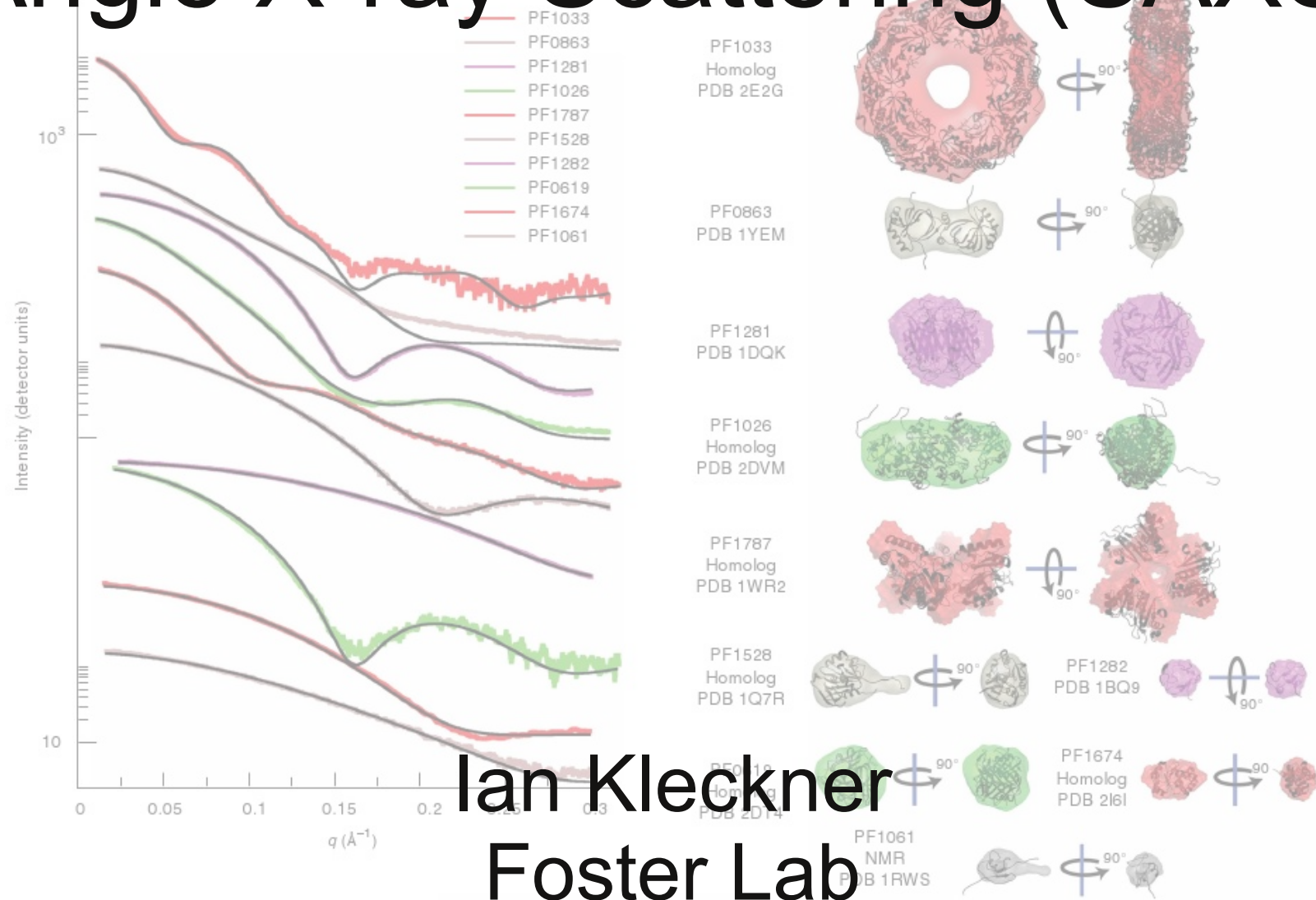


A practical introduction to Small Angle X-ray Scattering (SAXS)



Ian Kleckner
Foster Lab
January 26, 2010

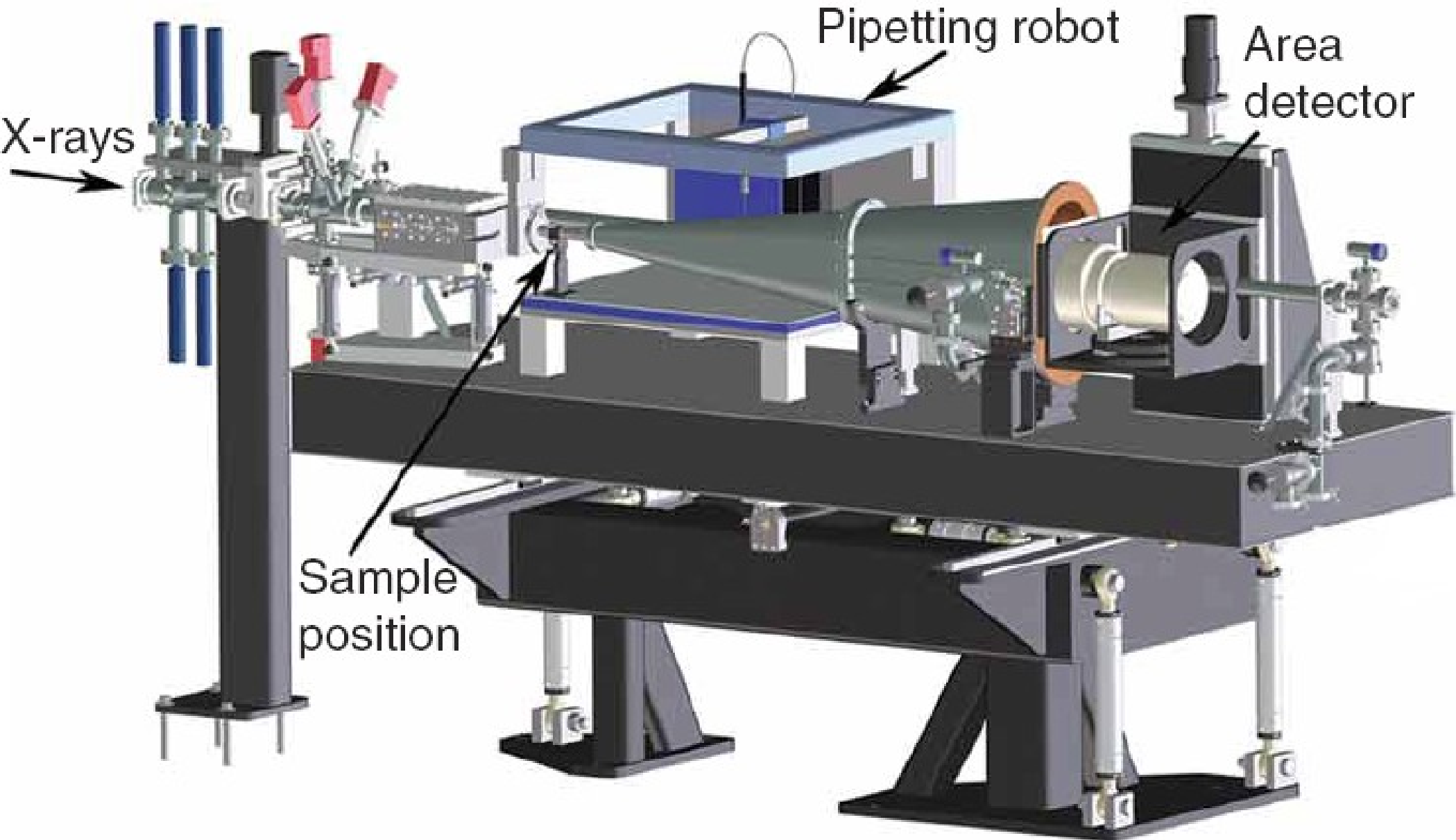
Outline

- X-ray scattering principles
- SAXS observable $I(q)$
- SAXS interpretables
- SAXS tutorial

