CHESS
CORNELL HIGH ENERGY SYNCHROTRON SOURCE

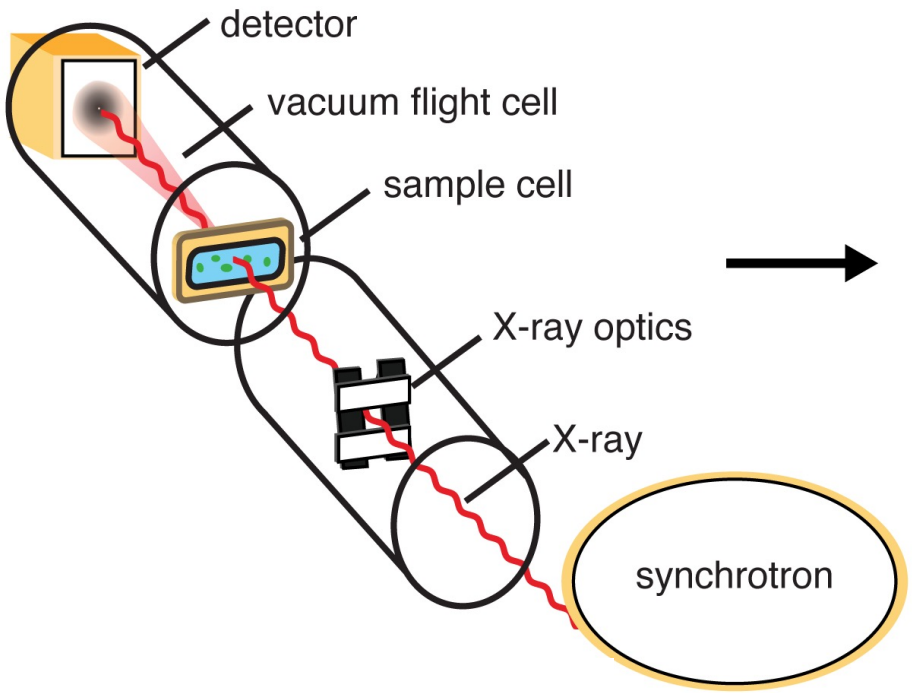
Extreme Biophysics RCN

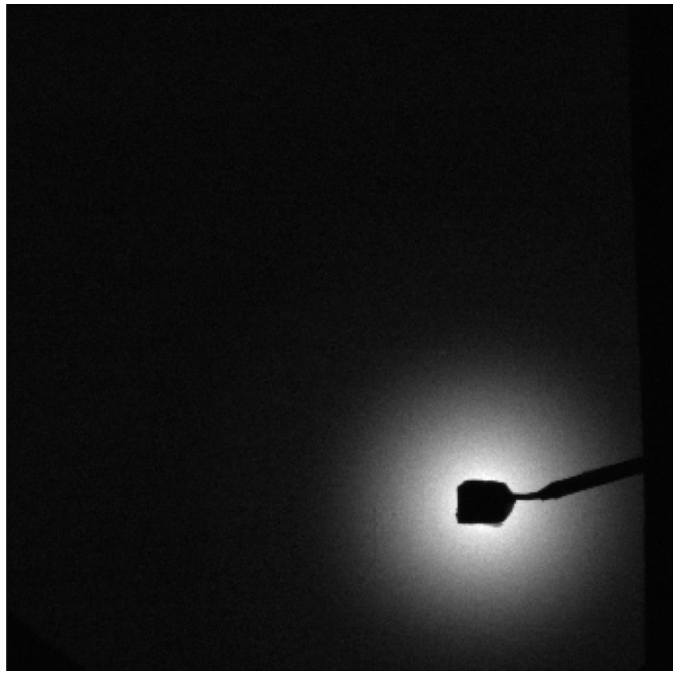
SAXS Fundamentals

Basic SAXS theory, experiments, and data interpretation

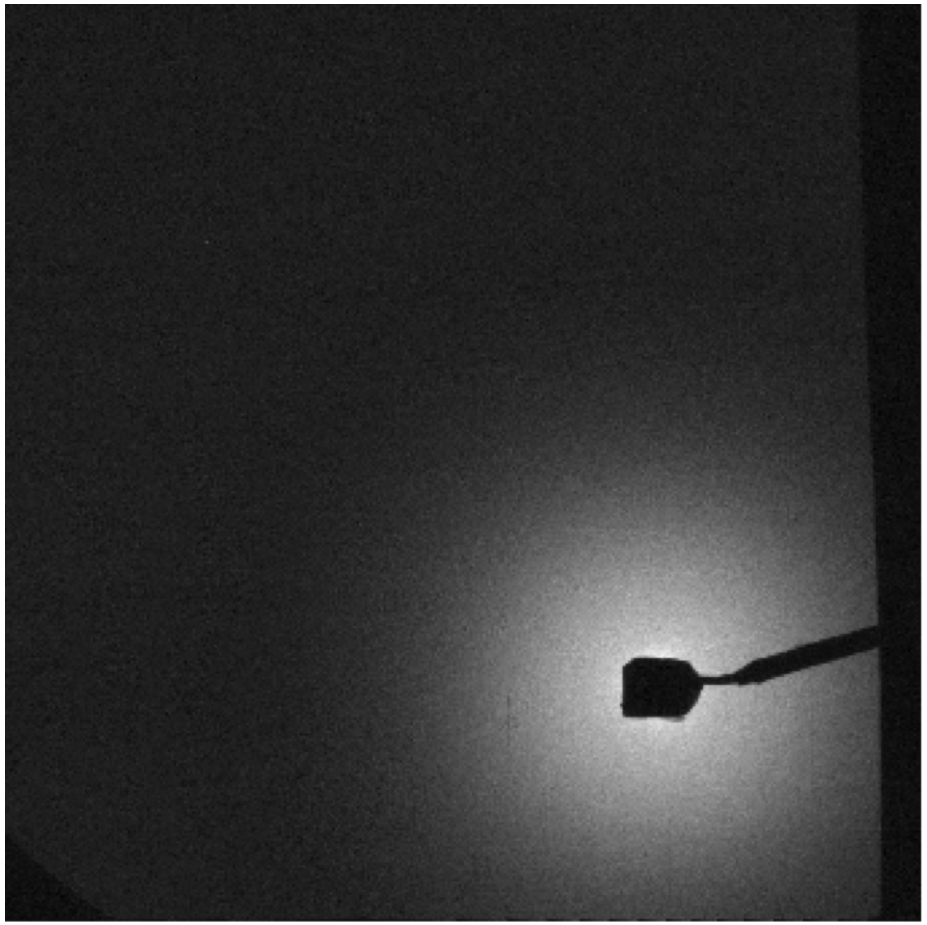
Will Thomas — Ando Lab @ Cornell — CHESS HP Bio Workshop 2021 — April 29 2021

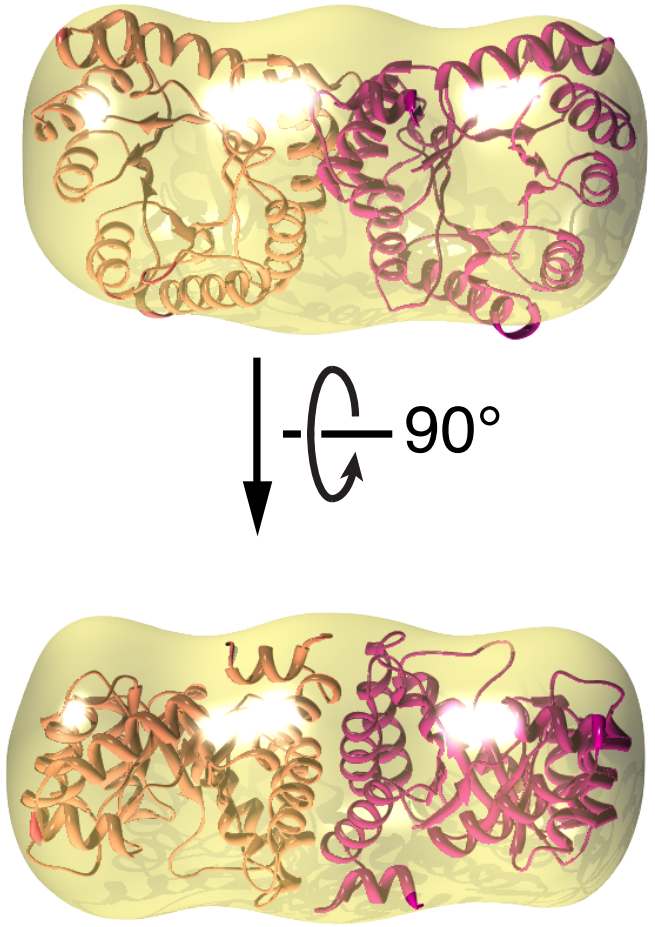
Beamline setup





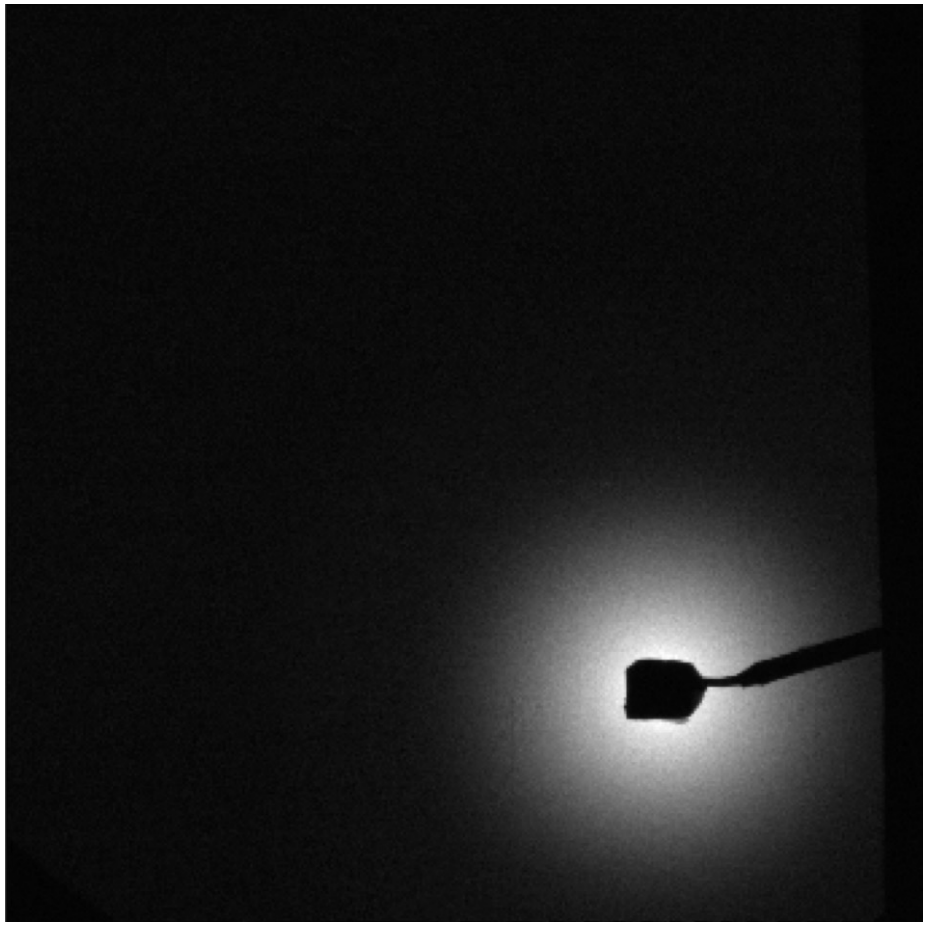
X-ray scattering of proteins

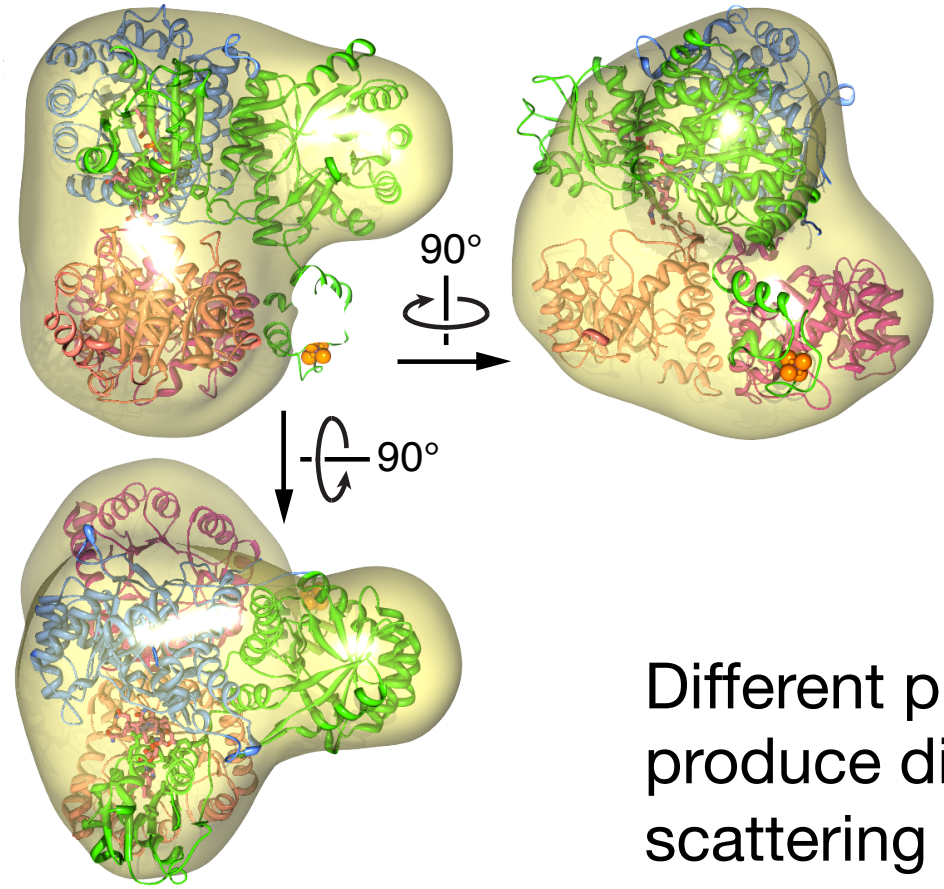




Ando, et al. (2012). JACS, 134(43), 17945–17954.

X-ray scattering of proteins





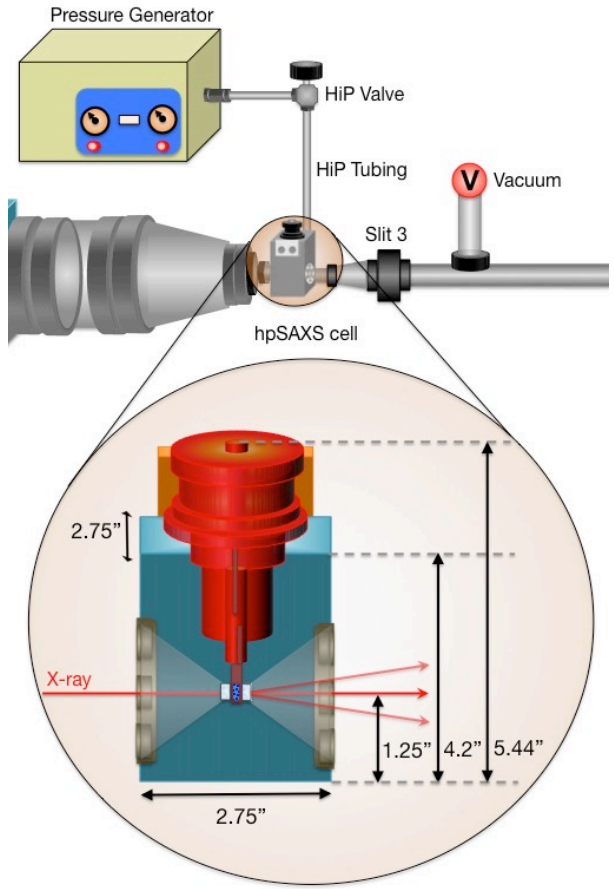
Different proteins produce different scattering

Ando, et al. (2012). JACS, 134(43), 17945–17954.

