Primary data reduction and evaluation of the overall geometrical parameters

Al Kikhney
EMBL Hamburg
Outline

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

Background subtraction
SAXS data from macromolecules in solution
SAXS data from macromolecules in solution

log I(s)

experimental SAXS pattern
calculated from model

nm⁻¹
SAXS data from macromolecules in solution

experimental SAXS pattern

calculated from model

log I(s) vs. nm$^{-1}$
Notations and units

$\text{X-ray} \rightarrow 2\theta$
Notations and units

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

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

\[ I(s), \text{ a.u.} \]
Exposure time

0.05 second
0.2 second
Exposure time

0.05 second
0.2  second
0.8  second
Multiple exposures
Multiple exposures

frame 1
frame 2
Multiple exposures

average
Sample and buffer

$I(s)$

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

$I(s)$

$s, \text{ nm}^{-1}$
3.2 mg/ml lysozyme + buffer + cell

Sample and buffer

I(s)

s, nm⁻¹
Background subtraction

Solution minus Solvent

I(s)

3.2 mg/ml lysozyme

s, nm\(^{-1}\)
Background subtraction

Solution minus Solvent

Normalization against:
- Concentration
Logarithmic plot

Log $I(s)$ vs. $s$, nm$^{-1}$
Dilution series

2 mg/ml
Dilution series

8 mg/ml
Dilution series

Log I(s)

32 mg/ml
Dilution series

2 mg/ml

32 mg/ml

Log I(s)
Inter-particle interactions

No interactions
Inter-particle interactions

Attractive interactions

Repulsive interactions
Merging data
Merging data

Log $I(s)$

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

Log $I(s)$

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

\[ \sim 6 \text{ nm} \quad 100 \text{ nm}^3 \]
Shape

Log I(s)

1dqi
58 kDa

3oux
58 kDa

3odt
66 kDa

3zx6
63 kDa
Size

Log I(s)

200 nm³

100 nm³

50 nm³

25 nm³
Radius of gyration

$R_g^2$ 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 ($R_g$)

$R_g$: 3.0 nm

$R_g$: 4.7 nm

Myomesin-1 My12-My13, SASBDB: SASDak5
Radius of gyration ($R_g$)
Radius of gyration ($R_g$)

Guinier approximation:

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

$$s \lesssim 1/R_g$$

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

Guinier plot

$\ln I(s)$ vs. $s^2$
Radius of gyration ($R_g$)

Guinier plot

\[ \ln I(s) \]

\[ s^2 \]
Radius of gyration \((R_g)\)

**Guinier plot**

\[ y = ax + b \]

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

\[ sR_g < 1 \]
Radius of gyration ($R_g$)

*Guinier plot*

$\ln I(s)$

$R_g \pm \text{stdev}$

Forward scattering $I(0)$

Data quality

Data range
Aggregation
Aggregation

Monodisperse sample
Aggregation

Aggregated sample
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

\[ \ln I(s) \]

- \( R_g = 2.3 \text{ nm} \)
- \( s_{\min}R_g = 1.01 \)
- \( s_{\max}R_g = 1.45 > 1.3 \)

\( 0.44 \text{ nm}^{-1} \)

\( 0.63 \text{ nm}^{-1} \)
\(I(0)\) and Molecular Mass

Assuming \(I(0)\) is normalized against concentration (mg/ml)

- \(I(s)\) on absolute scale (cm\(^{-1}\))

\[
\text{MM}_{\text{sample}} = 10^3 I(0)_{\text{sample}} N_A / (\Delta \rho v_{\text{sample}})^2
\]

\(N_A\) – Avogadro’s number
\(\Delta \rho\) – contrast in cm\(^{-2}\)
\(v_{\text{sample}}\) – partial specific volume in cm\(^3\)/g

- \(I(s)\) on relative scale (a.u.)

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

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

Excluded volume of the hydrated particle

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

For proteins: \( \text{MW [kDa]} \sim \frac{V_p [\text{nm}^3]}{1.6} \)
Porod volume

Excluded volume of the hydrated particle

69.2 nm$^3$

69.5 nm$^3$

~43 kDa

~43 kDa

*Simulated data*
Porod volume

Excluded volume of the hydrated particle

23.7 nm$^3$

770 nm$^3$

$\sim$14.8 kDa

$\sim$480 kDa
**SAXS MoW**

**Determination of protein MW from a SAXS measurement on a relative scale**


V. Piiadov et al. (2019) *Protein Sci.* 28(2), 454–463

DATMOW

saxs.ifsc.usp.br
Volume-of-correlation

Determination of the molecular mass of proteins or RNA ranging from 10 to 1000 kDa

Distance distribution function

\[ \gamma(r) \]

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

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

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

\[ p(r) \]

\[ D_{\text{max}} = 6 \text{ nm} \]
Maximum intra-particle distance $D_{\text{max}}$
Distance distribution function

$p(r)$

$r, \text{nm}$

Diagram showing a probability distribution function $p(r)$ with a peak at a certain distance in nanometers (nm).
Distance distribution function
Distance distribution function

Log $I(s)$

$p(r)$

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

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

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

*Distance distribution function*

![Graphs of $p(r)$ showing $D_{\text{max}}$ and $r, \text{nm}$ on the x-axis.](image)
Data quality

$s_{min} < \frac{\pi}{D_{max}}$
R_g and I(0) from p(r)

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

\[ R_g^2 = \frac{\int_0^{D_{max}} r^2 p(r) dr}{2 \int_0^{D_{max}} p(r) dr} \]
Data reduction and analysis steps

1. Radial averaging
2. Normalization
3. Radiation damage check
4. Background subtraction
5. Merge multiple concentrations
6. $R_g$, molecular weight
7. $D_{max}$, $p(r)$
8. Porod volume
9. Ab initio shape determination
Thank you!
Structural and biophysical methods for biological macromolecules in solution

14 – 20 October 2019 | Santiago, Chile