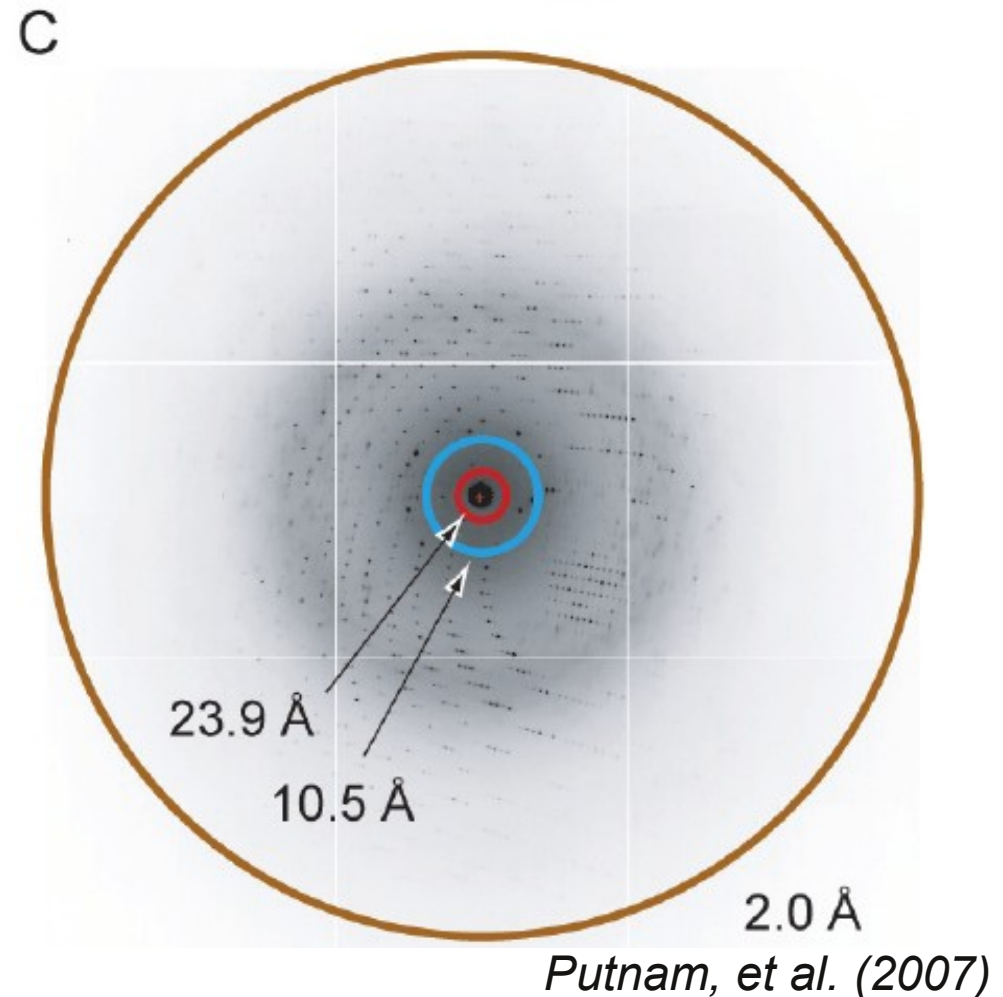
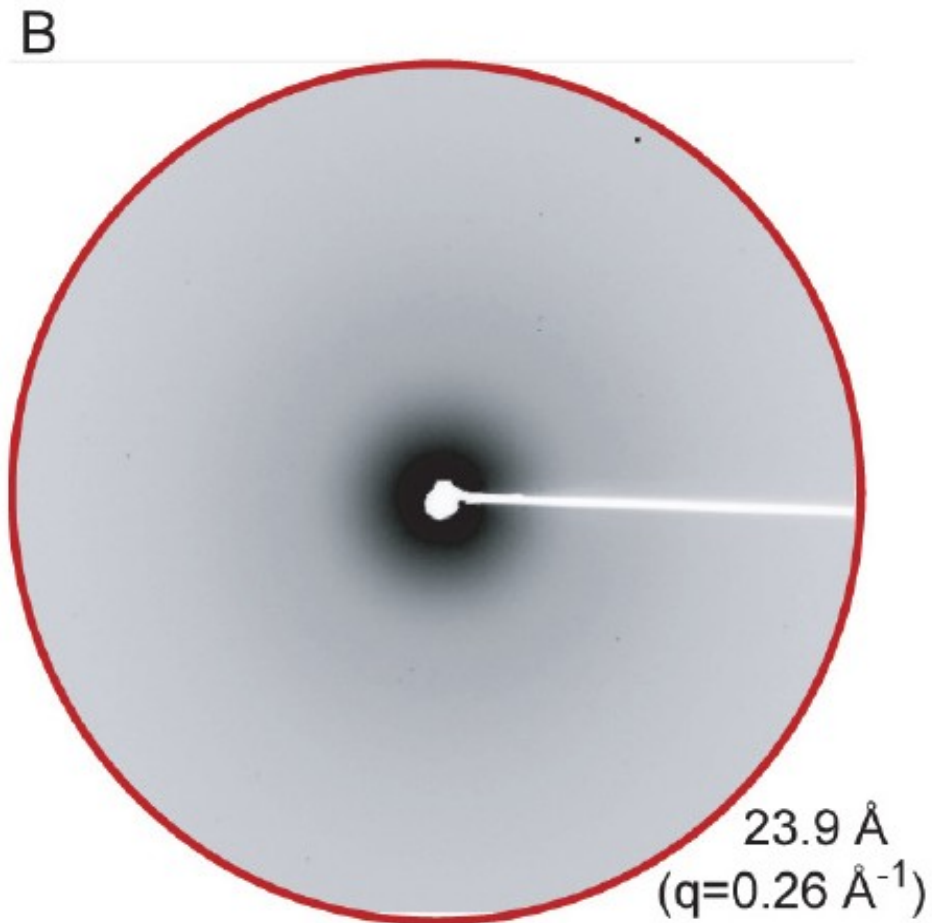
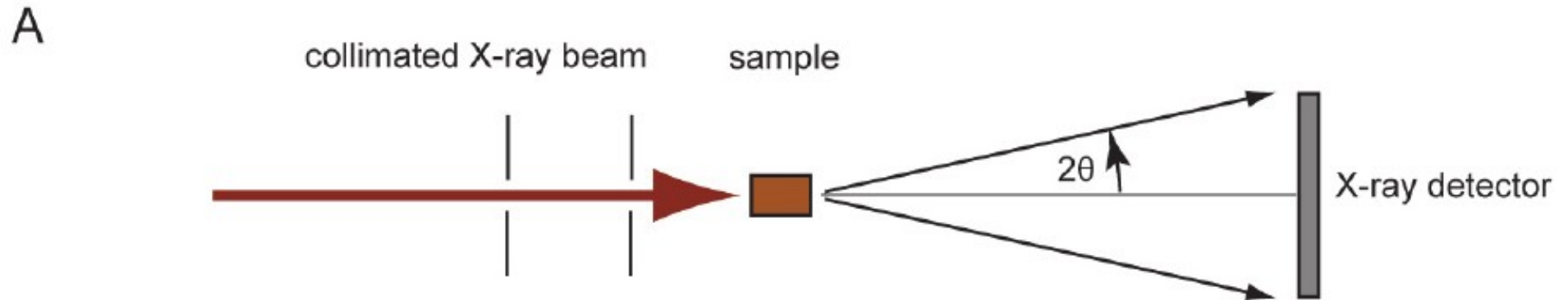
Putnam, CD; Hammel, M; Hura, G; Tainer, JA. **"X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution"**. *Quart. Rev. Biophys.* (2007).

Greg L Hura, Angeli L Menon, Michal Hammel, Robert P Rambo, Farris L Poole II, Susan E Tsutakawa, Francis E Jenney Jr, Scott Classen, Kenneth A Frankel, Robert C Hopkins, Sung-jae Yang, Joseph W Scott, Bret D Dillard, Michael W W Adams & John A Tainer. **"Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS)"**, *Nature Methods* 6:606-612 (2009).

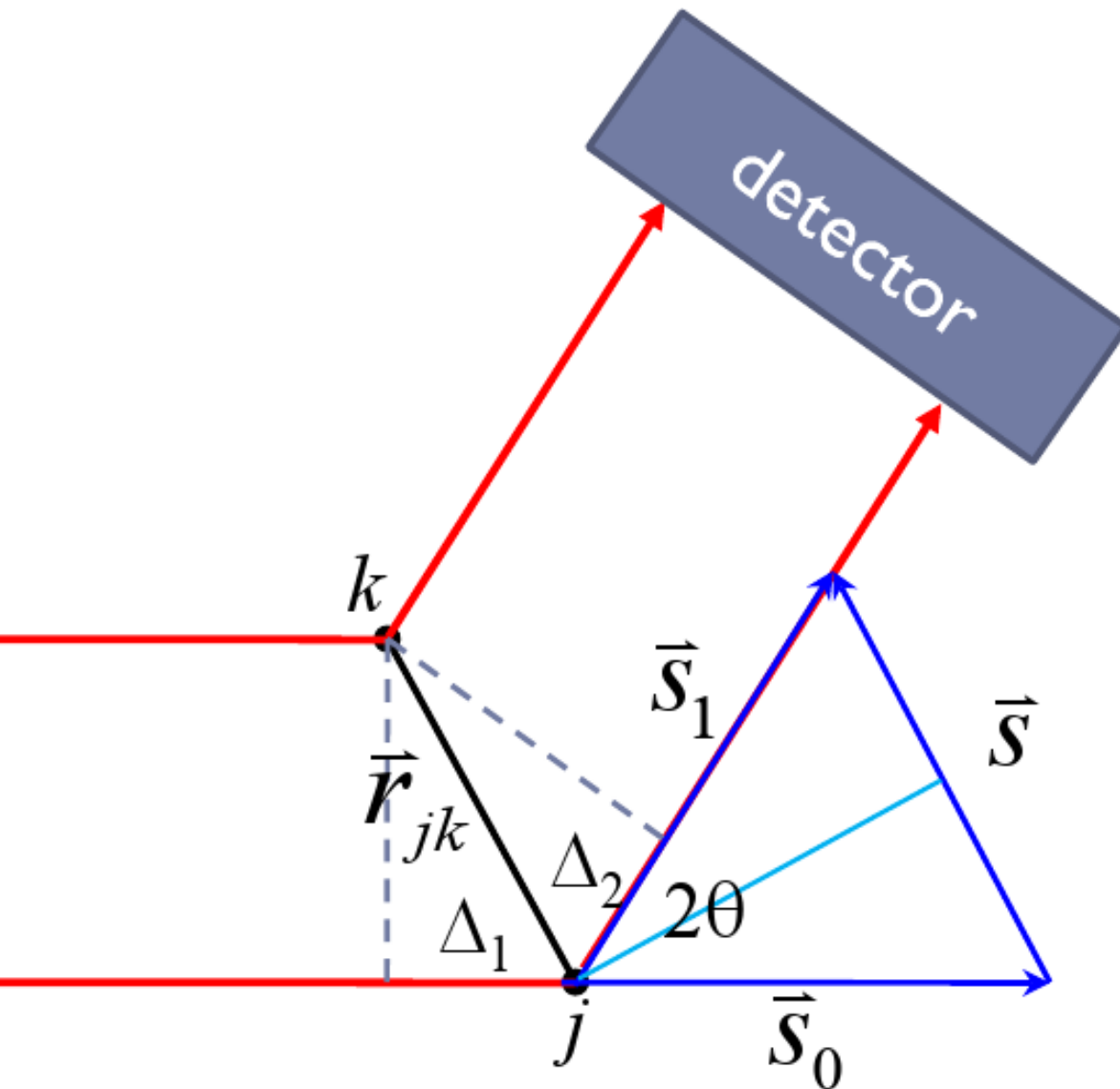
Beam line 12.3.1 setup (ALS)