SAXS can be adapted to many types of experiments

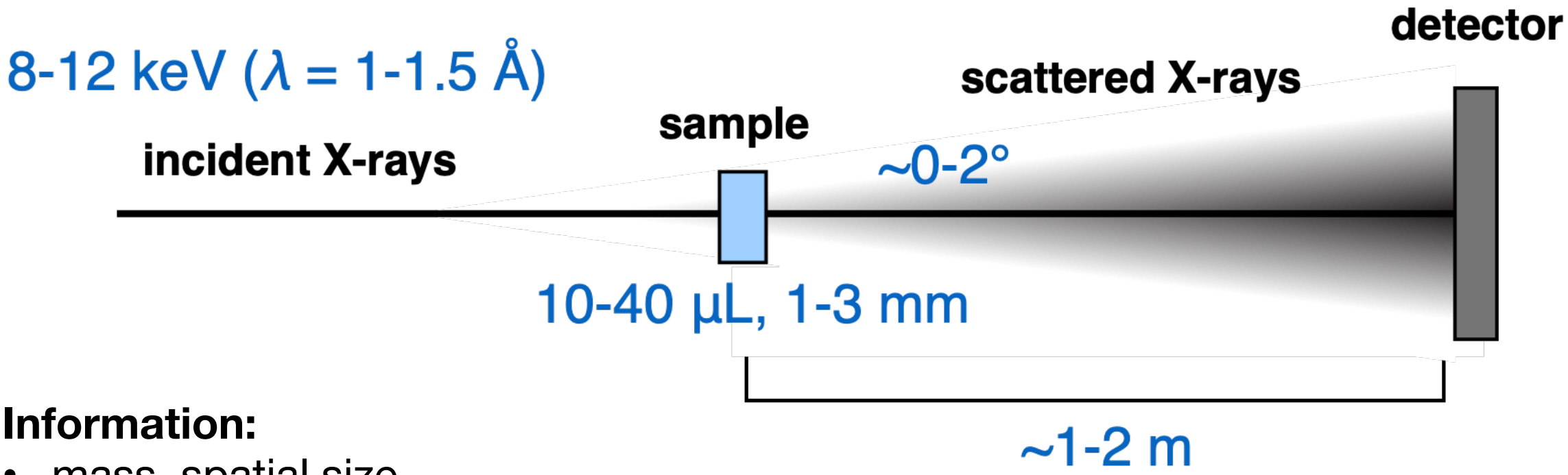
Examples of sample cells:

- Static cell
- Oscillating “flow” cell
- Continuous flow
- Co-flow
- Chromatography-coupled continuous flow
- High-pressure cell



Rai, D. et. Al. (2021). J. Appl. Cryst. 54(1).

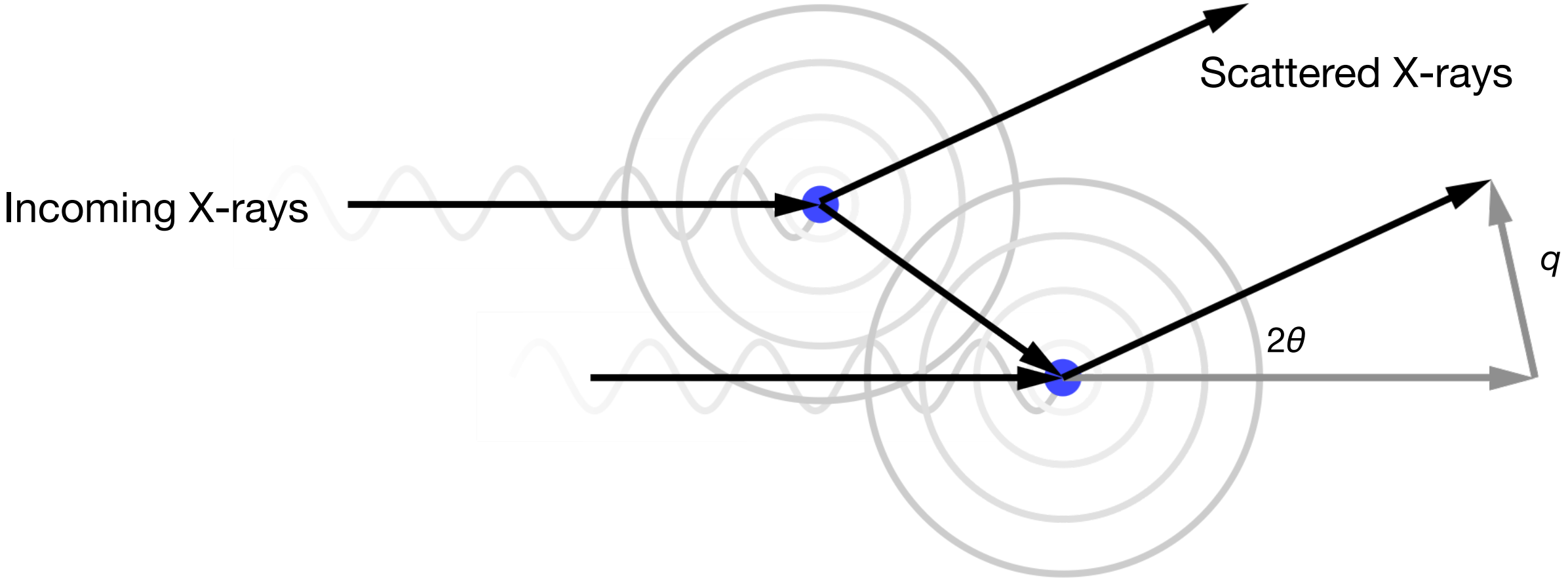
SAXS probes nanometer length scales



Information:

- mass, spatial size
- foldedness, packing, flexibility
- overall shape
- stoichiometry and shape of complexes
- conformational changes as functions of time, ligands, solution conditions

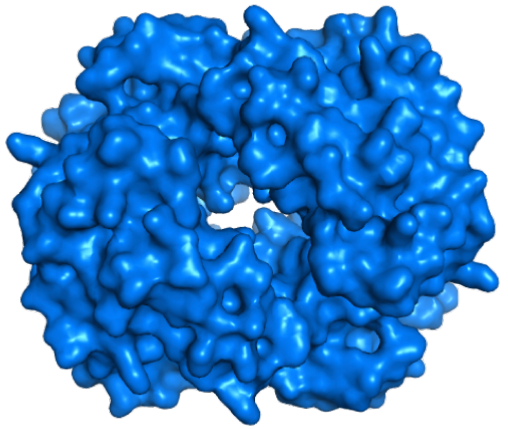
SAXS signal is the interference of scattering from all electrons



Momentum transfer vector: $q = (4\pi/\lambda)\sin\theta$

Scattering from a protein

incident X-ray



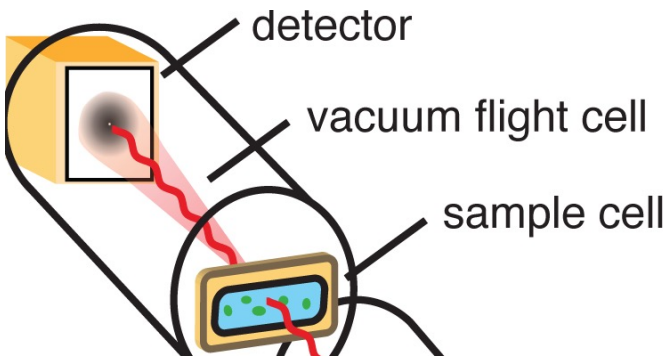
- The sum of the scattering from all electrons in a protein is mathematically equivalent to the Fourier Transform of its electron density, $\rho(r)$. In other words, scattering is directly related to **protein structure**.

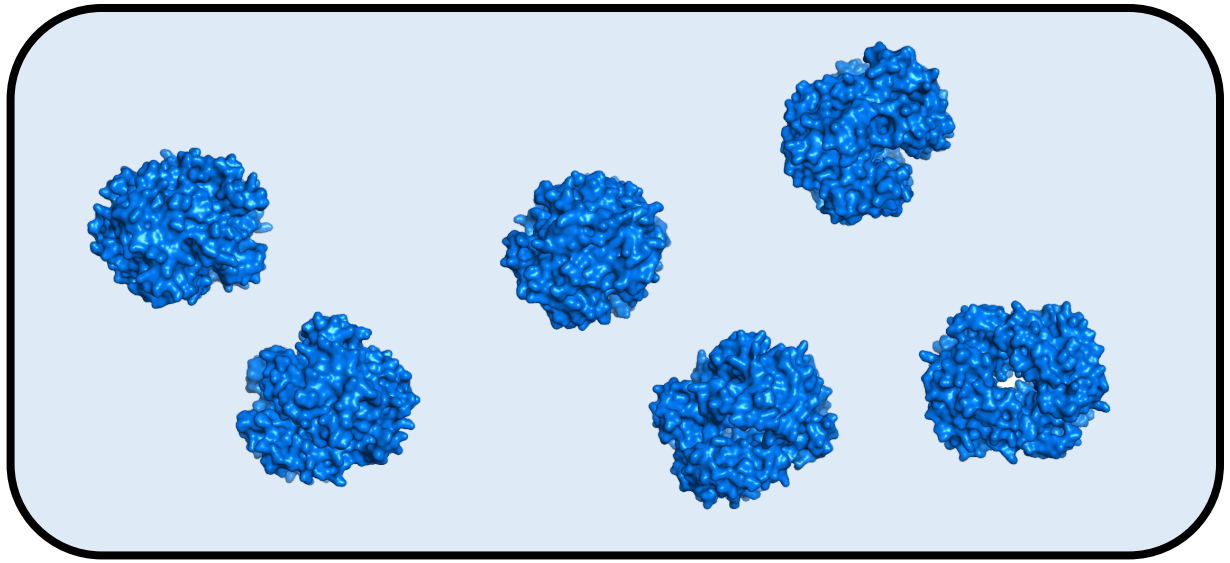
- Amplitude of the scattered wave:

Fourier Transform!

$$\text{Amplitude} = \int \rho(r) e^{i\boldsymbol{q}\cdot r}\, dV$$

Scattering of proteins in solution





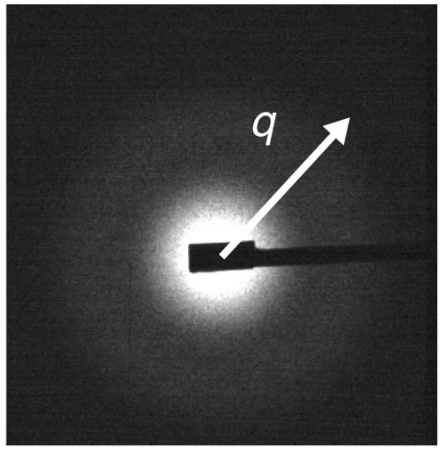
Scattering from *N* identical proteins under dilute conditions:

Scattering intensity $\rightarrow I(q) \propto N < \left|Amplitude_{particle}\right|^2 >_{\Omega}$

Therefore: SAXS gives info about rotational average of single protein **for a homogeneous sample**

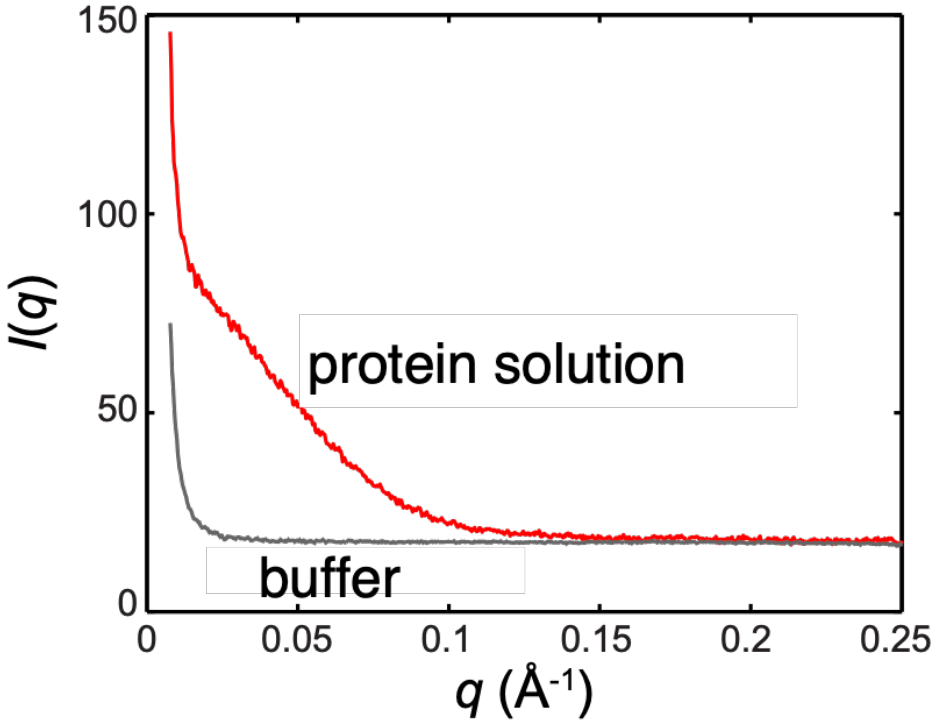
SAXS integration and buffer subtraction

scattering image



Integration

$q = (4\pi/\lambda) \sin\theta$

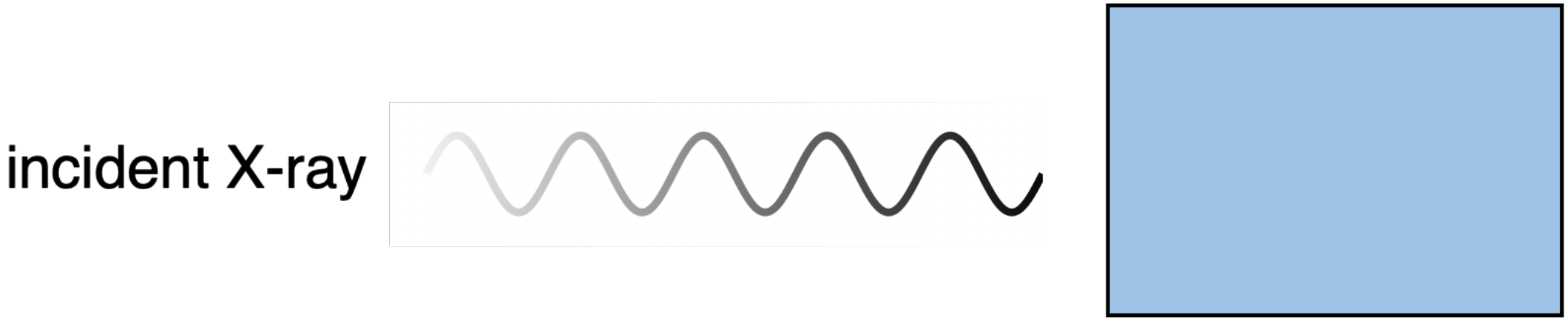


Integrate and normalize based on beamstop diode readings

$I_{protein} = I_{solution} - I_{buffer}$

Important:
Buffer must be matched **EXACTLY**, and the profiles must be accurately normalized for changes in X-ray intensity.

