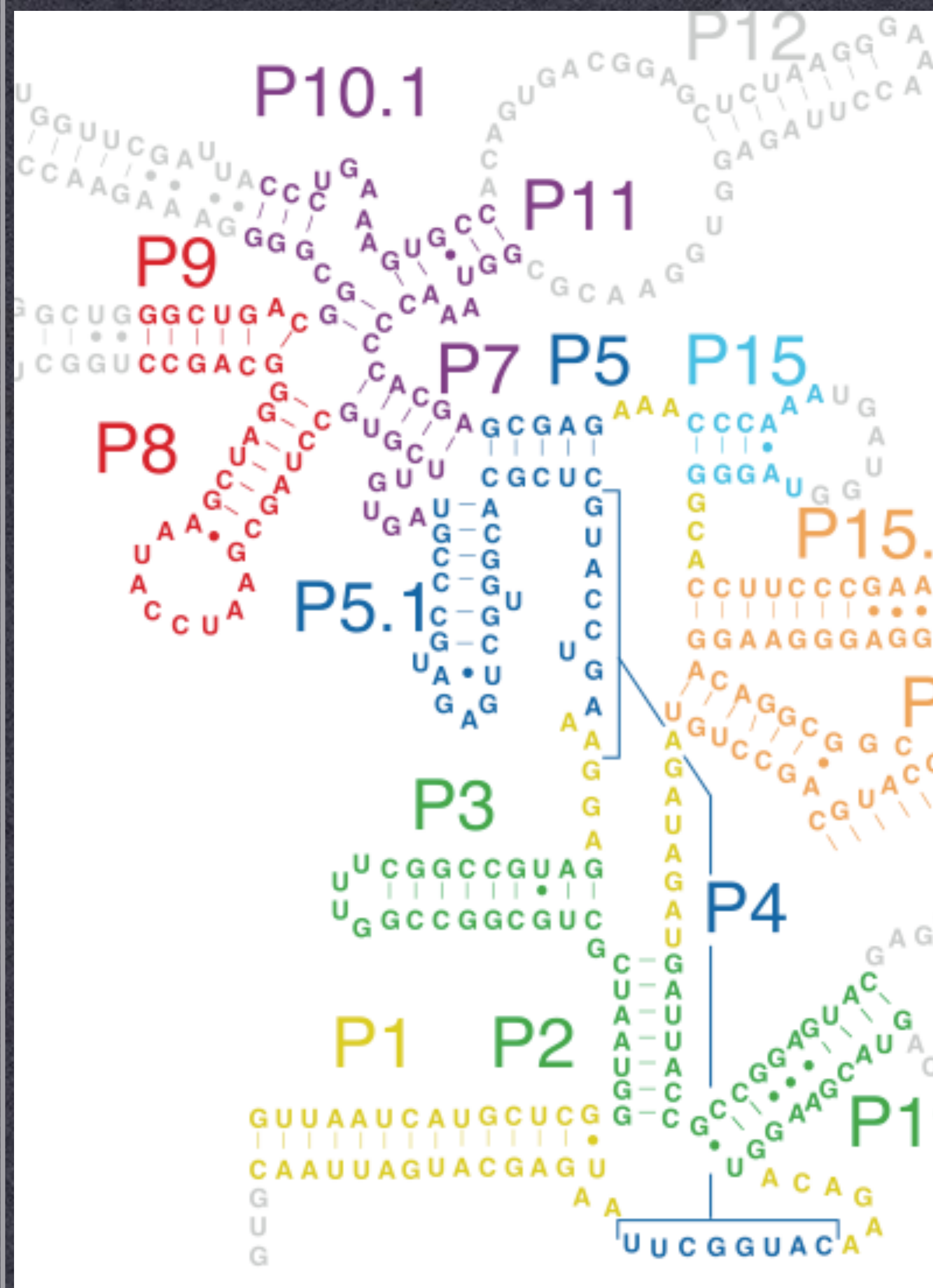
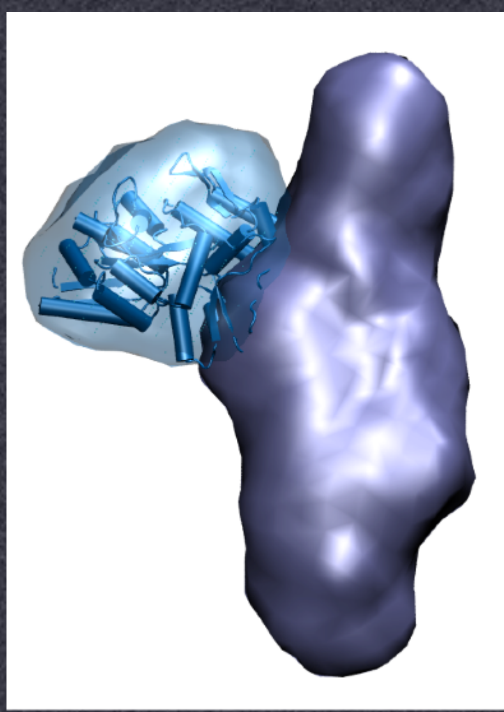
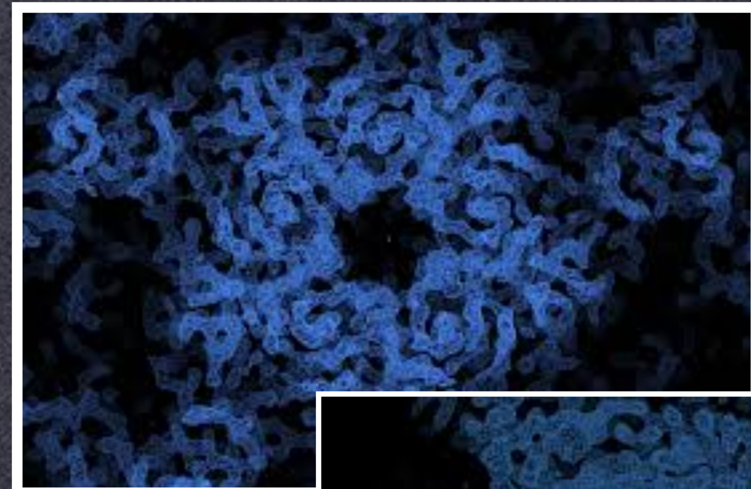
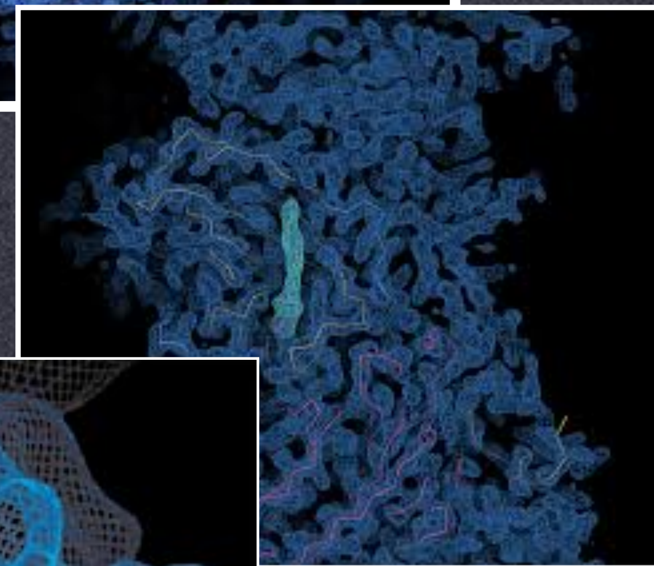
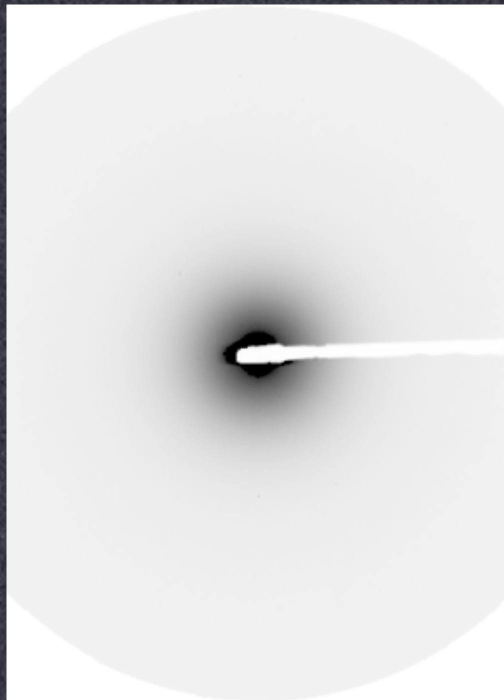
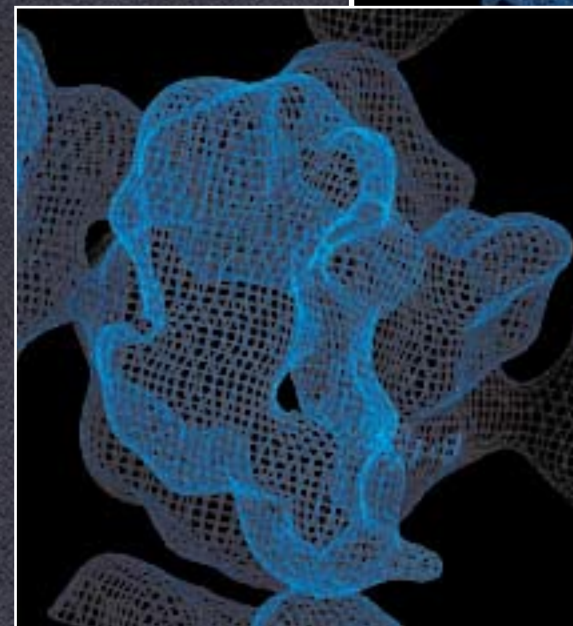












ALS

SAXS BASICS

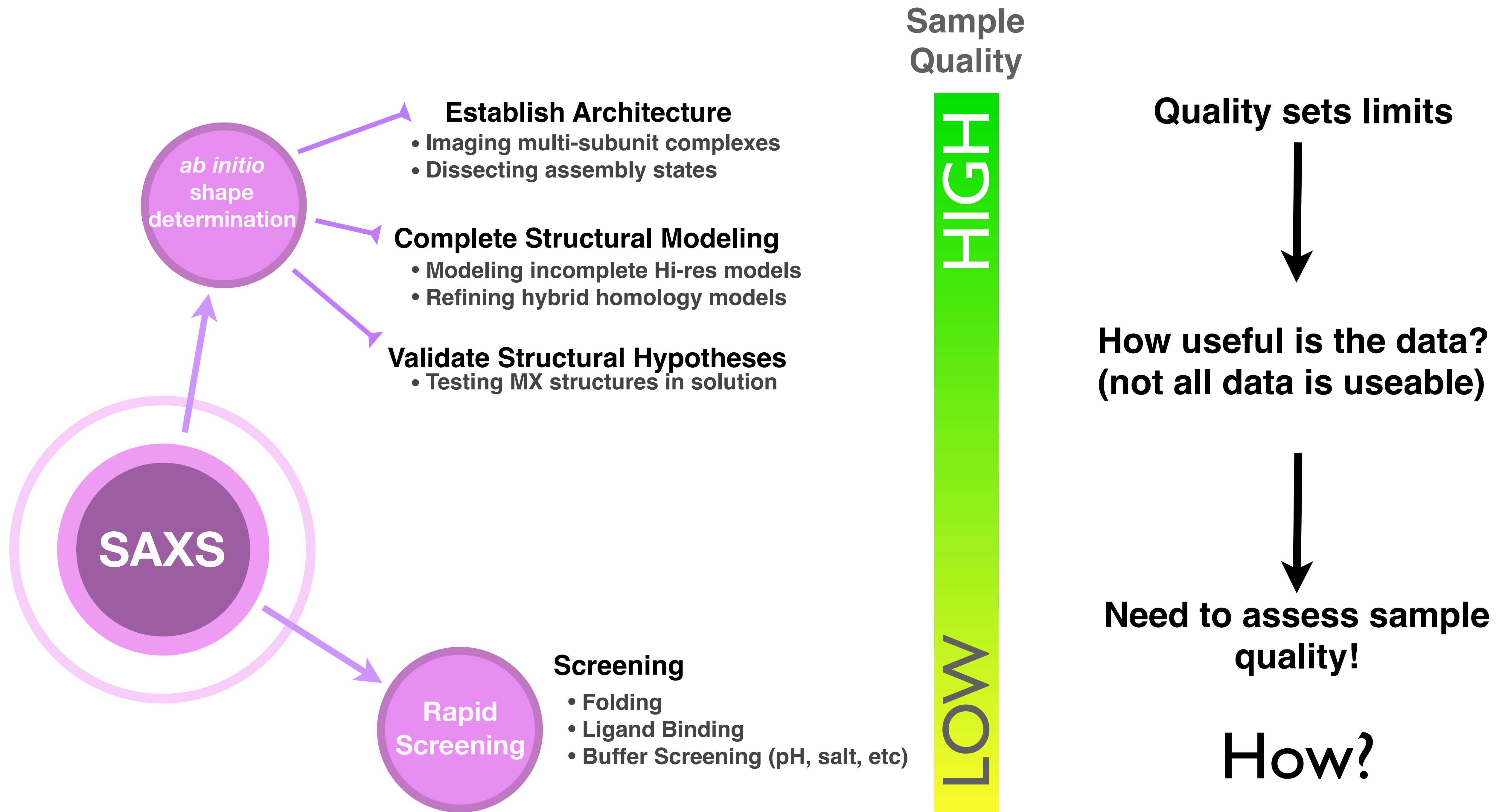
B21 DIAMOND LIGHT SOURCE

DATE OCT 22, 2018

ROBERT P RAMBO

Sample Quality

for SAXS



What do you want from SAXS?

What is a SAXS signal?

- Basics of a SAXS curve

Data quality

- instrumentation background
- sample cell
- radiation damage

Sample quality

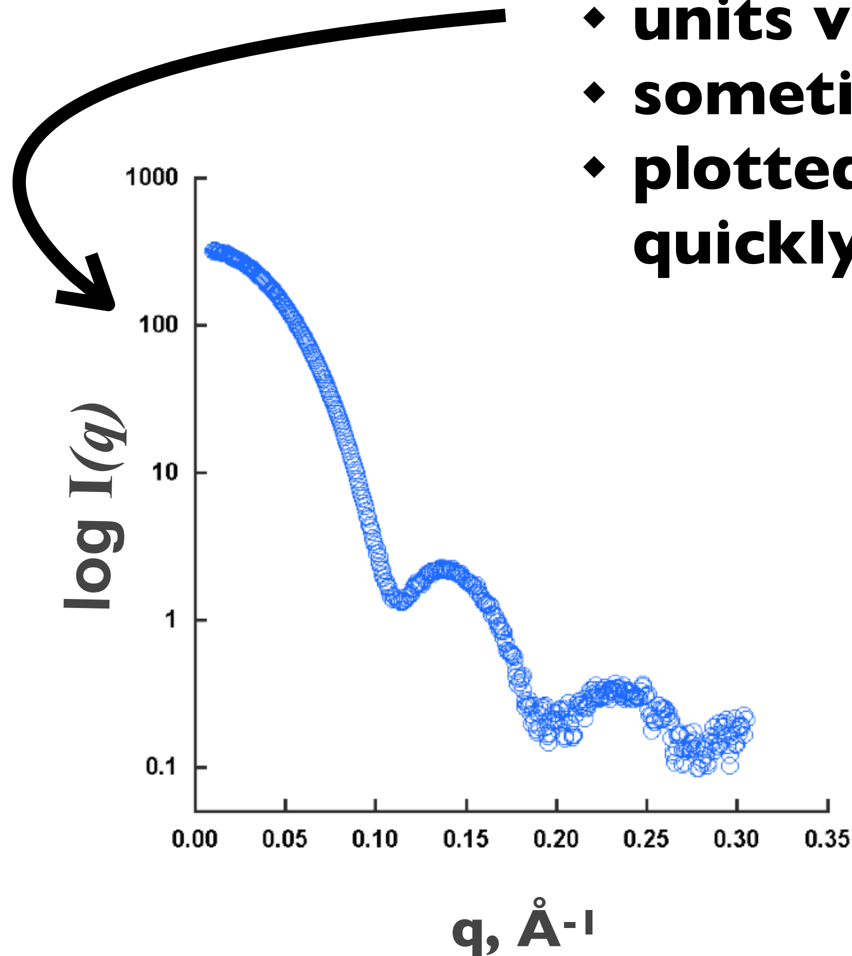
- heterogeneity (aggregation)
- buffer matching
- concentration dependent scattering
  - ★interparticle interference or multimerization

How to tell good from bad?

- instrumentation background (smiling in Guinier residuals)
- sample cell (positive contribution to high-q)
- radiation damage (smiling in Guinier residuals)

SAXS: Basics of a curve

- **I(q) measured intensities**
  - **units vary**
  - **sometimes calibrated to water**
  - **plotted as $\log_{10}$ of I(q), intensities decay quickly**

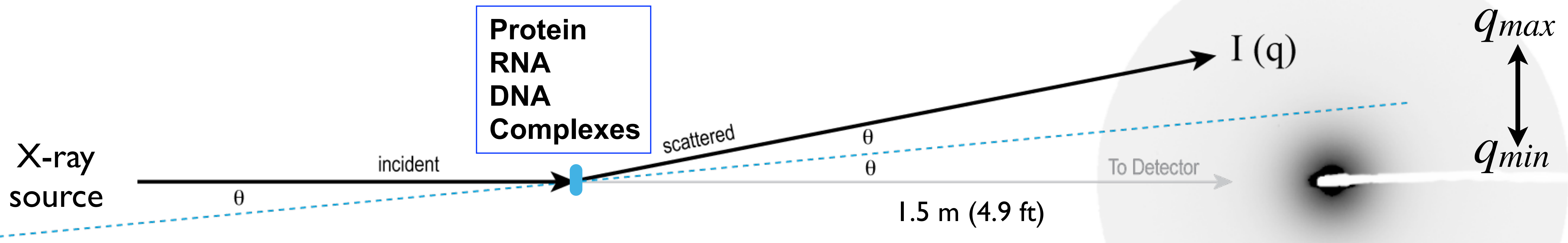


- ***q*** is a vector (*momentum transfer vector via elastic collision*)
- independent of distance to detector and wavelength (λ)
- units are $Å^{-1}$ (UK/US) or nm (EU^{-1})
- defines scattering curve in *reciprocal space*

low resolution ↔ high resolution

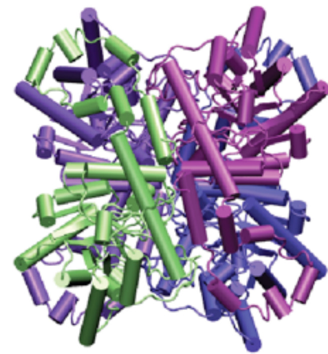
- **Features throughout the curve relate to shape**
- **At low resolution, can approximate particle as a homogenous body of electron density**
- **larger the object, the faster the I(q) decay**

Solution State SAXS



SAXS measures everything, nothing goes missing!

SAXS dilute systems measures particle form factor
SAXS at high conc measure structure and form factor and everything else

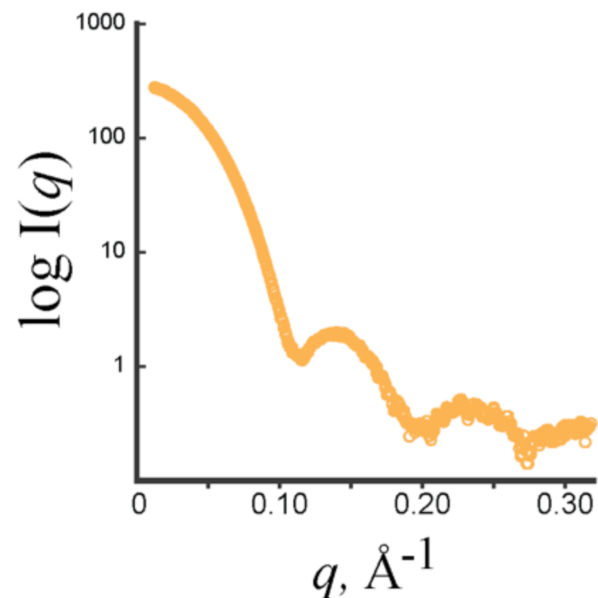


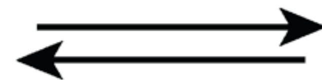
Set of all pairwise distances within particle
Resolution is seen as features in P(r)

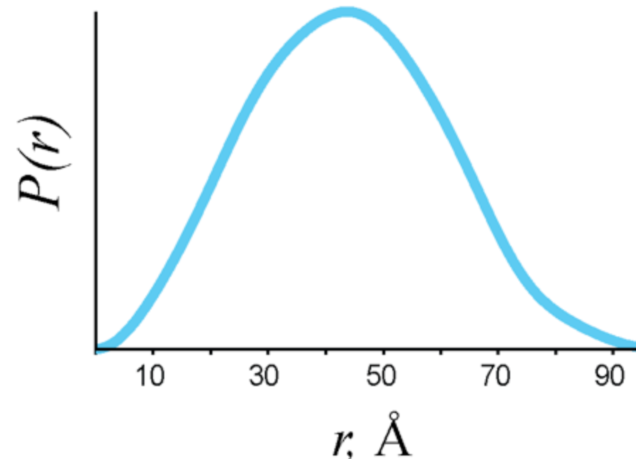
I(q) vs q
(Reciprocal Space)

P(r)-distribution
(Real Space)

$$I(q) = \int_0^{d_{max}} P(r) \frac{\sin q \cdot r}{q \cdot r} dr$$



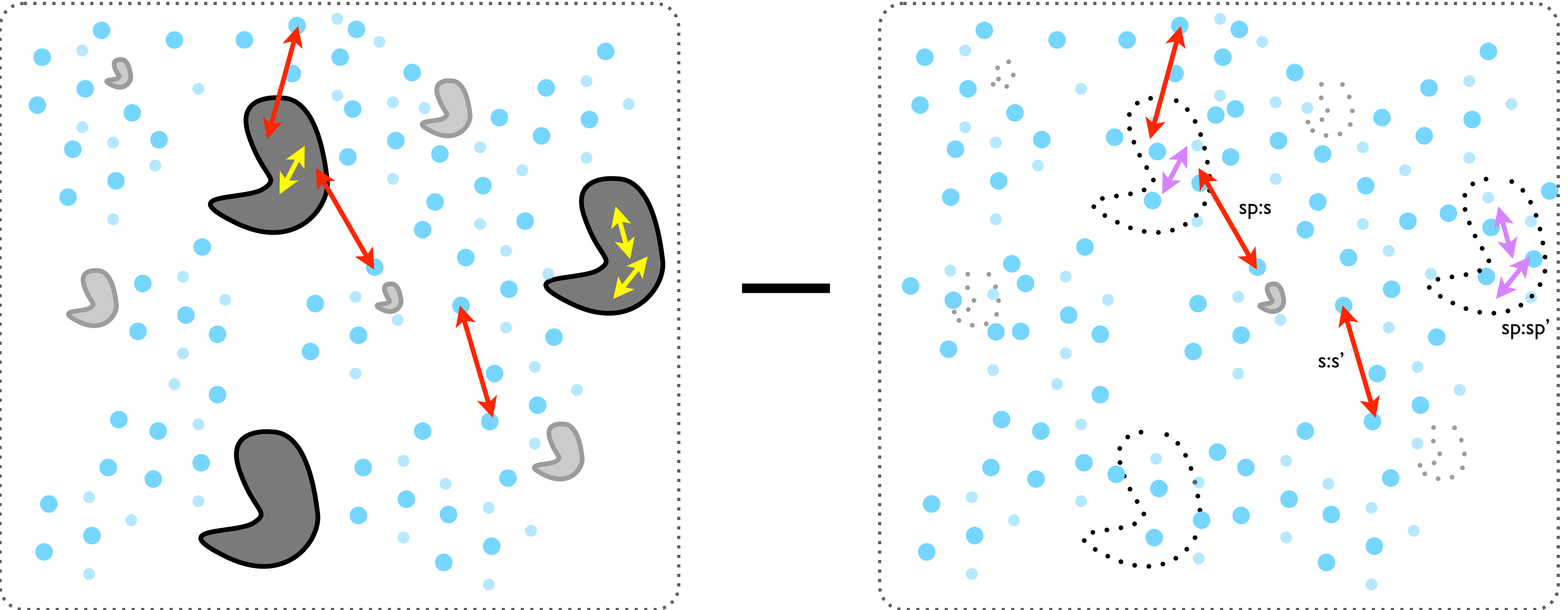




Signal from > 1,000 billion molecules

Conformational changes are thermodynamic changes of state and can be observed in P(r)

SAXS: Difference Measurement



$$I_{particle}(q) = I_{sample}(q) - I_{buffer}(q)$$

SAXS is a difference measurement

It is derived as a difference of two separate measurements!