Sample interactions with X-rays



q : magnitude of momentum transfer to incident X-ray photon



unit incident wavevector \vec{s}_0
 unit scattering wavevector \vec{s}_1
 wave vector change: \vec{s}

$$\vec{s} = \vec{s}_1 - \vec{s}_0$$

$$|\vec{s}| = 2 \sin \theta$$

Definition: momentum transfer

$$\vec{q} = (2\pi / \lambda) \vec{s}$$

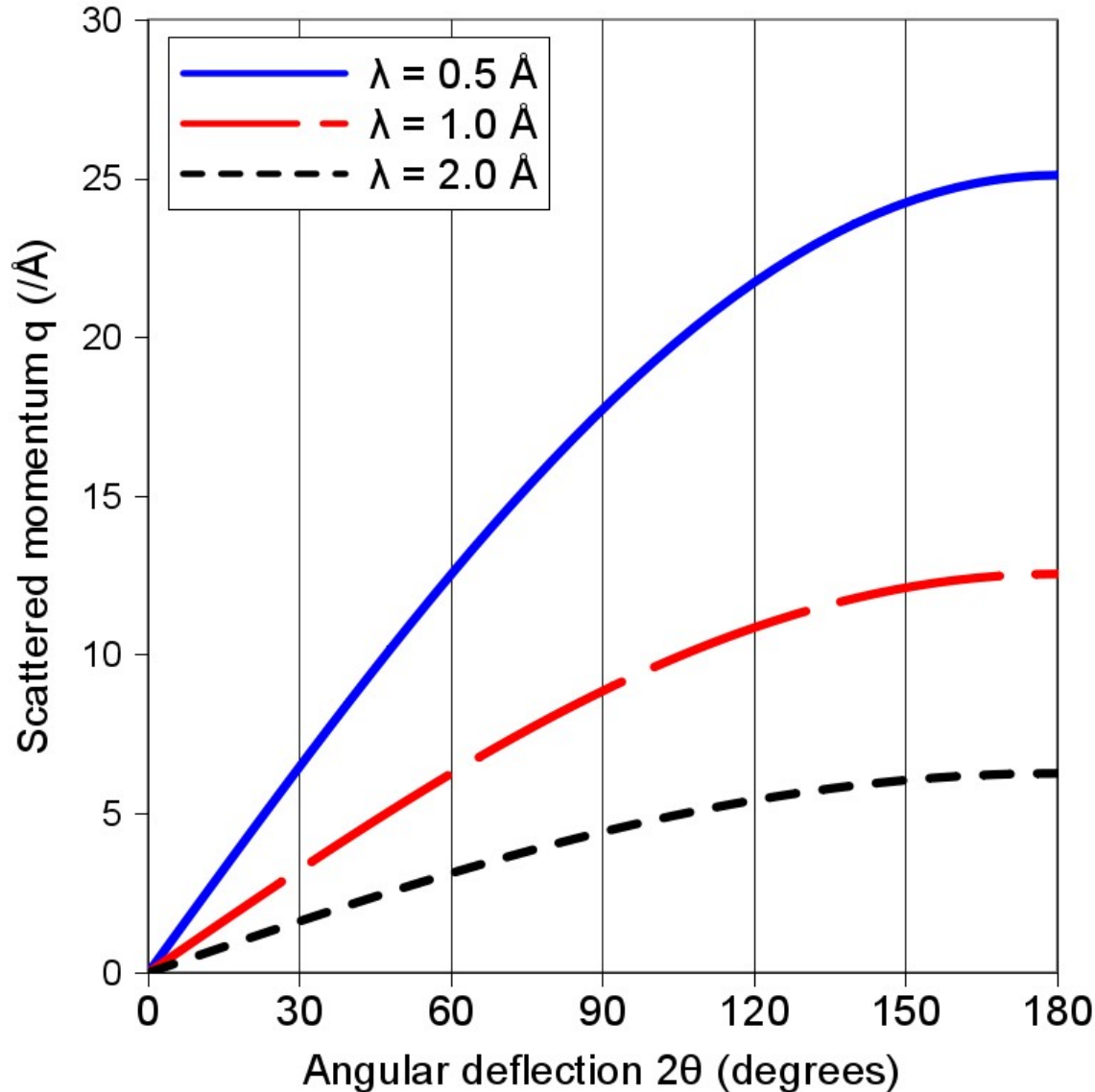
$$q = |\vec{q}| = (4\pi / \lambda) \sin \theta$$

Phase change: $\Delta\phi = 2\pi * \frac{\Delta}{\lambda} = \vec{q} \bullet \vec{r}_{jk}$
 (radian)

Incident X-ray photon momentum
 $p = h / \lambda$ $h = \text{Planck's const}$

q: magnitude of momentum transfer to incident X-ray photon

- q is increased by increasing incident photon momentum p (decreasing λ) and/or increasing angle of deflection up to $2\theta=180^\circ$

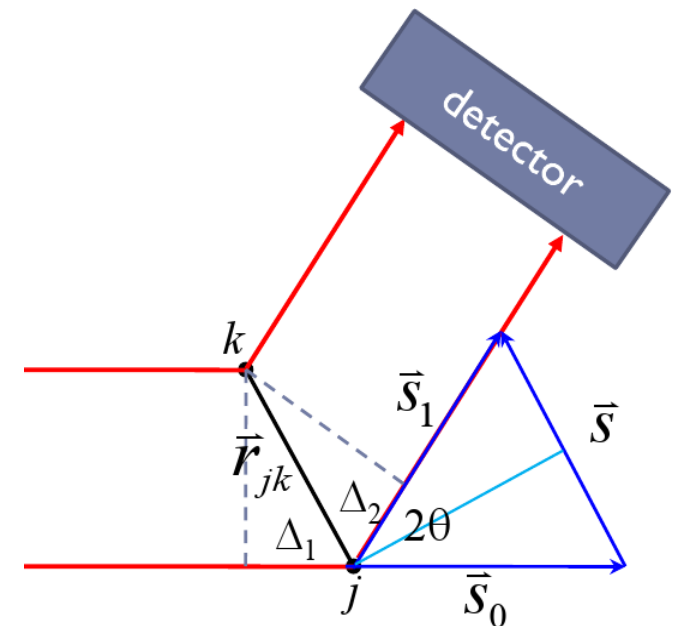


Incident X-ray photon momentum
 $p = h / \lambda$ $h = \text{Planck's const}$

Definition: momentum transfer

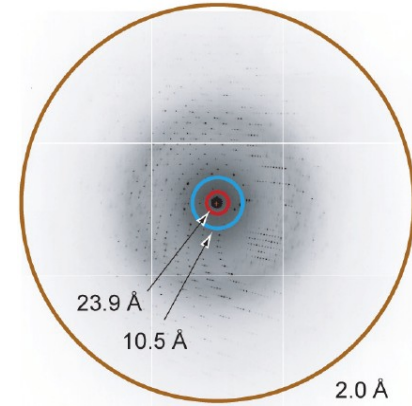
$$\vec{q} = (2\pi / \lambda) \vec{s}$$

$$q = |\vec{q}| = (4\pi / \lambda) \sin \theta$$



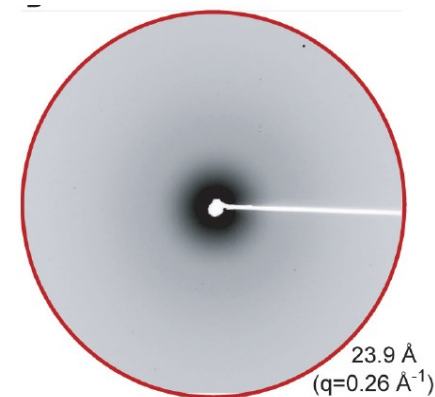
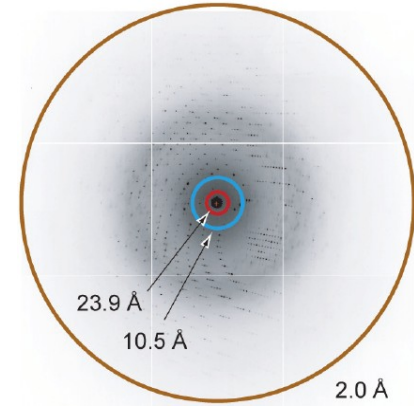
Crystals diffract and solution-particles scatter

- X-ray crystallography exploits repeating unit of the crystal
- Incident photons scatter (“reflect”), interfere and are detected
- Crystal lattice structure permits coherent interference to yield well-defined maxima