Scattering from a uniform material (e.g. buffer)



Buffer scatters some and absorbs X-rays. Affects contrast and needs to be subtracted!

Contrast is small: ρ_{water} = 0.334 electrons/Å^3 and $\rho_{protein}$ ~ 0.44 electrons/Å^3

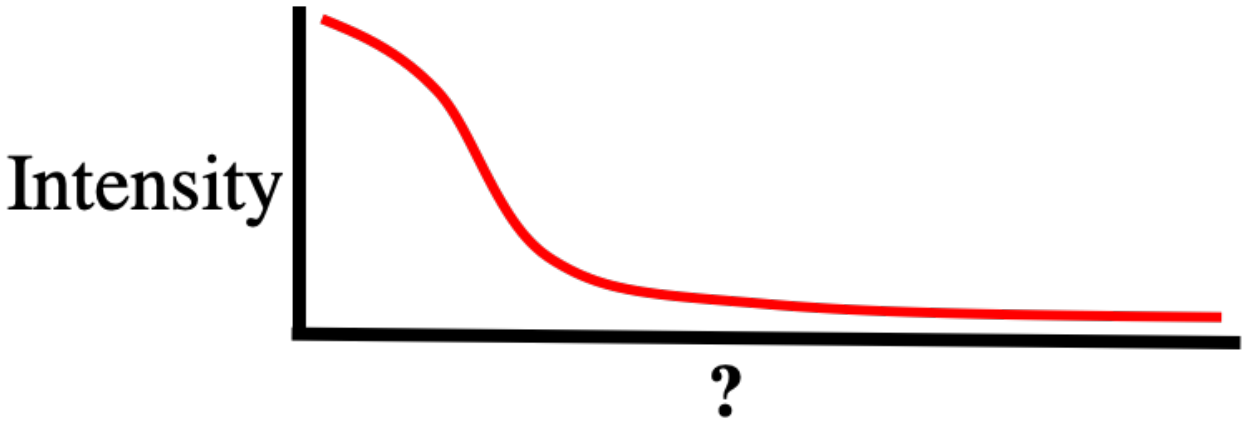
At standard temperature and pressure!

A word on sample prep

- Exactly matched buffers
- Perform buffer exchange with size exclusion chromatography (SEC) or dialysis and save excess exact buffer
- Perform SEC-SAXS at beamline

- Protein quality
- Optimize sample purity, stability, and conformational heterogeneity beforehand if possible
- (e.g. SDS-PAGE, chromatography, dynamic light scattering, multiangle light scattering)

Be careful of *x*-axis definitions and units!

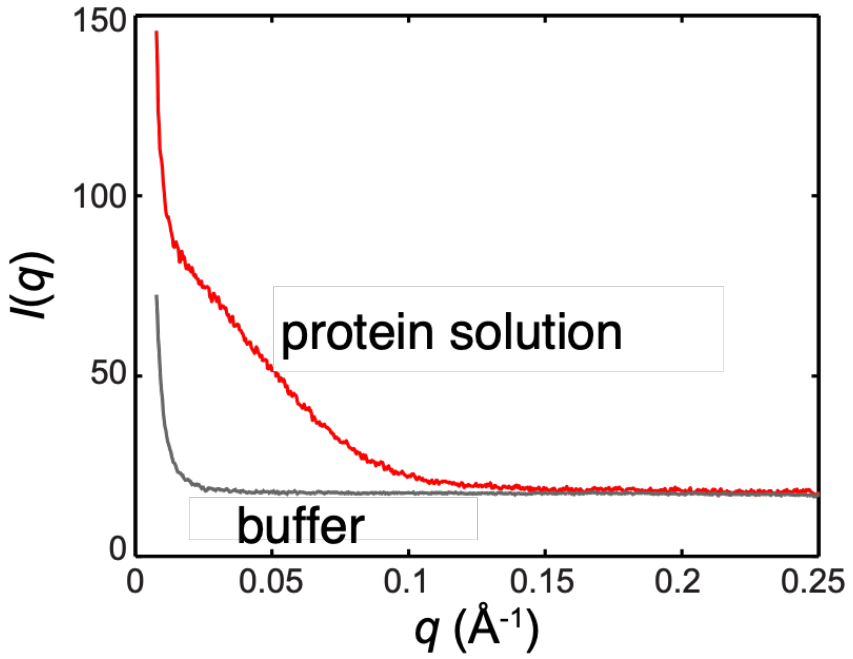


- In SAXS, this variable is sometimes called momentum transfer or scattering vector. Can have units of $Å^{-1}$ (biological) or sometimes nm^{-1} (materials).

- Usually...
- $q = (4\pi/\lambda)\sin\theta$
- $s = (2/\lambda)\sin\theta$

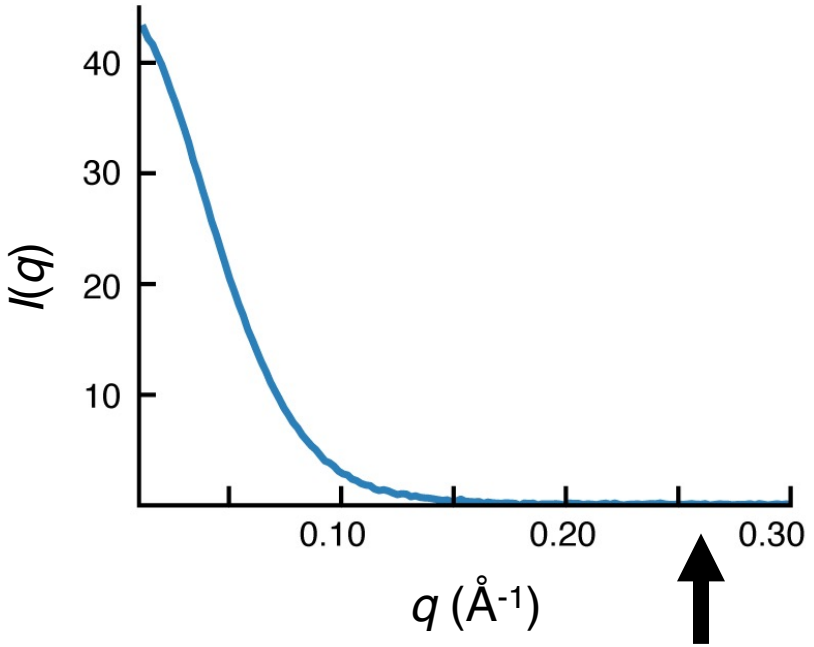
Always check that units in analysis programs match your input! Always define in methods!

Subtracted SAXS profile



Buffer subtraction

Background subtracted profile



Integrate and normalize based on beamstop diode readings

Approaches zero

$I_{protein} = I_{solution} - I_{buffer}$

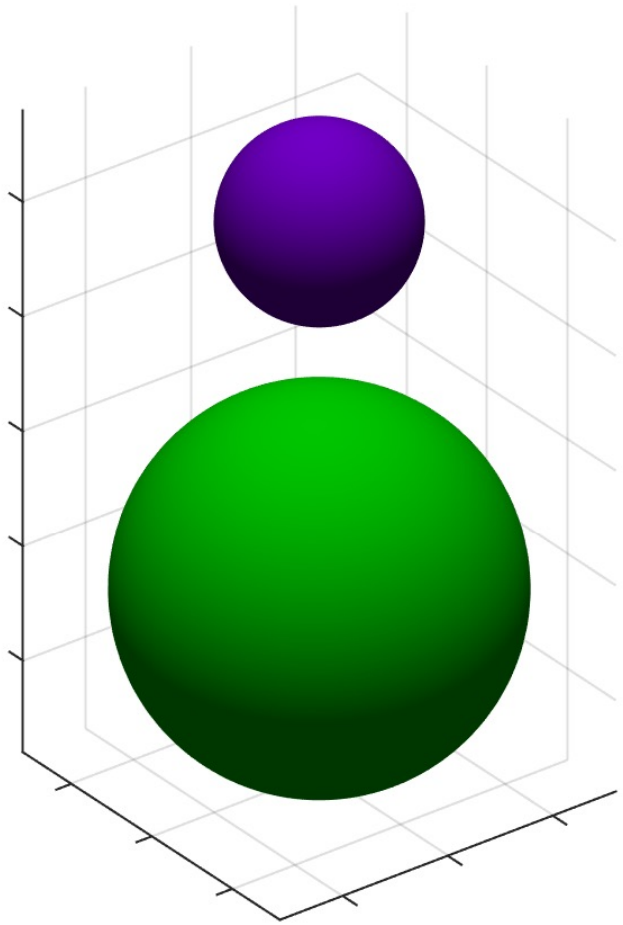
Background-subtracted intensity

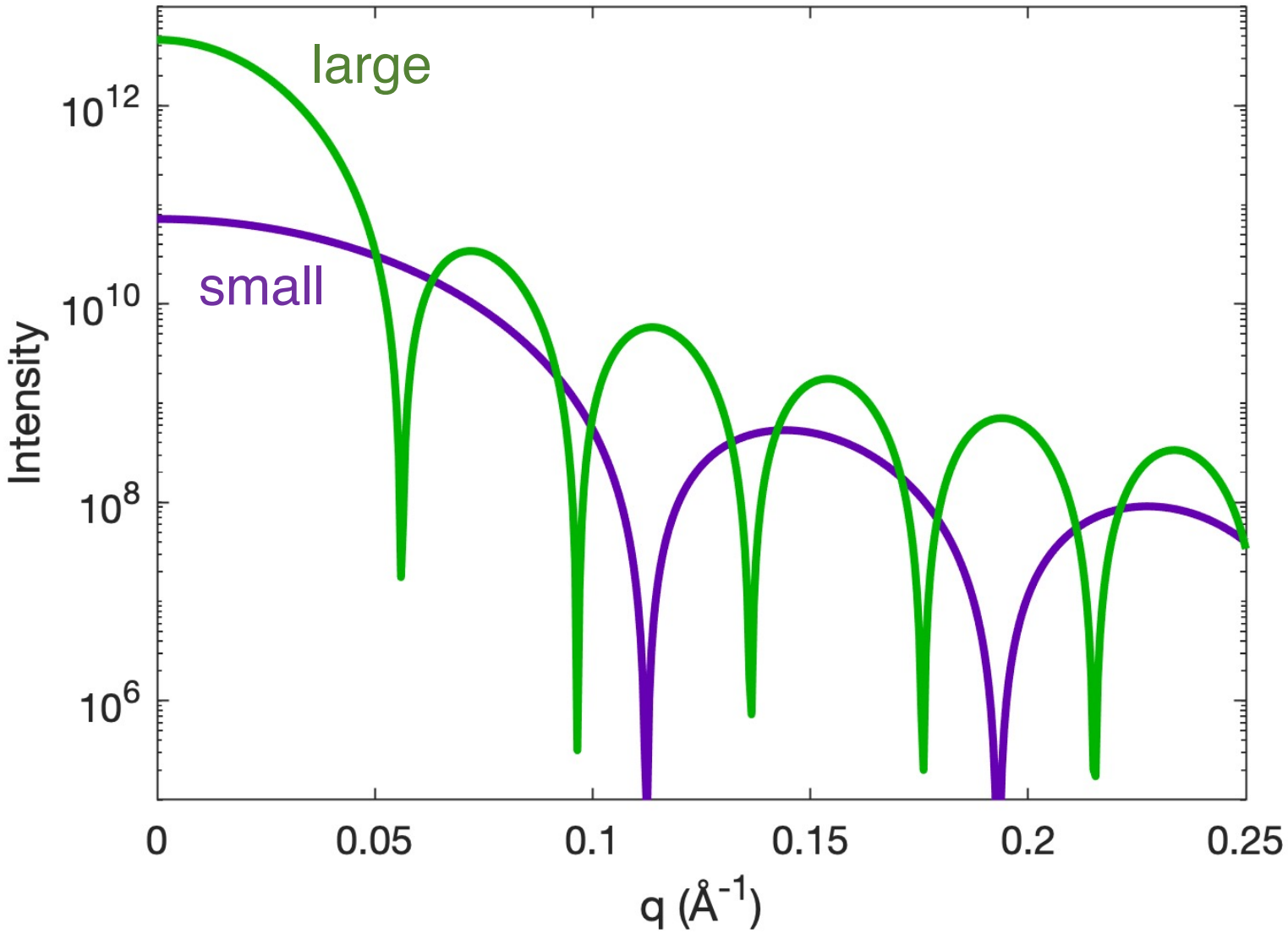
$$I(q) \approx S(q)F(q)$$

- $F(q)$ = form factor (scattering from single protein, rotationally averaged)
- $S(q)$ = structure factor. Describes interparticle interactions [Note: this is not the same as “structure factor” in crystallography]
- Ideally and in dilute solutions, $S(q) \approx 1$

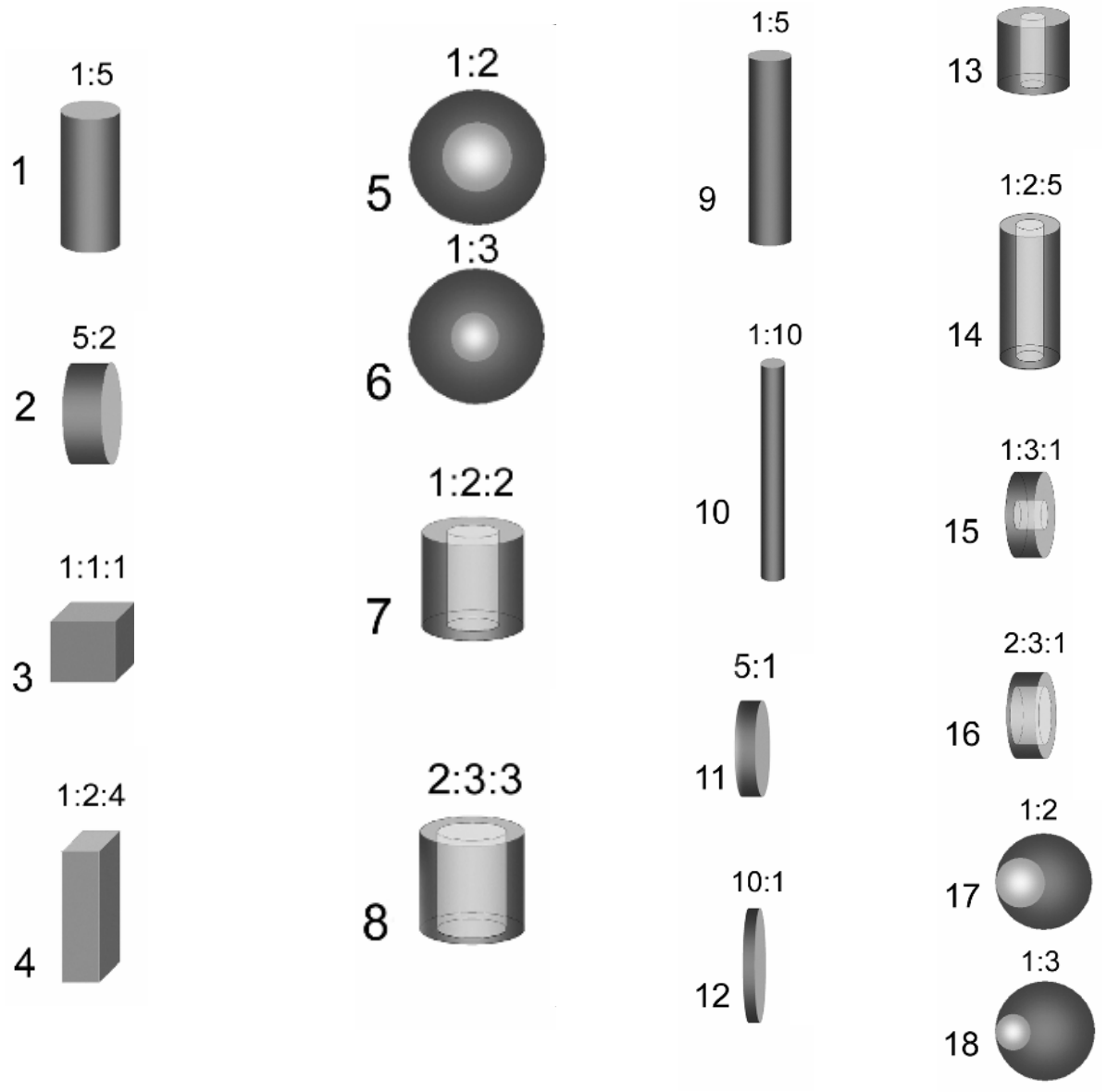
Stuhrmann, H. B. (1980). Synchrotron Radiation Research, edited by H. Winick, Doniach, S., pp. 513-531. New York: Plenum Press.

