Solution scattering from biological macromolecules

Primary data reduction and analysis

Al Kikhney, EMBL Hamburg
Outline

• 3D → 2D → 1D
• Experiment design and data reduction
  • Exposure time
  • Background subtraction
  • Dilution series
• Overall parameters:
  • Guinier analysis:
    \( R_g, I(0), \text{molecular mass} \)
  • Volume
  • \( p(r), D_{max} \)
X-ray → SAXS experiment

- Few kDa to GDa
- Monodisperse and homogeneous
- Concentration: 0.5–10 mg/ml
- Amount: 10–100 μl
2D → 1D
Normalization

- Transmitted beam
- Exposure time
Notations and units

X-ray →

solution

2θ

s

X-ray detector
Notations and units

\[ |s| = \frac{4\pi \sin\theta}{\lambda} \]

2θ – scattering angle  
\( \lambda \) – wavelength  
\( s \) – scattering vector  
\( I(s) \) – intensity
Notations and units

\[ |s| = \frac{4\pi \sin\theta}{\lambda} \]

- \(2\theta\) – scattering angle
- \(\lambda\) – wavelength
- \(s\) – scattering vector
- \(I(s)\) – intensity

\(I(s)\),
\[ cm^{-1} \]

\[ s, \text{ nm}^{-1} \]
Notations and units

\[ |q| = \frac{4\pi \sin\theta}{\lambda} \]

2\(\theta\) – scattering angle
\(\lambda\) – wavelength

\(I(q), a.u.\)

\(q, \text{ nm}^{-1}\)
$|s| = 4\pi \sin\theta / \lambda$

2$\theta$ – scattering angle
$\lambda$ – wavelength

Notations and units
Notations and units

\[ |s| = \frac{4\pi \sin \theta}{\lambda} \]

2\(\theta\) – scattering angle  
\(\lambda\) – wavelength
Notations and units

\[ |s| = \frac{4\pi \sin \theta}{\lambda} \]

2\(\theta\) – scattering angle  
\(\lambda\) – wavelength
Exposure time

0.05 second
0.2 second
0.8 second
Exposure time

I(s)

RADIATION DAMAGE!

0.05 second
0.2 second
0.8 second
1.6 second

s, nm⁻¹
Multiple exposures

frame 1
Multiple exposures

frame 1
frame 2
Multiple exposures
Multiple exposures

frame 1
frame 10 – discard
Sample and buffer

I(s)

3.2 mg/ml lysozyme + buffer + cell

s, nm$^{-1}$
Sample and buffer

3.2 mg/ml lysozyme
Background subtraction

Solution minus Solvent

$I(s)$

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

Solution minus Solvent

Normalization against:
- Concentration
Logarithmic plot

Log \( I(s) \)

\( s, \text{ nm}^{-1} \)
Dilution series

Log I(s)

2 mg/ml
Dilution series

8 mg/ml
Dilution series

32 mg/ml
Dilution series

Log $I(s)$

2 mg/ml
32 mg/ml

$s, \text{nm}^{-1}$
Inter-particle interactions

No interactions
Inter-particle interactions

Attractive interactions

Repulsive interactions
Log $I(s)$

Merging data
Merging data
Merging data

Log $I(s)$ vs. $s$, nm$^{-1}$
Data analysis
Radius of gyration ($R_g$)

Definition

Measure for the overall size of a macromolecule

Average of square center-of-mass distances in the molecule weighted by the scattering length density
Radius of gyration ($R_g$)

- Sphere: 2.2 nm
- Cylinder: 3.6 nm
- Oblate spheroid: 3.4 nm
- Rod: 6.4 nm
- Dumbbell: 4.8 nm

Volume: 100 nm$^3$
Radius of gyration ($R_g$)

Guinier approximation:

$$I(s) \approx I(0) \exp(s^2 R_g^2 / -3)$$

$$s \lesssim 1/R_g$$

André Guinier 1911-2000
Radius of gyration ($R_g$)

Guinier plot

$\ln I(s) \approx -\frac{1}{2} \frac{R_g^2}{s^2}$
Radius of gyration \((R_g)\)

*Guinier plot*
Radius of gyration ($R_g$)

Guinier plot

$\ln I(s) = a \cdot s + b$

$R_g = \sqrt{-3a}$

$sR_g < 1.0\sim1.3$
Radius of gyration ($R_g$)