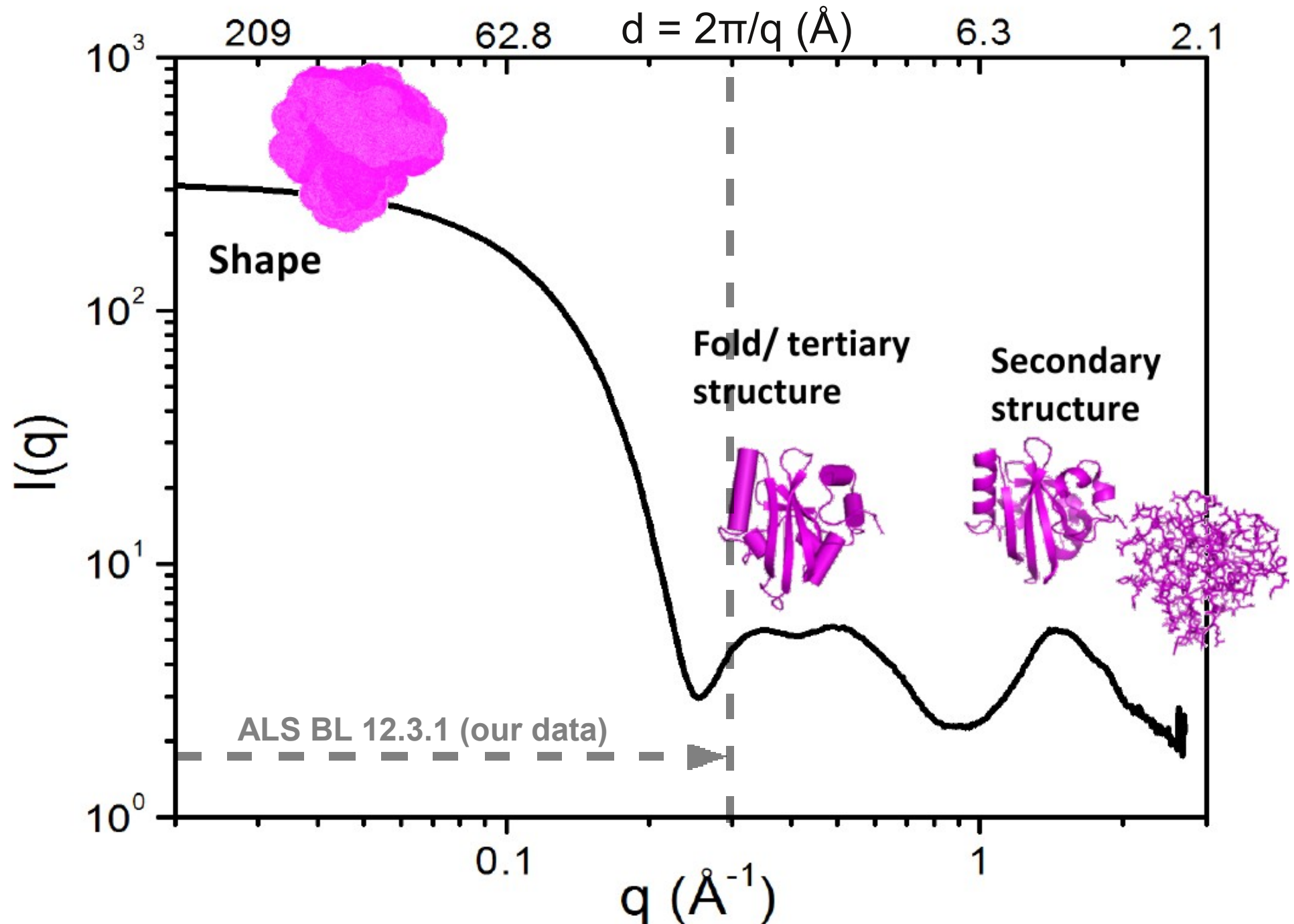


Crystals diffract and solution-particles scatter

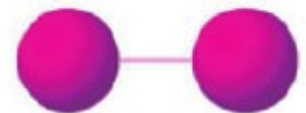
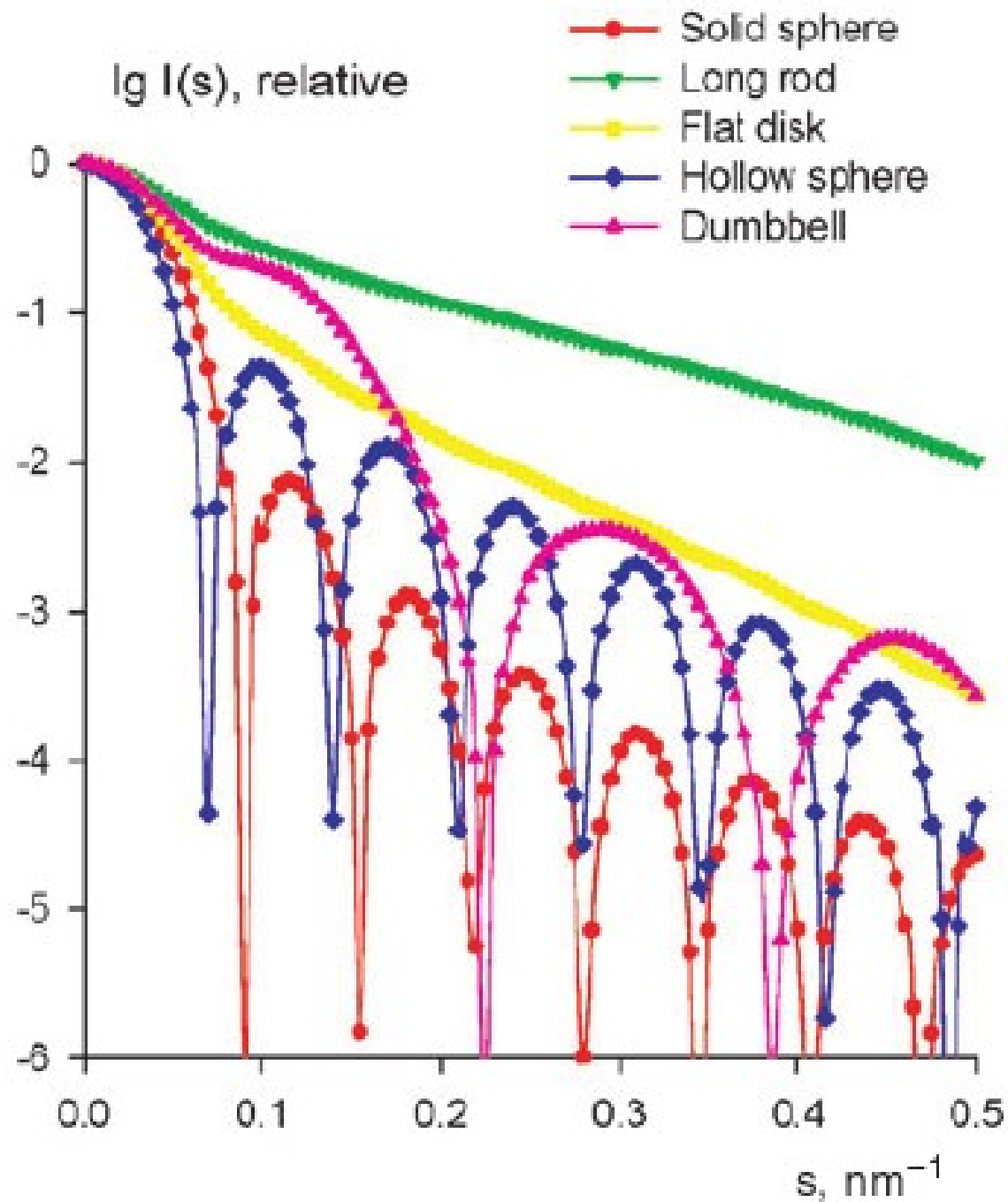
- X-ray crystallography exploits repeating unit of the crystal
 - Incident photons scatter (“reflect”), interfere and are detected
 - Crystal lattice structure permits coherent interference to yield well-defined maxima
- X-ray solution scattering has a disordered arrangement of particles
 - Incident photons scatter but phase is incoherent and therefore interference is not quantized
 - Therefore, radially averaged scattering profile (1D) is obtained
 - Buffer scattering must be subtracted



SAXS $I(q)$ reflects particle shape and size



SAXS $I(q)$ reflects particle shape and size

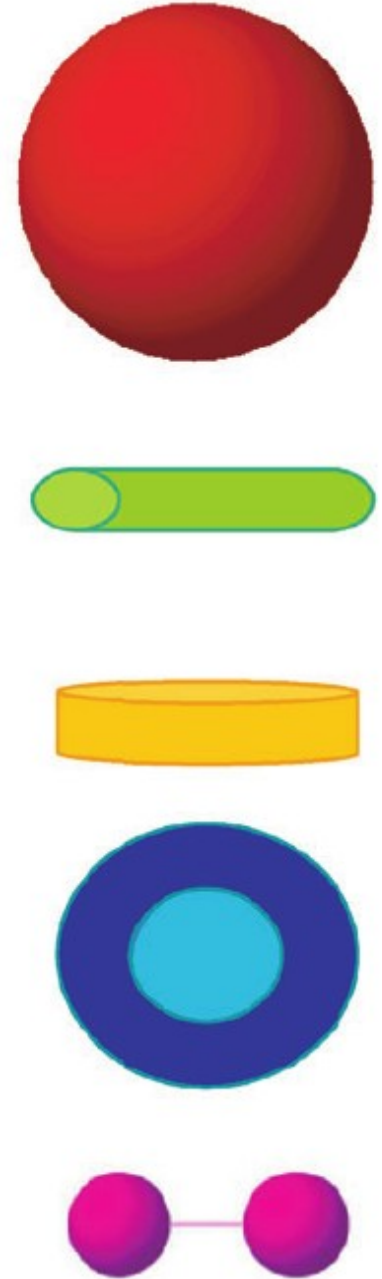
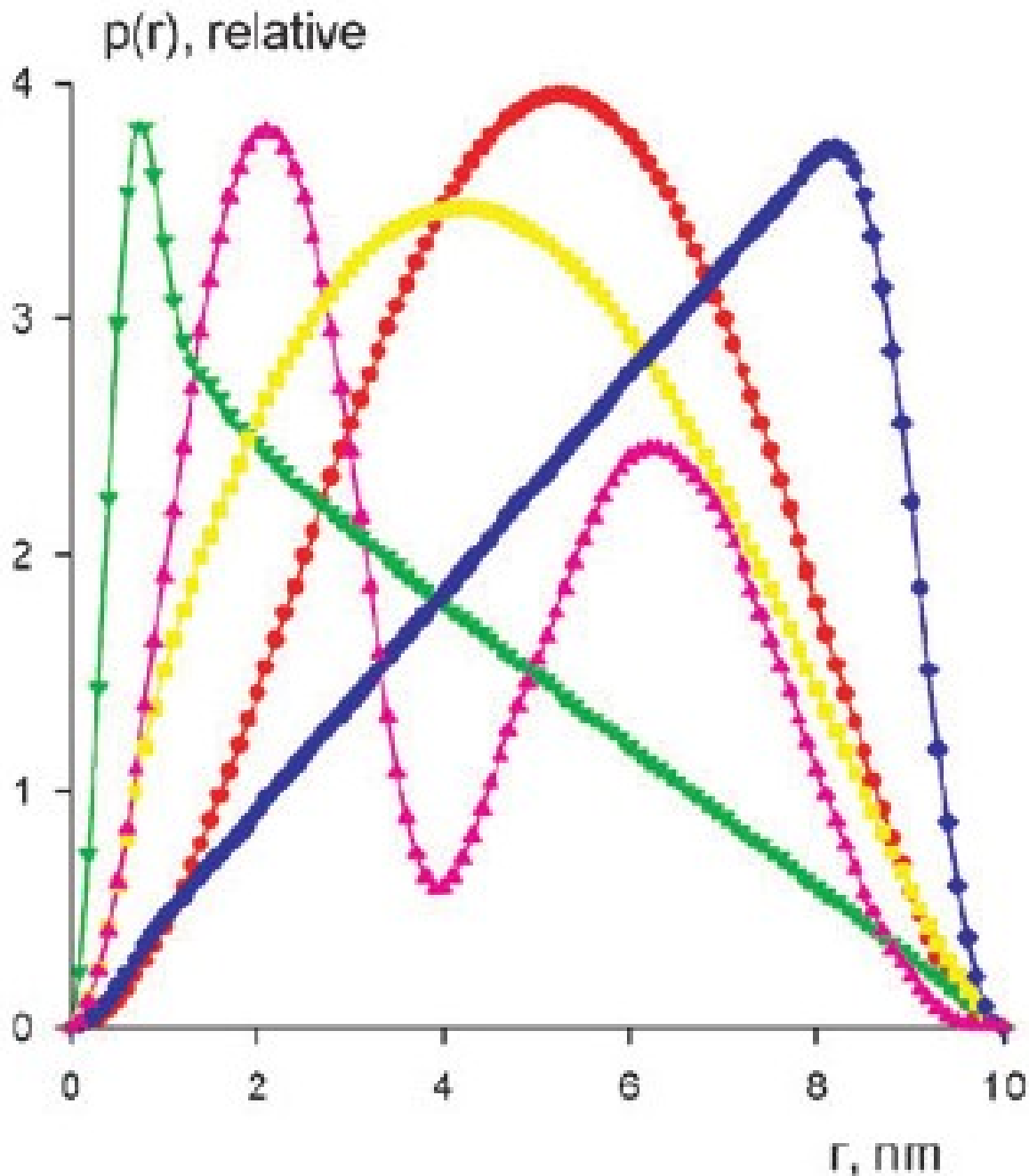


Pair distribution function $P(r)$

- $P(r)$ quantifies how well the electron density in the 3D structure “knows” its spatially-proximal density over space
- Characterized by the length over which the electron density is changed from current value
- This is a radially-averaged electron density auto-correlation function
- It depends only on the magnitude of the distance $|r|$ and not the orientation

$$P(r) = r^2 \left\langle \int_V \Delta\rho(\vec{r}) \Delta\rho(\vec{u} + \vec{r}) d\vec{r} \right\rangle_{\Omega}$$