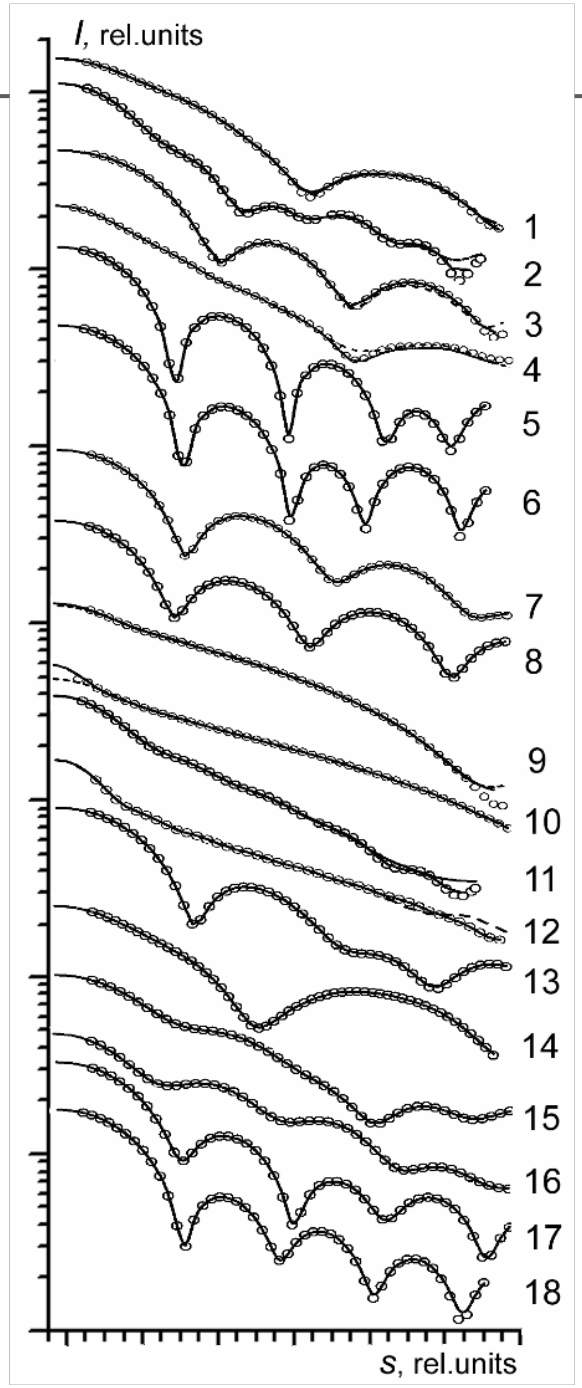
Form factor is related to a protein's shape and size





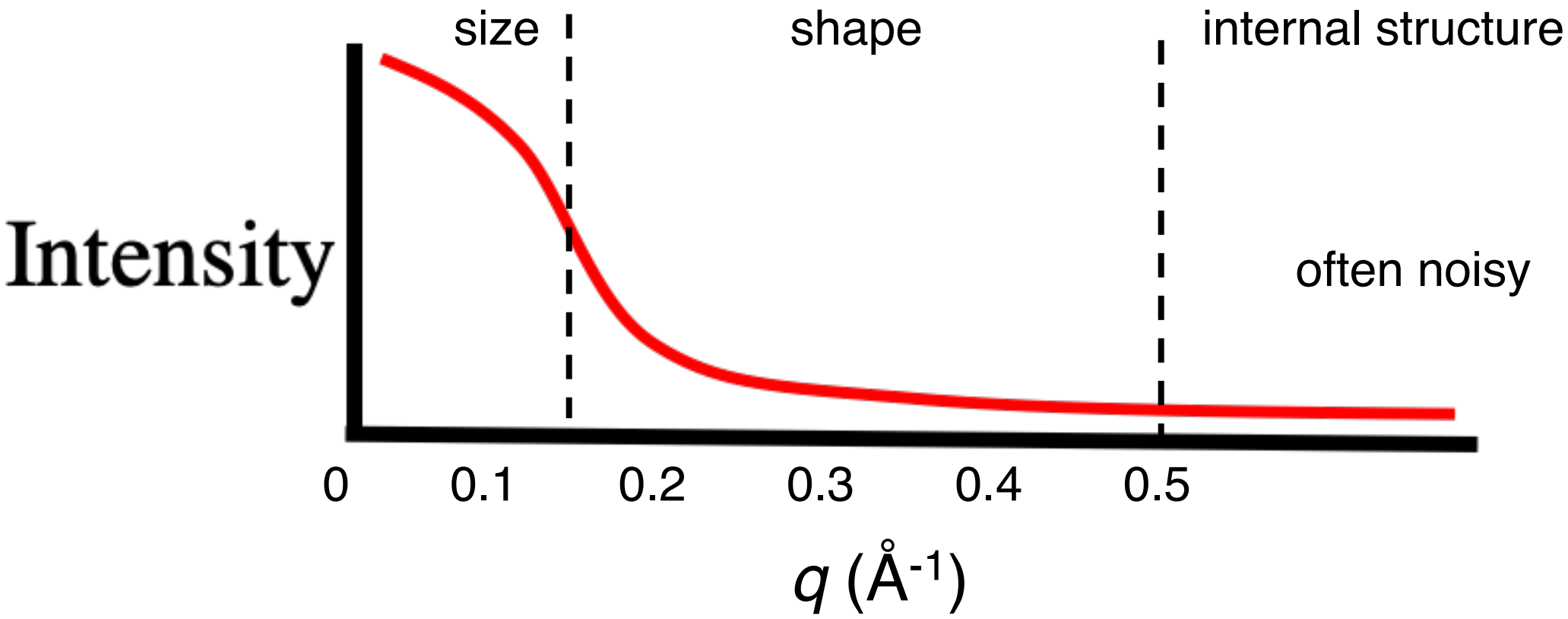
Form factors of different shapes





Volkov & Svergun (2003). J. Appl. Cryst. 36(3), 860–864.

SAXS profile regions

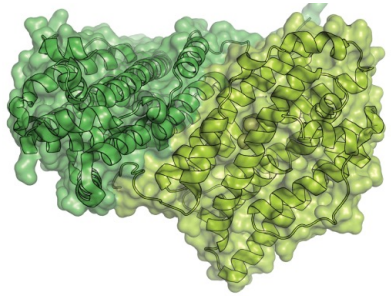


What SAXS can teach us about a protein

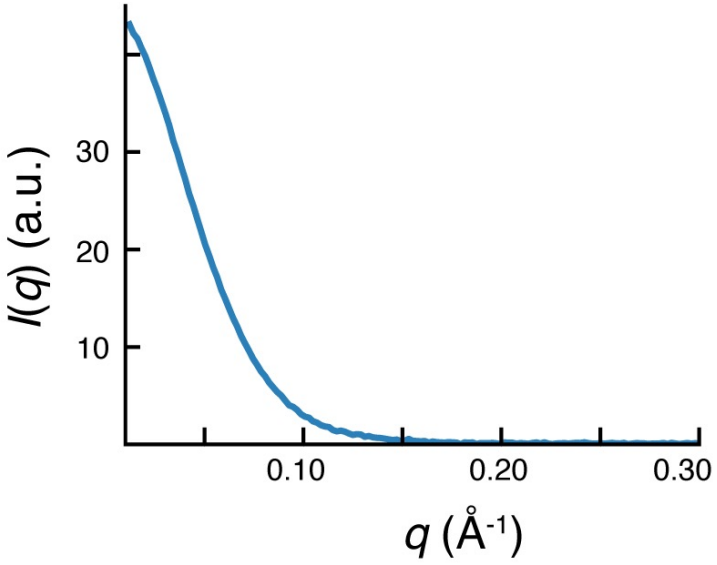
Sample characteristic	Parameter to study
Purity/monodispersity	Guinier plot
Oligomeric state	Volume and molecular weight
Shape/anisometry	Envelope, R_g, $P(r)$
Flexibility (folded/non-folded)	Kratky plot, Porod exponent
Interparticle interactions	Guinier, Concentration series
Conformation	Overall scattering profile

Different ways to inspect data by eye

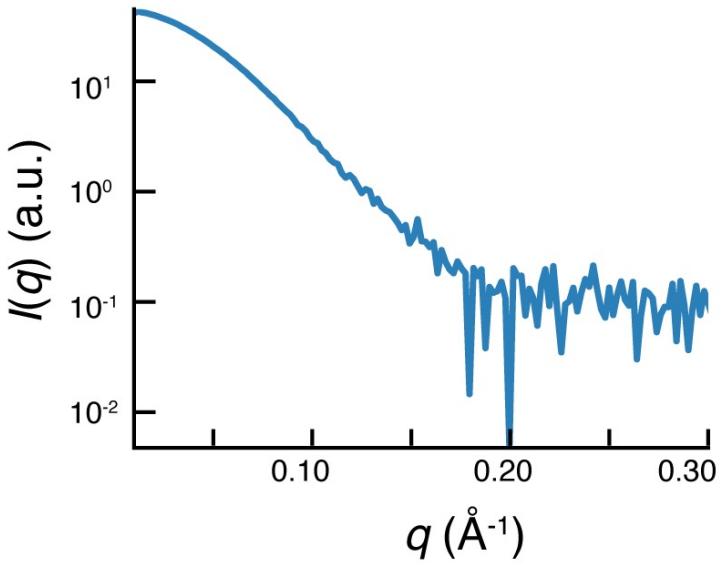
Protein sample (NrdF)