1. **particles in a buffer (often measure many times, i.e., dilution series)**
2. **buffer (often measured once, but should measure many times)**

Difference is taken in domain of *Real Numbers!*

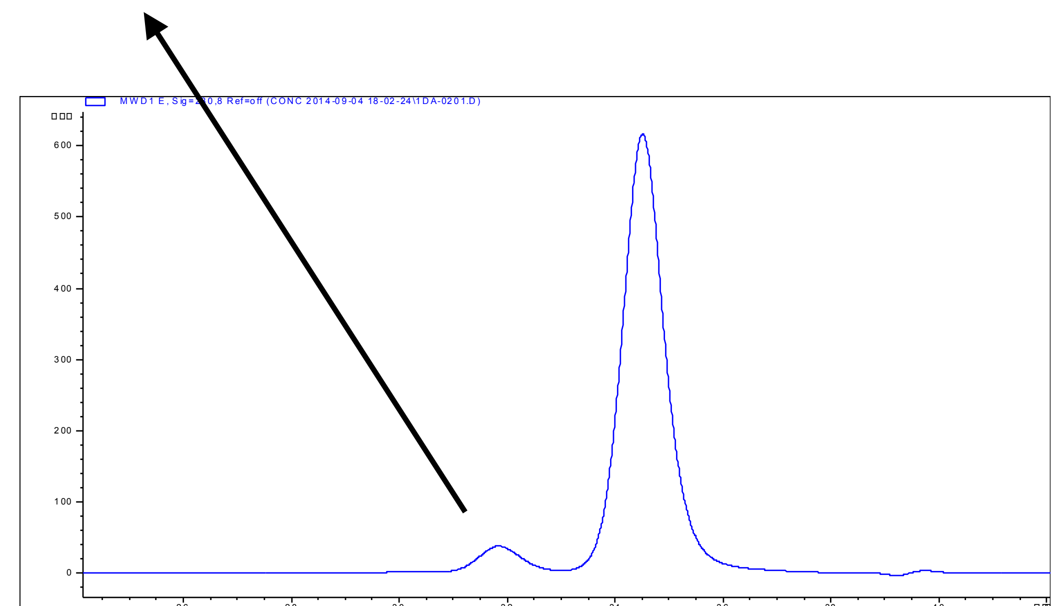
Data Collection Strategies

Batch Mode

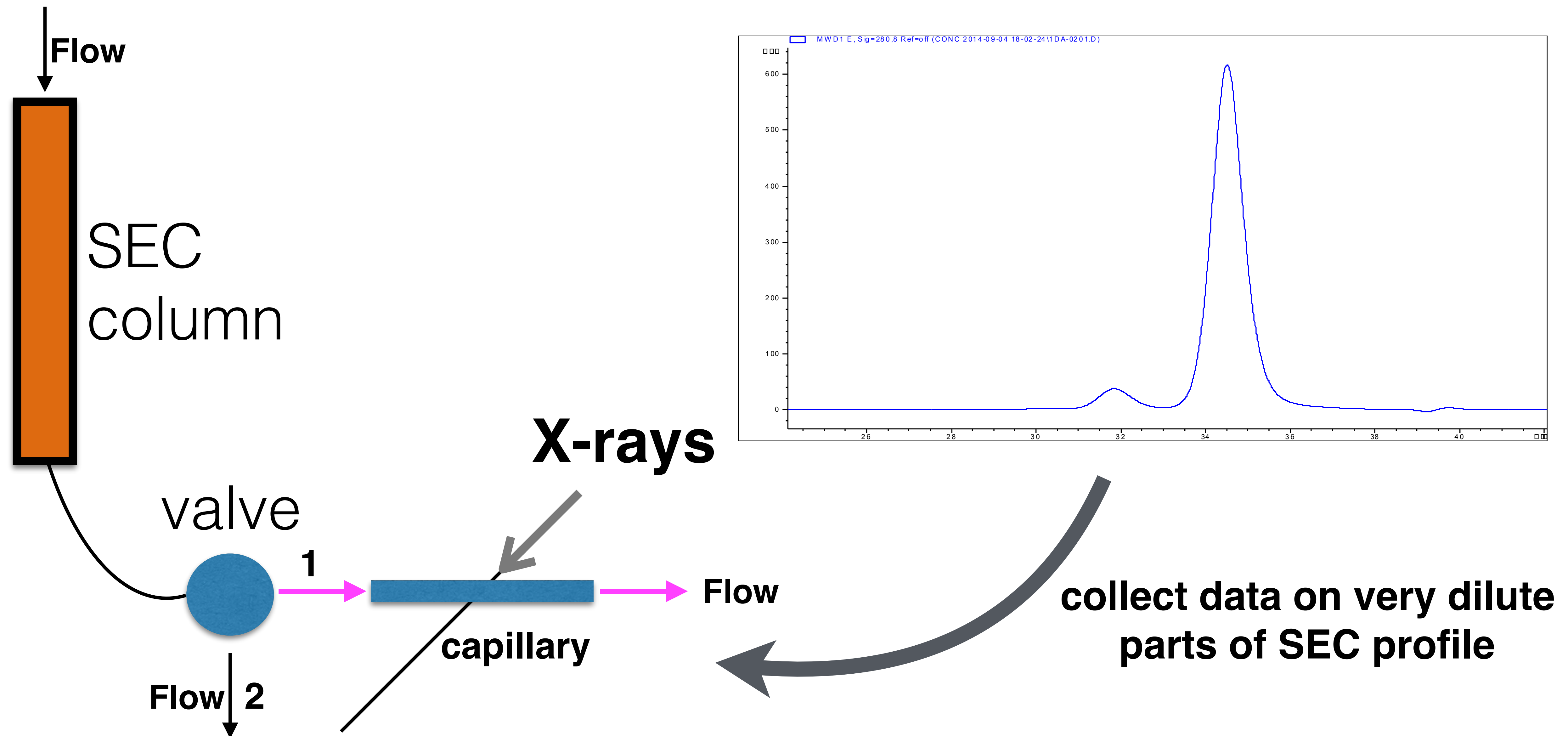
- Single Protein (20 to 35 uL per well)
- condition screening (effects of pH, TRIS, phosphate, sucrose, glycerol)
- test ligand binding (screening)
- complexation (does A and B make AB?)
- critical to buffer match

SEC SAXS (2.4 to 4.8 mL column)

- 30 to 50 uL sample
- expect 3.5x dilution of sample (inject 7 mg/ml => peak 2 mg/ml)
- get a great signal at 0.9 to 1 mg/ml
- if you can not concentrate your particle
  - try repeated runs at low concentration and average (like 10 times)



InLine SEC-SAXS



Inline SEC SAXS:

- run times (18, 30, or 60 minutes per sample)
- capture region of interest from eluting peak (~20 uL)
- can get data to high q with extended exposures
- must mitigate radiation damage

SAXS as a Structural Tool

SAXS is a solution state measurement:

- **everything in the sample contributes to scattering (no missing bits)**
- **sample quality determines information that can be derived from measurement**
- **structural assessment of the thermodynamic state (exposures 0.05 to 300 seconds)**

Easy to calculate SAXS profile from PDB (CRYSOL, FOXS, aquaSAXS, wetSAXS)

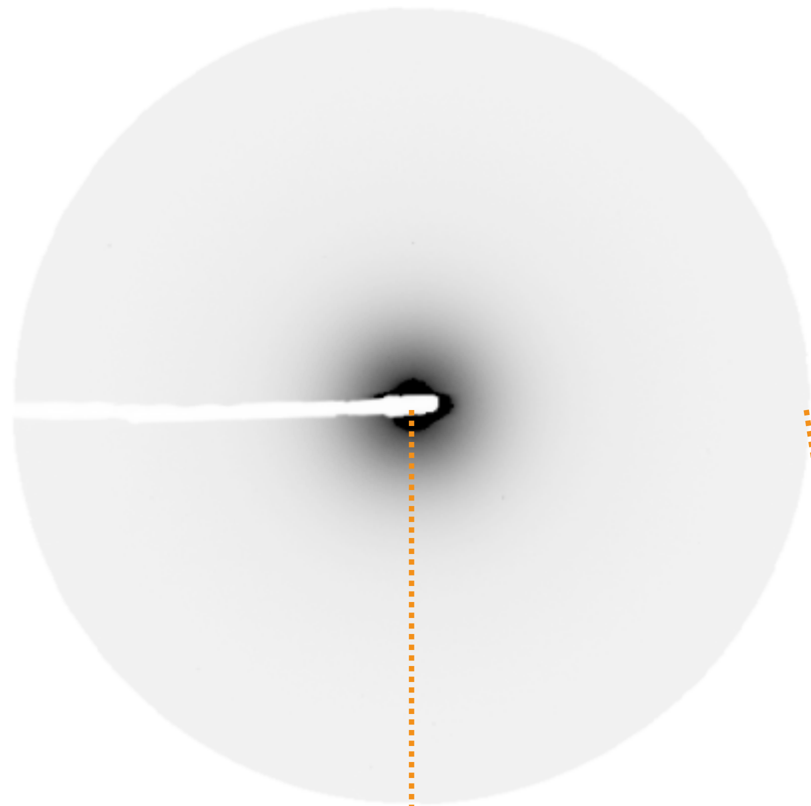
- **test structural hypothesis ⇒ Is the crystal structure the solution state?**
- **~40% of the time, MX structure explains SAXS data**

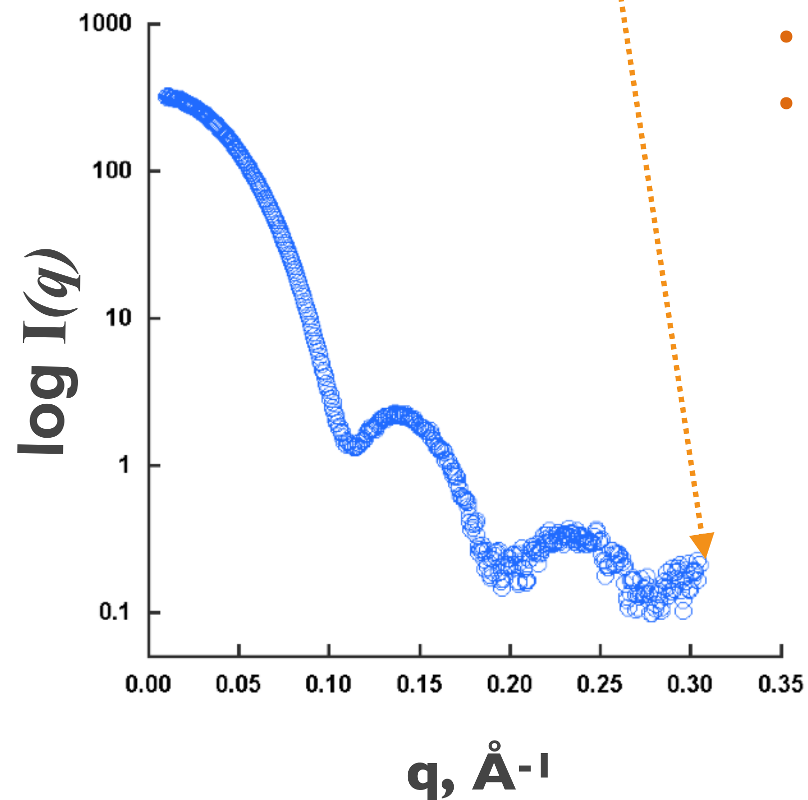
1. MX structure is incomplete
2. different oligomeric state
3. sample is a mixture of states

As an Experimental tool

- **sensitive to changes in thermodynamic state**
- **monitor conformational changes (resolution dependent)**
- **assess flexibility (conformational degrees-of-freedom)**
- **improve samples for MX or EM**
- **monitor fiber formation (SOD enzyme, microtubules)**
- **monitor gel formation**

SAXS: What is the Signal





Each point is the average around the beam centre

Error is counting error

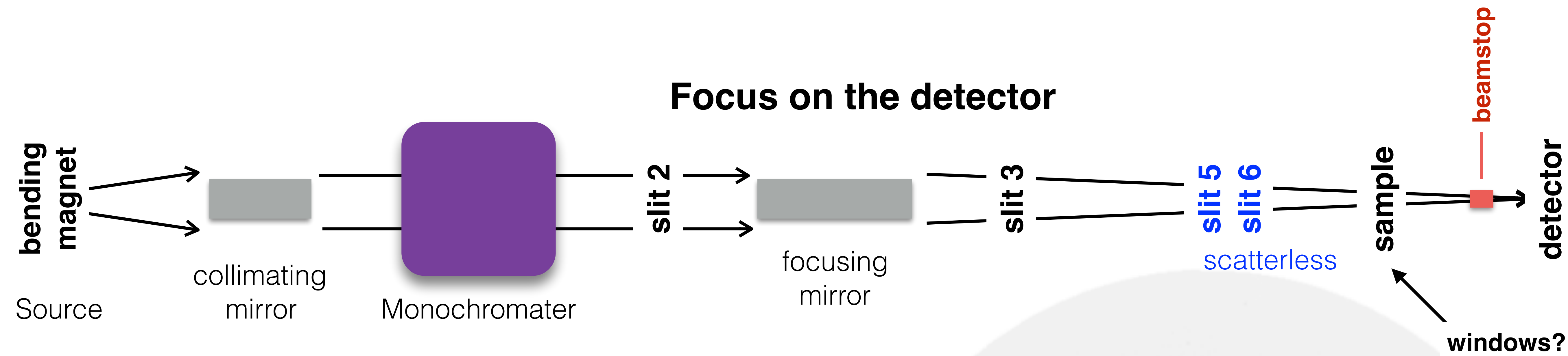
Instrumentation (systematic) errors are removed in subtraction

- Mis-alignment of beamstop
- Slit scatter
- Detector chips

Mask out parts of the detector to exclude from integration

- edge of detector
- beam stop shadow

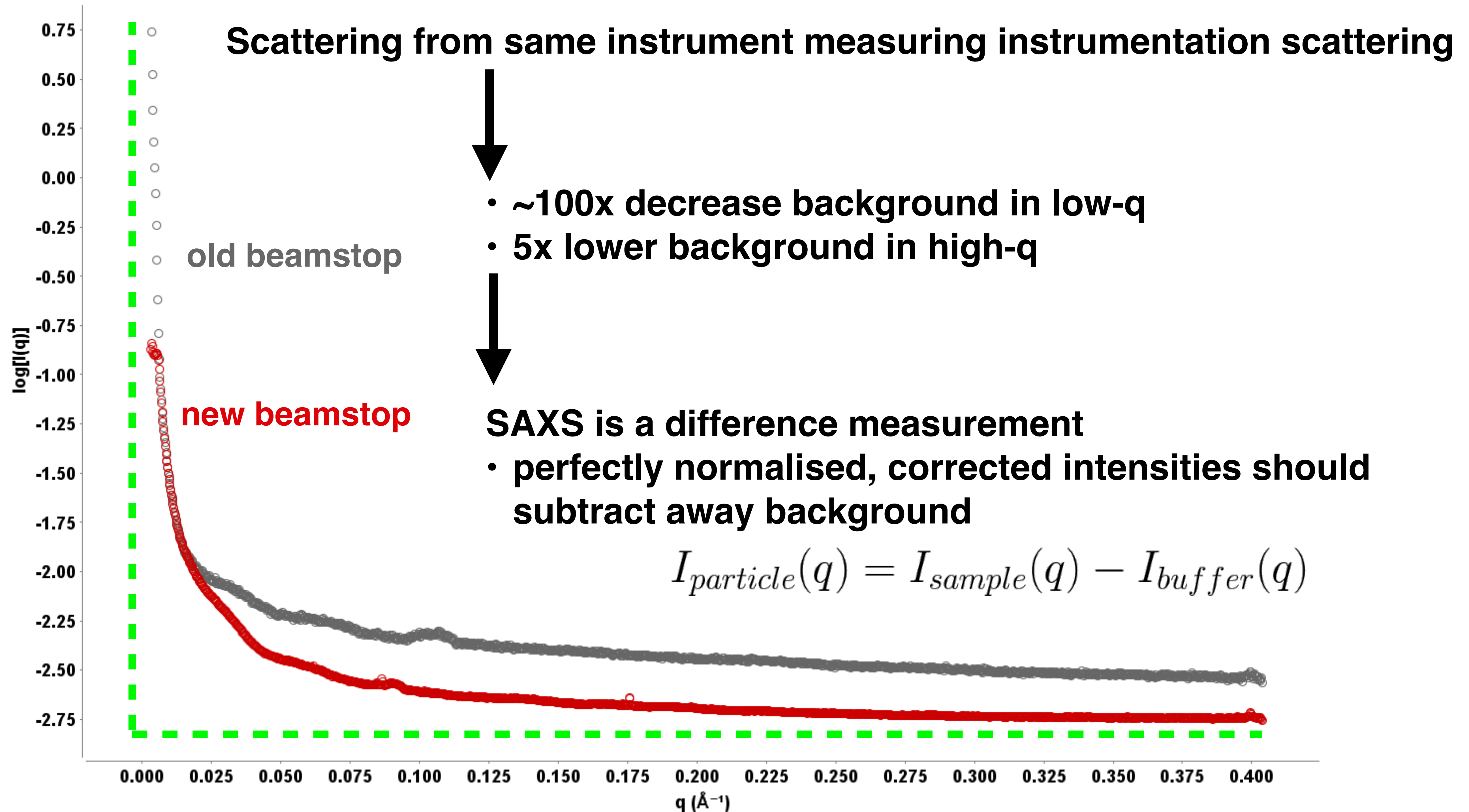
SAXS: Instrument Background



- Each of these components interact with the X-ray beam
- Each will contribute some level of scatter
  - Your sample I(q) is going to add to background I(q)
- At the detector, two types of observations
  1. focused scattering
  2. instrumentation background
- Intensity of background varies with the beam fluctuations
- Magnitude of background determines limit of detection

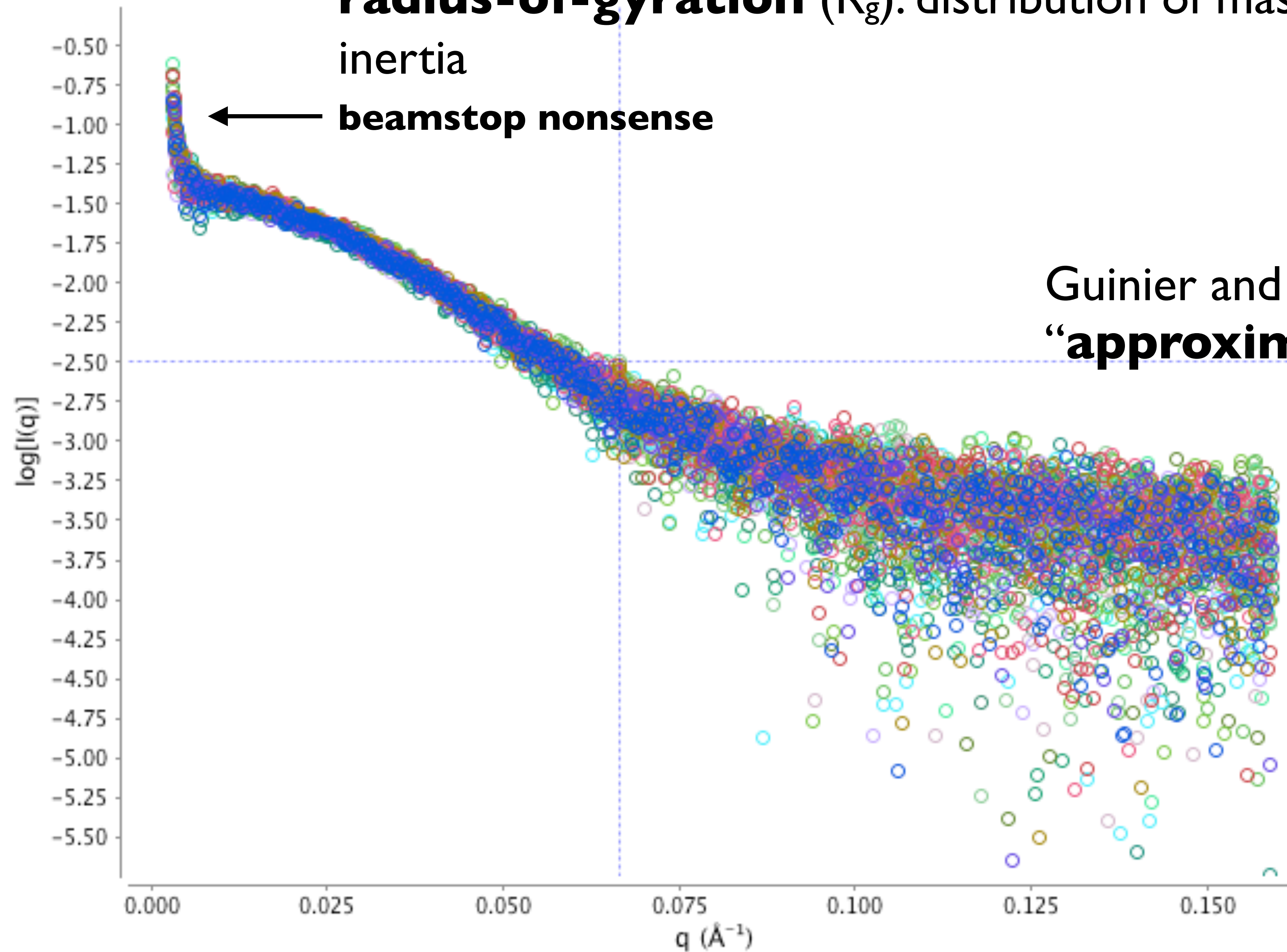
beamstop shadow

SAXS: Instrument Background



- Under 3x(very) dilute conditions, sample measurements will start to show scattering due to beamstop (positive uplift in Guinier region)
- Trim low-q data to remove it from further analysis

SAXS: Brief Diversion (Guinier)



radius-of-gyration (R_g): distribution of mass around particle's center-of-inertia

$$R_g{}^2 = \frac{1}{2} \cdot \frac{\int r^2 \cdot \rho(r)\, dr}{\int \rho(r)\, dr}$$

Guinier and Debye worked out methods which "**approximate**" R_g ⇒ It is not measured!