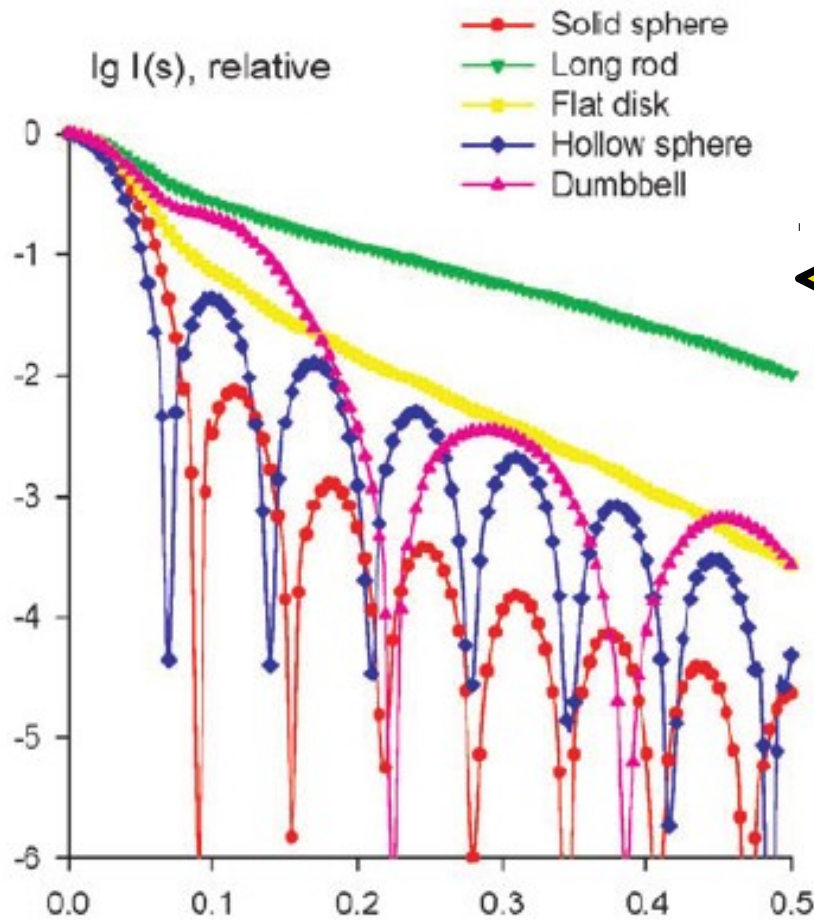
SAXS $P(r)$ reflects particle shape



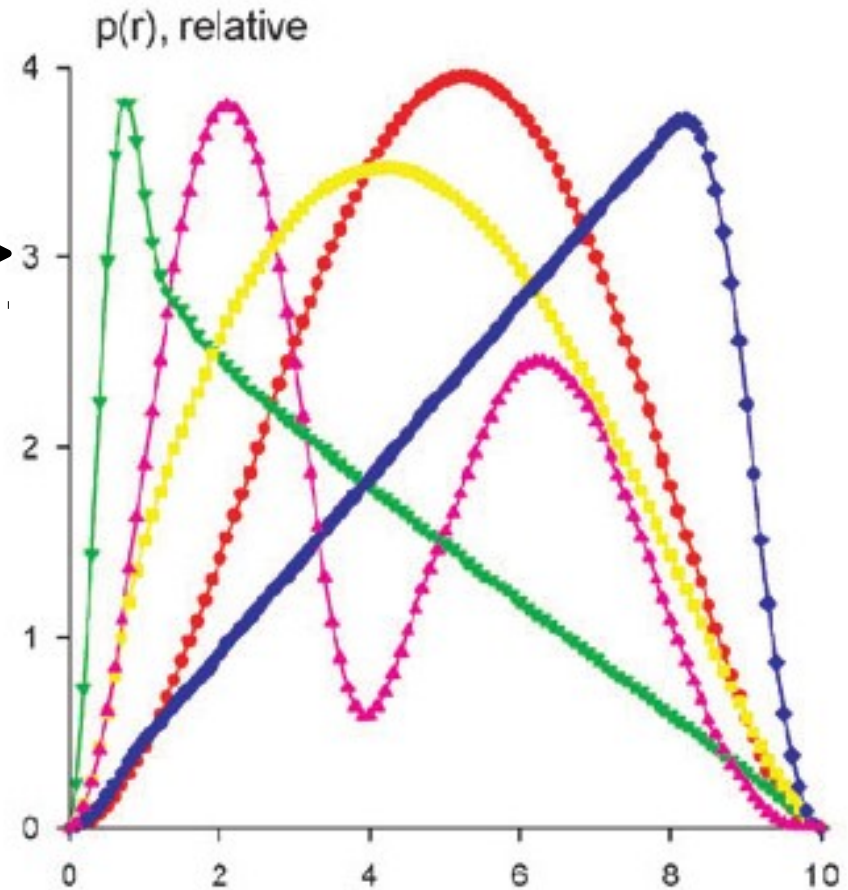
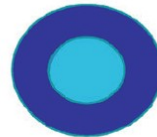
SAXS I(q) and P(r) inter-converted via FT

Momentum domain (/nm or /Å)
Intensity of scatter at given momentum

Position domain (nm or Å)
“Recollection” of position throughout space



Fourier transform



$$I(q) = 4\pi \int_0^{D_{\max}} P(r) \frac{\sin(qr)}{qr} dr$$

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

NMR correlation function and spectral density inter-converted via FT

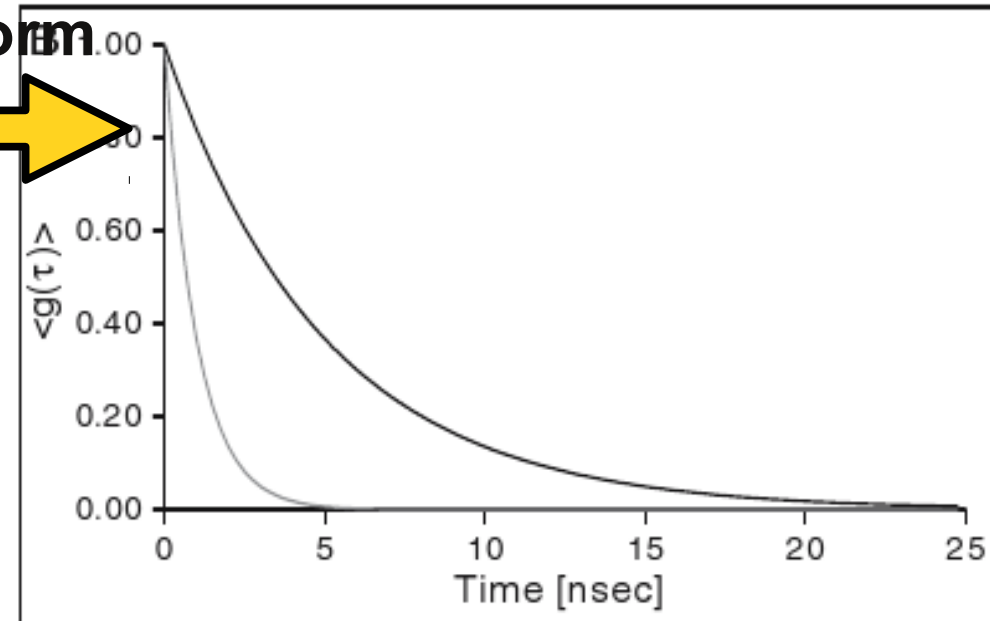
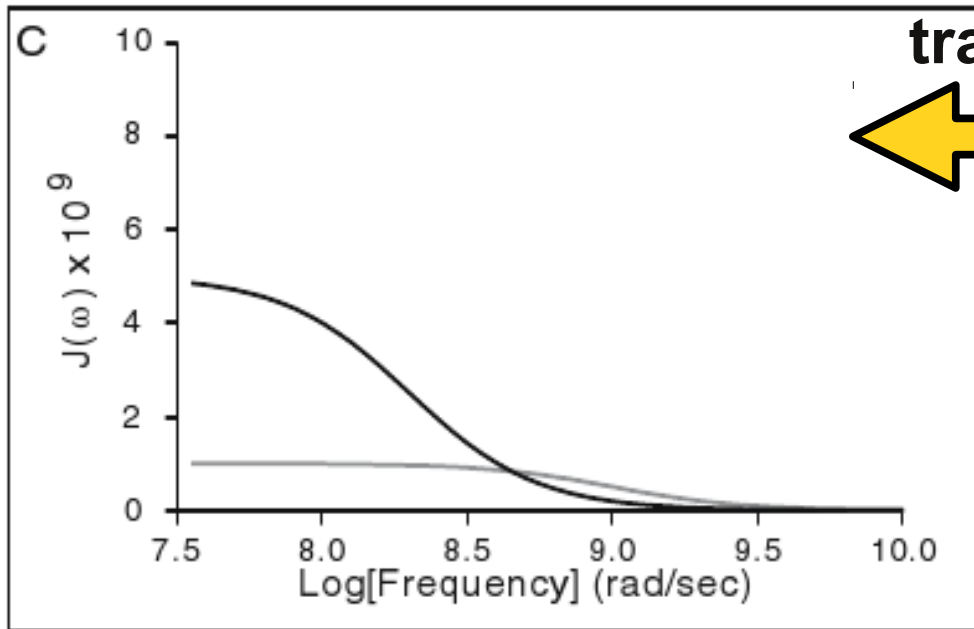
Frequency domain (/sec)

Intensity of motion at given frequency

Time domain (sec)

“Recollection” of position throughout time

**Fourier
transform**



$$g(\tau) = e^{-\tau/\tau_c}$$

$$J(\omega) = \int_0^{\infty} g(\tau) e^{-i\omega\tau} d\tau$$

$$J(\omega) = \frac{\tau_c}{1 + \omega^2 \tau_c^2}$$

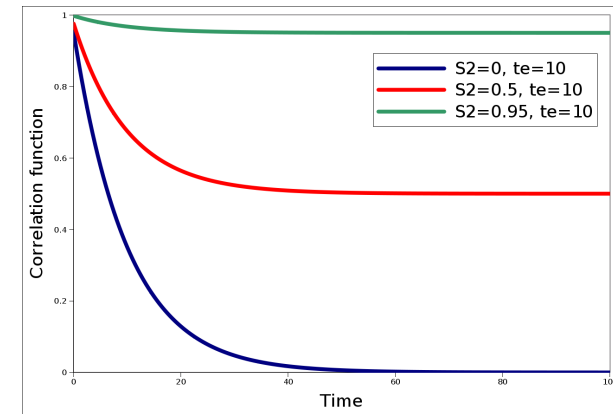
SAXS $P(r)$ and NMR $C_I(\tau)$ are somewhat* analogous