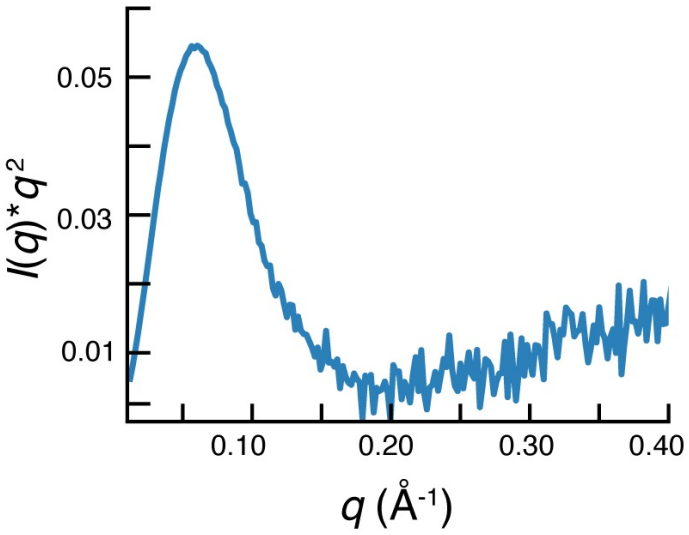
Subtracted profile



Semilog plot

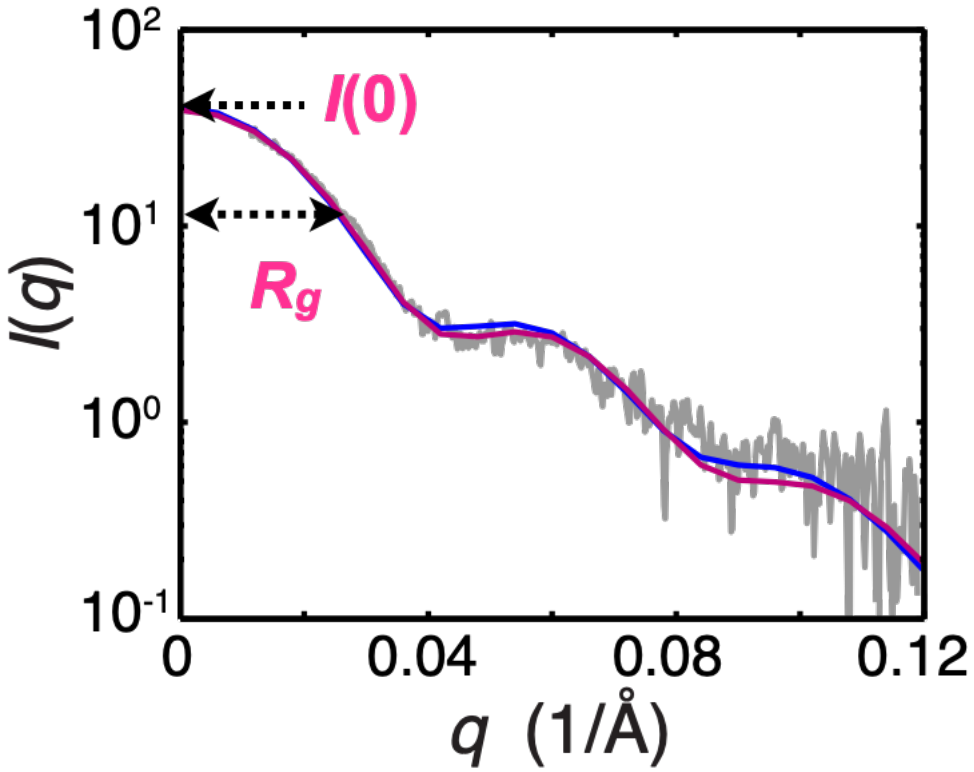


Kratky plot



Guinier analysis

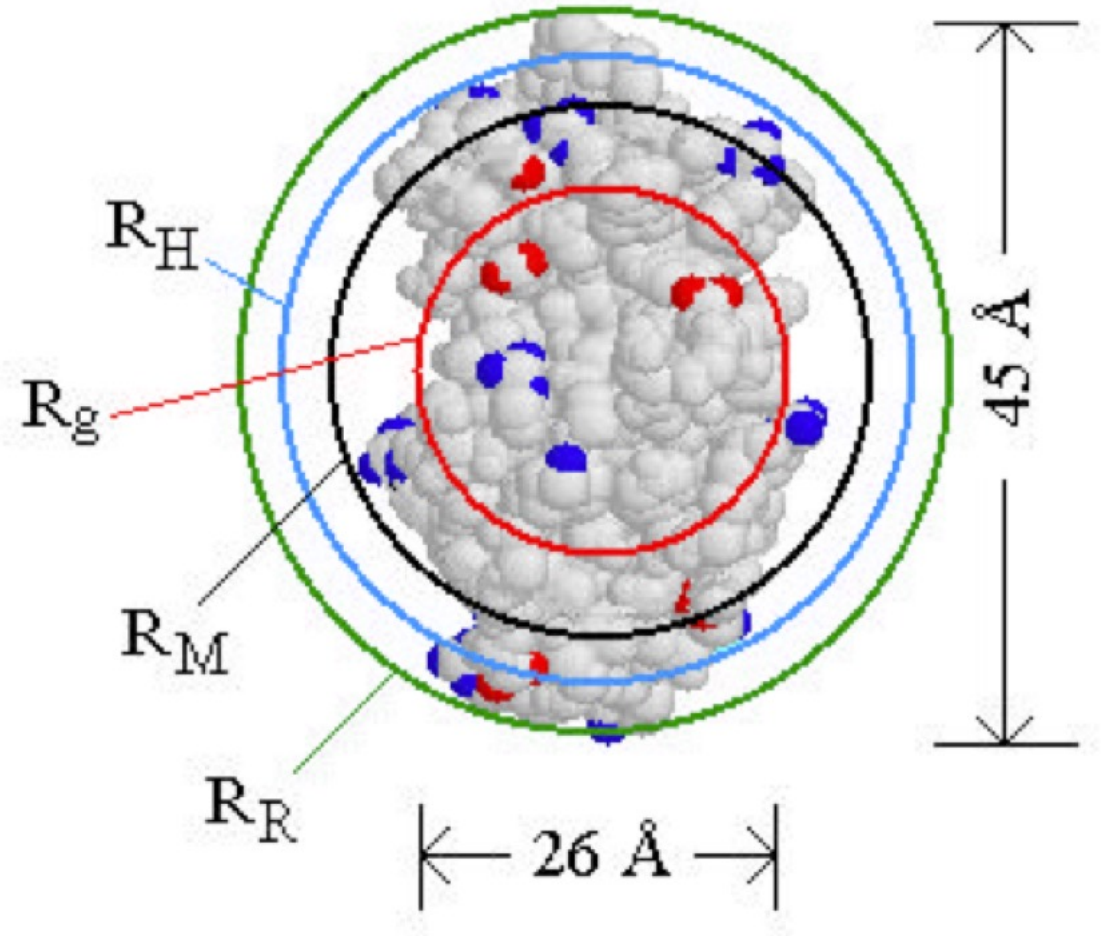
Can be used to estimate R_g and $I(0)$, which tell us about protein shape and size



$I(0)$: forward scattering
R_g: Radius of gyration

Ando, *et al.* (2011) PNAS.

What is R_g?



Useful definitions of R_g:

Radius of gyration is the root mean square distance of a particle from the center of its electron density

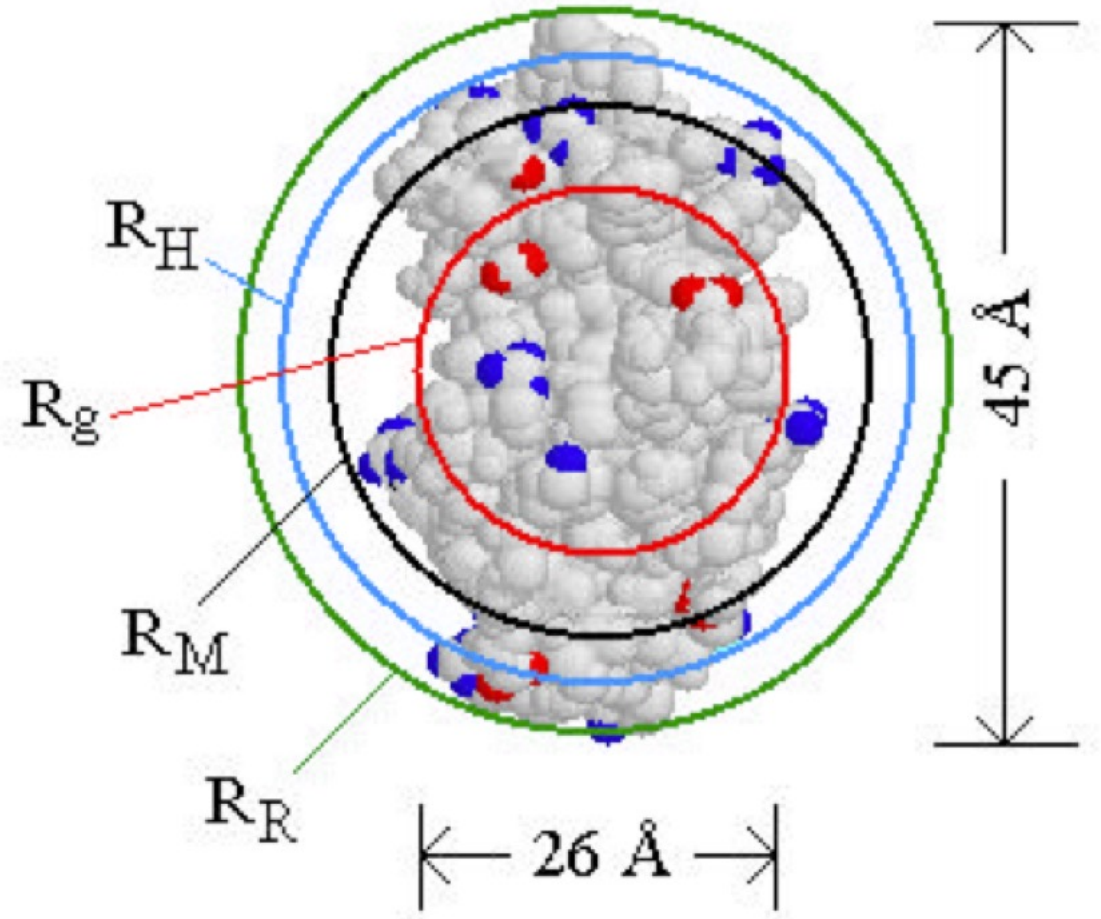
$$R_g^2 = \frac{1}{N}\sum \left\| \vec{r}_i - \overrightarrow{r_{COM}} \right\|^2 \qquad \text{by atoms}$$

$$R_g^2 = \int_V \rho(r) r^2 \, dr \Big/ \int_V \rho(r) \, dr \qquad \text{by electron density}$$

R_g – radius of gyration
R_M – radius of mass-equivalent sphere
R_H – hydrodynamic radius (not always > R_g)
R_R – maximum hard sphere radius

Slide adapted from Richard Gillilan
https://www.silver-colloids.com/Papers/hydrodynamic-radius.pdf

What is R_g?



Experimental lysozyme parameters

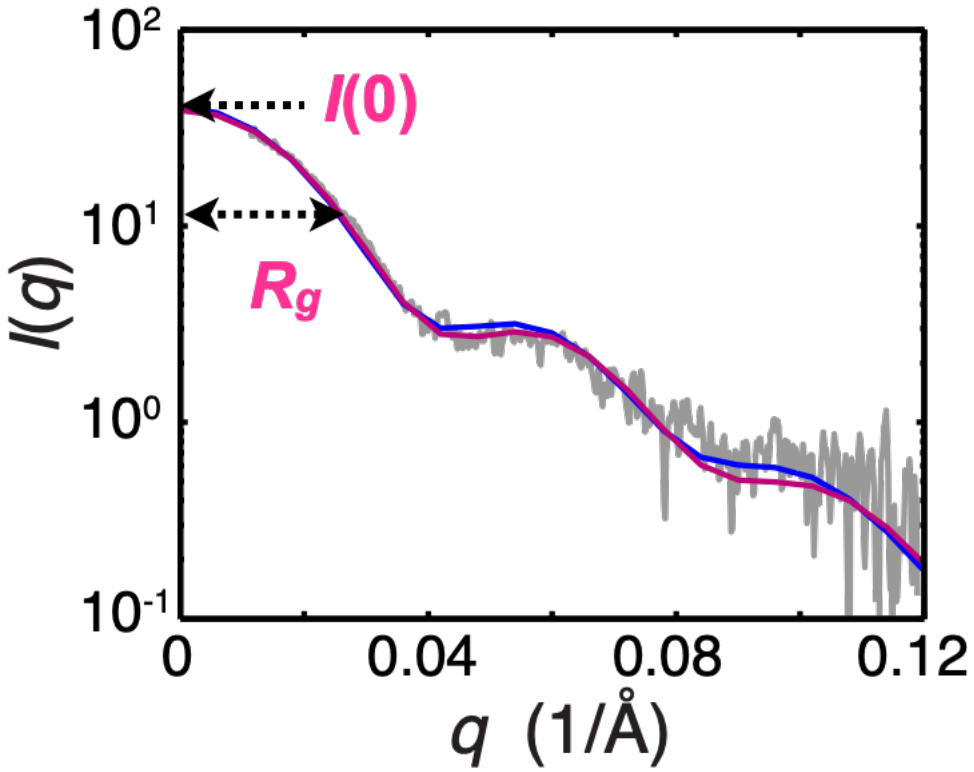
MW = 14.4 kDa
R_R = 23 Å
R_g = 15 Å

R_g (pressure unfolded) = 32 Å

Slide adapted from Richard Gillilan
https://www.silver-colloids.com/Papers/hydrodynamic-radius.pdf

Guinier analysis continued

Can be used to estimate R_g and $I(0)$, which tell us about protein shape and size



$I(0)$: forward scattering
R_g: Radius of gyration

Guinier approximation at small angles ($qR_g < 1.3$):

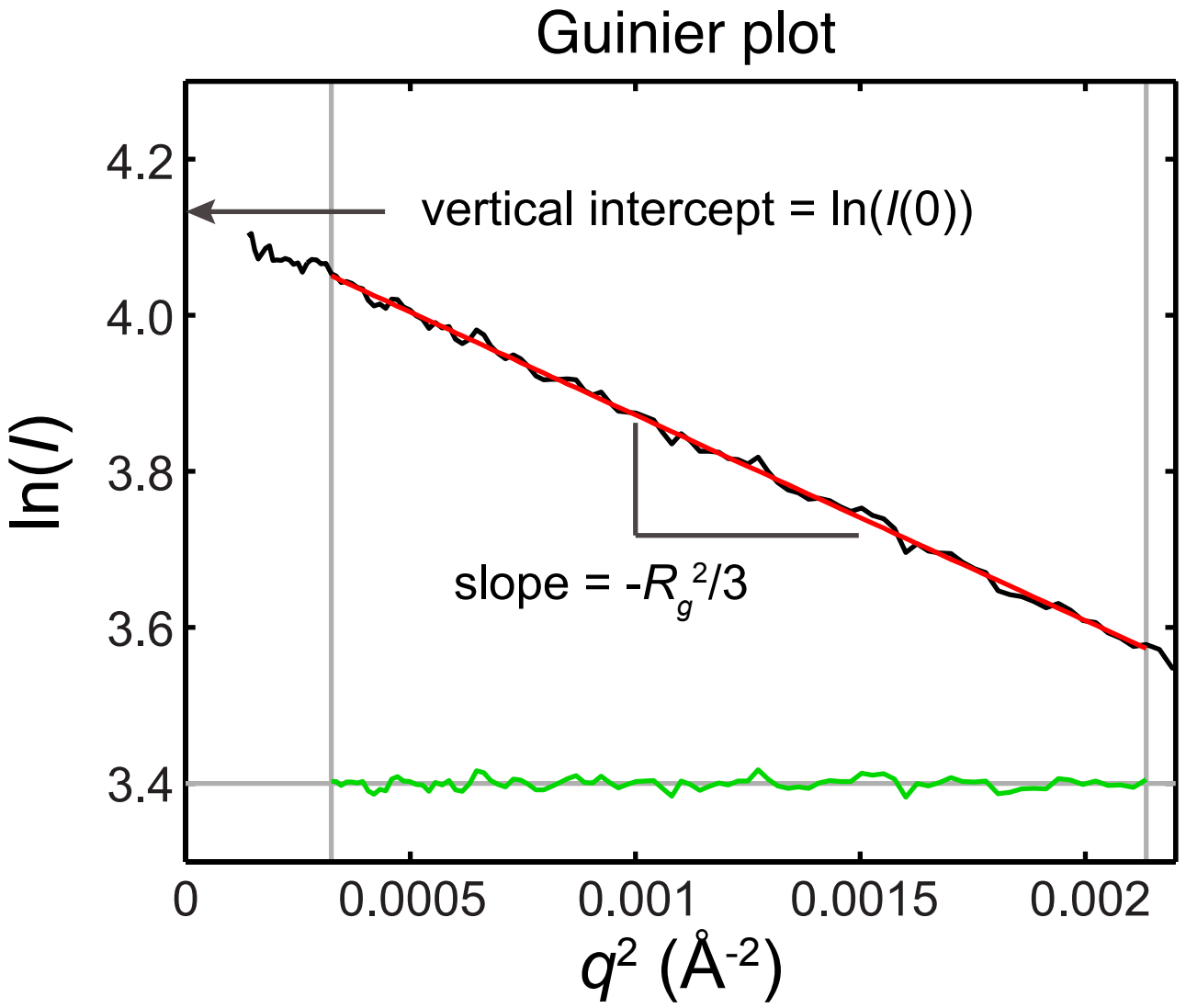
$$I(q) = I(0)e^{-R_g^2 q^2/3}$$

$$\ln[I(q)] = \ln[I(0)] - \frac{R_g^2}{3}q^2$$

$y = mx + b$

Ando, *et al.* (2011) PNAS.

