Data reduction and processing

Al Kikhney
Outline

- SAXS experiment setup
- $3D \rightarrow 2D \rightarrow 1D$
- Background subtraction
- High vs. low concentration
- $R_g$, MM
- Volume
- Distance distribution function $p(r)$
Small Angle X-ray Scattering

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

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

1-2 mg purified material
concentration from 0.5 mg/ml,
exposure times: a few seconds/minutes
Small Angle X-ray Scattering

Exposure

beamstop

Beam

lower angle

higher angle

X-ray detector
Small Angle X-ray Scattering

Exposure

beamstop

Beam

X-ray detector
Small Angle X-ray Scattering

Exposure

X-ray detector
Small Angle X-ray Scattering

Radial averaging

Normalization against:
- data collection time,
- transmitted sample intensity
Small Angle X-ray Scattering

Notations and units

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

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

Log \( I(q) \),
a.u.
Small Angle X-ray Scattering

*Notations and units*

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

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

\[ \text{Log } I(s), \quad a.u. \]

\[ s, \ \text{nm}^{-1} \]
Small Angle X-ray Scattering

Notations and units

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

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

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

\(s, \text{Å}^{-1}\)
Small Angle X-ray Scattering

Notations and units

$|s| = 2 \sin \theta / \lambda$

2$\theta$ – scattering angle
$\lambda$ – wavelength
Small Angle X-ray Scattering

Notations and units

$|s| = 4\pi \sin \theta / \lambda$

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

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

Radiation damage

Log I(s), a.u.

sample
same sample again

RADIATION DAMAGE!
Data quality

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

sample
same sample again
average

$s$, nm$^{-1}$
Buffer and sample

Solution and Solvent
Buffer and sample

Solution and Solvent

Looking for protein signals less than 5% above background level…
Background subtraction

Solution minus Solvent

Normalization against:
- concentration
Data quality

“Can I use this data for further analysis?”

Log $I(s)$

lysozyme

AGGREGATED!
Dilution series

Low and High Concentration

Log I(s)

LytA protein, 60 kDa, 1 mg/ml
10 mg/ml
Dilution series

Low and High Concentration

Log I(s)

s, nm\(^{-1}\)
Merging data

Low and High Concentration
Merging data

*Low and High Concentration*

![Graph showing Log I(s) vs. s, nm^{-1}](image-url)
Merging data

Low and High Concentration

Log $I(s)$

$s, \text{ nm}^{-1}$
Data analysis
Shape and size

Log $I(s)$

lysozyme

apoferitin
Crystal vs. solution
Data range
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 ($R_g$)

Guinier plot
Radius of gyration ($R_g$)

*Guinier plot*

$\ln I(s)$

$S^2$
Radius of gyration ($R_g$)

**Guinier plot**

- Estimate of the overall size of the particles

Guinier approximation:

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

$$ sR_g \leq 1.3 $$

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

*Guinier plot*

- Estimate of the overall size of the particles

Guinier approximation:

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

$$sR_g \leq 1.3$$

- Quality of the data
  - aggregation
  - polydispersity
  - improper background substraction

- Zero angle intensity $I(0)$

- First point to use
Radius of gyration ($R_g$)

Log $I(s)$ vs. $s$, 1/nm

Bovine serum albumin (BSA)
Radius of gyration ($R_g$)

Guinier plot
Radius of gyration (R_g)

Guinier plot

\[ y = ax + b \]

\[ R_g = \sqrt{-3a} \]

R_g ± stdev
Forward scattering I(0)
Data quality
Data range
Molecular mass

Guinier approximation

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

Log \( I(0)_{\text{apo}} \)

Log \( I(0)_{\text{lys}} \)

lysozyme

apoferitin

s, nm\(^{-1}\)
$I(0)$ and Molecular Mass

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

$$\text{MM}_{\text{sample}} = I(0)_{\text{sample}} \times \frac{\text{MM}_{\text{BSA}}}{I(0)_{\text{BSA}}}$$

**BSA**

$R_g = 3.1 \text{ nm}$

$I(0) = 11.7$

$\text{MM}_{\text{BSA}} = 66 \text{ kDa}$

$R_g = 1.46 \text{ nm}$

$I(0) = 3.66$

$\text{MM} = 20.6 \text{ kDa}$

$R_g = 6.81 \text{ nm}$

$I(0) = 79.45$

$\text{MM} = 448.2 \text{ kDa}$
Porod law

\[ I(s) \sim s^{-4} \]

Intensity decay is proportional to \( s^{-4} \) at higher angles
(for globular particles of uniform density)
Porod law

\[ I(s) \propto s^4 \]
Porod law

Excluded volume of the hydrated particle

\[
V_P = \frac{2\pi^2 I(0)}{\int_0^\infty \left[ I(s) - K_4 \right] s^2 ds}
\]

\(K_4\) 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 plot

Excluded volume of the hydrated particle

Primus

Porod plot

Excluded volume of the hydrated particle

Primus
Kratky plot

Patterns of globular and flexible proteins

Natively unfolded
Globular
Multidomain with flexible linkers
Distance distribution function

\[ \gamma(r) \]
Distance distribution function

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

\[ p(r) \]
$p(r)$ function

Distance distribution function
\( p(r) \) function

Distance distribution function

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

\( \Leftrightarrow \)

\[
p(r) = \frac{r^2}{2\pi^2} \int_0^\infty s^2 I(s) \frac{\sin(sr)}{sr} \, ds
\]

Indirect Fourier Transform
$p(r)$ plot

Distance distribution function

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

\[ s_{\text{min}} \leq \frac{\pi}{D_{\text{max}}} \]
Yeast bleomycin hydrolase 3GCB

- Compact dimer
- Extended dimer
- 50 kDa Monomer
- Hexamer
$p(r)$ plots

Yeast bleomycin hydrolase 3GCB
Summary

- Exposure 3D → 2D
- Radial averaging → 1D
- Normalization
- Background subtraction
- Analysis

- Log plot
- Guinier plot ($R_g$, MM)
- Porod plot
- Kratky plot (flexibility)
- $p(r)$ plot
Thank you! 

After questions: group photo!

www.saxier.org/forum