Basics of X-ray scattering by solutions

D.Svergun, EMBL-Hamburg
Small-angle scattering in structural biology

**Data analysis**

**Shape determination**

**Rigid body modelling**

**Missing fragments**

**Oligomeric mixtures**

**Hierarchical systems**

**Flexible systems**

**Complementary techniques**

MS  EM  NMR  Bioinformatics  AUC  EPR

Crystallography  Biochemistry

Additional information

Homology models  Atomic models  Distances

Orientations  Interfaces

Resolution, nm:

- 3.1
- 1.6
- 1.0
- 0.8

Scattering curve $I(s)$

Radiation sources:

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

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

Sample

Incident beam

Scattered beam, $k_1$

$2\theta$

Sample

Solvent

Detector

2\theta

Scattered beam, $k_1$

Resolution, nm:

- 3.1
- 1.6
- 1.0
- 0.8

Scattering curve $I(s)$

$0$  $2$  $4$  $6$  $8$

$s, \text{ nm}^{-1}$

$\lg I, \text{ relative}$

**Additional information**

$2\theta$

Sample

Solvent

Incident beam

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

Detector

Scattered beam, $k_1$

Radiation sources:

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

**Small-angle scattering in structural biology**

**Data analysis**

**Shape determination**

**Rigid body modelling**

**Missing fragments**

**Oligomeric mixtures**

**Hierarchical systems**

**Flexible systems**

**Complementary techniques**

MS  EM  NMR  Bioinformatics  AUC  EPR

Crystallography  Biochemistry

Additional information

Homology models  Atomic models  Distances

Orientations  Interfaces

Resolution, nm:

- 3.1
- 1.6
- 1.0
- 0.8

Scattering curve $I(s)$

$0$  $2$  $4$  $6$  $8$

$s, \text{ nm}^{-1}$

$\lg I, \text{ relative}$

**Additional information**

$2\theta$

Sample

Solvent

Incident beam

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

Detector

Scattered beam, $k_1$

Radiation sources:

- X-ray tube ($\lambda = 0.1 - 0.2 \text{ nm}$)
- Synchrotron ($\lambda = 0.05 - 0.5 \text{ nm}$)
- Thermal neutrons ($\lambda = 0.1 - 1 \text{ nm}$)
General principles of solution SAXS
Small-angle scattering: experiment

Monochromatic beam

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

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

Scattering vector $s = k_1 - k$

$2\theta$

Log (Intensity)

Detector

Sample

$\theta$

$s = 4\pi \sin(\theta)/\lambda$, nm$^{-1}$
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.
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>$b_X, 10^{-12}$ cm</td>
<td><strong>0.282</strong></td>
<td><strong>0.282</strong></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>$b_N, 10^{-12}$ cm</td>
<td><strong>-0.374</strong></td>
<td><strong>0.667</strong></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.
Solution of particles

\[
\Delta \rho(r) \ast F(c,s) = \Delta \rho_p(r) \ast d(r) \ast \delta(c,s)
\]
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

*For most of the following derivations of the structural parameters, we shall make the double assumption 1 and 3*
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.

\[
A(s) = \mathcal{I}[\Delta \rho(r)] = \int_V \Delta \rho(r) \exp(isr) \, dr
\]

Ideality and monodispersity $I(s) = Ni_1(s)$
For an ideal crystal,
\[ I(\mathbf{s}) \] is the three-dimensional scattering intensity from the unit cell
\[ S(\mathbf{s}) = \delta(\mathbf{s}) \] is a sum of \( \delta \)-functions along the directions of the reciprocal space lattice
\[ s = (h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^*) \]

For an ideal dilute solution,
\[ I(\mathbf{s}) = I(0,\mathbf{s}) \times S(\mathbf{c},\mathbf{s}) \]
\[ I(s) = I(s) \] is the orientationally averaged intensity of the single particle
\[ S(s) = 1 \] 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 ($d >> \lambda$).
Main equations and overall parameters
Relation between real and reciprocal space

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

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

or, taking into account that \( \left\langle \exp(isr) \right\rangle_\Omega = \sin(sr)/sr \) and integrating in spherical coordinates,

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

where

\[
\gamma(r) = \left\langle \int \Delta \rho(u) \Delta \rho(u+r) du \right\rangle_\omega
\]
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 $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$.