- NMR correlation function quantifies how well the bond vector “knows” its temporally-proximal (recent) orientation over time
- Characterized by the time over which the bond vector is changed from current orientation

$$C_I(\tau) = \langle P_2(\vec{u}(0) \circ \vec{u}(\tau)) \rangle$$

\vec{u} = N-H Bond vector

$$C_I(\tau) = S^2 + (1 - S^2)e^{-\tau/\tau_e}$$



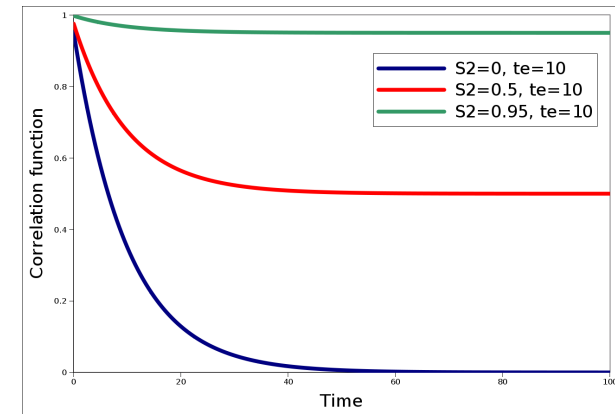
SAXS $P(r)$ and NMR $C_I(\tau)$ are somewhat* analogous

- NMR correlation function quantifies how well the bond vector “knows” its temporally-proximal (recent) orientation over time
- Characterized by the time over which the bond vector is changed from current orientation

$$C_I(\tau) = \langle P_2(\vec{u}(0) \circ \vec{u}(\tau)) \rangle$$

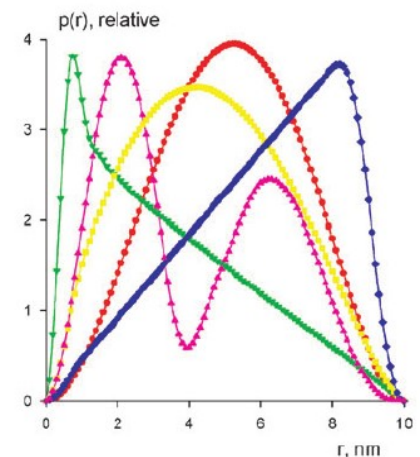
\vec{u} = N-H Bond vector

$$C_I(\tau) = S^2 + (1 - S^2)e^{-\tau/\tau_e}$$



- SAXS pair-distribution function quantifies how well the electron density “knows” its spatially-proximal density over space
- Characterized by the length over which the electron density is changed from current value

$$P(r) = r^2 \left\langle \int_V \Delta\rho(\vec{r}) \Delta\rho(\vec{u} + \vec{r}) d\vec{r} \right\rangle_{\Omega}$$



*Since $P(r)$ is radially averaged, only the magnitude of distance $|r|$ is important. This would be akin to the NMR correlation function using only the magnitude of time $|t|$ representing both prior and future orientations of the bond-vector

SAXS interpretables are numerous and redundant

- There are many ways to compute these interpretables
- Estimates of accuracy are sometimes ill-defined
- Therefore it is important to seek consistency
 - E.g., R_G from Guinier should agree with R_G from $P(r)$
 - Discrepancies may indicate a flawed interpretation

Parameter	Formula	Range of Data Used and Variable Definitions	Comments
Radius of Gyration (R_G): Guinier Approximation	$\ln[I(q)] = \ln[I(0)] - \frac{q^2 R_G^2}{3}$	$qR_G < 1.3$ globular, $qR_G < 0.8$ elongated. $I(0)$: Intensity of the scattering profile extrapolated to $q=0$	Most common method of estimating R_G . Measured via the slope of the plot $\ln[I(q)]$ vs. q^2 .
Radius of Gyration (R_G): Debye Approximation	$I(q) = \frac{2I(0)}{q^4 R_G^4} (q^2 R_G^2 - 1 + e^{-q^2 R_G^2})$	$qR_G < 1.4$ for elongated	Particularly useful for elongated proteins where the Guinier approximation is valid over narrower range.
Radius of Gyration (R_G): defined by $P(r)$	$R_G^2 = \frac{\int_0^{D_{max}} r^2 P(r) dr}{\int_0^{D_{max}} P(r) dr}$	Entire q range. D_{max} : Maximum dimension of particle	Good consistency check for R_G , D_{max} , and $P(r)$
Pair Distribution Function ($P(r)$)	$P(r) = \frac{r}{2\pi^2} \int_0^\infty I(q) q \sin(qr) dq$	Entire q range.	Indirect Fourier transform methods have been developed for calculating $P(r)$.
Maximum Dimension (D_{max})	D_{max} is the value of r at $P(r)=0$ for large r	Requires data $q \leq \pi/D_{max}$	Assignment of D_{max} may be complicated by flexibility or multimerization.
Particle Volume (V): Porod Invariant	$V = 2\pi^2 I_{exp}^2(0) / \left(\int_0^\infty I(q) q^2 dq \right)$	Entire q range. $I_{exp}(0)$ is the experimental intensity at $q=0$ and does not require an absolute scale.	The integral portion of this equation is known as the Porod invariant. Accuracy varies for shape and size; however absolute scale and concentration information are unnecessary.
$I(0)$: Intensity at $q=0$ which is also proportional to Mass and Volume	$I(0) = 4\pi \left(\int_0^{D_{max}} P(r) dr \right)$	Entire q range.	Calculation of M and V using this version of $I(0)$ is less susceptible to aggregation and inter-particle correlations than extrapolation of low q data.

Parameter	Formula	Range of Data Used and Variable Definitions	Comments
$I(0)$: Intensity at $q=0$ which is also proportional to Mass and Volume	$I(0) = 4\pi \left(\int_0^{D_{max}} P(r) dr \right)$	Entire q range.	Calculation of M and V using this version of $I(0)$ is less susceptible to aggregation and inter-particle correlations than extrapolation of low q data.
Mass (M)	$M = \frac{I(0) \mu^2}{N_A (1 - (\rho_s/\rho_p))^2}$	μ : Average mass per number of electrons. ρ_s : Solvent electron density. ρ_p : Particle electron density. N_A : Avogadro's number.	$I(0)$ must be on an absolute scale and normalized by mass/volume and not molar concentration.
Formulas for elongated or flexibly linked linear macromolecules			
Radius of Gyration of cross-section (R_{XC})	$\ln[qI(q)] = \ln[qI(0)] - \frac{q^2 R_{XC}^2}{2}$	Intermediate q values.	The slope of the linear portion of a plot of $\ln[qI(q)]$ vs. q^2 is R_{XC}^2 ; however, R_{XC}^2 goes to 0 as q goes to 0 in regimes where scattering is dominated by R_G .
Length (L)	$L = (12(R_G^2 - R_{XC}^2))^{1/2}$	See R_G and R_{XC} .	The co-axial length rather than the hypotenuse (D_{max}).

<i>Parameters</i>	<i>Assessment</i>
<i>Experimental</i>	
q -range	Range must be suitable through the entire spatial resolution required for determined models
Guinier plot	Non-linear behavior indicates aggregation or inappropriate q -range.
R_G	Consistency of extracted R_G with multiple methods (Table 1) increases confidence in not only R_G but also assigned D_{max} .
$I(0)$	Should correlate with molecular weight and concentration
D_{max}	Proper description of the range of D_{max} for well behaved $P(r)$ functions
$P(r)$	High frequency oscillations or discontinuities in $P(r)$ may indicate problematic Fourier transform process.

Radius of gyration R_G is not radius of particle

R_g : radius of gyration

$$R_g^2 = \frac{\int \Delta\rho(r)r^2 dV}{\int \Delta\rho(r)dV}$$

$$R_g^2 \approx \frac{\sum_j \Delta n_j r_j^2}{\sum_j \Delta n_j} \quad \text{excess electrons}$$

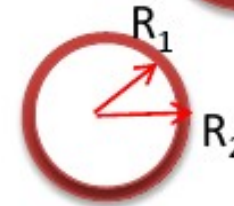
Sphere
(radius R)

$$R_g^2 = (3/5)R^2$$



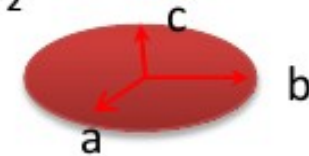
Hollow sphere
(radii $R_1 < R_2$)

$$R_g^2 = (3/5) \frac{R_2^5 - R_1^5}{R_2^3 - R_1^3}$$



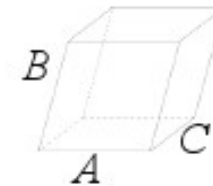
Ellipsoid
(semi-axes a, b, c)

$$R_g^2 = (1/5)(a^2 + b^2 + c^2)$$



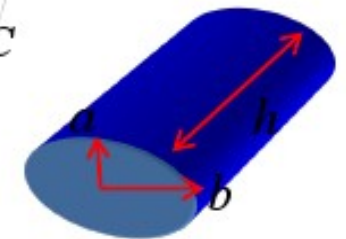
Parallelepiped
(edge length A, B, C)

$$R_g^2 = (1/12)(A^2 + B^2 + C^2)$$



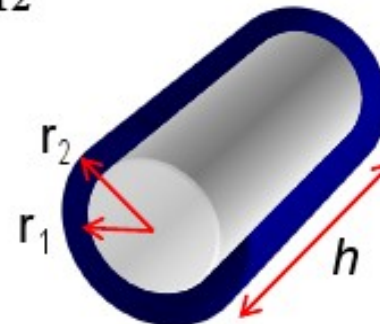
Elliptic cylinder
(semi-axes a, b ; height h)

$$R_g^2 = \frac{a^2 + b^2}{4} + \frac{h^2}{12} = R_c^2 + \frac{h^2}{12}$$



Hollow cylinder
(height h , radii r_1, r_2)

$$R_g^2 = \frac{r_1^2 + r_2^2}{2} + \frac{h^2}{12}$$

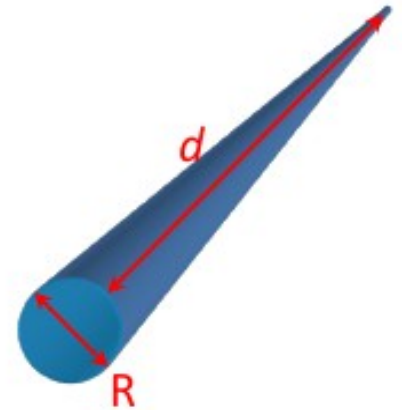


Radius of gyration R_G is not radius of particle

Rod-like and lamellar particles

Elliptic cylinder
(semi-axes a, b ; height h)

$$R_g^2 = \frac{a^2 + b^2}{4} + \frac{h^2}{12} = R_c^2 + \frac{h^2}{12}$$



Rod/needle-like particle: $d \gg R$

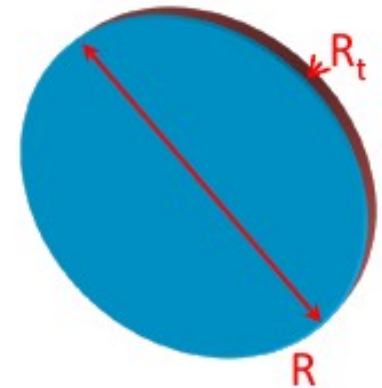
$$qI(q) = I_c(q) = I_c(0) \exp\left(-\frac{R_c^2 q^2}{2}\right)$$