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

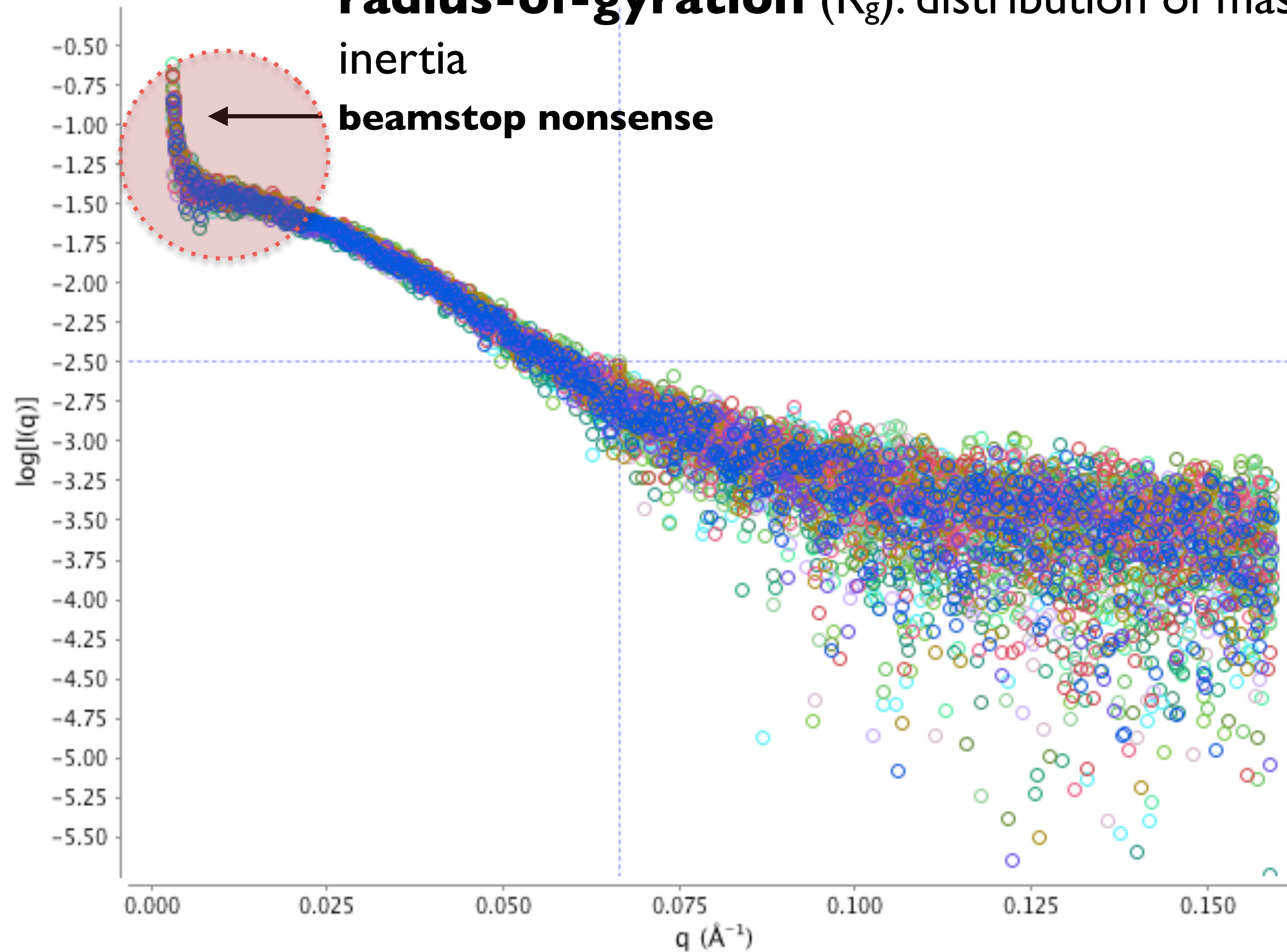
$$y \quad = \quad b \quad + \quad m \quad x$$

- thus a plot of ln*I(q)* vs q^2 will have a linear region
- I(0) is *concentration* x (Volume · $\Delta\rho$)^2 ⇒ total number of electrons^2

effective number of electrons per Å^3

SAXS: Brief Diversion (Guinier)

radius-of-gyration (R_g): distribution of mass around particle's center-of-inertia



$$R_g{}^2 = \frac{1}{2} \cdot \frac{\int r^2 \cdot \rho(r)\, dr}{\int \rho(r)\, dr}$$

- **Fitting a line with beamstop noise will over-estimate Rg**
- **Must inspect residuals of the fit**

Guinier

Small Angle X-ray Scattering

How valid is the approximation?

How well does the Guinier R_g approximate $R_g^{real\ space}$?

$q \bullet R_g < 1.0$	$q \bullet R_g < 1.3$	$q \bullet R_g < 1.5$
Particle type = 0 Points 10 to 34 fidel = 1.00 sRg limits : 0.587 to 1.01 Rg = 28.8 +- 0.673 I0 = 201.19 +- 2.23	Particle type = 0 Points 10 to 49 fidel = 1.00 sRg limits : 0.599 to 1.30 Rg = 29.4 +- 0.196 I0 = 202.94 +- 1.10	Particle type = 0 Points 10 to 68 fidel = 4.33e-2 sRg limits : 0.573 to 1.56 Rg = 28.1 +- 9.03e-2 I0 = 196.86 +- 0.714
R_g : 28.8	R_g : 29.4	R_g : 28.4
residuals	residuals	residuals
0.0005, 0.0010; s**2	0.0005, 0.0010, 0.0015, 0.0020; s**2	0.0010, 0.0020, 0.0030; s**2

16 fewer data points

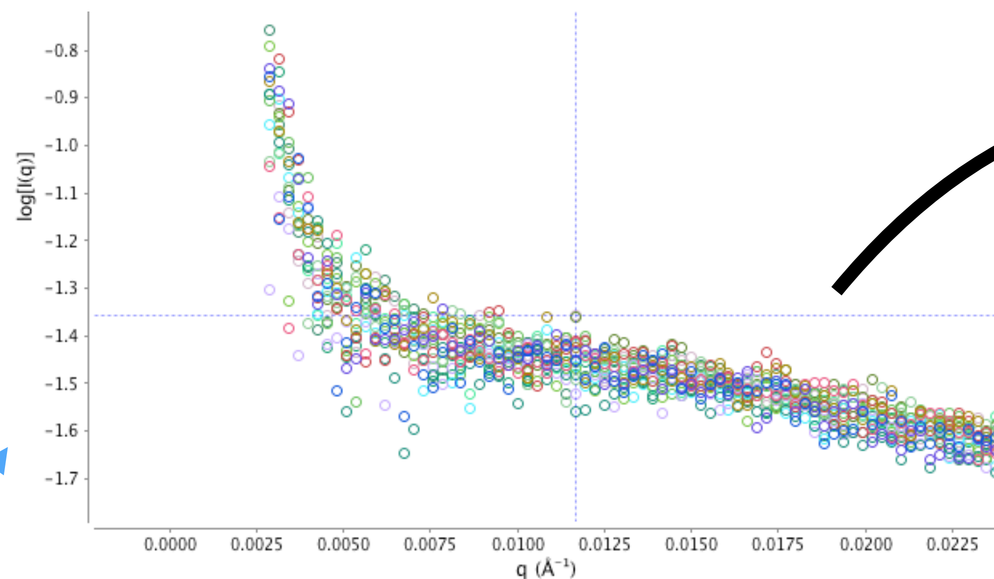
We recommend determining using data where $R_g < 1.3$

SAXS: Beamstop Noise

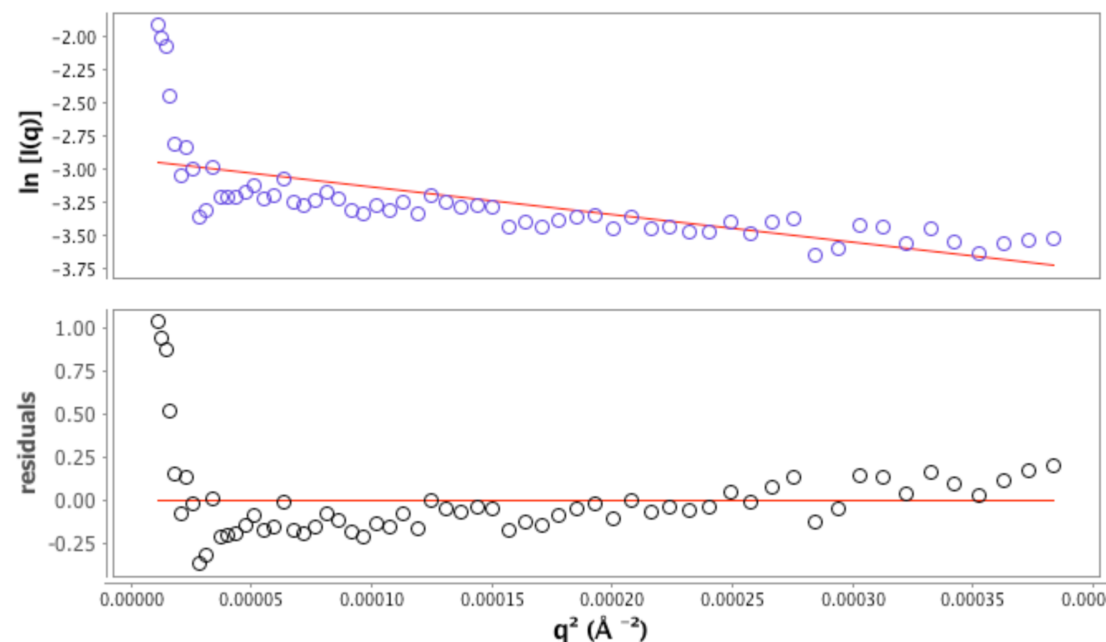
- **time-independent unlike radiation damage**

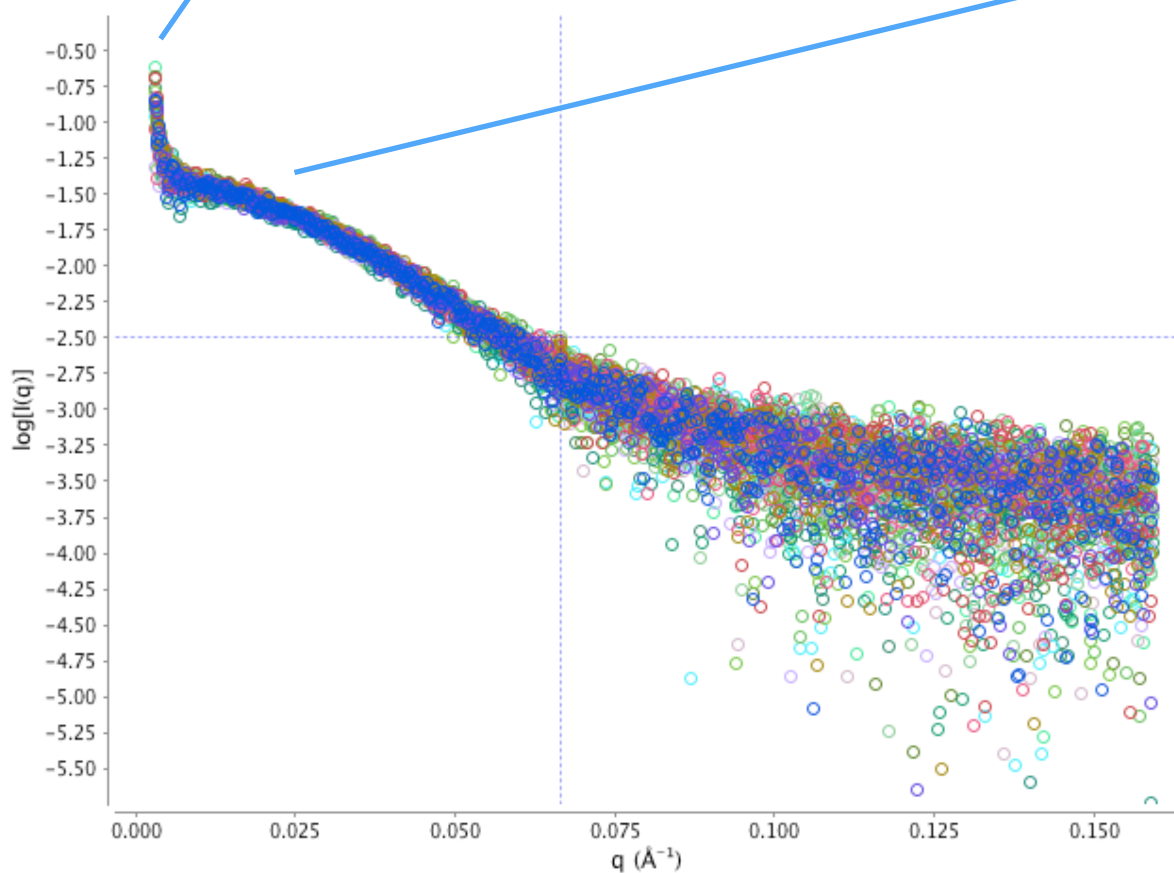
$$\ln I(q) = \ln I(0) - \frac{R_g^2}{3} \cdot q^2$$

$$y = b + m x$$

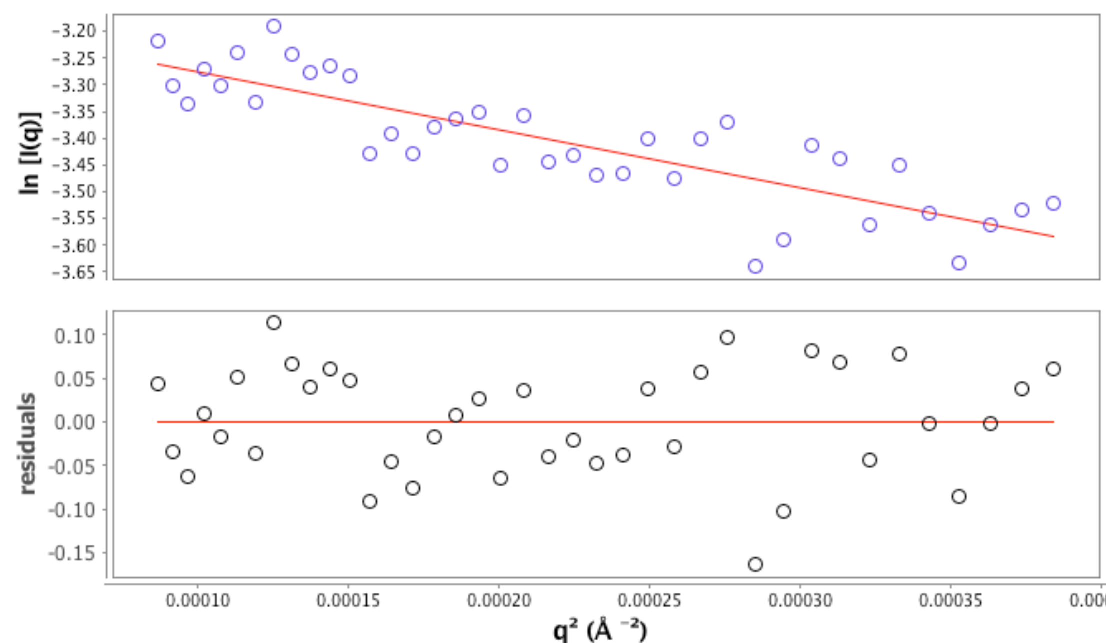


transform





- trim low-q points away
- maintain Guinier region $q_{max} \cdot R_g < 1.3$



SAXS of Different Windows

Measured at Diamond B21

log[I(q)]

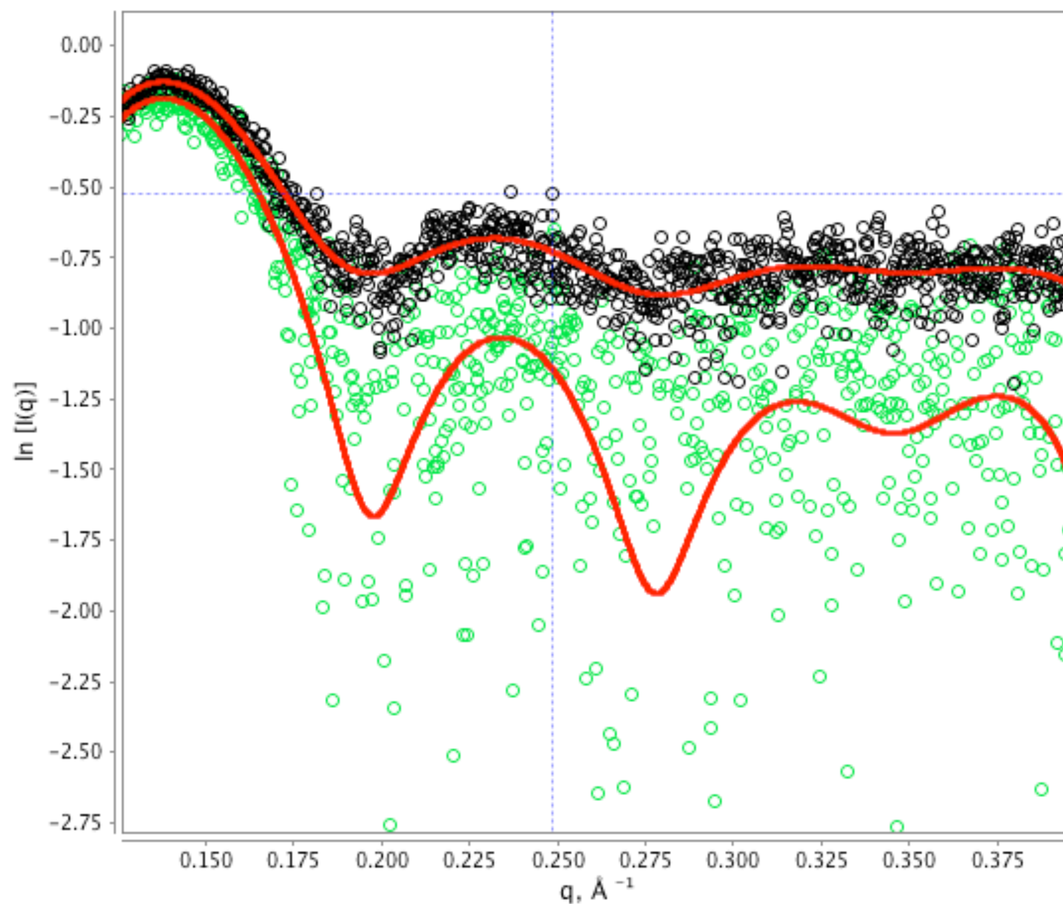
kapton (75 um)

quartz capillary (~200 um)

natural mica (~20 um)

synthetic mica (~25 um)

~100x



SEC SAXS glucose isomerase (173 kDa d_{max} 96 Å)

Path length 1.6 mm

14 x 1 sec exposures

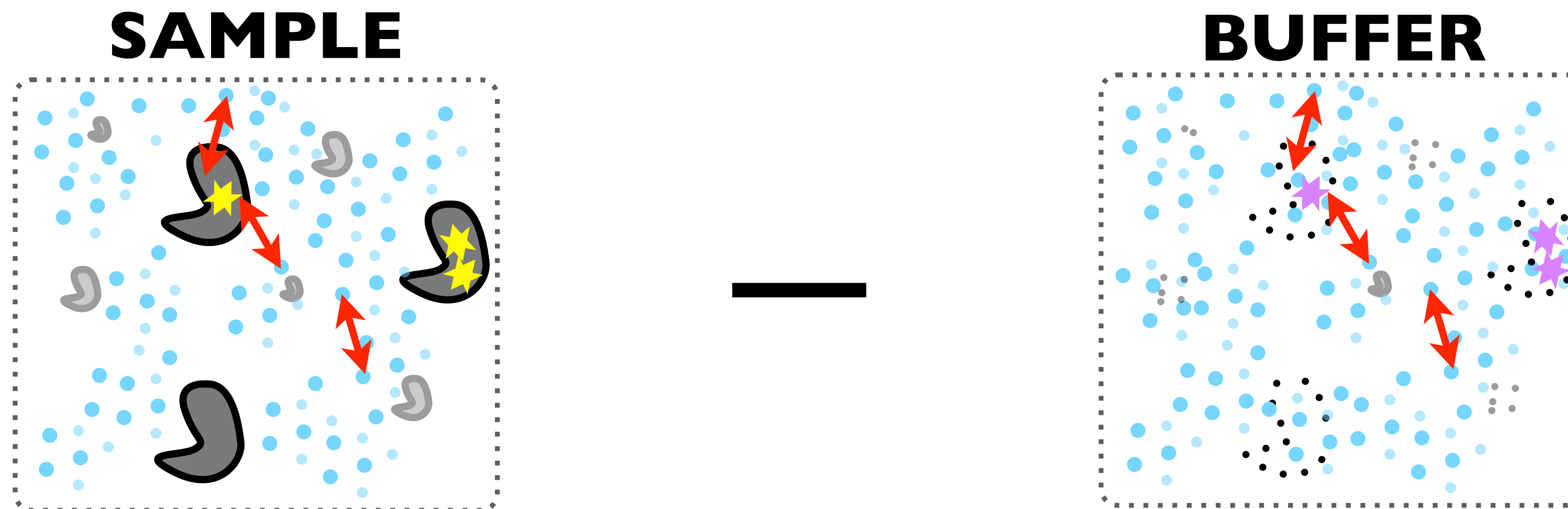
SiN cell (2 um total)

Quartz capillary (400 um total)

Changing window means more of the observed scattering is due to sample not cell

SAXS: Difference Measurement

Contributions from window materials



$$I_{particle}(q) = I_{sample}(q) - I_{buffer}(q)$$

$$I_{sample}(q) = \|A_{cell}(q) + A_{particle}(q) + A_{hydration}(q) + A_{solvent}(q)\|^2_{complex\ norm}$$
$$I_{buffer}(q) = \|A_{cell}(q) + A_{ex-volume}(q) + A_{bulk}(q) + A_{solvent}(q)\|^2_{complex\ norm}$$

at I(0): **quartz cell, 400 um thickness**
1.6 mm solvent (sample thickness)
1 mg/ml protein at 50,000 Å³

cell 1.07×10^{21} e_n^-

solvent 1.84×10^{21} e_n^-

protein 2.2×10^{17} e_n^-

- **At low q, solvent and SAXS cell will dominate scattering**
- **Ideally, want a windowless cell with low solvent content**
- **Change to SiN, cell drops to $6.8 \times 10^{18} e_n^-$ (mainly due to thick**

10^4 e_n^- ⟶ **X-ray intensities are proportional to $(e_n^-)^2$**

Summarize

- SAXS is a measurement of two
- buffer subtraction more reliable via SEC SAXS

- if low-q is important (i.e, Rg is large)
  - check q_{min}*Rg limit => $q_{min} < 1.3/R_g$
  - beamstop noise determines minimum protein concentration for reliable Guinier region
-

SAXS: What is the Signal

$$A(q) = \int \rho(r) \cdot e^{-iqr} \, dV$$

Integrated over particle

$\rho(r)$ **is electron density**

r **is a position in space**

q **is scattering vector**

e^{-iqr} **is Fourier term calculated with particle in fixed orientation**

Interaction of photon with electron described using amplitude, A(q)

Amplitude is over the entire space of the particle, call this *Molecular Form* factor

Amplitude is a complex number: norm (squared) of A(q) relates to intensity, I(q)

In SAXS, all particles are randomly oriented so, must average over all orientations

$$I(q)_{molecule} = \int_0^{d_{max}} P(r) \cdot \frac{sin(q \cdot r)}{q \cdot r} dr$$

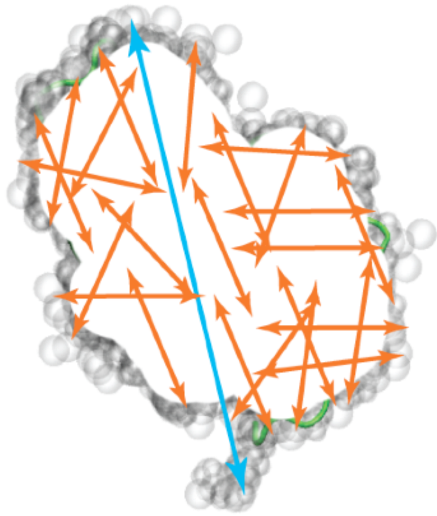
P(r) is the pair-distance distribution function

Set of all pairwise distances within particle

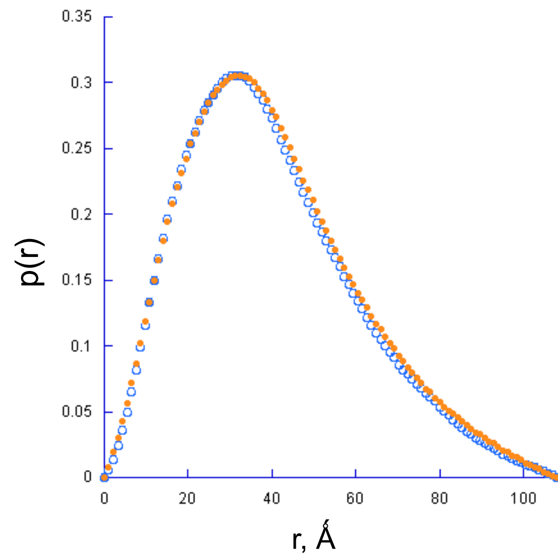
- SAXS : electron pair distances
- SANS : nuclei pair distances
- PDB : nuclei pair distances

Distance Distribution Function

Interatomic Vectors



P(r) **Function**



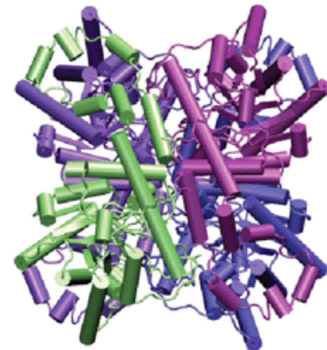
Defined on $0 < r < d_{max}$

Set of all pairwise distances within particle

- SAXS : electron pair distances
- SANS : nuclei pair distances
- PDB : nuclei pair distances