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

Log I(s) vs. s, nm\(^{-1}\)

lysozyme

apoferitin
Guinier law

Near \(s=0\) one can insert the McLaurin expansion
\[
\frac{\sin(sr)}{sr} \approx 1 - \frac{(sr)^2}{3!} + \ldots
\]
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

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

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

\[ i_1(0) = \int_{V_r} \int_{V_r} \Delta \rho(\mathbf{r}) \Delta \rho(\mathbf{r}') dV_r dV_{r'} \]

\[ i_1(0) = \Delta m^2 = (m - m_0)^2 = \left[ \frac{M}{N_A} \bar{v}_p (\rho - \rho_0) \right]^2 \]

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

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

ideal monodisperse
If: the concentration \( c \) (w/v), the partial specific volume \( \nu_p \), the intensity on an absolute scale, i.e. the number of incident photons are known,
Then, the **molecular mass** 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: 

$$R_g^2 = \frac{\int_{V_r} \Delta \rho(r)r^2dV_r}{\int_{V_r} \Delta \rho(r)dV_r}$$

$R_g$ 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 $(a, b)$ 

$$R_g = \sqrt{\frac{2a^2 + b^2}{5}}$$

Cylinder $(D, H)$ 

$$R_g = \sqrt{\frac{D^2}{8} + \frac{H^2}{12}}$$

ideal monodisperse
Repulsion versus attraction

Computed scattering:
(1), from a solid sphere with $R = 5$ nm,
(2), from a solution of non-interacting hard spheres with $\nu = 0.2$,
(3), from a dumbbell (divided by a factor of two)

At higher angles, interactions are not visible; at small angles, repulsion diminishes the scattering, attraction increases the scattering
The structure factor at a given concentration \( c \) can be obtained from the ratio of the experimental intensity at this concentration to that obtained by extrapolation to infinite dilution or measured at a sufficiently low concentration \( c_0 \) were all correlations between particles have vanished:

\[
S(q, c) = \frac{c_0 I_{\text{exp}}(q, c)}{c I(q, c_0)}
\]
For non-interacting hard spheres model, based on the statistical-mechanical treatment of hard-sphere fluids (Percus and Yevick 1958), $S(s)$ was expressed as an analytical function of two parameters, the sphere radius $R$ and the volume fraction $f$, $S(s, R, v) = (1 - C(s, R, f))^{-1}$ (Ashcroft and Leckner 1966). Here, $C(s, R, f)$ is the Fourier transformation of the interparticle correlation function

$$C(s, R, f) = -\frac{24\phi}{x^6} \left\{ \alpha x^3 [\sin x - x \cos x] + \beta x^2 \left[ 2x \sin x - (x^2 - 2) \cos x - 2 \right] + \gamma \left[ (4x^3 - 24x) \sin x - (x^4 - 12x^2 + 24) \cos x + 24 \right] \right\}$$

where $x = 2sR$, $\alpha = (1 + 2f)^2/(1-f)^4$, $\beta = -6f(1+f/2)^2/(1-f)^4$ and $\gamma = f(1+2f)^2/[2(1-f)^4]$. 
Structure factor

Structure factors computed for systems of hard spheres with the radius 5 nm (curves 1-4 correspond to the volume fractions $\nu = 0.05, 0.1, 0.2, 0.3$ respectively).
Virial coefficient

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

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

Where \( A_2 \) is the second virial coefficient which represents pair interactions and \( I(0)_{\text{ideal}} \) is proportional to \( c \).
\( A_2 \) is evaluated by performing experiments at various concentrations \( c \).
\( A_2 \) is proportional 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:

$$0 < sR_g < 1.3$$
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
**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:

\[ R_g^2 = \frac{\int_0^{D_{\text{max}}} r^2 p(r) dr}{2 \int_0^{D_{\text{max}}} p(r) dr} \quad \text{and} \quad I(0) = 4\pi \int_0^{D_{\text{max}}} p(r) dr \]