Guinier analysis: fitting region, picking it, and linearity



Guinier approximation at small angles ($qR_g < 1.3$):

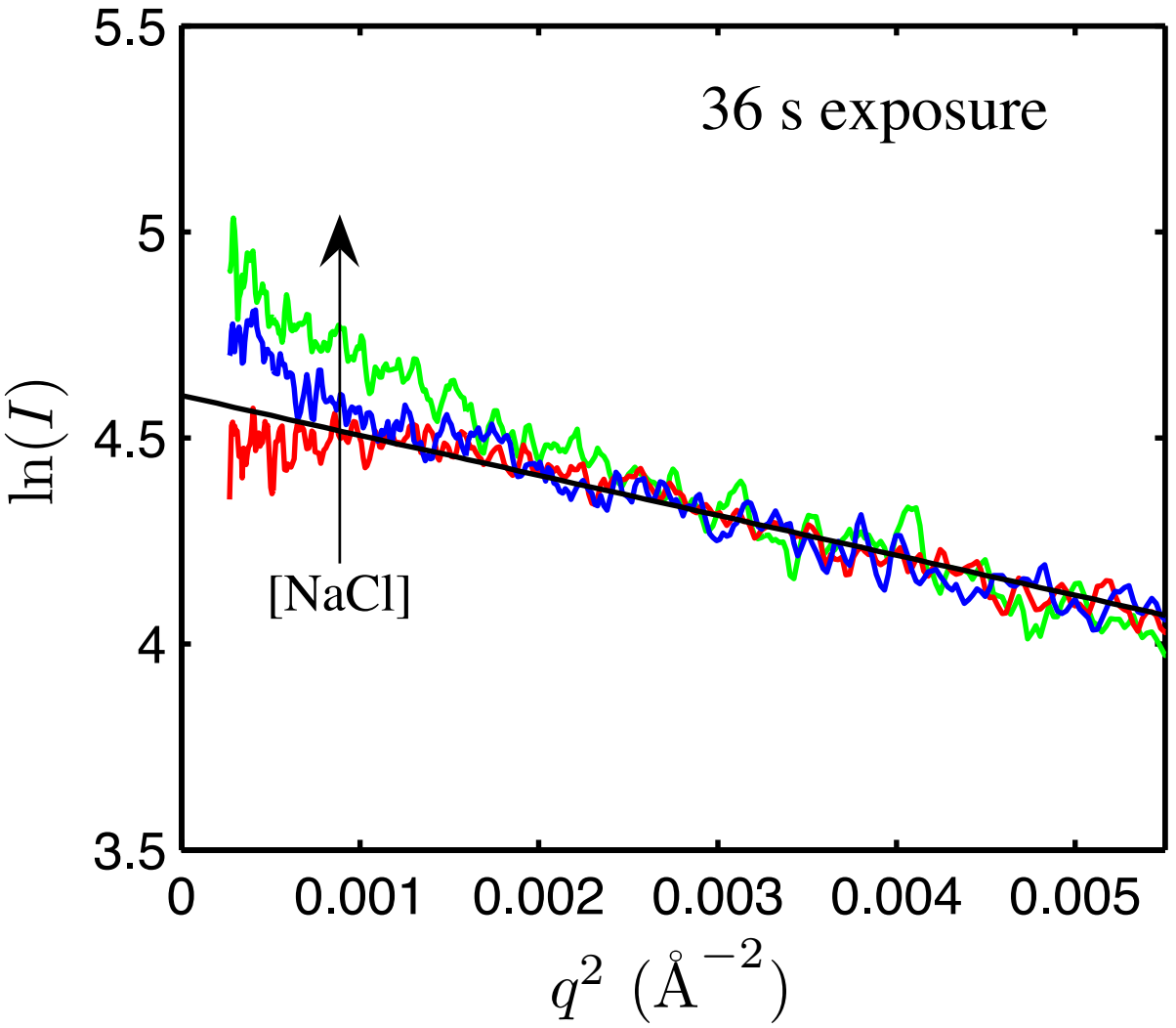
$$I(q) = I(0)e^{-R_g^2 q^2/3}$$

$$\ln[I(q)] = \ln[I(0)] - \frac{R_g^2}{3}q^2$$

Your Guinier plot should be **linear.** Upturns or downturns may reflect aggregation and/or interparticle effects!

Skou, *et al.* (2014) Nature Protocols.

Guinier region can show aggregation



Guinier approximation at small angles ($qR_g < 1.3$):

$$I(q) = I(0)e^{-R_g^2 q^2/3}$$

$$\ln[I(q)] = \ln[I(0)] - \frac{R_g^2}{3}q^2$$

- Aggregation leads to an upturn at low-q.
- Radiation damage often causes aggregation.

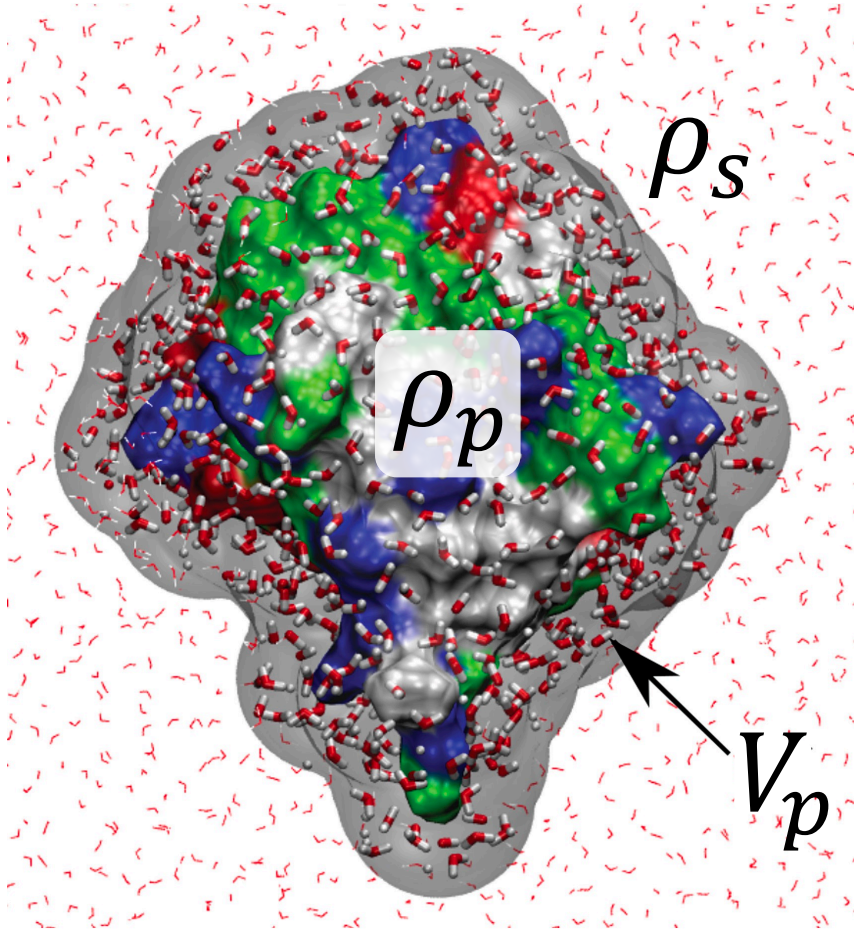
Skou, *et al.* (2014) Nature Protocols.

I(0) depends on contrast, molecular weight, concentration

$$I(0) \propto c_{molar}(\Delta\rho \cdot V_p)^2 \propto c_{molar}M^2 \propto c_{mass}M$$

- M = molecular weight (kDa)
- V_p = particle volume (Å^3)
- c_{molar} = molar concentration (μM)
- c_{mass} = concentration in mg/ml
- $\Delta\rho = \rho_p - \rho_s$ = electron density contrast between the hydrated particle (ρ_p) and solvent (ρ_s) (e$^-$/Å^3)

I(0) is related to a protein's molecular weight.

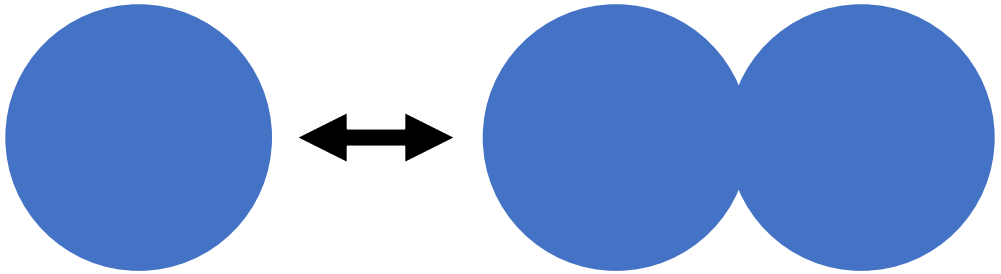


https://www.genengnews.com/magazine/295/new-peaks-in-hydrophobic-bioprocessing/

Stuhrmann, H. B. (1980). Synchrotron Radiation Research, edited by H. Winick, Doniach, S., pp. 513-531. New York: Plenum Press.

Relationship between MW, concentration, and I(0)

I(0) is proportional to MW^2. $I(0) \propto c_{molar} M^2 \propto c_{mass} M$



monomer-dimer transition

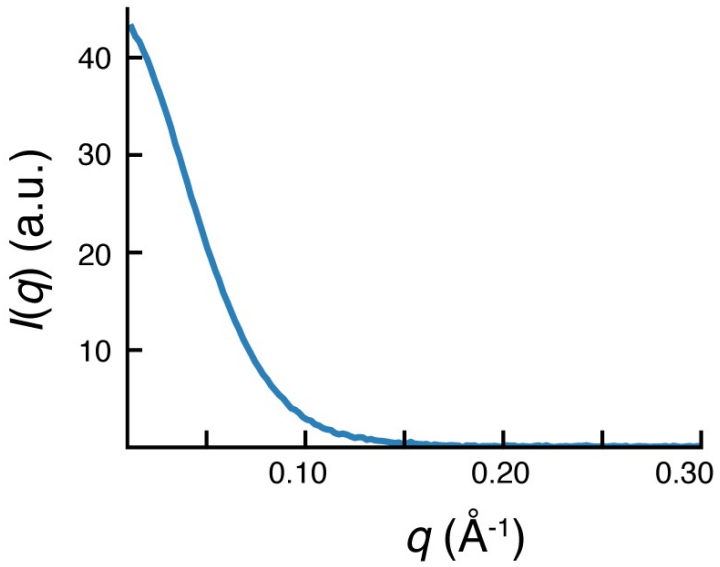
- At the same molar concentration, a sample of dimers has an *I*(0) that is 4x that of a sample of monomers.
- However, in a monomer-dimer transition, the molar concentration is halved due to dimerization, so the net change in *I*(0) is a factor of 4 x 1/2 = 2.

Content from bioxtas-RAW tutorial

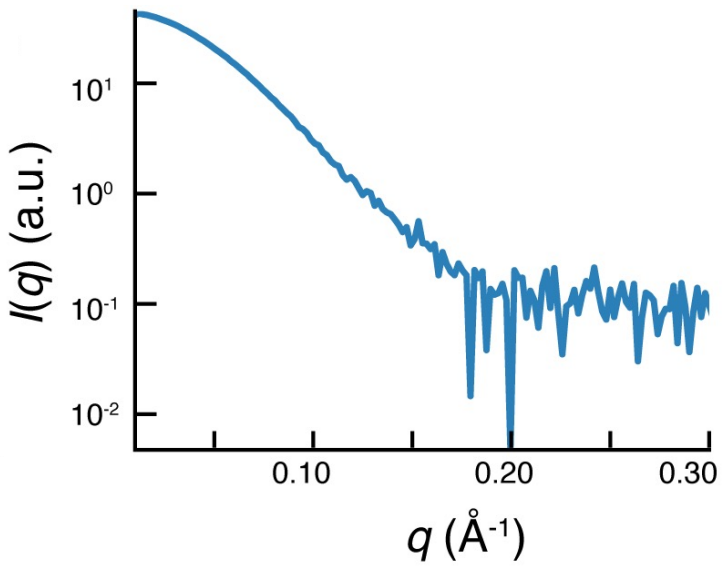
Kratky plots are sensitive to conformation

- Replot as Iq^2 vs q
- Emphasizes the power-law dependence in mid-q region (~q^{-n}) that contains information about foldedness, compactness, flexibility.

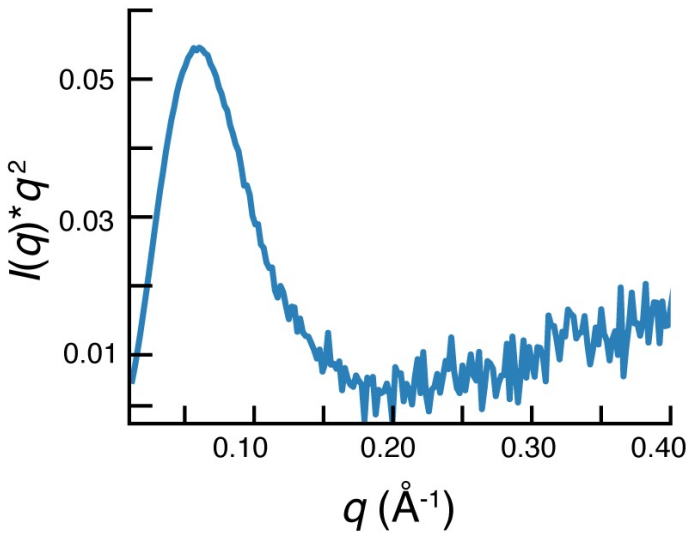
Subtracted profile



Semilog plot

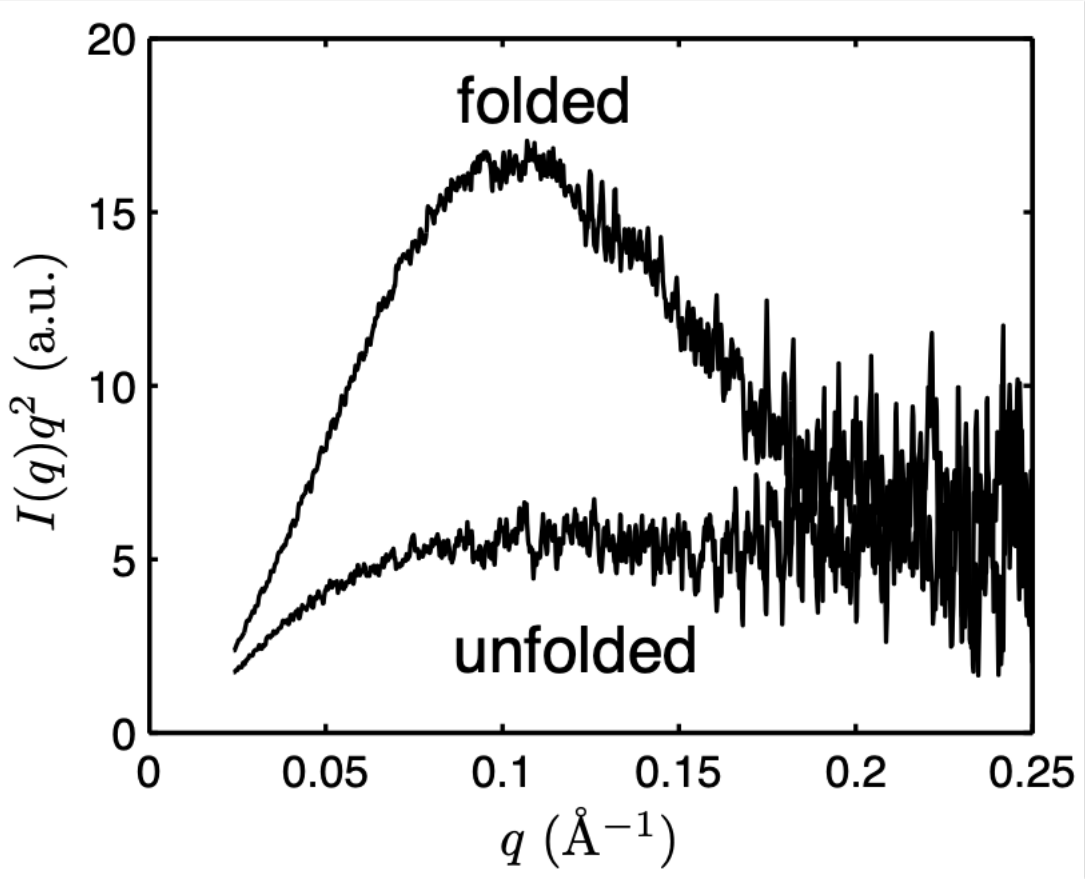


Kratky plot



Kratky plots show protein compactness and foldedness

Kratky plot



At high angles, $I(q)$ decays with a power-law dependence, q^n.