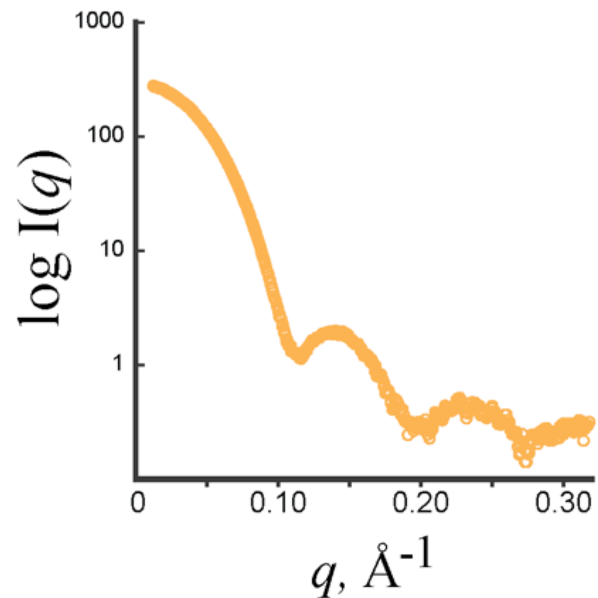
Resolution is seen as features in P(r)

In SAXS, $I_{obs}(q)$ is a sampling of the signal in reciprocal space

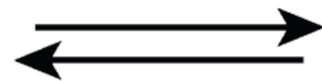


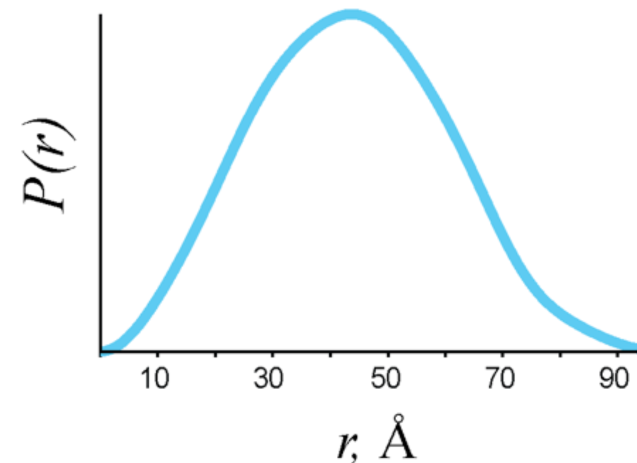
I(q) vs q
(Reciprocal Space)

P(r)-distribution
(Real Space)



$$I(q) = \int_0^{d_{max}} P(r) \frac{\sin q \cdot r}{q \cdot r} dr$$

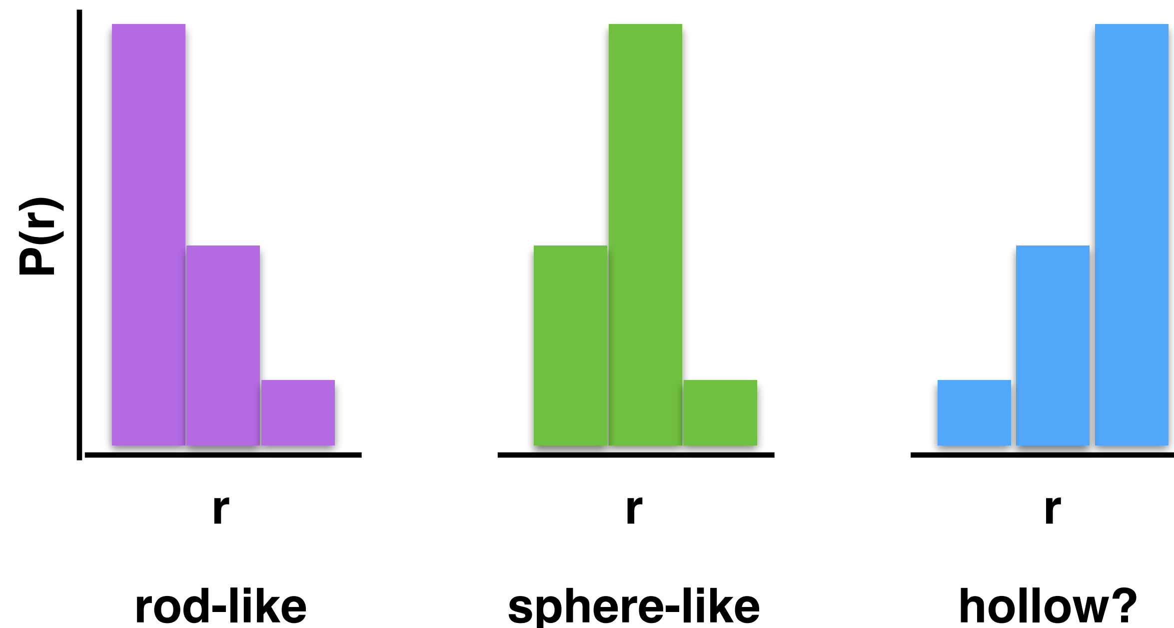




SAXS: P(r)-distribution

For a given object, make a histogram of distances found within macromolecule

- all electron-electron pair distances
- have to choose a bin size
- in this example, 3 bins (very coarse approximation of shape)



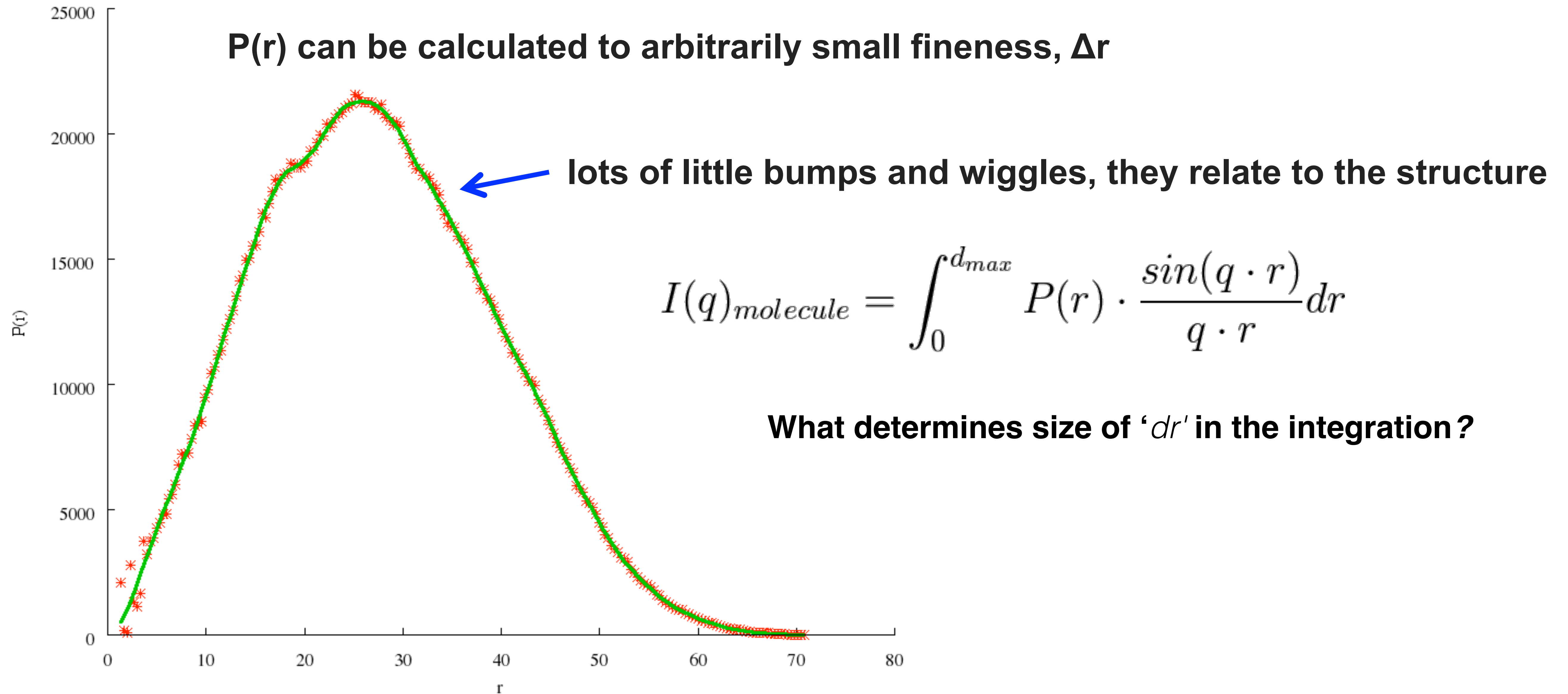
- Integral sine transform of P(r)-distribution gives I(q)

Number of bins is proportional to the q-max of the measured SAXS curve

$$n_S = \frac{q_{max} \cdot d_{max}}{\pi}$$

Resolution

Simulated *in vacuo* atomic scattering profile of P4P6 RNA domain

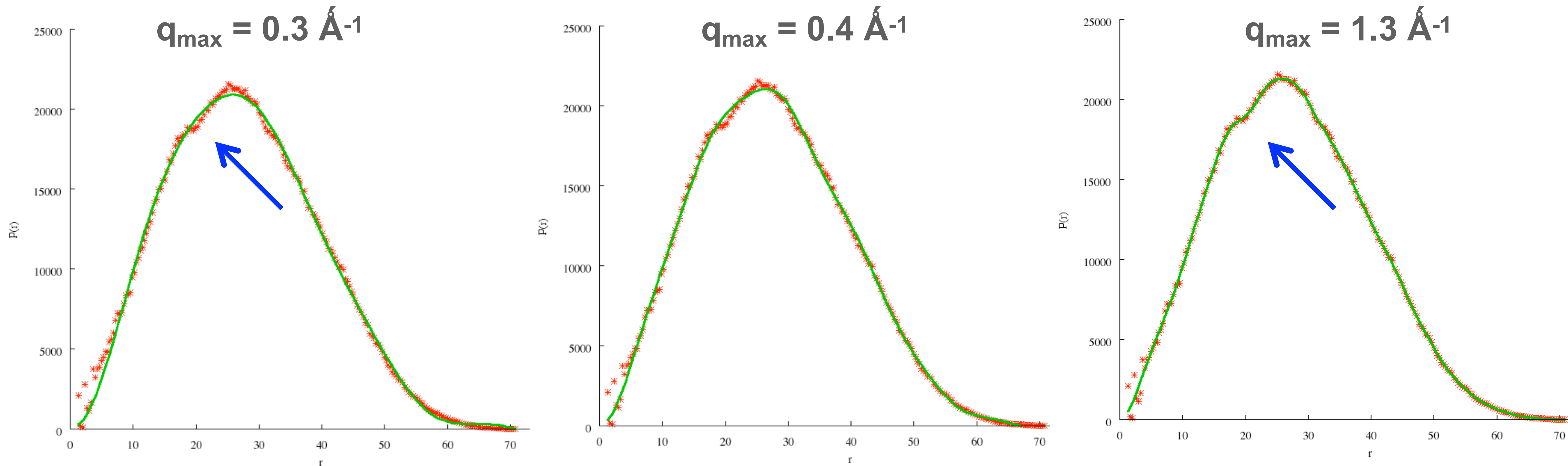


- A SAXS measurement is a resolution limited sampling of the molecules P(r)-distributon
- Resolution is a real phenomenon in SAXS, observed as "features" in P(r).
- Low resolution (green) the measured P(r) is very smooth
- Increasing q_{max} increases observed information content, start to fit more of the bumps (blue arrow)
- The fineness of the sampling is determined by q_{max} and d_{max}

Resolution

Simulated *in vacuo* atomic scattering profile of P4P6 RNA domain

P(r) curves at increasing q_{max} (⇒ increasing Shannon Number)



Resolution is a real phenomenon in SAXS, observed as “features” in P(r).

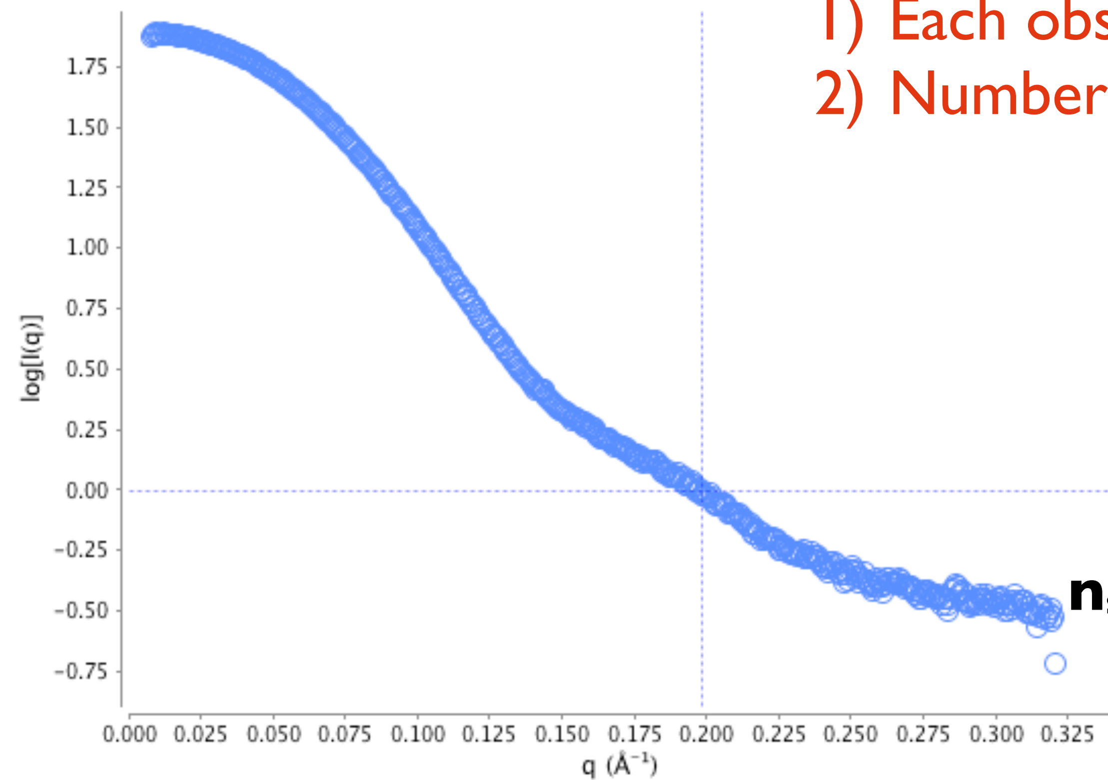
Low resolution (green) P(r) is very smooth

Increasing q_{max} increases observed information content, start to fit more of the bumps (blue arrow)

In terms of a PDB model, resolution is the RMSD variance of the set of models that best fit the data.

Information Theory and SAXS

Shannon-Nyquist Sampling Theorem



Given a SAXS dataset:

1) Each observed I(q) is not necessarily independent
2) Number of independent points, $n_s << N$

N: observed data points

$$n_S = \frac{q_{max} \cdot d_{max}}{\pi}$$

Shannon Number

Moore P. *SAS: Information content and error analysis* (1980) J. Appl. Cryst.

n_s: number of evenly distributed points needed to fully represent the observed scattering curve

q_{max}	d_{max}	n_s
0.42	43	6
0.42	71	10
0.42	240	32

We collect ~1100 data points

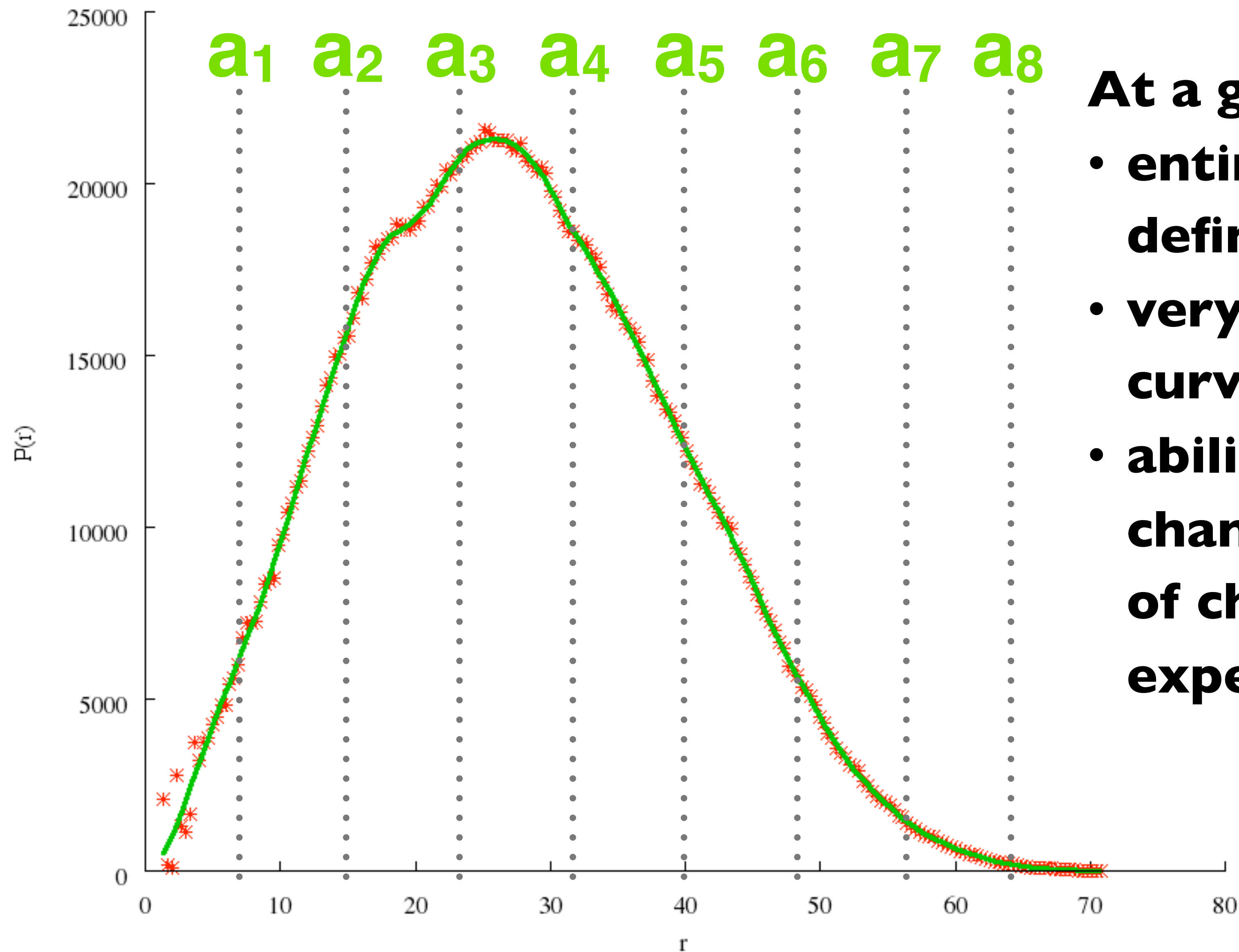
180 fold-redundancy

SAXS data are highly redundant but correlated
Few independent data points.

SAXS curve is a sampling of the P(r)-distribution determined at discrete points

Sampling

Simulated *in vacuo* atomic scattering profile of P4P6 RNA domain



At a given q-max

- **entire scattering experiment defined by these points**
- **very few points, expect smooth curve**
- **ability to detect conformational changes depends on magnitude of change and resolution of experiment**

$$I(q) = \sum_{n}^{m=1} P(r = a_i) \cdot \frac{sin(qr)}{qr}$$

Increasing q-max, increases the number of sampling points (bin-width gets smaller)

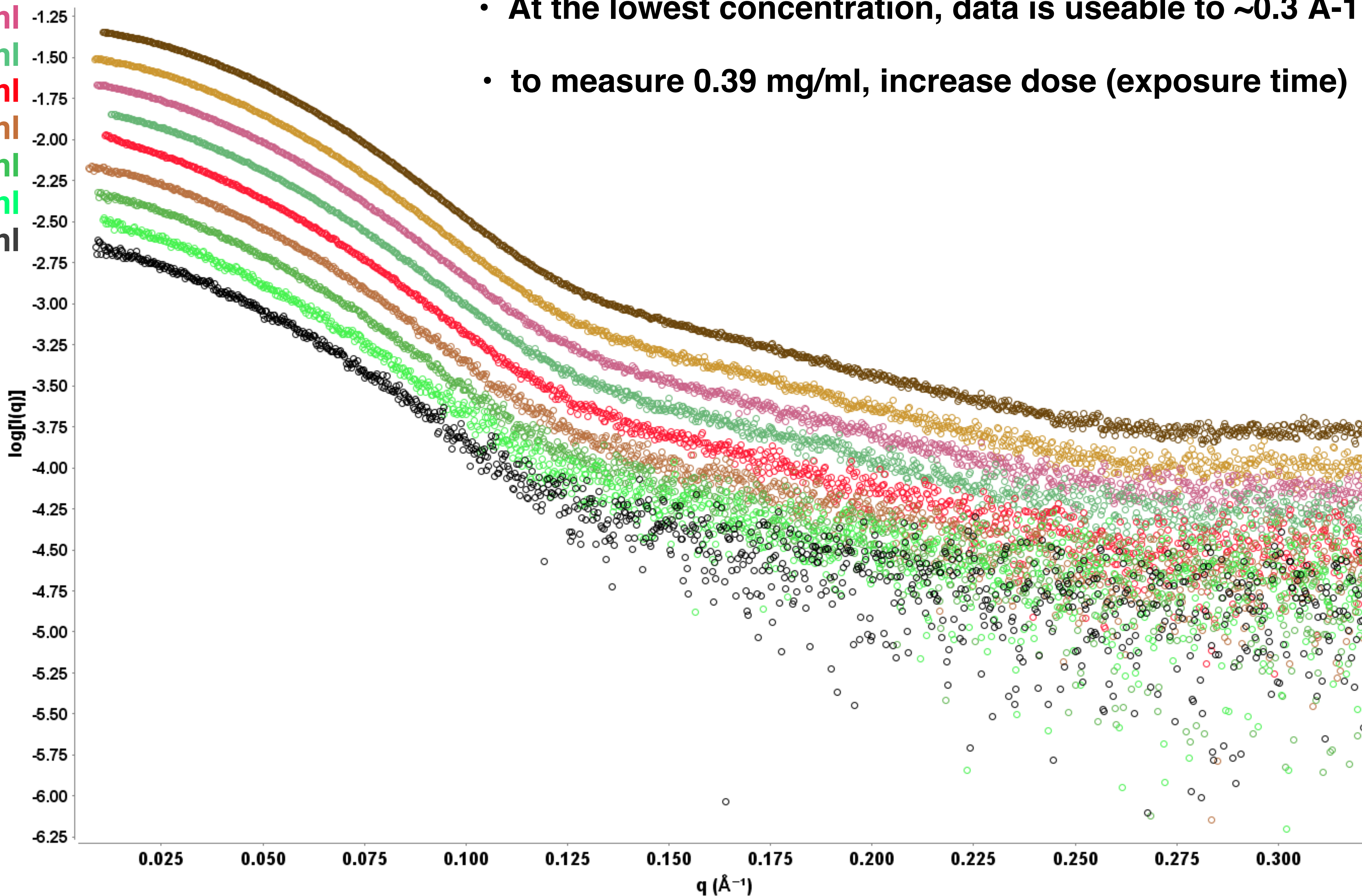
$$I(q) = P(a_1) \cdot \frac{sin(q \cdot a_1)}{q \cdot a_1} + ... + P(a_8) \cdot \frac{sin(q \cdot a_8)}{q \cdot a_8}$$