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
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 (Porod) volume of hydrated particle in solution (Porod, 1952) is

\[ V = \frac{2\pi^2 I(0)}{Q} . \]
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

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

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.
Correlation volume

One can also calculate 'correlation volume' using the (better converging) integral from the intensity weighted with s, not $s^2$

$$V_c = \frac{I(0)}{\int_0^\infty sI(s)ds} = \frac{V_p}{2\pi l_c}$$

where $l_c$ is the particle 'correlation length'. This quantity has no direct physical meaning (having units of surface). Rambo & Tainer (2013) showed that the ratio $(V_c^2)/R_g$ can be used to estimate molecular weight of particles from an empirical power law dependence.
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
Scattering from three 60 kDa proteins: globular (dark blue), half- (light blue) and fully disordered (gray). A, B: scattering intensity $I(s)$ and $p(r)$ functions in (inverse) nanometres. C: Kratky plot $s^2 I(s)$ vs. $s$. D: Normalized (or “dimensionless”) Kratky plot $(sR_g)^2 I(s)/I(0)$ vs. $sR_g$. Reference:
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)*

$R_g$ 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)*

Molecular mass estimate from correlation volume $V_c$ *(empirical)*

Specific surface $S/V$ *(from I(0), Porod invariant and asymptotics)*

Globularity/ unfodedness *(from (dimensionless) Kratky plot)*
Small-angle scattering: experiment

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

Monochromatic beam

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

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

Log (Intensity)
Crystal

VERSUS

solution
Crystal versus solution

- Thousands of reflections
- 3D, high resolution

- A few Shannon channels
- 1D, low resolution

- Data undersampled, 
  \[ \Delta s = \frac{2\pi}{D} \]

- Data oversampled, 
  \[ \Delta s \ll \frac{\pi}{D} \]
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)

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.
Example of strong attractive interactions

\[ \gamma\text{-crystallins c=160 mg/ml in 50mM Phosphate pH 7.0} \]

A. Tardieu et al., LMCP (Paris)
Example of strong repulsive interactions

ATCase in 10mM borate buffer pH 8.3

P.Vachette et al., CNRS-Université Paris-sud, Orsay
In dilute solutions, scattering is related to the shape (or low resolution structure)

Solid sphere

Hollow sphere

Flat disc

Long rod

Dumbbell
When biologists go for SAS

Care for a shape?

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

Data processing and manipulations
Rigid body refinement

Ab initio modeling suite
Analysis of mixtures

Employed by over 12,000 users worldwide

“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
Scattering from mixtures

\[ I(s) = \sum_{k} v_k I_k(s) \]

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. For equilibrium and non-equilibrium mixtures, solution scattering permits to determine the number of components and, given their scattering intensities \( I_k(s) \), also the volume fractions.
Complicated systems: mixtures and processes

Equilibrium oligomeric mixtures

Stoichiometry and complex formation

Natively unfolded proteins and multidomain proteins with flexible linkers

Protein folding/unfolding kinetics

Assembly/disassembly processes
A roadmap of biological SAS data analysis

Polydisperse systems
Volume fractions of components
Unfolded or flexible proteins

Databases of computed and experimental scattering from atomic models (X-rays and neutrons)

Ab initio analysis

High resolution models
Databases of computed and experimental

Ab initio analysis

Bead modelling
Multiphase bead modelling
Dummy residues modelling
Rigid body modelling
Manual and local refinement
Multisubunit complexes modelling
Multidomain proteins modelling

Data regularization and overall structural parameters
Add missing fragments to high resolution models
Models superposition, averaging and clustering

ATSAS 2.7 tutorials, practical sessions, SASQuest and remote data collection at the P12 beamline in Hamburg by M.Petoukhov, A.Kikhney, C.Blanchet, M.Graewert
A word of caution

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


Recent reviews on solution SAS