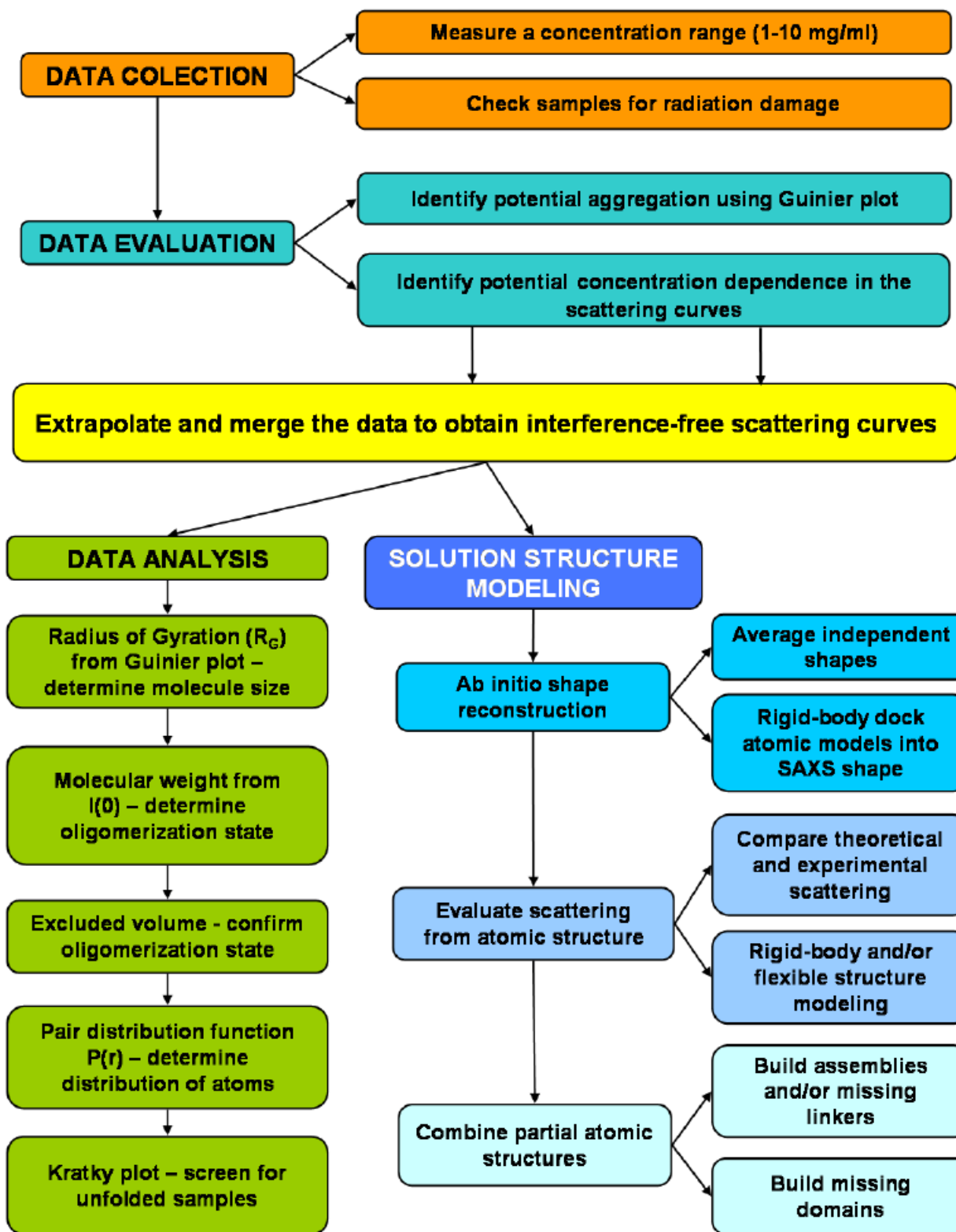
2-d analog of R_g : cross-section
$$R_c^2 = \frac{\int \Delta\rho_c(r) r^2 dr}{\int \Delta\rho_c(r) dr}$$

Lamellar/disk/sheet-like particle: $R \gg R_t$

$$q^2 I(q) = I_t(q) = I_t(0) \exp(-R_t^2 q^2)$$

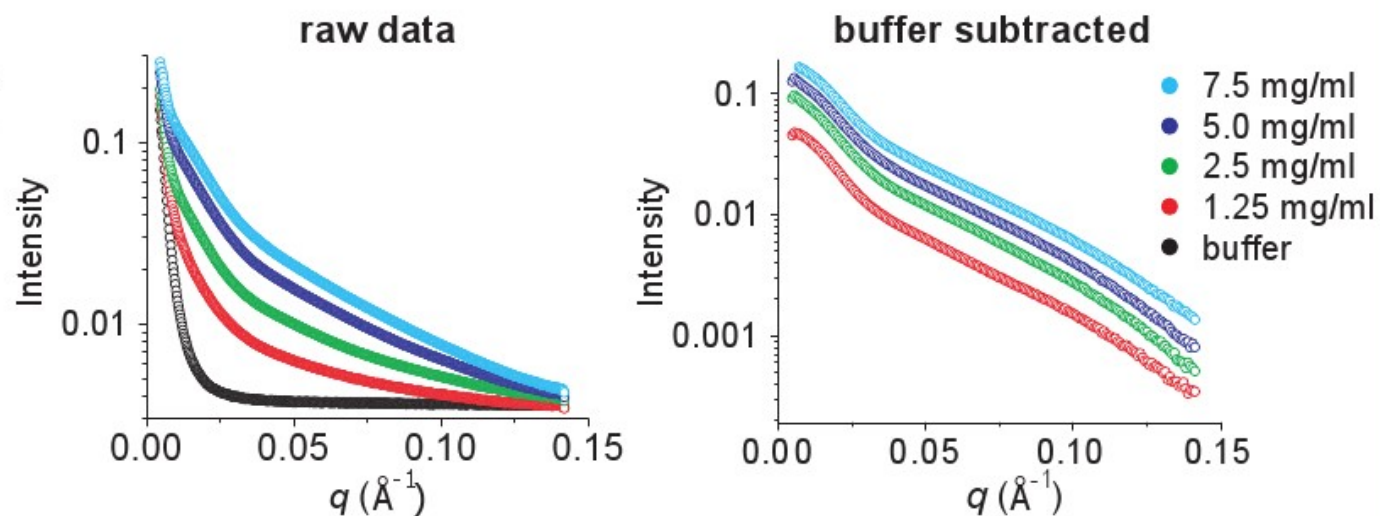
1-d analog of R_g : thickness
$$R_t^2 = \frac{\int \Delta\rho_t(r) r^2 dr}{\int \Delta\rho_t(r) dr}$$





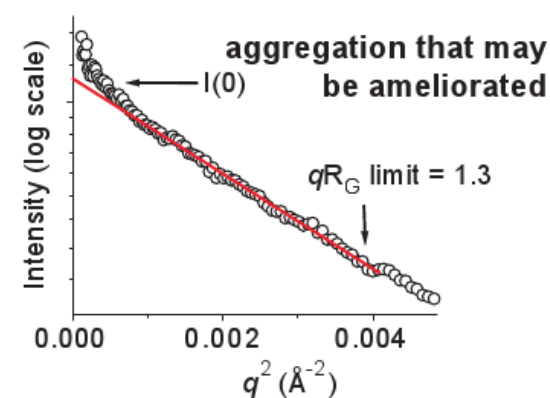
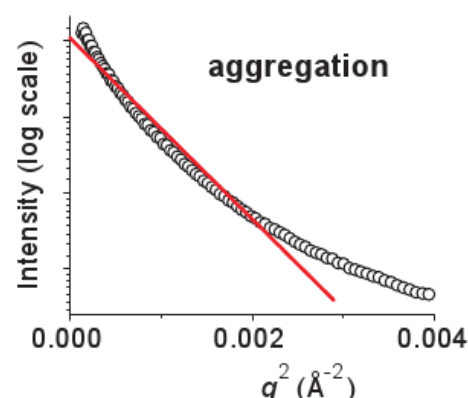
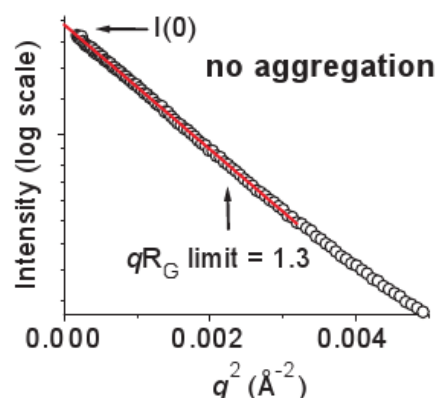
DATA COLLECTION

Measure four different concentrations in the range 1-10 mg/ml. X-ray sensitivity can be detected by changes in scattering observed in repeated exposures and typically lead to increases in R_G and $I(0)$. X-ray damage can be minimized by adding 5-10% glycerol.

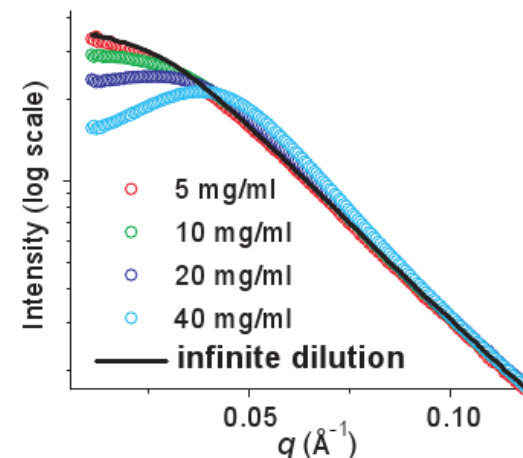


DATA EVALUATION

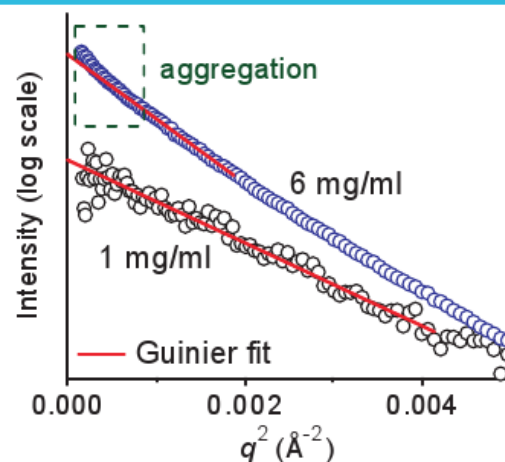
Calculate R_G and $I(0)$ from the Guinier plot. A non-linear dependence of $\log(I(q))$ vs. q^2 indicates the presence of aggregation. **Scattering from aggregated samples strongly influences the entire data set and no further processing should be performed!** Aggregation in the sample can be reduced or eliminated by varying buffer conditions, centrifugation, and filtration.