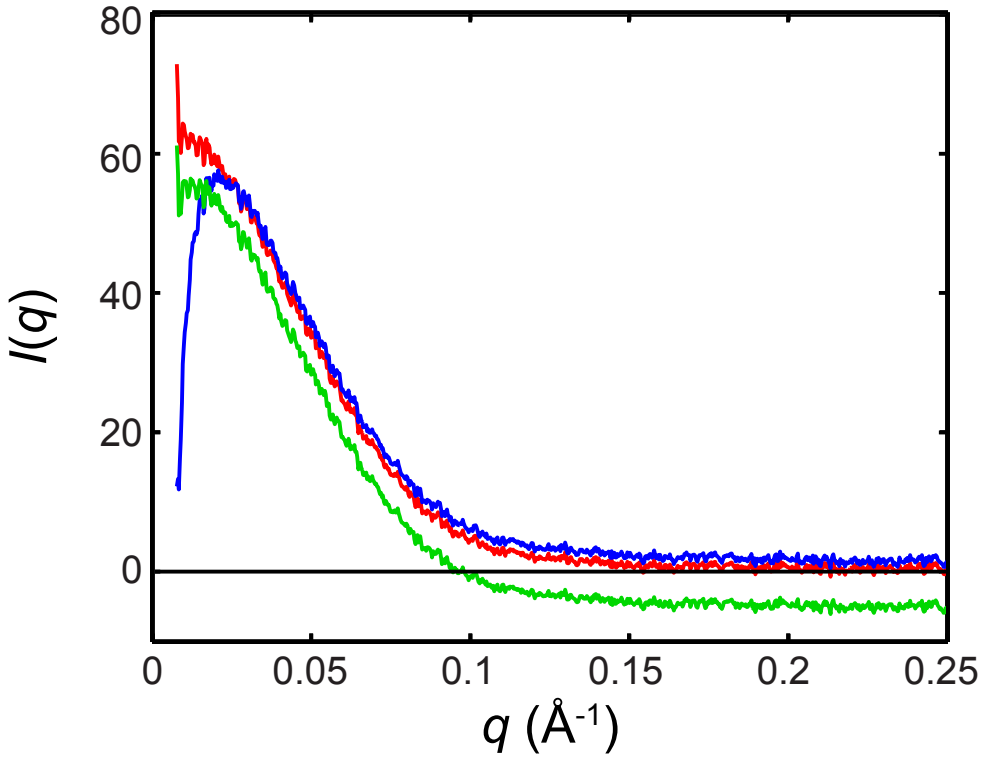
folded, compact: n ~ -4 ⟹ peak

random polymer chain: n ~ -2 ⟹ plateau

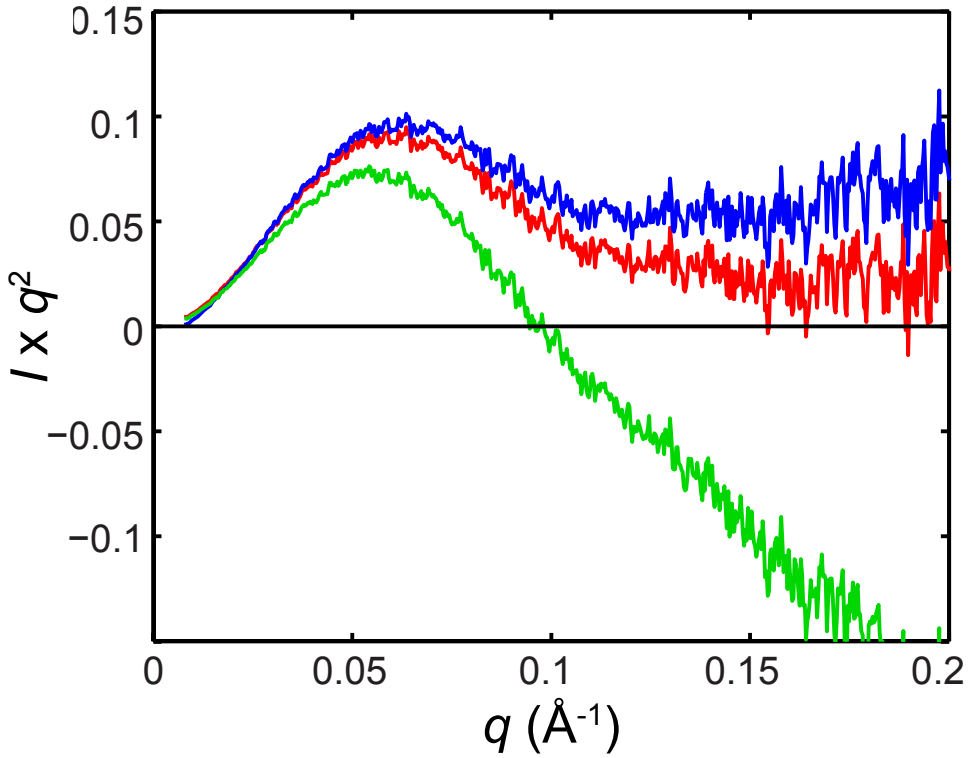
extended polymer chain: n ~ -1 ⟹ rise

Ando, et al. (2008). Biochemistry, 47(42), 11097–11109.

Kratky plots are sensitive to background subtractions



Subtracted profiles



Kratky plot

- properly matched buffer
- similar buffer with glycerol added
- different buffer

Skou, *et al.* (2014) Nature Protocols.

Porod Volume

- The Porod volume is the ratio of $I(0)$ and the area under the Kratky plot (Q_p). For compact particles, it gives an estimate of the volume of the macromolecule in solution.

$$Q_p = \int_0^\infty q^2 I(q) dq$$

$$V_p = \frac{2\pi^2 I(0)}{Q_p}$$

where Q_p is the Porod invariant (area under the plot)
V_p is the Porod volume

- Can be estimated in RAW or Atsas

Content from bioxtas-RAW tutorial

MW estimation

Can be estimated from SAXS data in multiple ways:

- From Porod Volume
- From absolute scaled *I*(0)
- From a reference standard
- From volume of correlation

- Important: As a rule of thumb, assume at least ~10% uncertainty. Better to combine with other techniques or use for things like checking oligomeric state.

Content from bioxtas-RAW tutorial

MW from Porod volume

- Can be calculated from a SAXS profile using a Porod invariant approach implemented in RAW.
- Porod volume (in $Å^3$) is ~1.6-1.7x the MW (in Da).
- **Pros**: Fairly accurate and does not need standard, concentration, or absolute calibration.
- **Cons**: Sensitive to subtraction errors, does not work for non-protein or extended structures.

Piiadov, V., de Araújo, E. A., Oliveira Neto, M., Craievich, A. F. & Polikarpov, I. (2018). Protein Sci. 2–22. DOI: 10.1002/pro.3528

Content from bioxtas-RAW tutorial

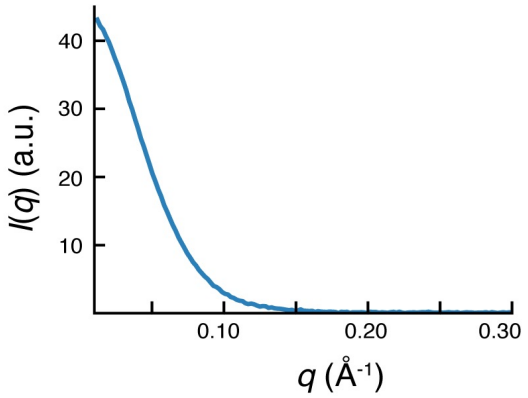
MW from reference standard

- Scattering at $I(0)$ is proportional to MW of the macromolecule. If a reference sample is known, it can be used to calibrate.

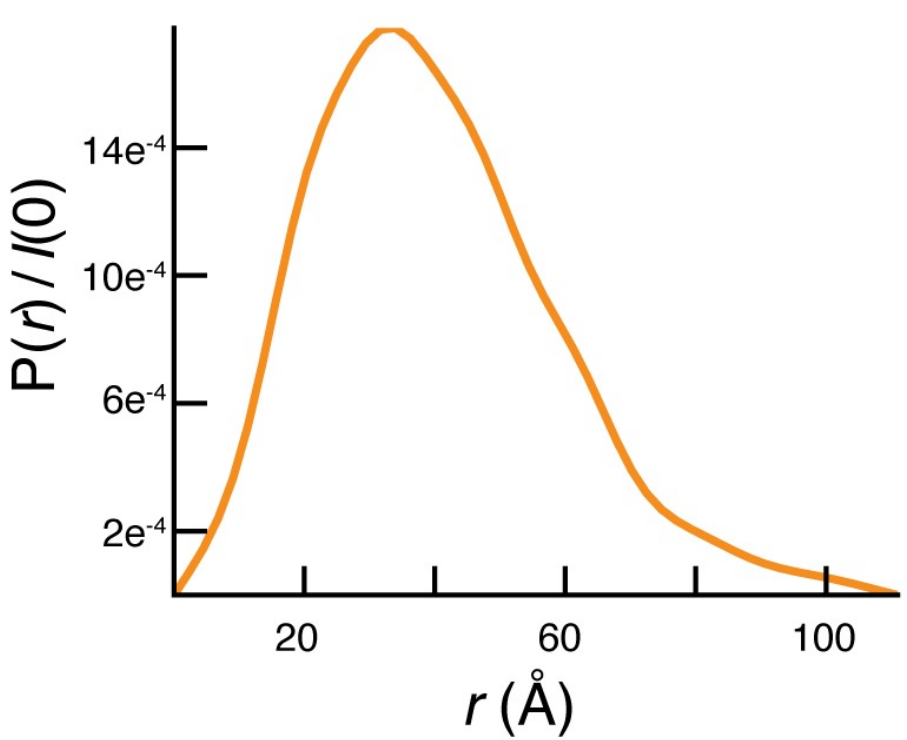
- **Pros**: Can be accurate if standard and sample are similar and conditions are similar
- **Cons**: Requires accurate concentration, similar standard and standard conditions (e.g. same buffer).

Content from bioxtas-RAW tutorial

Pair-distance distribution function



P(*r*) is the inverse Fourier Transform of *I*(*q*) and represents the histogram of electron-pair distances in a protein.



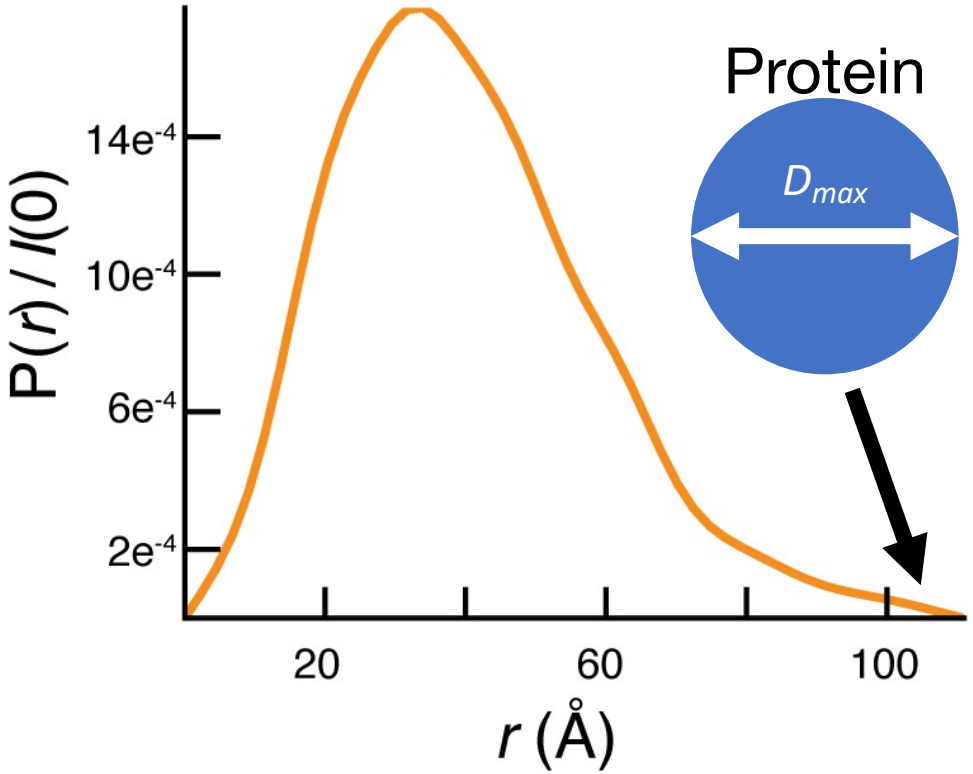
$$P(r) = \frac{r^2}{2\pi^2} \int_0^\infty q^2 I(q) \frac{\sin(qr)}{qr} dq$$

Svergun & Koch (2003). Reports on Progress in Physics, 66, 1735–1782.

Pair-distance distribution function

- $P(r)$ is the inverse Fourier Transform of $I(q)$ and represents the histogram of electron-pair distances in a protein.

- When D_{max} is well-defined, provides R_g, $I(0)$, shape information



$$R_g^2 = \frac{\int_0^{D\max} r^2 P(r)\,dr}{2\int_0^{D\max} P(r)\,dr}$$

Svergun & Koch (2003). Reports on Progress in Physics, 66, 1735–1782.

P(r) is determined numerically

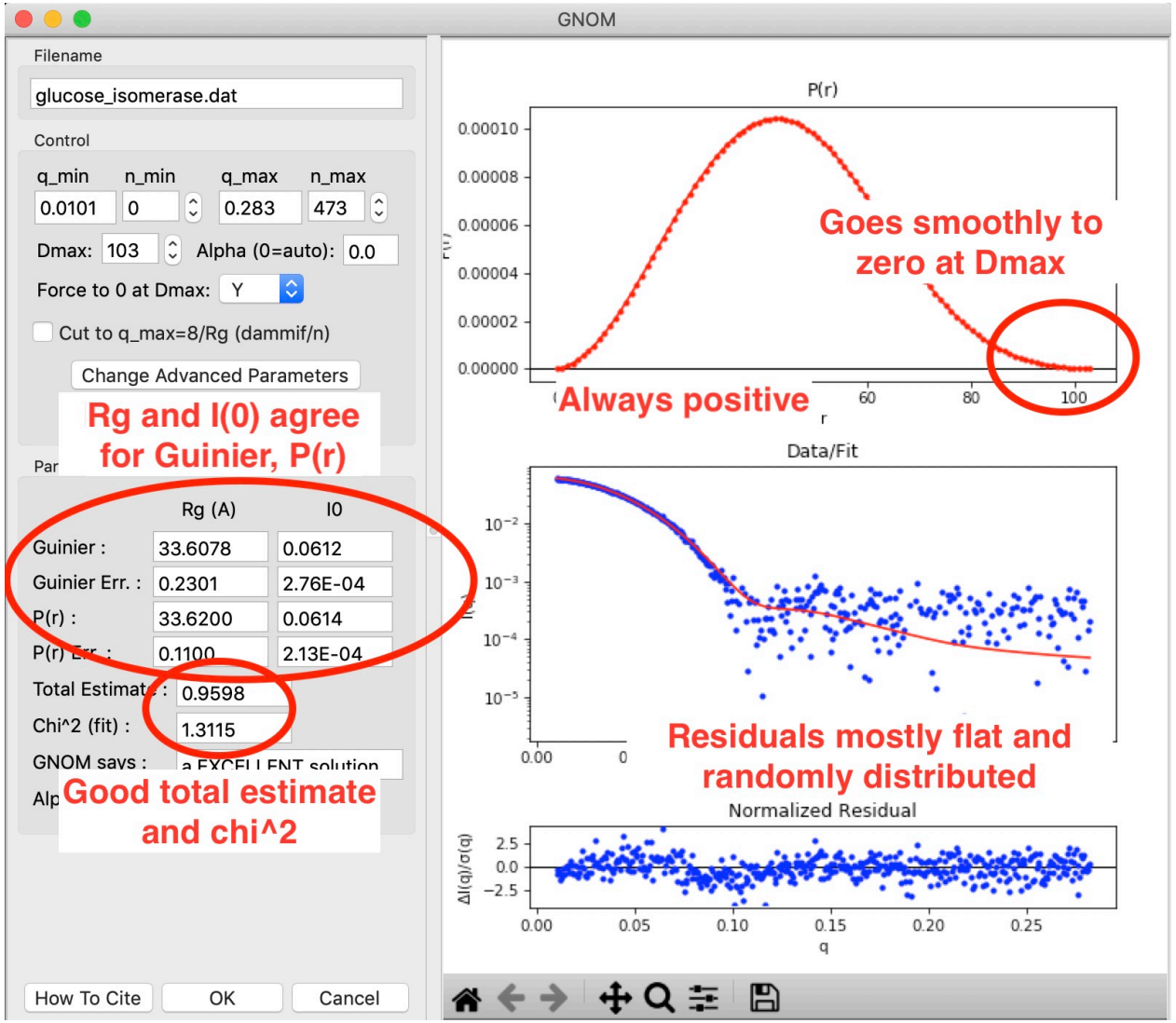
In RAW:

- Bayesian Indirect Fourier Transform (BIFT)
- GNOM from ATSAS package (if ATSAS is installed)

Important: Methods require iterative adjustment of a few parameters: D_{max} and α (a fit weighting parameter).