SAXS curve is a sampling of the P(r)-distribution determined at discrete points

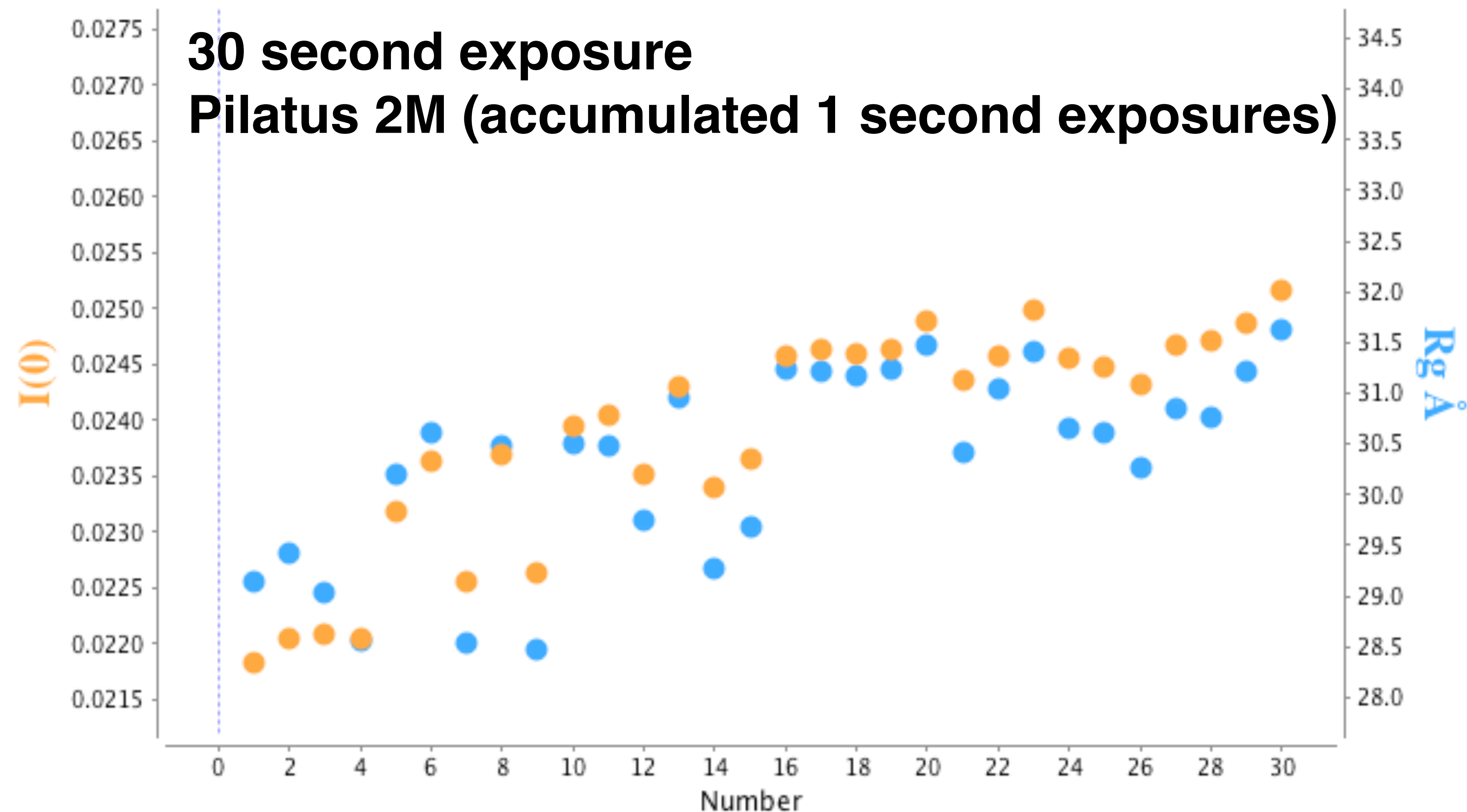
Lower background enables more reliable measurements at low-q

- 2/3rds dilution series BSA (66 kDa) starting at 10 mg/ml

6.67 mg/ml
4.44 mg/ml
2.96 mg/ml
1.97 mg/ml
1.32 mg/ml
0.88 mg/ml
0.58 mg/ml
0.39 mg/ml



SAXS and Radiation Damage



Radiation sensitive, samples shows in an increase in Rg and I(0) as exposure time increases

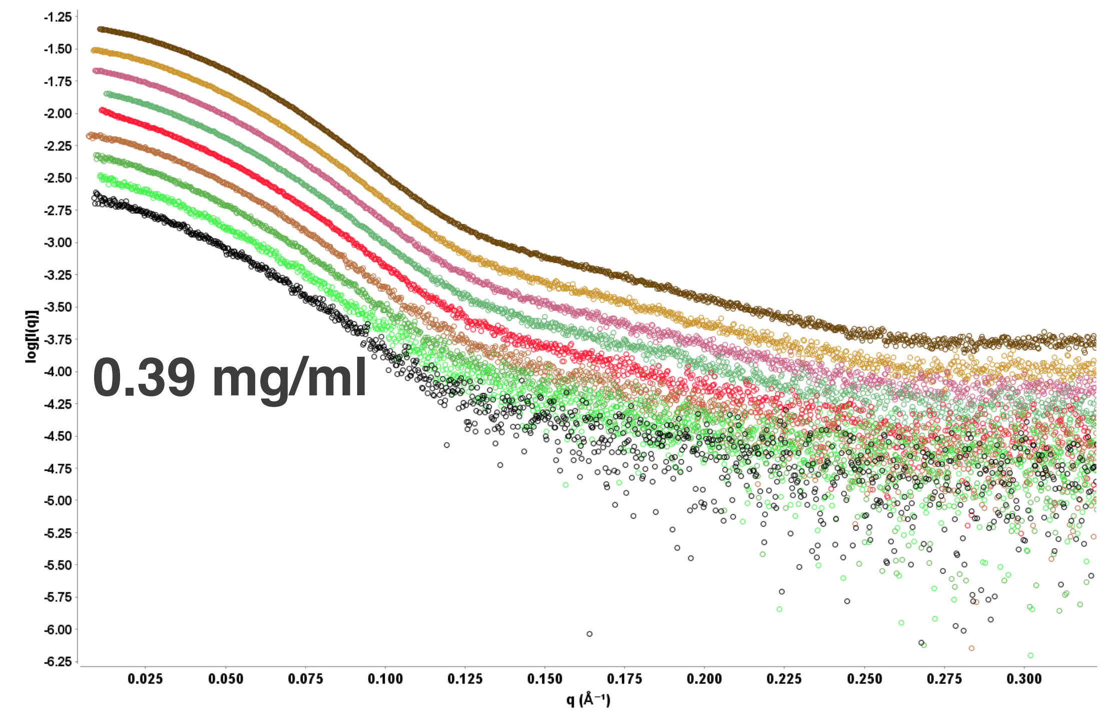
If you need long exposure for high-q information, must mitigate radiation damage

SAXS and Radiation Damage

Increasing dose (exposure time) can be damaging

Buffer composition can mitigate radiation damage

- **1% sucrose (very good)**
- **glycerol (increases viscosity)**
- **K/Na-nitrate (best)**
- **HEPES (good)**



Changes in R_g of BSA in PBS buffer

Rg (Å)

Dose (kGy)

~10 sec at 10^{11} photons per sec

PBS alone

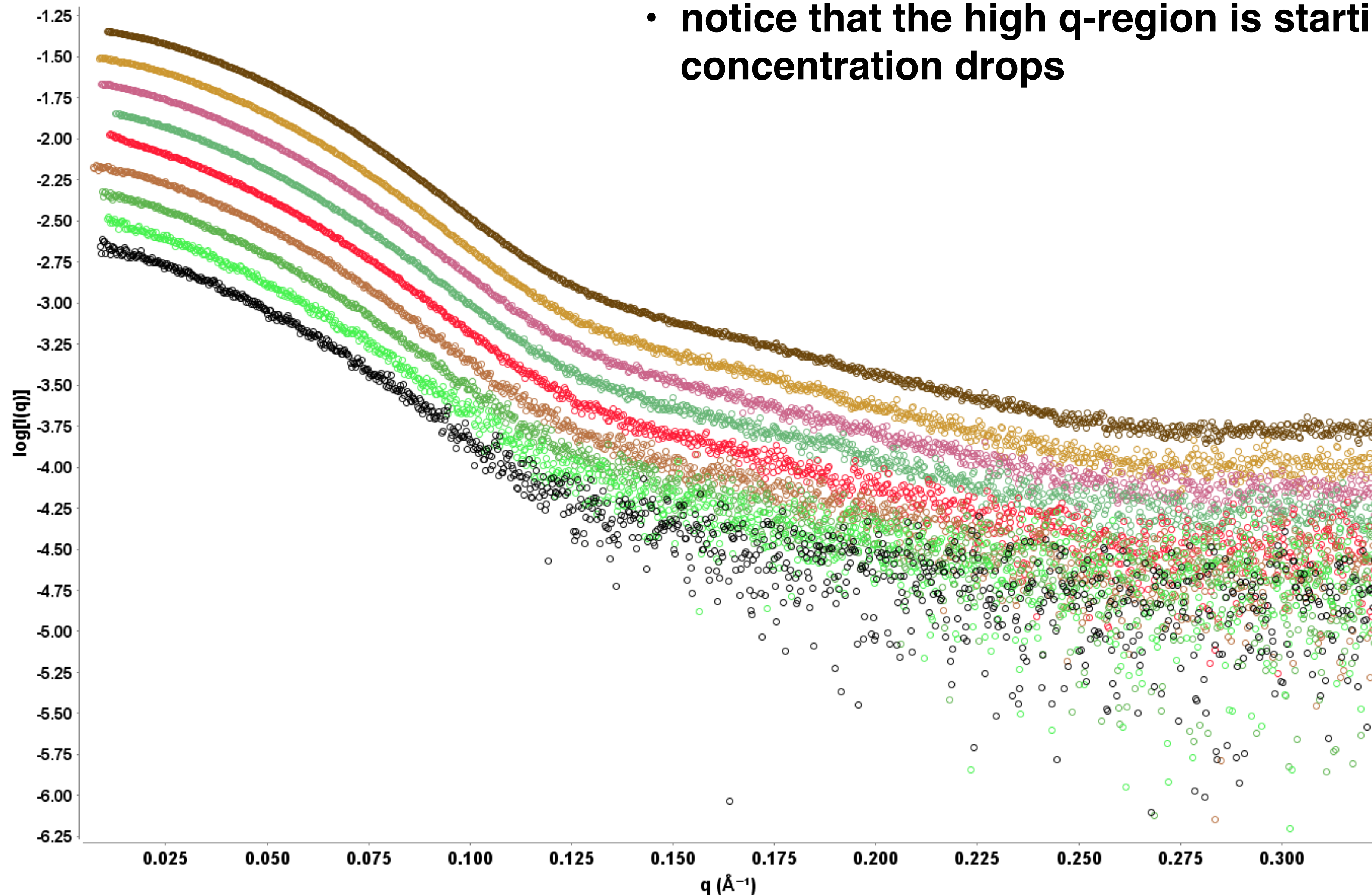
PBS with 10 mM K-Nitrate and 1% sucrose

High-q measurements require longer exposures to measure enough counts to provide a reliable subtraction

Must be done while maintaining sample integrity

SAXS Where does it End?

- 2/3rds dilution series BSA (66 kDa) starting at 10 mg/ml



- notice that the high q-region is starting to flatten out as concentration drops

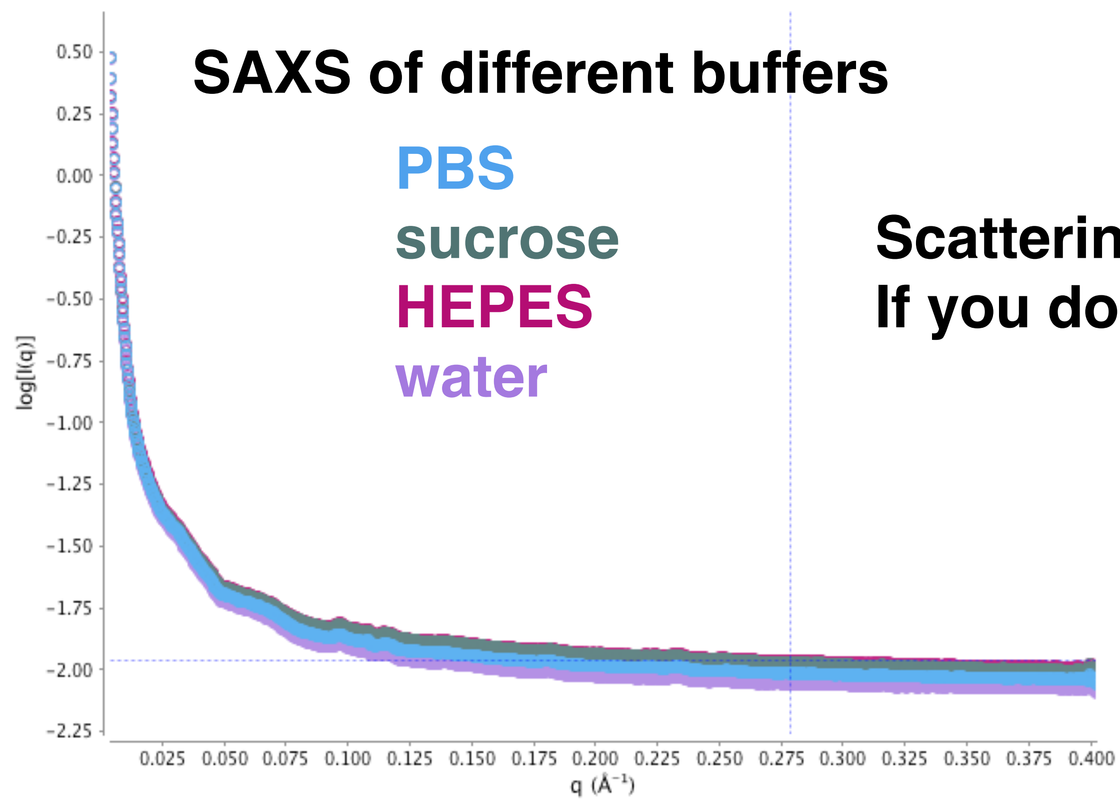
James Holton (ALS) circa 2013

If you are trying to measure a difference to better than 1% with measurement errors of 1%, you need around 20,000 counts in each (buffer and sample)

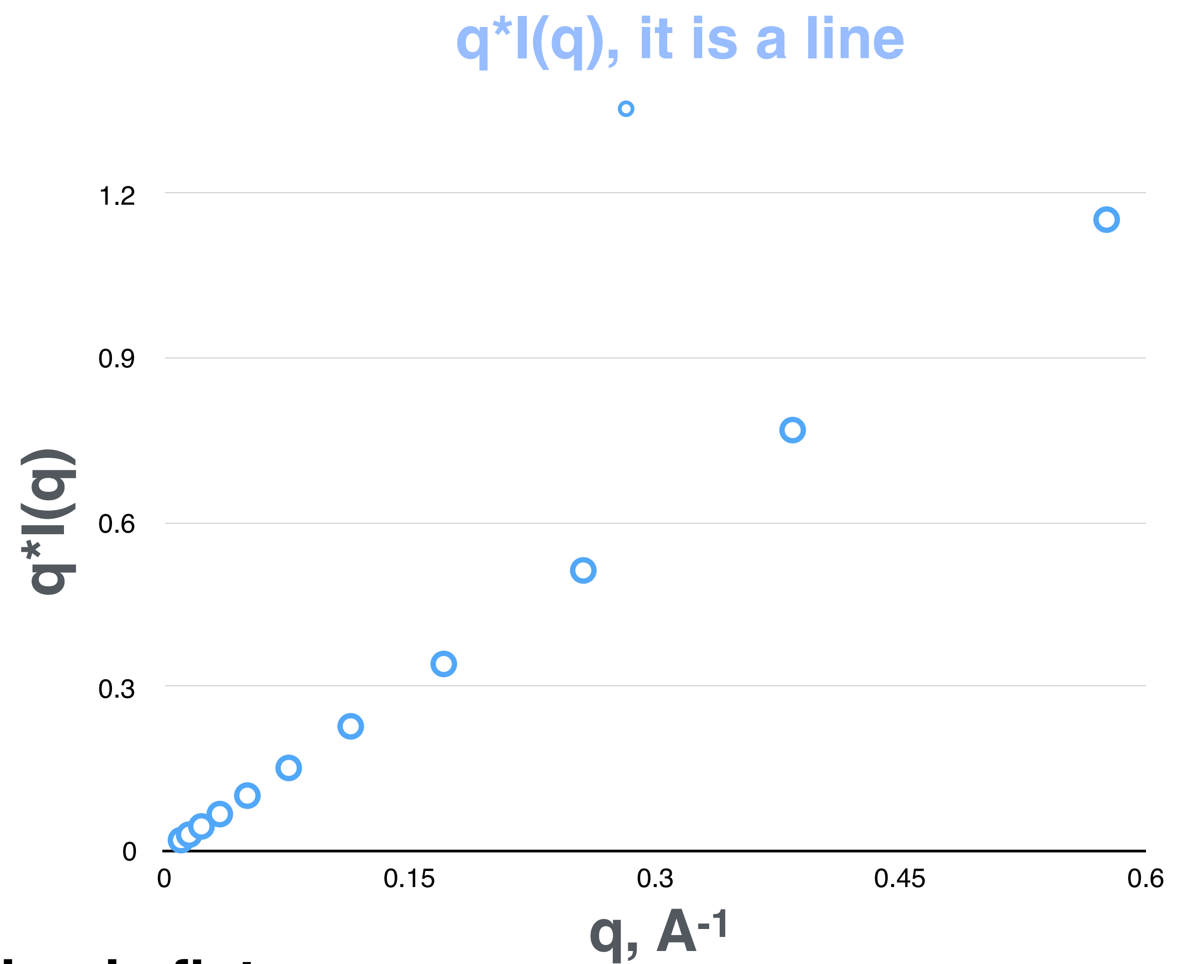
Typically at high q, counts < 100

- **means sample/buffer scattering is poorly measured**
- **statistics are in the Poisson regime (no longer Gaussian)**

SCÅTTER ≡ Intensity Plot



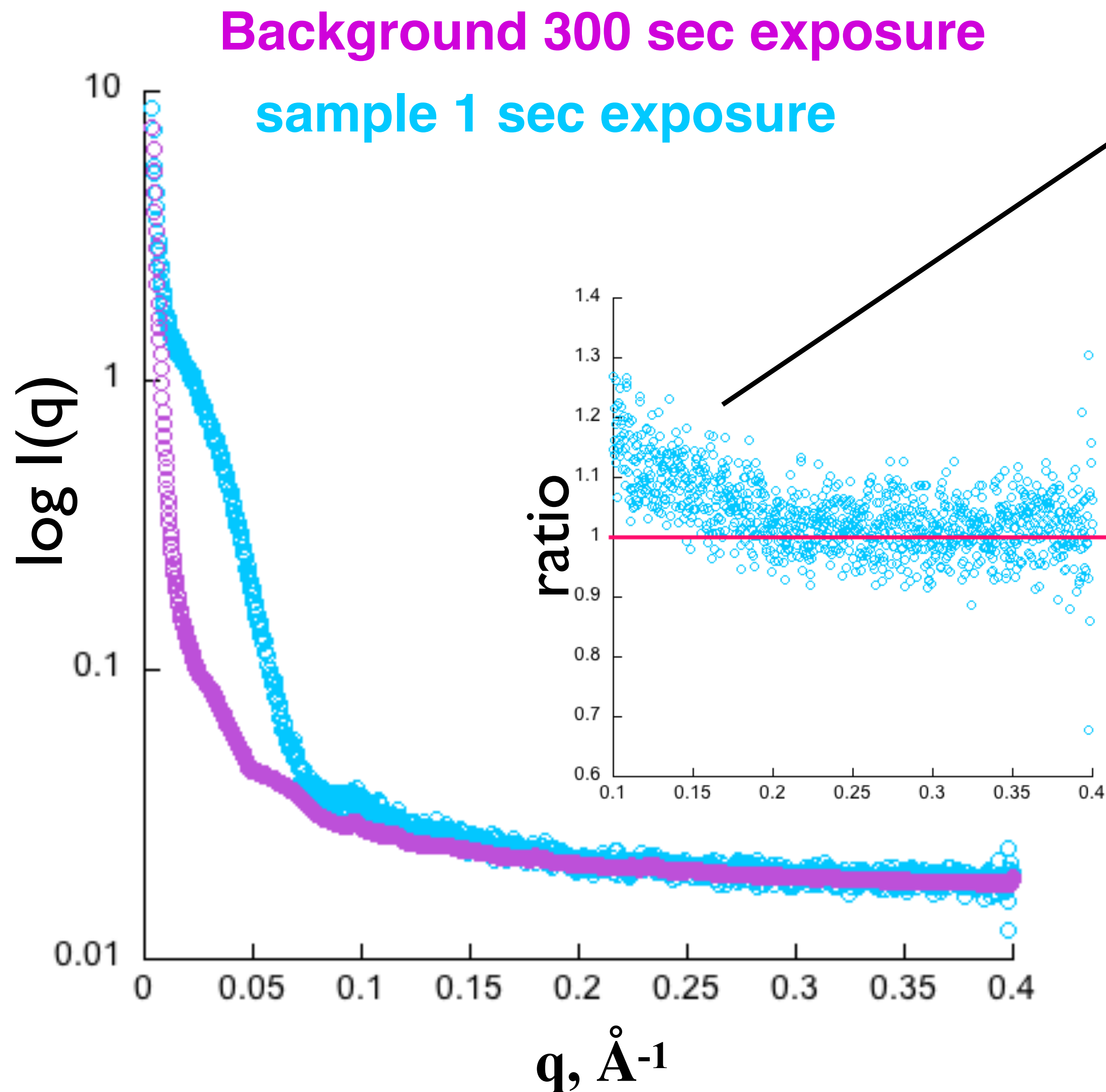
Scattering at high q is flat
If you do a q*I(q), it is a line

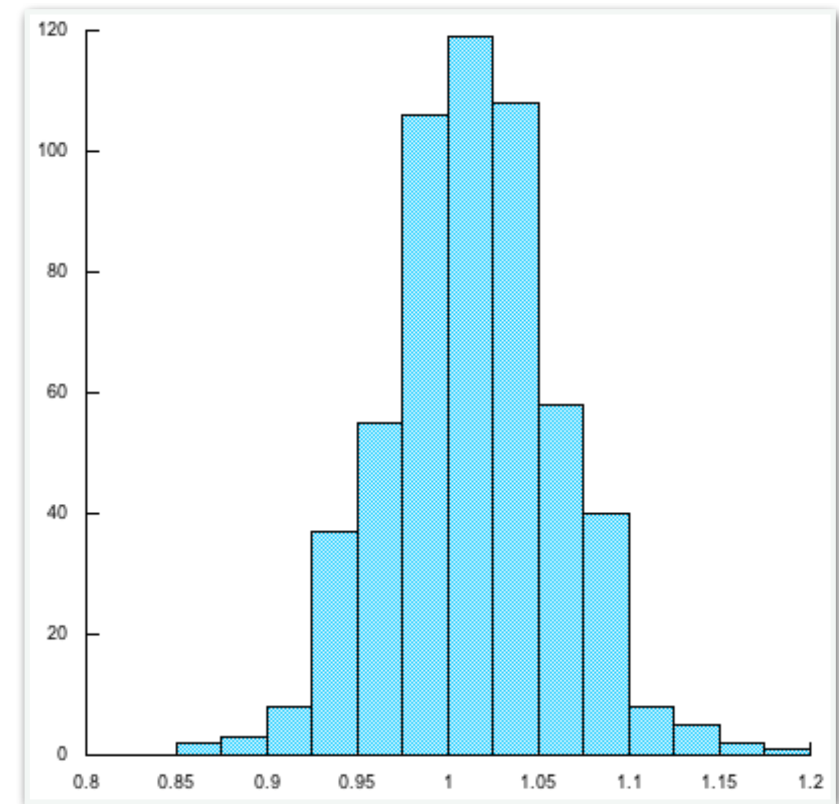


As measurement is mostly the cell, scattering is flat

- **Causes a linear uplift in a q*I(q) plot**
- **easily to visualize in V_c-plot**

Optimal Signal Extraction





average 1.01(0.05)
kurtosis 4.9

If no signal, ratio of two curves will be Gaussian

Kurtosis is a measure of Gaussianity (k = 3)

Implication for k > 3, signal is present

Well-determined background with a short sample exposure?