**Guinier plot**

$\ln I(s)$ vs $s^2$

- $R_g \pm \text{stdev}$
- Forward scattering $I(0)$
- Data quality
- Data range
Sample quality

Log I(s)

s, 1/nm
Aggregation

Monodisperse sample
Aggregation

Aggregated sample
Logarithmic plot

Log $I(s)$
Guinier plot

$\ln I(s)$

$s^2$
Guinier plot

\[ \ln I(s) \]

\[ s^2 \]

\[ R_g = 2.0 \text{ nm} \]

\[ s_{\text{min}} R_g = 0.52 \]

\[ s_{\text{max}} R_g = 1.26 < 1.3 \]
Guinier plot

\[ \text{Ln } I(s) \]

\[ R_g = 2.3 \text{ nm} \]

\[ s_{\text{min}} R_g = 1.01 \]

\[ s_{\text{max}} R_g = 1.45 > 1.3 \]
Molecular mass

Guinier approximation

Log $I(s)$, a.u.

$\log I(0)_{\text{apo}}$

$\log I(0)_{\text{lys}}$

$\text{lysozyme}$

$\text{apoferitin}$
I(0) and Molecular Mass

\[
\frac{\text{MM}_{\text{sample}}}{\text{MM}_{\text{BSA}}} = \frac{I(0)_{\text{sample}}}{I(0)_{\text{BSA}}}
\]

\[
\text{MM}_{\text{sample}} = I(0)_{\text{sample}} \cdot \frac{\text{MM}_{\text{BSA}}}{I(0)_{\text{BSA}}}
\]

**BSA**
- \(R_g = 3.1\) nm
- \(I(0) = 11.7\) a.u.
- \(\text{MM}_{\text{BSA}} = 66\) kDa

**Sample**
- \(R_g = 1.46\) nm
- \(I(0) = 2.68\) a.u.
- \(\text{MM} = 15.1\) kDa

**Other**
- \(R_g = 6.81\) nm
- \(I(0) = 79.45\) a.u.
- \(\text{MM} = 450\) kDa
Porod volume

Excluded volume of the hydrated particle

\[ V_P = \frac{2\pi^2 I(0)}{\int_0^\infty I(s)s^2 ds} \]
Porod volume

*Excluded volume of the hydrated particle*

\[
V_P = \frac{2\pi^2 I(0)}{\int_{0}^{\infty} [I(s) - K_4] s^2 ds}
\]

\(K_4\) is a constant determined to ensure the asymptotical intensity decay proportional to \(s^{-4}\) at higher angles following the Porod's law for homogeneous particles.
Porod law
Excluded volume of the hydrated particle

21 nm$^3$
~13 kDa

974 nm$^3$
~610 kDa (?)
Distance distribution function
Distance distribution function

\[ \gamma(r) \]

\[ r, \text{ nm} \]
Distance distribution function
Distance distribution function

\[ p(r) = r^2 \gamma(r) \]
Distance distribution function

\[ p(r) = r^2 Y(r) \]
Distance distribution function

Distance distribution function $p(r)$ with $D_{\text{max}} = 6 \text{ nm}$ and volume $100 \text{ nm}^3$. The graph shows the distribution of distances $r$, with a peak at $D_{\text{max}} = 6 \text{ nm}$.
Distance distribution function

\[ p(r) \]

\[ r, \text{ nm} \]
Distance distribution function

\[ p(r) \]

\[ r, \text{ nm} \]
Distance distribution function

Log $I(s)$

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

$p(r)$

$r, \text{nm}$

Distance distribution function
\[ I(s) = 4\pi \int_{0}^{D_{\text{max}}} p(r) \frac{\sin(sr)}{sr} dr \]
$p(r)$ plot

Distance distribution function

$D_{\text{max}}$
Data quality

\[ s_{\text{min}} \leq \pi / D_{\text{max}} \]
Data range

$s_{\text{min}} < \pi/D_{\text{max}}$
Beamline P12

Data range can be adjusted by changing the wavelength $\lambda$ or the sample-detector distance.
Detector closer to the sample – collect wider angles (for smaller particles)
Beamline P12

Detector further from the sample – collect smaller angles (for larger particles)
Data reduction and analysis steps

Radial averaging

Normalization

Radiation damage check

Background subtraction

Merge multiple concentrations

\( R_g, \) molecular weight

\( D_{\text{max}}, p(r) \)

Porod volume

\textit{Ab initio} shape determination

\( p(r) \)