Content from bioxtas-RAW tutorial

P(r) determination



Criteria:

1. The *P*(*r*) function smoothly fits the measured scattering profile.
2. The *P*(*r*) function goes to zero at $r=0$ and $r=D_{max}$

Additional usual criteria:

1. The R_g and $I(0)$ from the Guinier fit and the *P*(*r*) function agree well.
2. The *P*(*r*) function is always positive.

Content from bioxtas-RAW tutorial

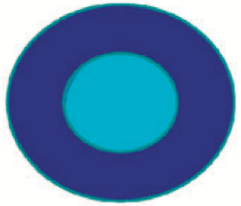
Example P(r) functions

Sphere

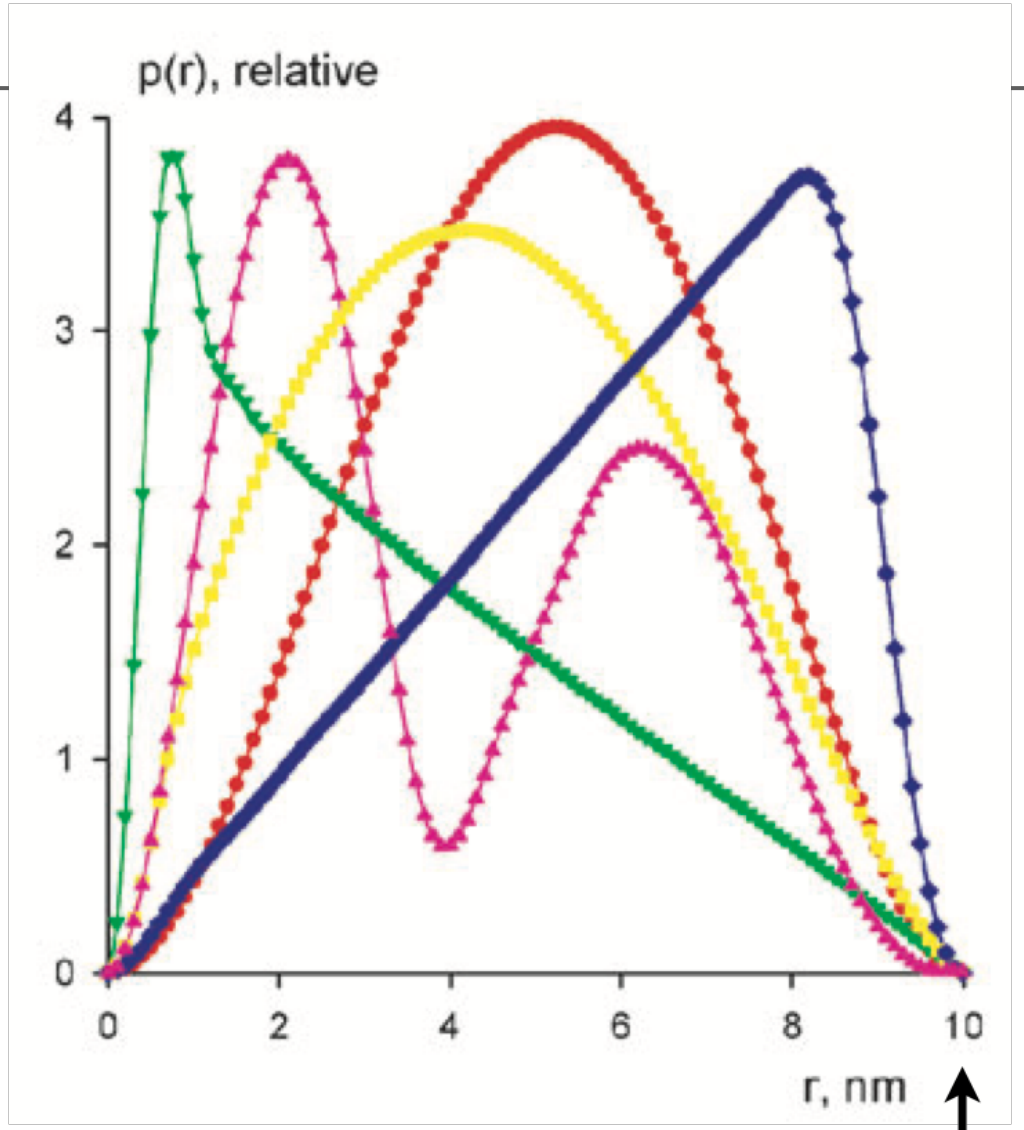
Cylinder

Disc

Hollow sphere



Barbells



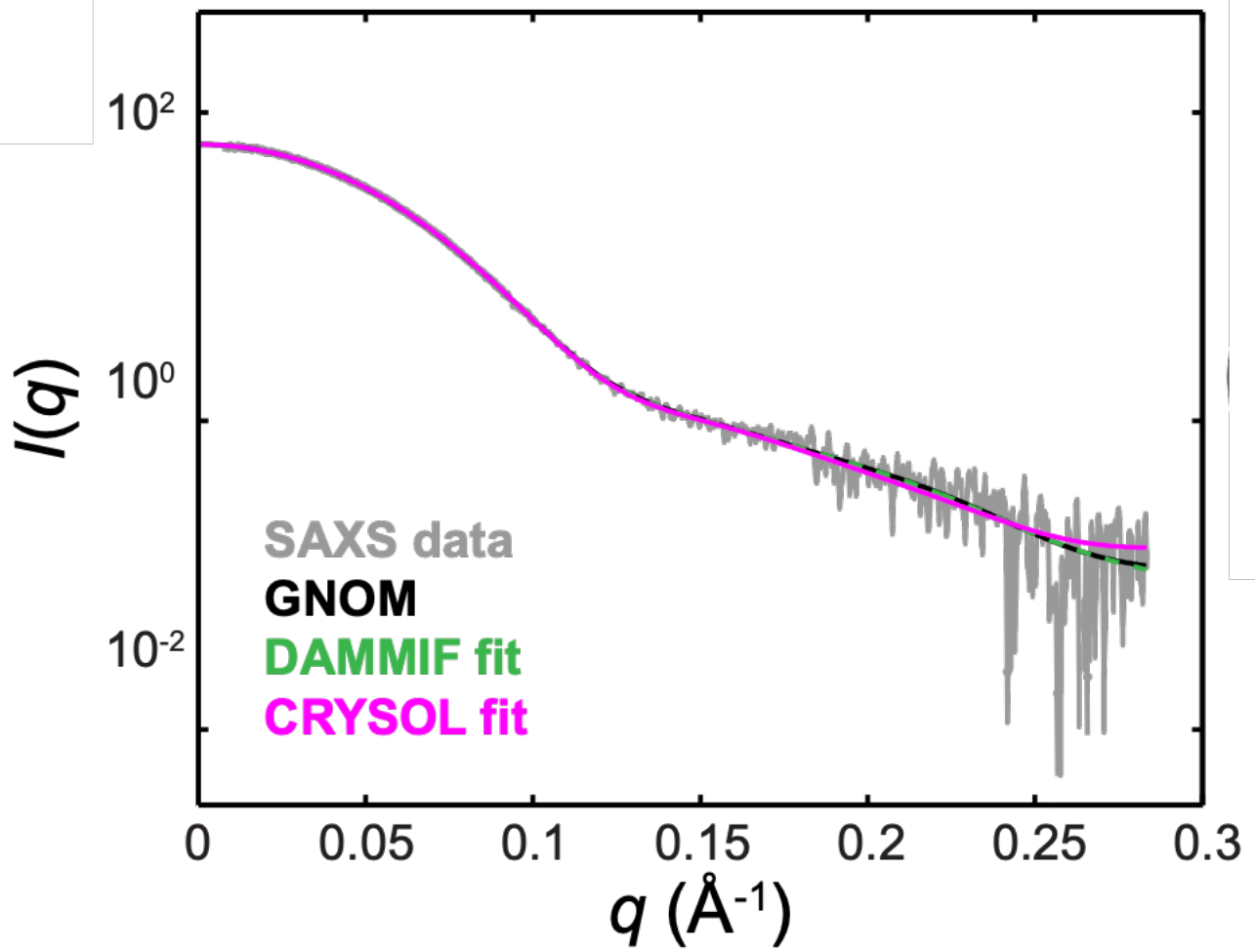
D_{max} = maximum dimension

Svergun & Koch (2003). Reports on Progress in Physics, 66, 1735–1782.

Fitting data to theoretical scattering

- CRYSOL from ATSAS package (Svergun group, EMBL)
- FoXS server (Sali group, UCSF)

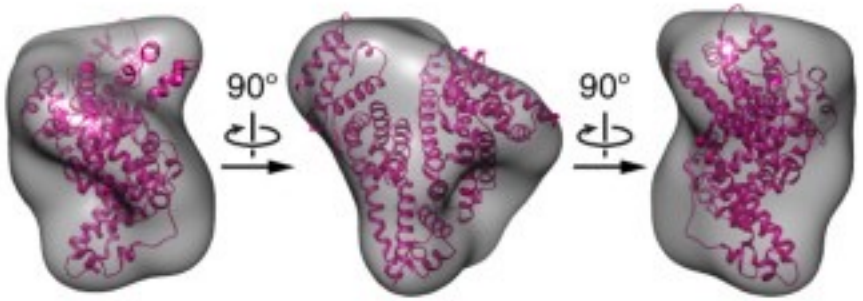
- More on this with Steve



Skou, *et al.* (2014) Nature Protocols.

Ab initio shape reconstructions

- Following *P*(*r*) analysis, a low-resolution 3D model can be reconstructed from SAXS data.
- Options in ATSAS package
- More on this later, use carefully!

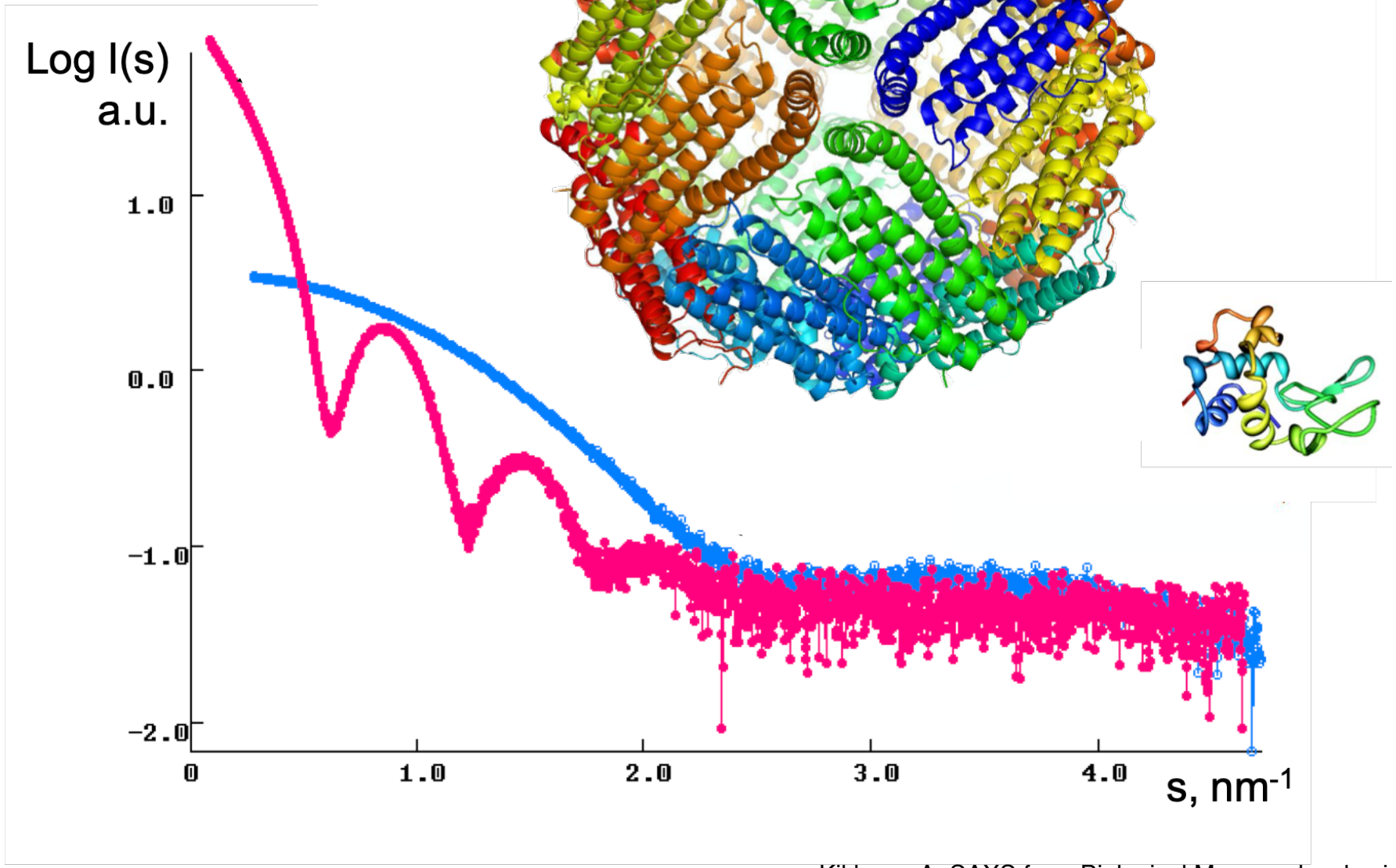


Skou, *et al.* (2014) Nature Protocols.

What SAXS can teach us about a protein

Sample characteristic	Parameter to study
Purity/monodispersity	Guinier plot
Oligomeric state	Volume and molecular weight
Shape/anisometry	Envelope, R_g, $P(r)$
Flexibility (folded/non-folded)	Kratky plot, Porod exponent
Interparticle interactions	Guinier, Concentration series
Conformation	Overall scattering profile

Example profiles



Kikhney, A. SAXS from Biological Macromolecules in Solution (2013).