The Volume-of-Correlation

$$V_c = \frac{I(0)}{\int q \cdot I(q)\, dq} = \frac{c \cdot V^2 \cdot (\Delta\rho)^2}{c \cdot V \cdot (\Delta\rho)^2 \cdot 2\pi\, l_c} = \frac{V}{2\pi\, l_c}$$

independent of:
1. contrast
2. concentration

l_c is the expected correlation length

1. substitute for *I(q)*

$$c \cdot V \cdot (\Delta\rho)^2 \int q \int P(r)\, \frac{sin(q \cdot r)}{q \cdot r}\, dr\, dq$$

2. collect like terms

$$c \cdot V \cdot (\Delta\rho)^2 \iint \frac{P(r)}{r}\, sin(q \cdot r)\, dr\, dq$$

3. integrate by parts

$$-c \cdot V \cdot (\Delta\rho)^2 \int \frac{P(r)}{r^2}\, cos(q \cdot r)\, dr \Bigg|_0^\infty \rightarrow c \cdot V \cdot (\Delta\rho)^2 \int \frac{P(r)}{r^2}\, r\, dr$$

4. substitute $P(r) = 4\pi\, r^2\, \gamma(r)$

correlation function

5. collect like terms

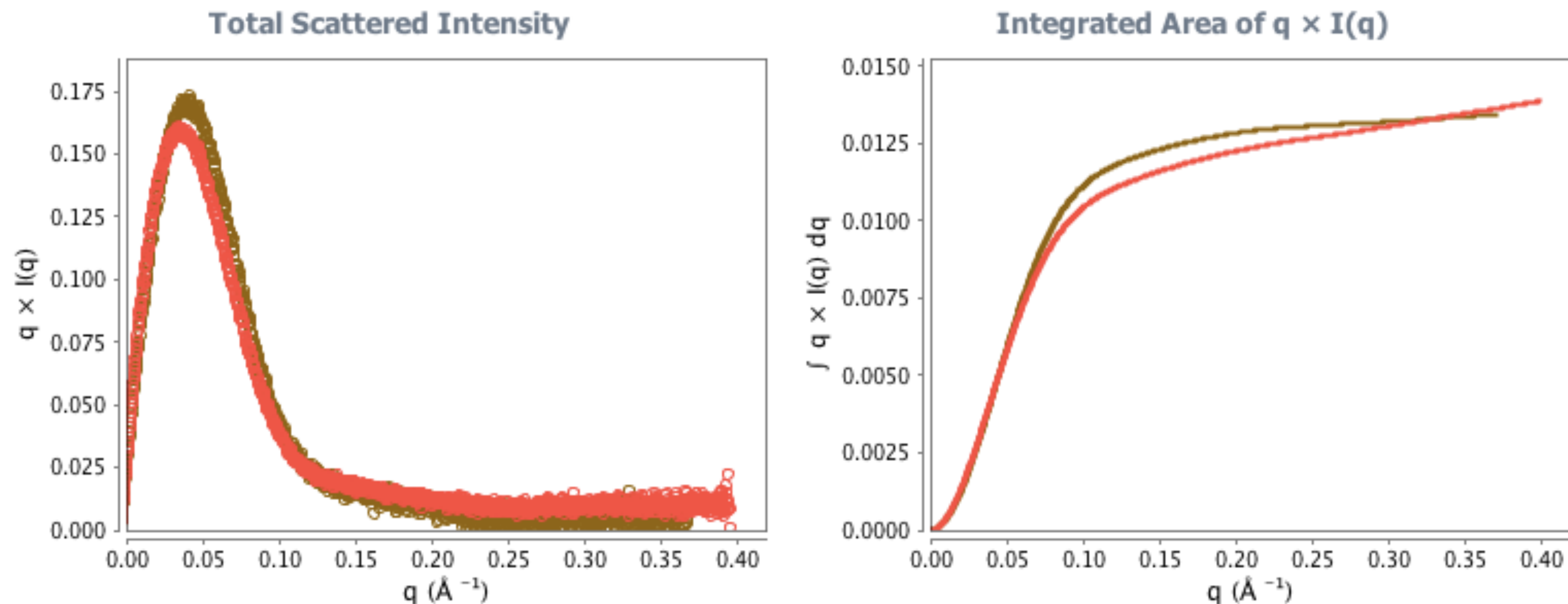
$$c \cdot V \cdot (\Delta\rho)^2 \int \frac{4\pi \cdot r^2\, \gamma(r)}{r^2}\, r\, dr$$

$$c \cdot V \cdot (\Delta\rho)^2 \int 4\pi\, r \cdot \gamma(r)\, dr$$

As an asymptotic limit, a plot of V_c vs q should approach a constant value.

Integrated Intensity Plot

$$V_c = \frac{I(0)}{\int q \cdot I(q)\, dq} = \frac{c \cdot V^2 \cdot (\Delta\rho)^2}{c \cdot V \cdot (\Delta\rho)^2 \cdot 2\pi\, l_c} = \frac{V}{2\pi\, l_c}$$



Use the integrated intensity plot to detect poor background subtraction. The plot should approach a constant value at high q.

Brown (SEC) matched background
Red protein dissolved into buffer

- **salts from lyophilised powder cause buffer mismatch**
- **cut data back to ~0.2**

Summarize

low-q

- beamstop noise and aggregation bias Guinier region
- Can be mitigated via truncation
- critical to modeling : determine where q-min should start

high-q, useable data determined by:

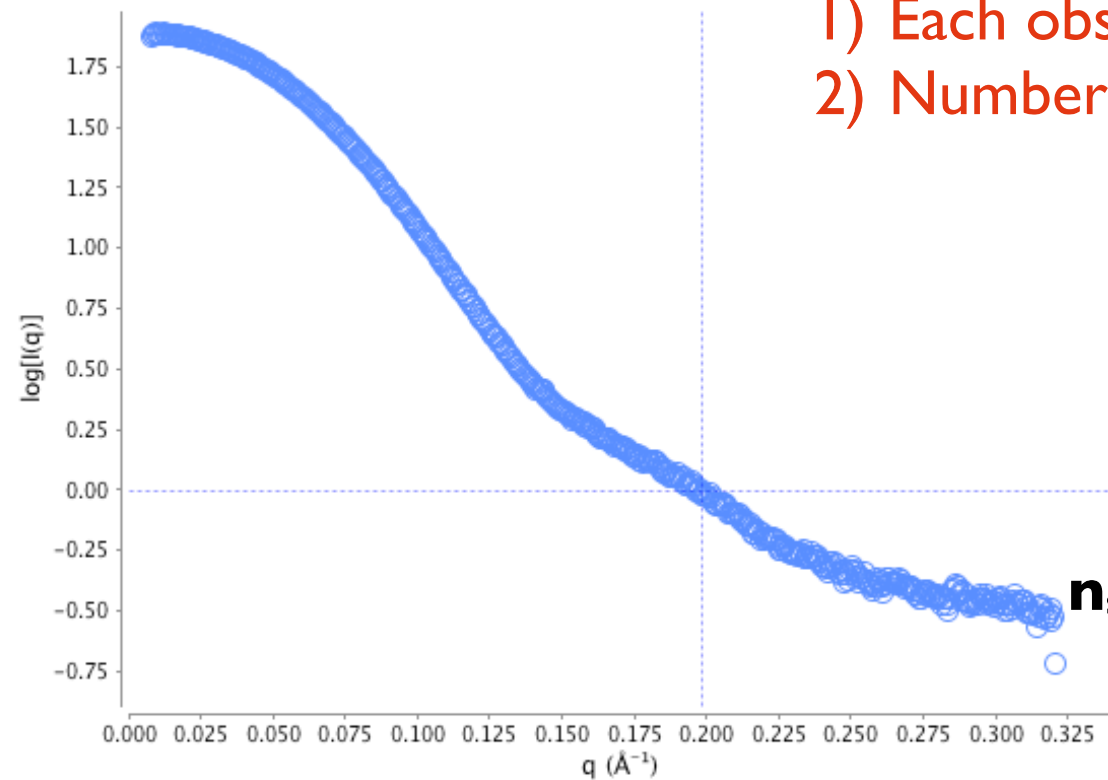
- how well (counts) were measured
- buffer matching
- inspect Integrate Intensity Plot
- if difficult to determine P(r), truncate data until you get a smooth curve

How much of this matters:

- ab initio modeling?
  1. not much, typical DAMMIN/GASBOR/DENFERT use 1/q^4 weighting
- atomistic modeling?
- prone to over-fitting if you do not determine q-max
- programs (FOXS/CRYSOL) do not account for poor background subtraction

Information Theory and SAXS

Shannon-Nyquist Sampling Theorem



Given a SAXS dataset:

1) Each observed I(q) is not necessarily independent
2) Number of independent points, n_s << N   **N: observed data points**

$$n_S = \frac{q_{max} \cdot d_{max}}{\pi}$$ **Shannon Number**

Moore P. *SAS: Information content and error analysis* (1980) J. Appl. Cryst.

n_s: number of evenly distributed points needed to fully represent the observed scattering curve

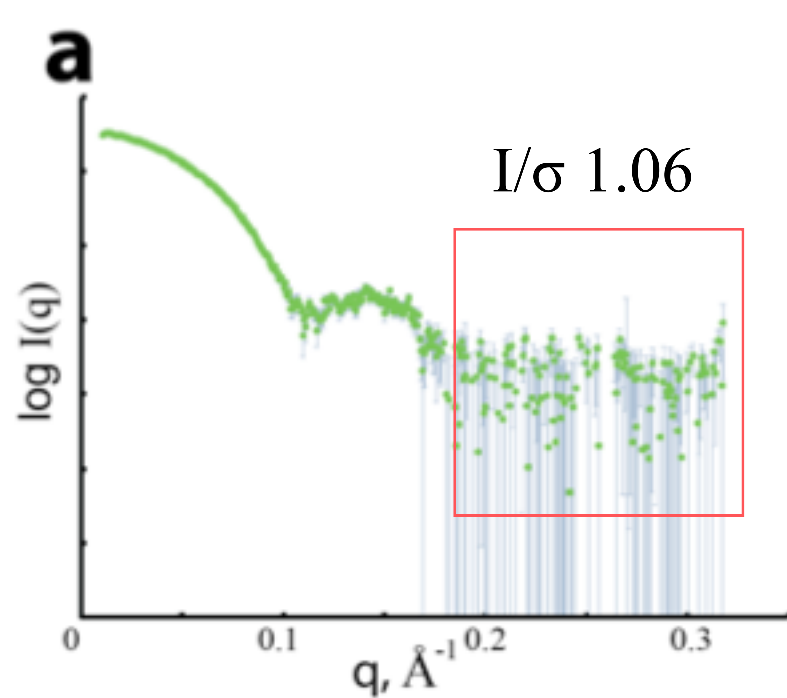
q_{max}	d_{max}	n_s
0.42	43	6
0.42	71	10
0.42	240	32

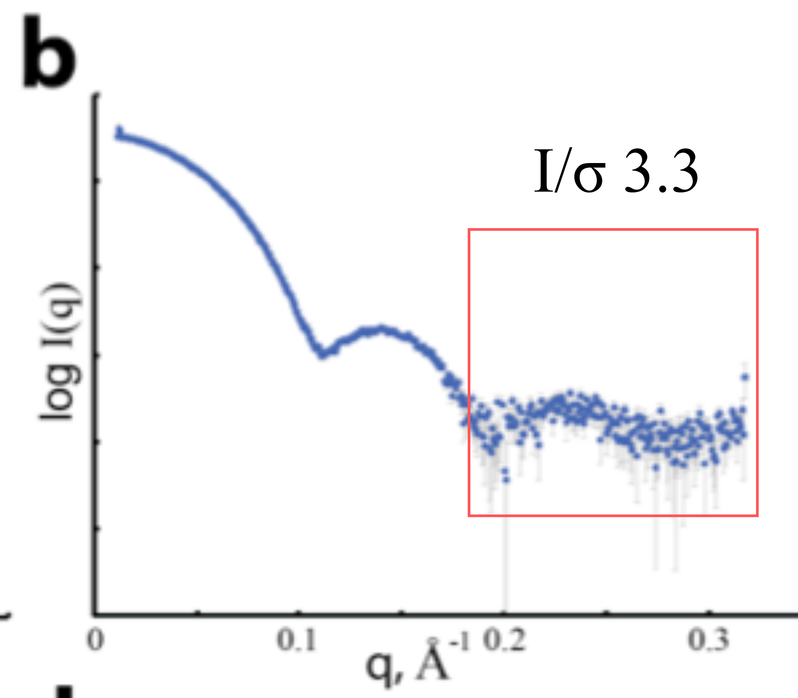
We collect ~1100 data points

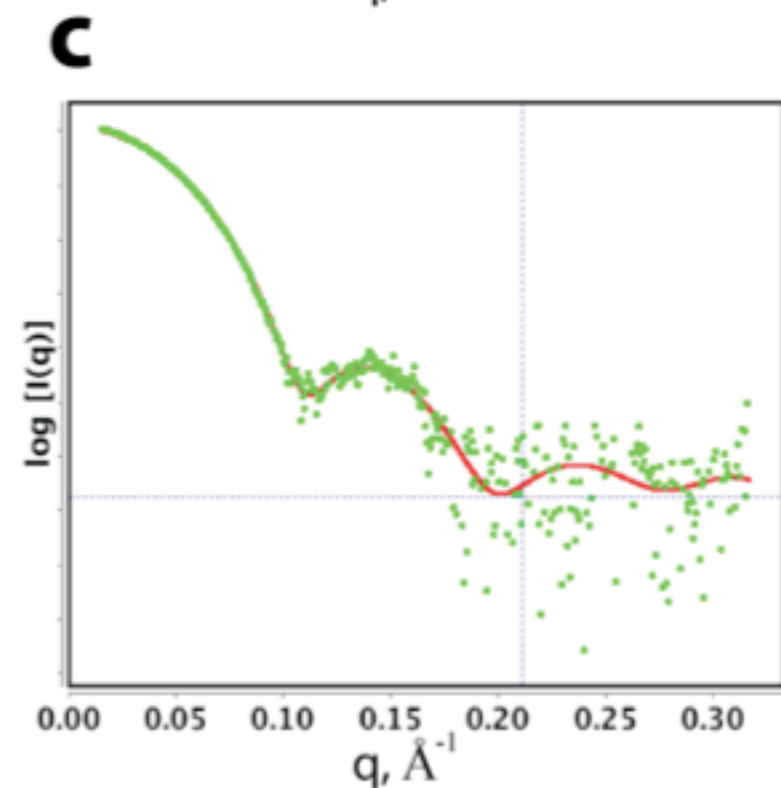
180 fold-redundancy

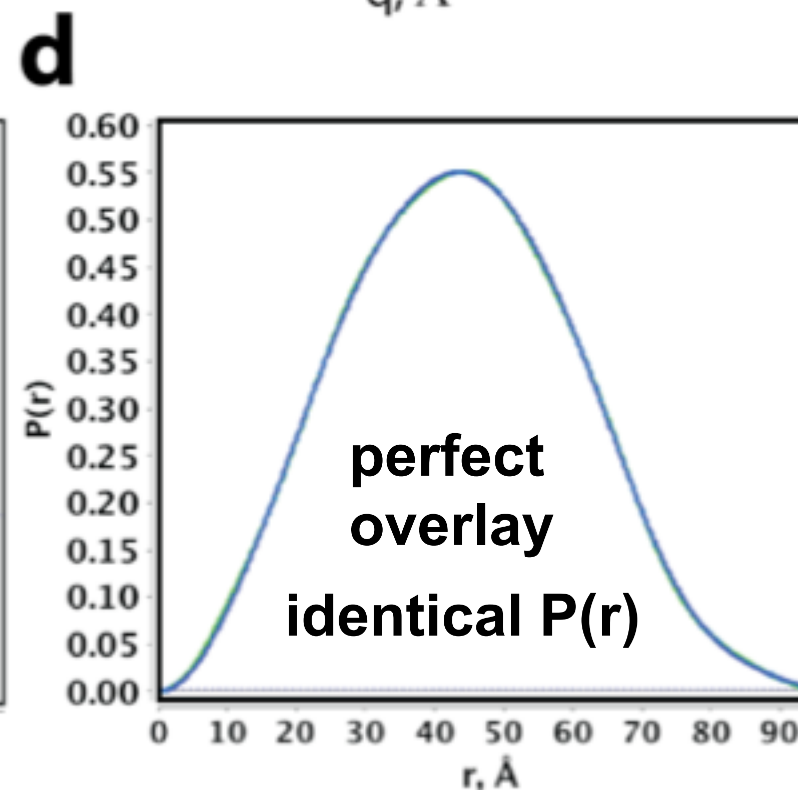
SAXS data are highly redundant but correlated
Few independent data points.

Fine sampling of the SAXS curve (redundancy) helps recover signal in high noise, but how?









(Shannon-Hartley Theorem)

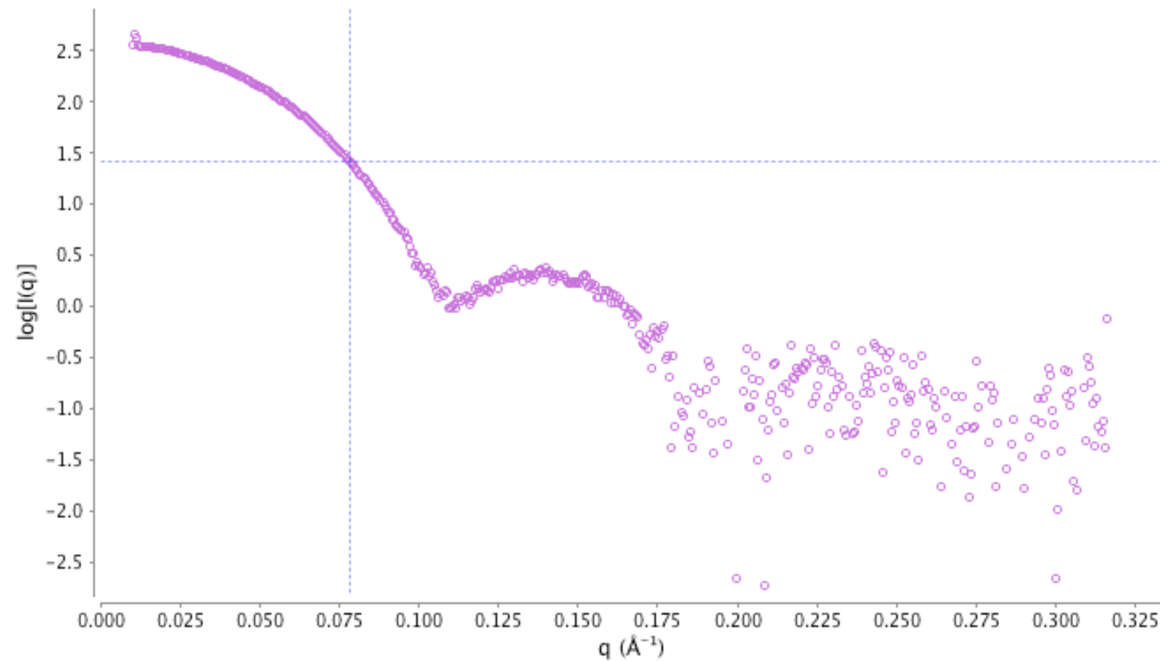
Noisy Coding Channel Theorem:
Guarantees error free recover of the signal as long as the sample rate, $R < C$

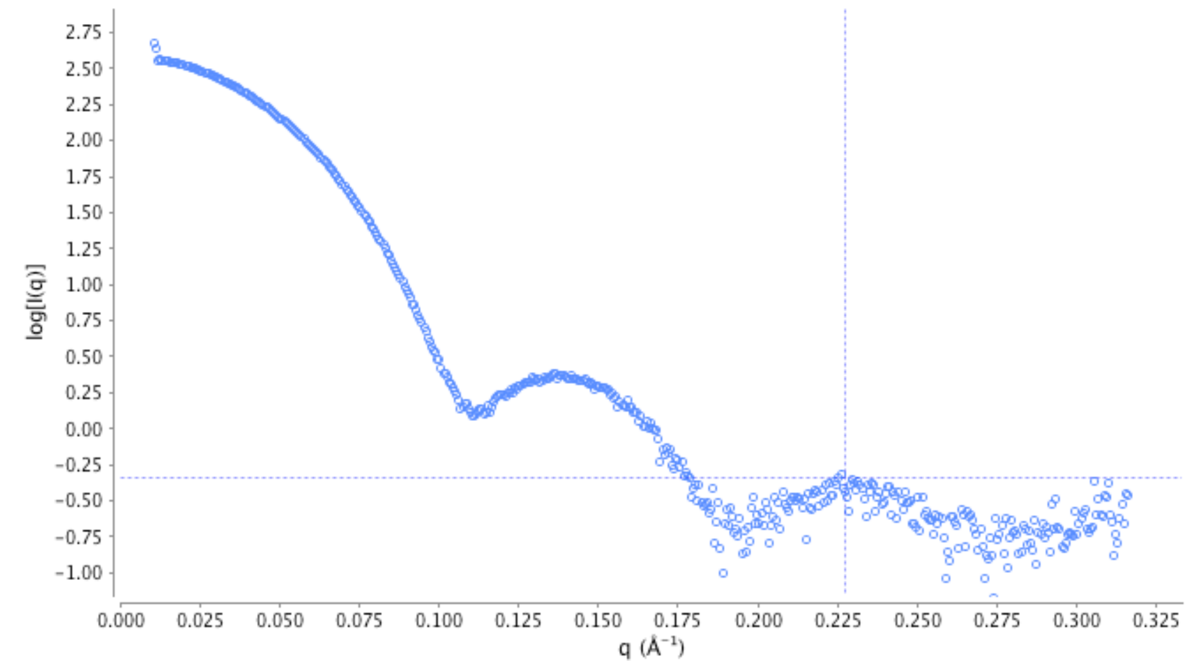
$$R < C = \frac{2\pi}{d_{max}} \cdot log_2\left(1 + \frac{S}{N}\right)$$

sampling (Δq)

We can recover our “signal” error free in high noise environments
Dependence on the algorithm and scoring function. i.e., GNOM, SCATTER, FOXS, CRYSOL

Take Median





Single measurement of a protein using 5 independent buffer measurements

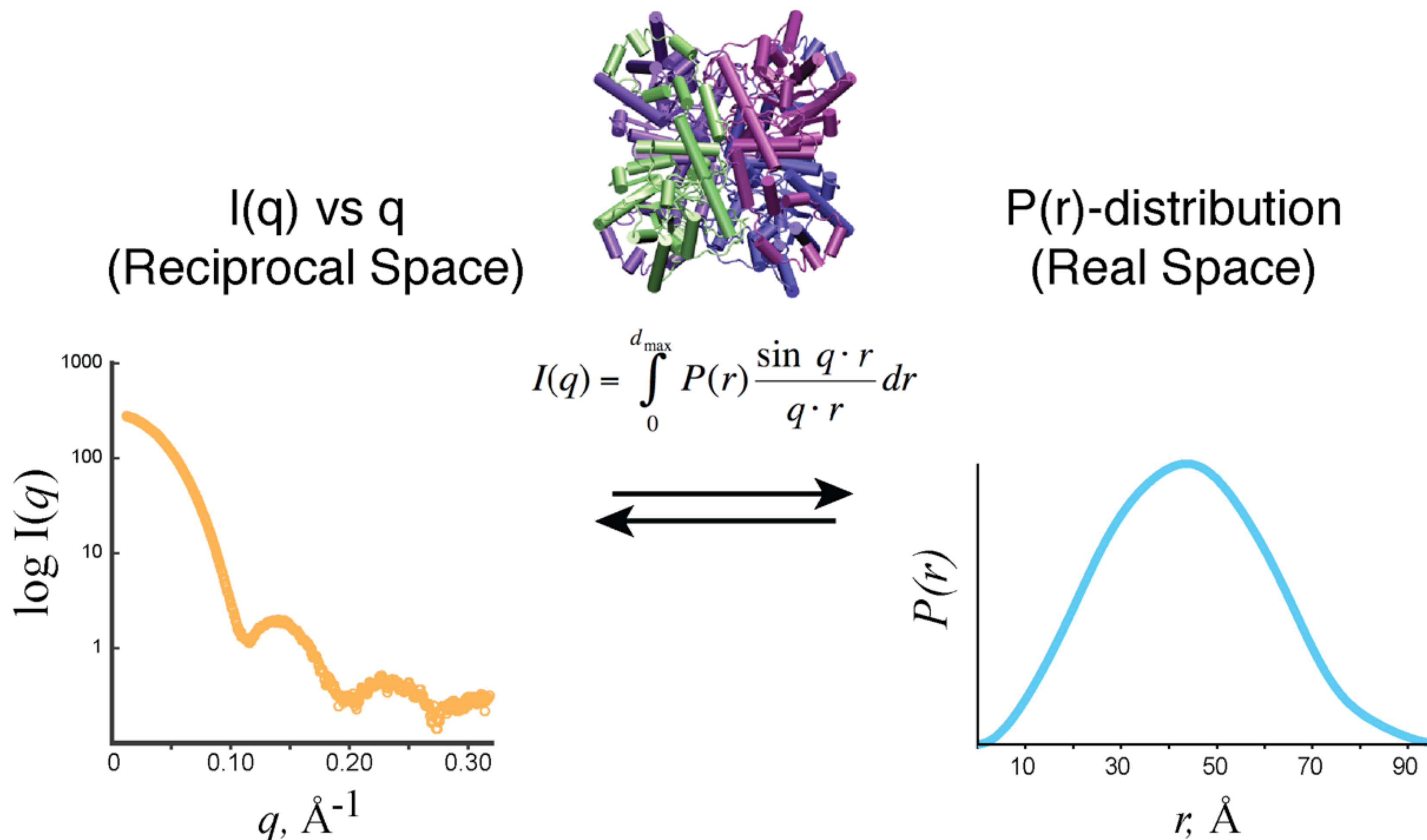
Preserve error for a single exposure

Don't get the benefit of signal averaging (increase S-to-N ratio)

- **You can if median used as a filter with a scaling statistic**
- **median provides protection from outliers**
- **median and average should overlay, if they don't got a problem!**
  - **capillary fouling**
  - **aggregation/radiation damage**
  - **bubble?**

Indirect Fourier Transform

A measured SAXS curve determines a unique P(r)-distribution.
A P(r) distribution (from a model) can be used to determine a scattering curve.