Evaluate the effects of interparticle interference $F(q)$ with the scattering of individual particles $S(q)$. Superposition of scaled scattering curves from multiple concentrations can reveal concentration-dependent scattering. Decreases in intensity at very small q with increasing protein concentration indicates interparticle repulsion. Repulsion can be dealt with by dilution, increasing ionic strength, and/or extrapolating the data to infinite dilution. Scaling the data to extrapolated scattering is only required in the range where the interference is visible, typically $q < 0.1 \text{\AA}^{-1}$.

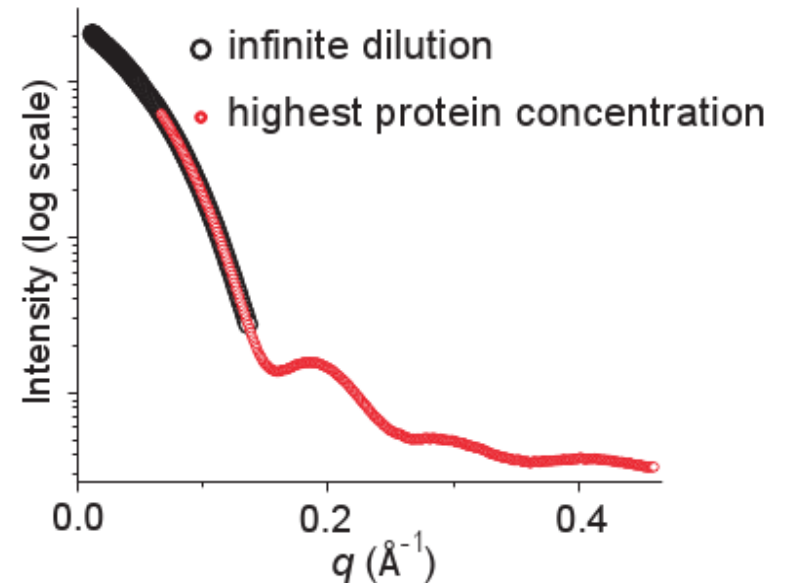


Watch for increasing $I(0)/c$ and R_G with increasing concentration. These can be due to changes in multimerization state or increasing amounts of aggregation in the sample.



sample	R_G	$I(0)/c$
6 mg/ml	30 \AA	94
1 mg/ml	22 \AA	40

To obtain an ideal scattering curve for the entire q range, the scattering profile must be extrapolated to infinite dilutions at low resolution ($q < 0.1 \text{ \AA}^{-1}$) and merged with the scattering profile for larger angles. Accurate large angle data can be obtained by measuring higher concentrations, using longer exposure times, and/or decreasing the sample-to-detector distance.

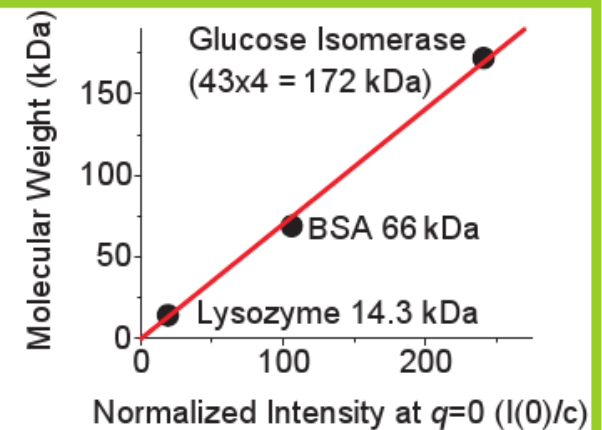


If SAXS data are collected from monodisperse samples without interparticle interference, then reconstruction of the solution structure can proceed.

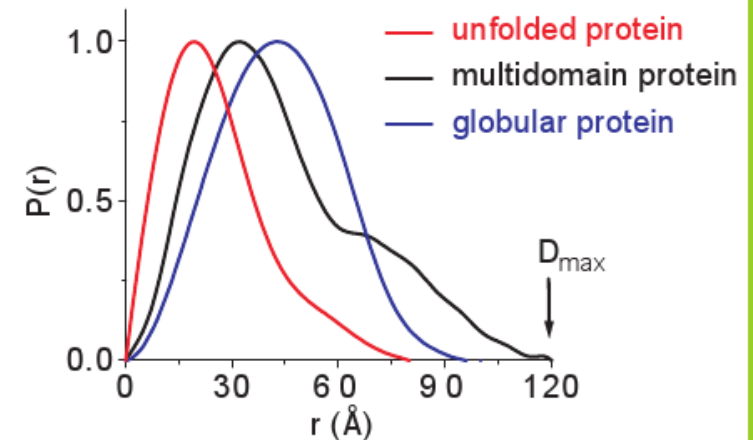
DATA ANALYSIS

R_G , a measure of the overall size of the molecule, can be determined from the Guinier plot.

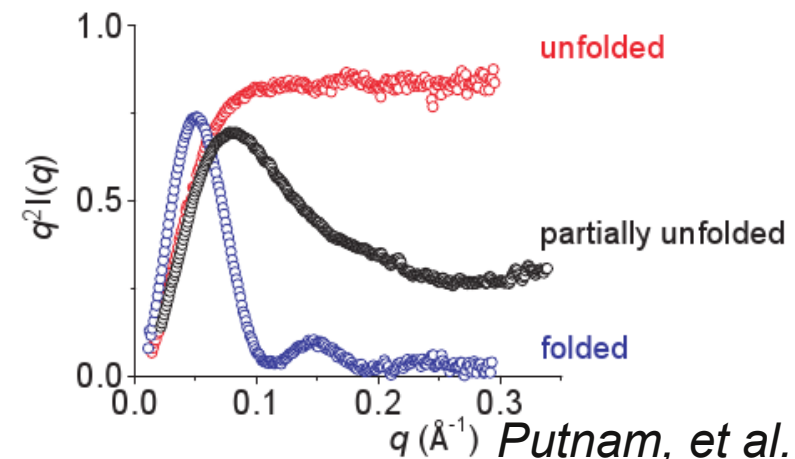
Molecular weight from $I(0)$ requires a calibration curve and can be used to determine **oligomerization state**. This can be validated using calculation of the **excluded volume**.



Globular macromolecules have a $P(r)$ function with a **single peak**, while elongated macromolecules have a longer tail at large r and can have multiple peaks. The maximum length in the particle, D_{max} , is the position where the $P(r)$ function returns to zero at large values of r . **Disagreements for values of R_G and $I(0)$** calculated from the $P(r)$ function and from the Guinier plot can indicate small amounts of aggregation that primarily affect the low resolution data and the accuracy of the Guinier plot.



The **Kratky plot identifies unfolded samples**. Globular macromolecules follow Porod's law and have bell-shaped curves. Extended molecules, such as unfolded peptides, lack this peak and have a plateau or are slightly increasing in the larger q range.

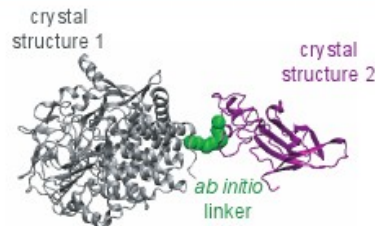


Putnam, et al. (2007)

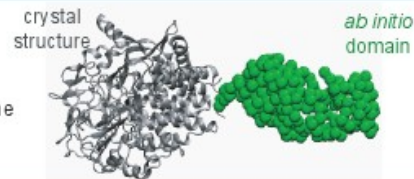
SOLUTION STRUCTURE MODELING

**Partial atomic structure:
rigid body modelling +
ab initio reconstruction**

Rigid body modeling does not identify candidate assemblies with appropriate scattering and/or linkers between domains are long. Investigate building multidomain models with BUNCH using individual domain structures and potentially combining multiple scattering curves when SAXS data from deletion constructs are available.

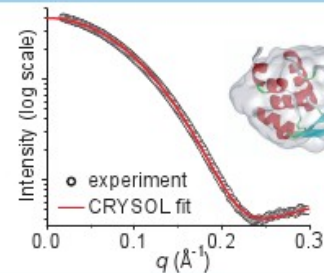


Structure of one domain is known, but others are missing. Construct missing domains using CREDO in combination with the known fragment.

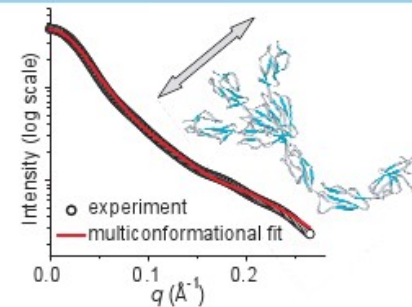


**Available atomic structure:
evaluate theoretical scattering
reconstruction**

Compare experimental SAXS scattering to SAXS scattering calculated from atomic structures. Perfect matches indicate that the solution and crystal structures are essentially identical.

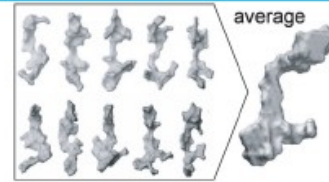


Theoretical and experimental scattering can disagree. a) Investigate fitting with alternate assembly states. If symmetry is suspected, try GLOB SYMM. b) Investigate fitting with mixtures of different assembly states. c) Investigate the possibility of flexibility between domains and missing loops that may need to be modeled in a multi-conformational way.



**No atomic structure:
ab initio shape
reconstruction**

Average *ab initio* envelopes from multiple runs. To avoid problems due to over fitting and to ensure a reproducible solution, calculate multiple individual envelopes and use the averaged solution.



Determine the low resolution envelope with *ab initio* calculations. When the number of residues are known at the data extends to reasonably large values of q ($q_{max} > 0.3 \text{ \AA}^{-1}$), GASBOR is recommended as it provides some advantages for anisotropic shapes.