resolution structures

Addition of missing fragments to high resolution models
The scattering is proportional to that of a single particle averaged over all orientations, which allows one to determine size, shape and internal structure of the particle at low (≤ 10 nm) resolution, given their scattering intensities \( I_k(s) \), also the volume fraction:

\[
I(s) = \sum_k v_k I_k(s)
\]
A roadmap of biological SAS data analysis

ATSAS 2.5 tutorials by M.Petoukhov, P.Konarev, C.Blanchet

A word of caution

- Sample preparation
- Experiment
- Data processing
- Unambiguous interpretation
- Changing conditions
- Relation to function
### Books on SAXS


The Proceedings of the SAS Conferences held every three years were published in the Journal of Applied Crystallography. The latest proceedings are in the J. Appl. Cryst., 40, (2007).

### Recent reviews on solution SAS