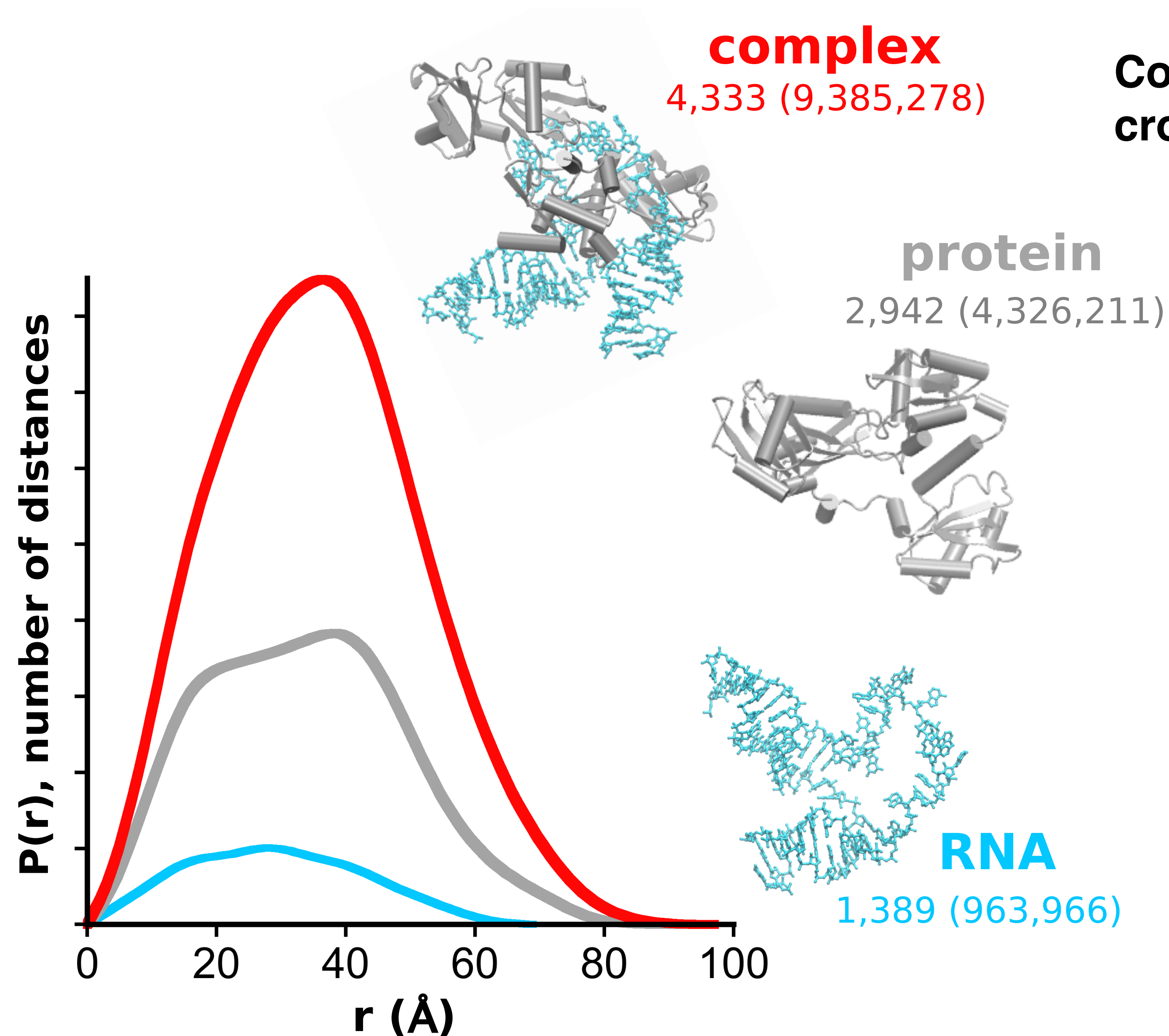
Gnom (Svergun)
- Invert Ill-conditioned Matrix
- L2 regularization

GIFT (Glatter's method)
- use cubic splines

My Method
- use Shannon theorem (Moore)
- L1 regularisation

Expect a smooth curve
Minimize oscillations
No negative values
Iterative process in determining d_{max}
These methods assume a single d_{max} (mixtures cause problems)

Complexes always contain an additional cross-term



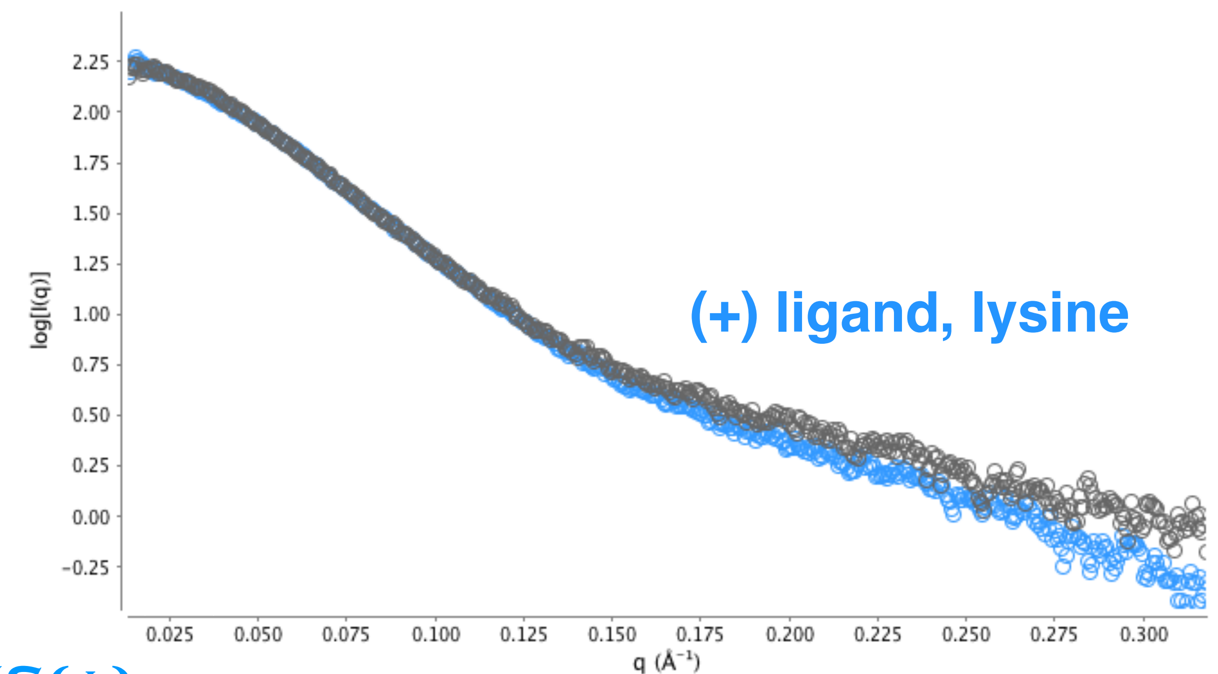
$$P(r)_{\text{complex}} = P(r)_{\text{RNA}} + P(r)_{\text{protein}} + P(r)_{\text{cross-terms}}$$

Presence of the cross-term grossly effects I(q)
If you had structures of components, cross-term contains all the information for reassembly
SAXS could be used to orient particles in a complex

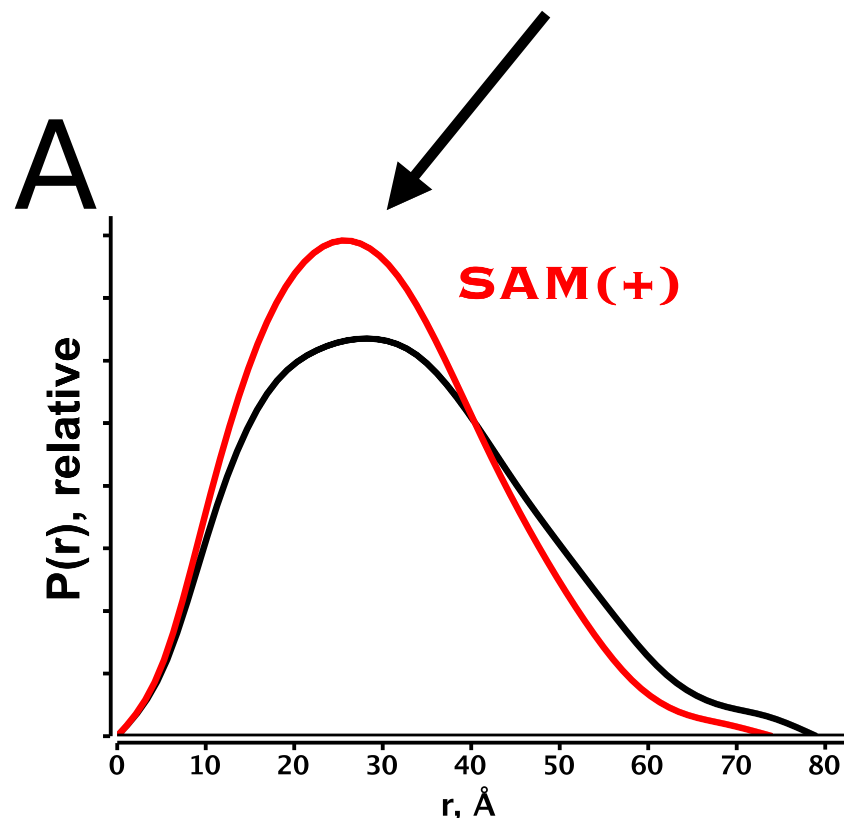
P(r) is the best method to detect and assert conformational changes between conditions because the distribution utilises all the data but is resolution limited.

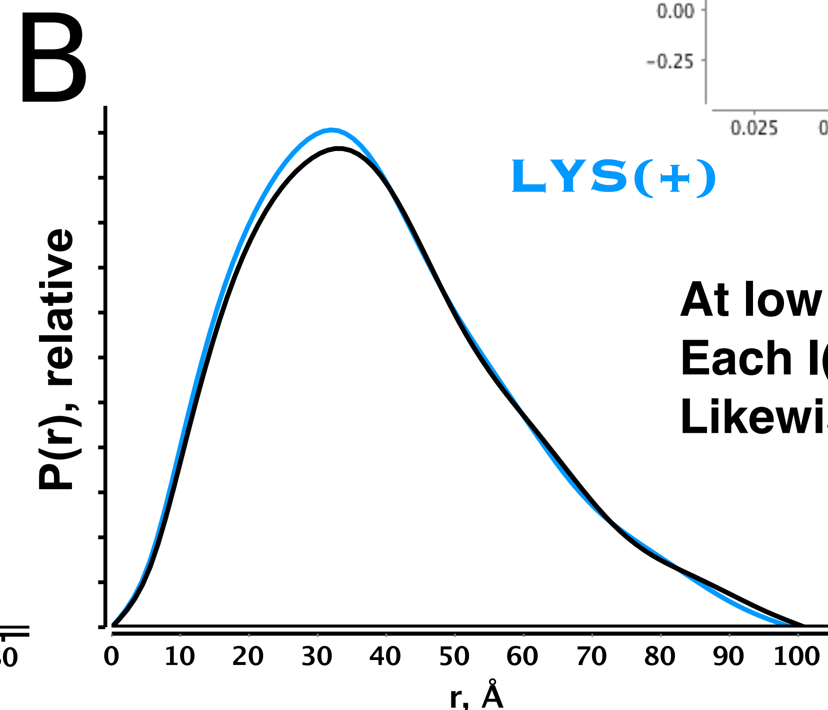
Small changes may not produce changes in R_g, need data beyond Guinier

Magnitude of the change determines required resolution limit of the SAXS data.



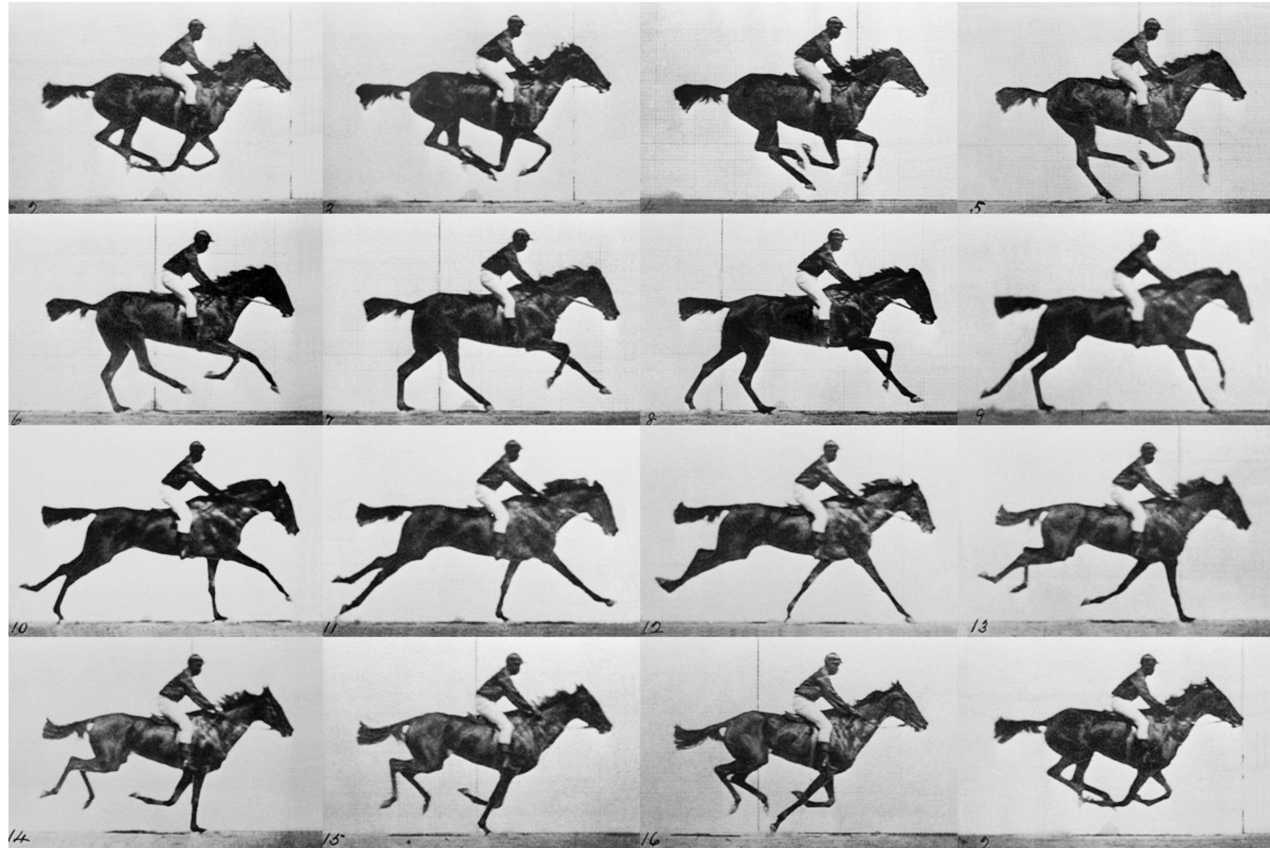
Large conformational change upon binding





At low res ($q < 0.125$) structures look the same
Each I(q) is the FT of the structure
Likewise, each point in P(r) is FT of the entire I(q)

The Thermodynamic State



$$I_{obs}(q) = I_1(q) + I_2(q) + .. + I_n(q)$$

Solid-state techniques provide information about individual frames

- typically get one frame per grad student
- dynamics produces incomplete structures

NMR gives information on observable parts, dynamics reduces available NOEs

- may not see the legs

SAXS signal is the accumulated sum over all the observed molecules in their various structural states

- at best, get blurred image
- but, image is rotationally averaged
- every part of molecule is observed

$$I_{obs}(q) = I_1(q) + I_2(q) + .. + I_n(q)$$

$I_i(q) \neq I_j(q)$ => structurally resolvable microstates at a given SAXS resolution, q_{max}

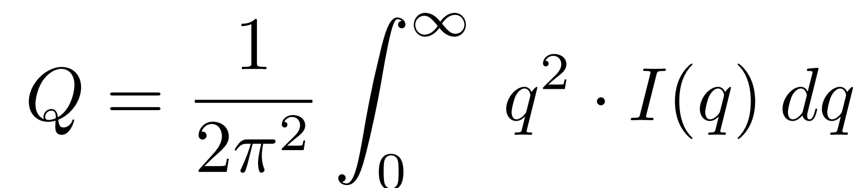
Porod Invariant

Assessing flexibility

G. Porod deduced an integral constant contained within a SAXS curve:

Assumption: defined $\Delta\rho$ between particle and solvent and scatterer has homogenous electron density

Integration of data transformed as $q^2 \cdot I(q)$ should be constant



$$Q = \frac{1}{2\pi^2} \int_0^\infty q^2 \cdot I(q)\, dq$$

$$Q = 2\pi^2 \cdot (\Delta\rho)^2 \cdot V$$

Q is the direct product of the excess scattering electrons of the particle and $V_{particle}$

$$Q = 2\pi^2 \cdot c \cdot (\Delta\rho)^2 \cdot V$$

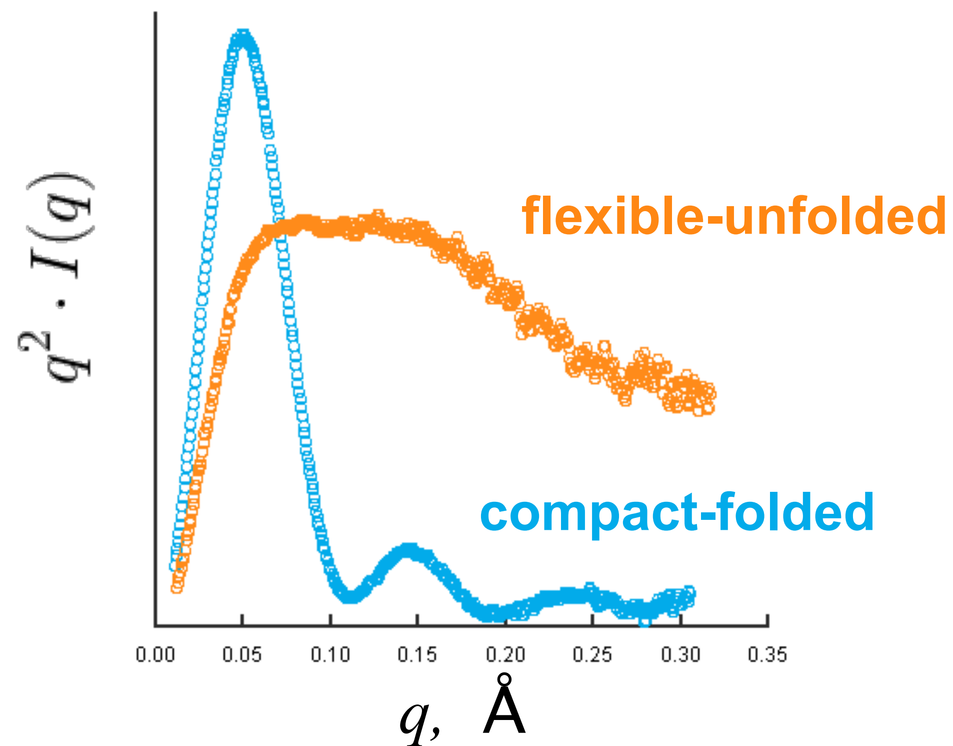
Regardless of beamline, source, or wavelength;
Data should have the same constant with the same sample at the same concentration.

Porod Invariant

Assessing flexibility

Kratky Plot

- visualization of Q
- used to interpret samples with flexibility



$$Q = \frac{1}{2\pi^2} \int_0^\infty q^2 \cdot I(q)\, dq$$

A plot of $q^2 \cdot I(q)$ should show a curve that captures an area
Define area means transformed data converges.

SAXS Invariants

(structural parameters derived directly from SAXS)

Q, Porod Invariant

$$Q = \int_0^\infty q^2 \cdot I(q)\, dq$$

Directly related to mean square electron density of scattering particle.
Requires convergence in Kratky plot ($q^2 I(q)$ vs q).

V_p, Porod Volume

$$V_p = 2\pi \cdot \frac{I(0)}{Q}$$

l_c, correlation length

$$l_c = \pi \cdot \frac{\int_0^\infty q \cdot I(q)}{Q}$$

Requires a folded particle, otherwise Q won't converge properly.
Q acts as a normalization constant and corrects for:

1. concentration
2. contrast, $(\Delta\rho)^2$

R_g, radius-of-gyration

$$R_g{}^2 = \frac{1}{2}\frac{\int r^2 \cdot P(r)\, dr}{\int P(r)\, dr}$$

Does not require Q
Concentration independent
Contrast independent (as long as structure does not change)
Essentially normalized to I(0)

Porod's Law

Asymptotic behaviour

for $q \cdot R_g$ > 1.3, the scattering decays as $1/q^4$

Assumption: defined Δρ between particle and solvent

$$I(q) \approx c \cdot (\Delta\rho)^2 \cdot \frac{2\pi}{q^4} \cdot S$$

$$q^4 \cdot I(q) = c \cdot (\Delta\rho)^2 \cdot 2\pi \cdot S = constant$$

remove c and Δρ by dividing by Q

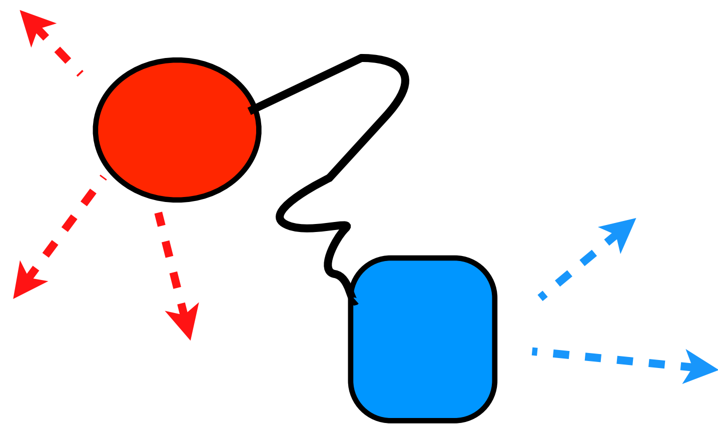
Graphically, a plot of q4 I vs q should approach a constant

- Works well for nearly spherical particles
- Hard to see with elongated or flexible particles

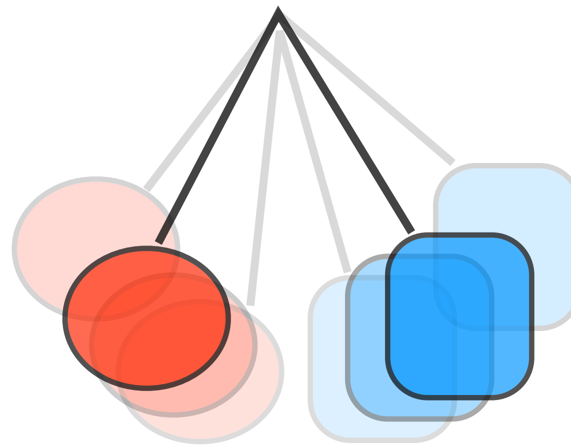
$$\frac{S}{V} = \pi \cdot lim \frac{I(q) \cdot q^4}{Q}$$

Types of Flexibility

uncorrelated flexibility



correlated flexibility



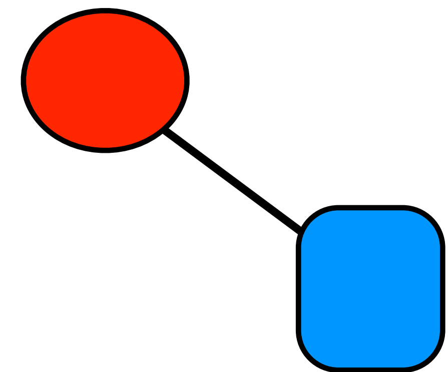
domains act independent of each other

- Observed particle volume is > the sum of the individual domains
- Flexibility has the effect of reducing the particle's contrast.
- Protein is smeared throughout a larger volume of space

V_p is larger than expected, $P_E \approx 4$

V_p is larger than expected, $P_E < 4$

discrete state (rigid)



$V_p \Rightarrow$ density near 1.37

$P_E \approx 4$

P_E ~ Porod-Debye Exponent
What is it?

Rambo and Tainer Biopolymers SAXS Special Issue 2011

Detecting Flexibility

Debye P. *Molecular-weight Determination by Light Scattering* (1947) J. of Physical and Colloid Chemistry

Scattering by a Gaussian Coil

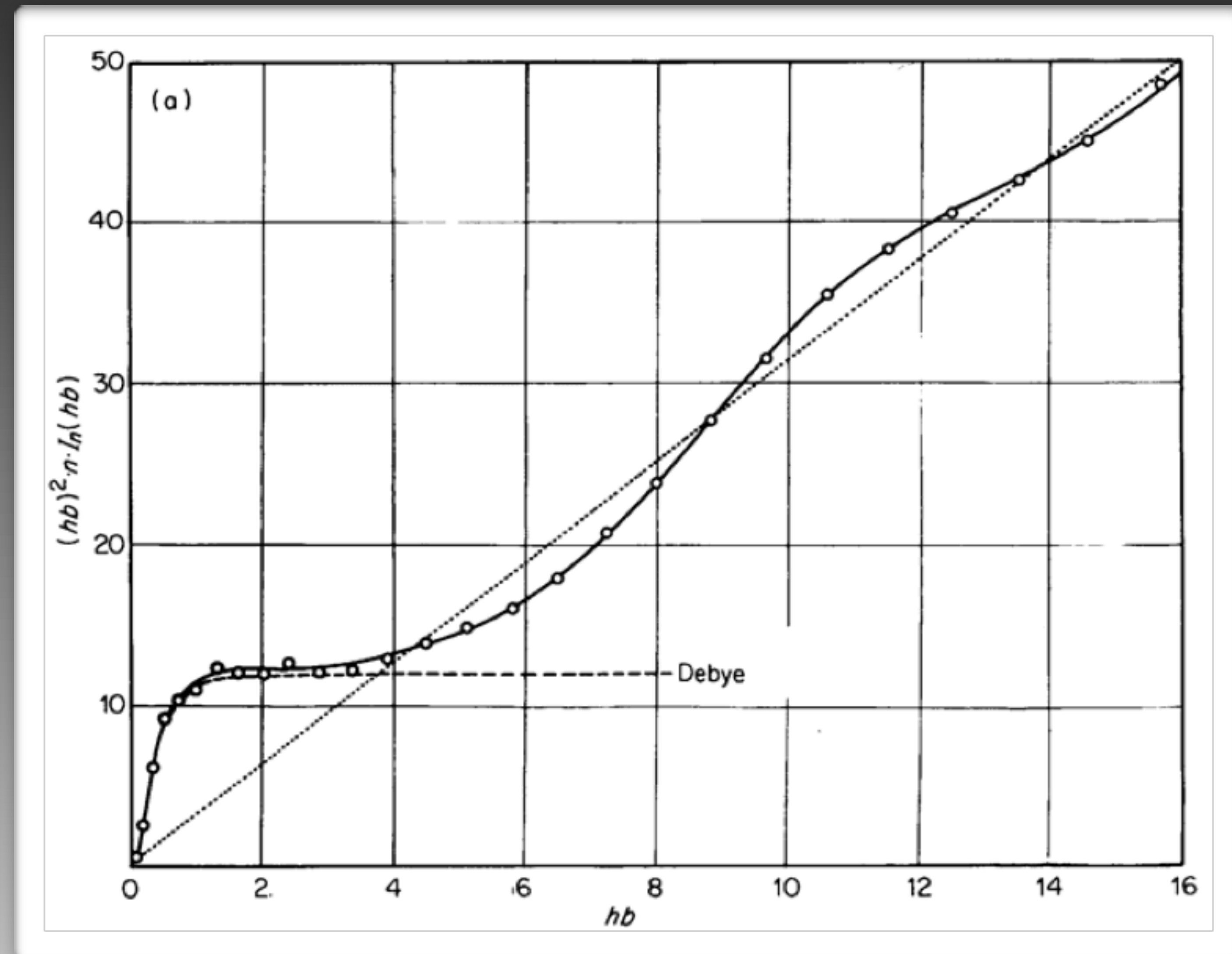
$$I(q) = \frac{2(e^{-R_g^2 \cdot q^2} + R_g^2 \cdot q^2 - 1)}{(R_g^2 \cdot q^2)^2}$$

ASYMPTOTIC CHARACTERISTIC

$$\lim_{q \to \infty} I(q) \cdot q^2 = \frac{2}{R_g^2}\left(1 - \frac{1}{q^2 \cdot R_g^2}\right)$$

within a limited q range where $q^2 \cdot R_g^4 << 1$

$$q^2 \cdot I(q) \approx K$$

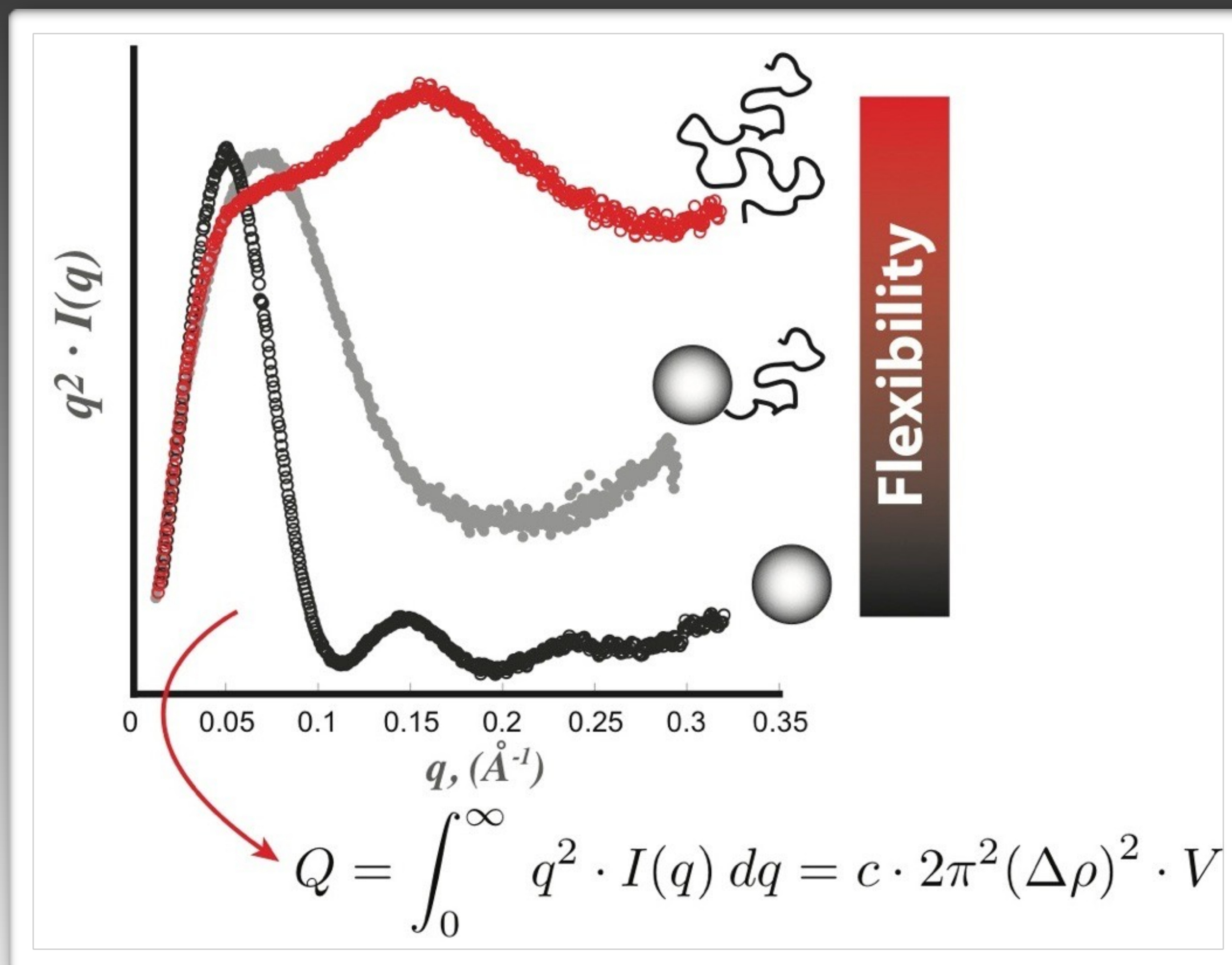


$q^2 \cdot I(q)$ becomes constant at high q
Creates hyberbolic curve
Basis for Kratky Plot ($q^2 \cdot I(q)$ vs q)

Kratky Plot

Qualitative Assessment of flexibility

for $q \bullet R_g > 1.3$, the scattering decays as $1/q^2$



A plot of $q^2 \cdot I(q)$ vs. q should approach a constant

Data must be collected to sufficiently high q with good S-to-N ratio

Porod's Law

Porod, G. (1951). Kolloid-Z. 124, 83

Fourth Power law (Porod's Law)

$$I_{particle}(q) = V \cdot \int_0^{d_{max}} \rho(r) \cdot \frac{sin(q \cdot r)}{q \cdot r} dr$$

ASSUMING:
- compact particle
- discrete e_n- contrast

$$I(q) = \Delta\rho^2 V \cdot \frac{1}{l} \cdot \frac{8\pi}{q^4}$$

$$I(q) = k \cdot \frac{1}{q^4}$$

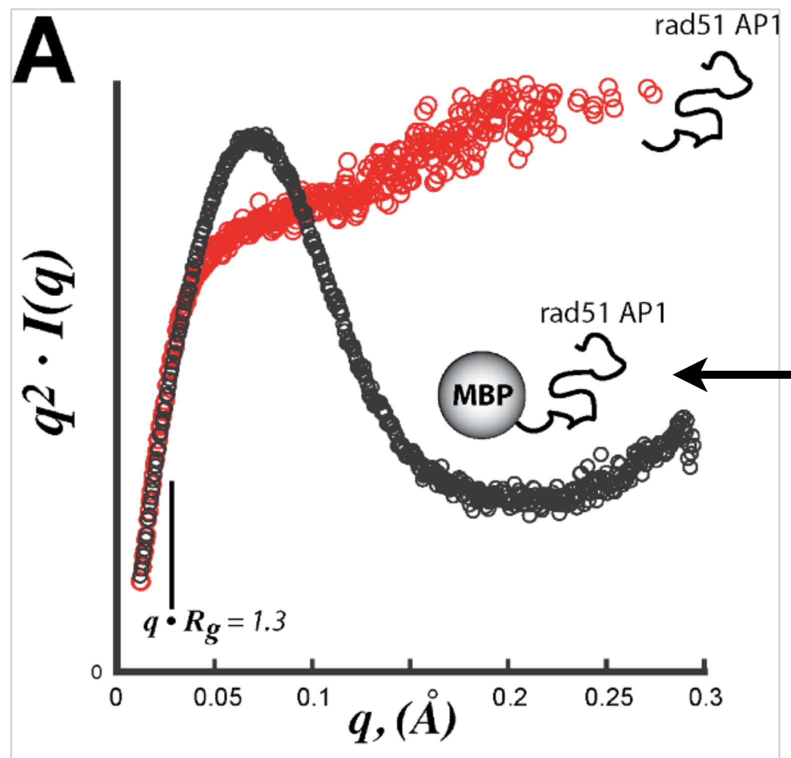
$$q^4 \cdot I(q) = constant$$

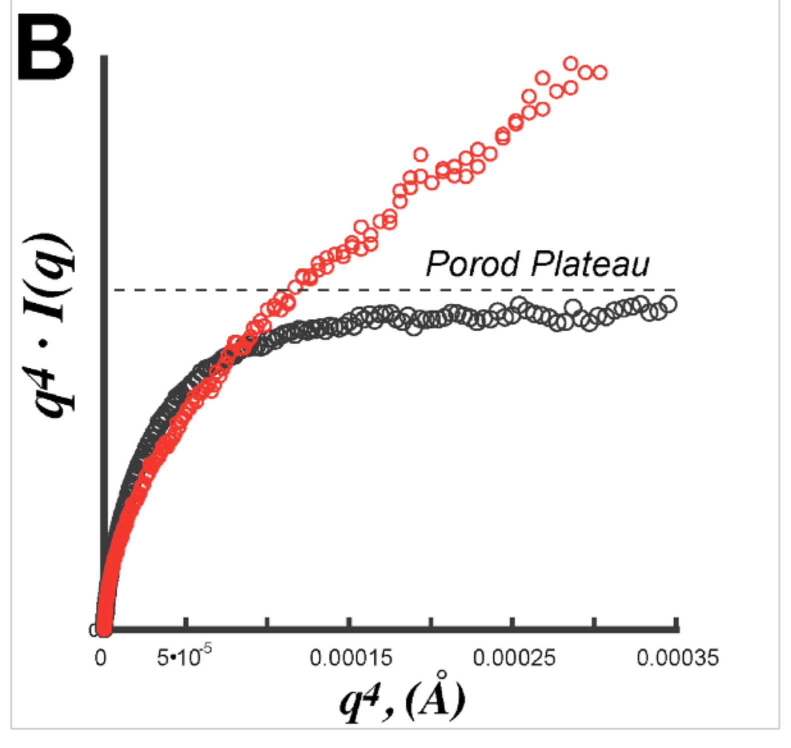
$$\frac{S}{V} = \pi \cdot lim \frac{I(q) \cdot q^4}{Q}$$

I(q) **decays as** q^{-4} **scaled by a constant value**

$q^4 \cdot I(q)$ **becomes constant at high** q

k **proportional to surface area (V/l)**





80 kDa

V_P : 145,000 Å³ $\xrightarrow{d=1.37}$ 120 kDa protein

- Flexible region reduces protein density: more like d =0.97
- Scattering contrast dominated by MBP
- Presence of the Porod plateau suggests discrete electron density contrast exists

Power Law Relationship

log vs log plot... quantitating flexibility?

if particle is flexible, should see a plateau in $q^2 \bullet I(q)$ vs. q

Porod

$$q^4 \cdot I(q) = constant$$

Debye

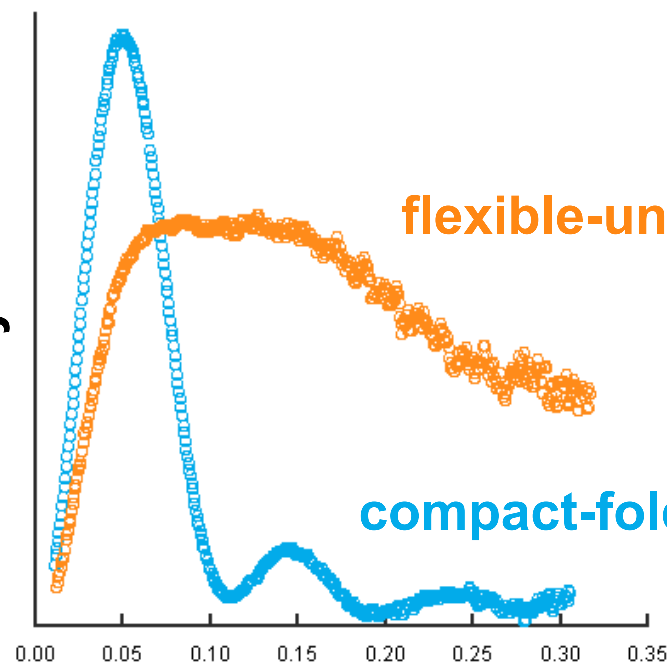
$$constant \approx q^2 \cdot I(q)$$

if particle is compact, should see a plateau in $q^4 \bullet I(q)$ vs. q and $q^4 \bullet I(q)$ vs. q^4

Defines a power law relationship!

$$I(q) = \frac{1}{q^{P_E}} \cdot S' \quad \text{where } 2 \leq P_E \leq 4$$

$$lnI(q) = -P_E \cdot ln(q) + ln(S')$$



Porod-Debye exponent quantifies the power-law decay

Low-resolution SAXS

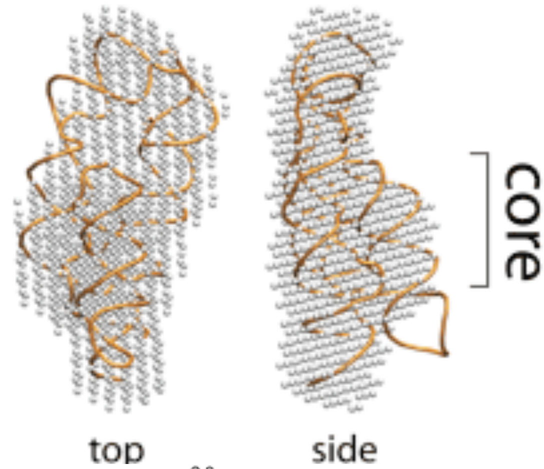
Requires small amounts of sample

Quantifying Flexibility

example Lysine Riboswitch

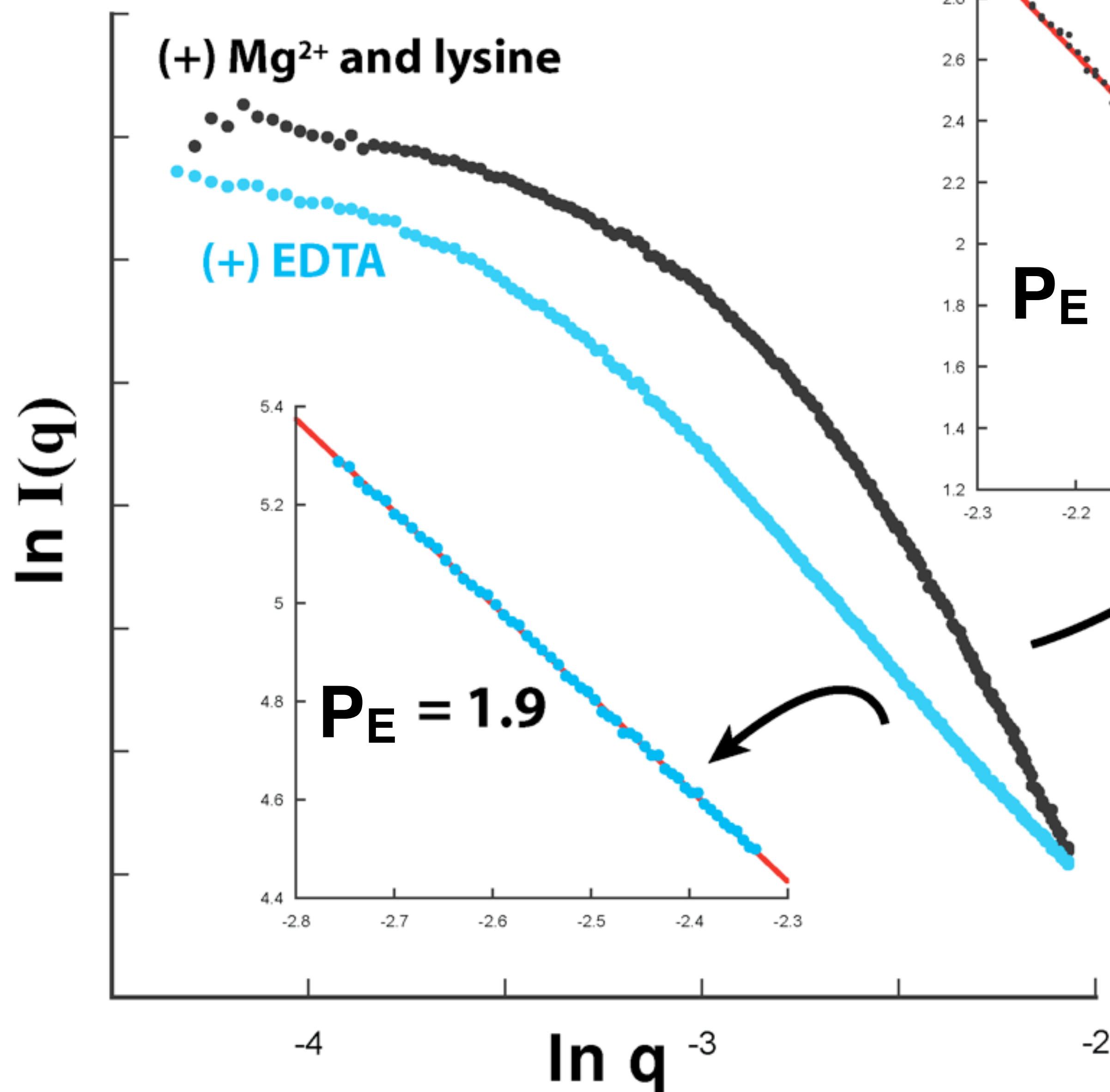
Lysine riboswitch requires:

- Mg^{2+} for folding
- binds lysine



Initial slope defines the Porod-Debye region

$$lnI(q) = -P_E \cdot ln(q) + ln(S')$$

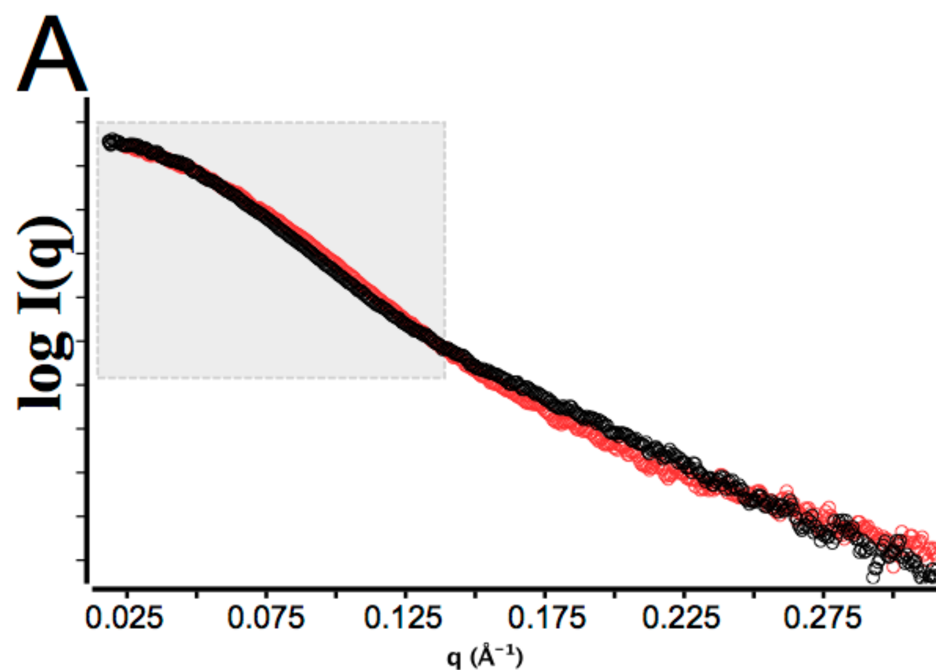


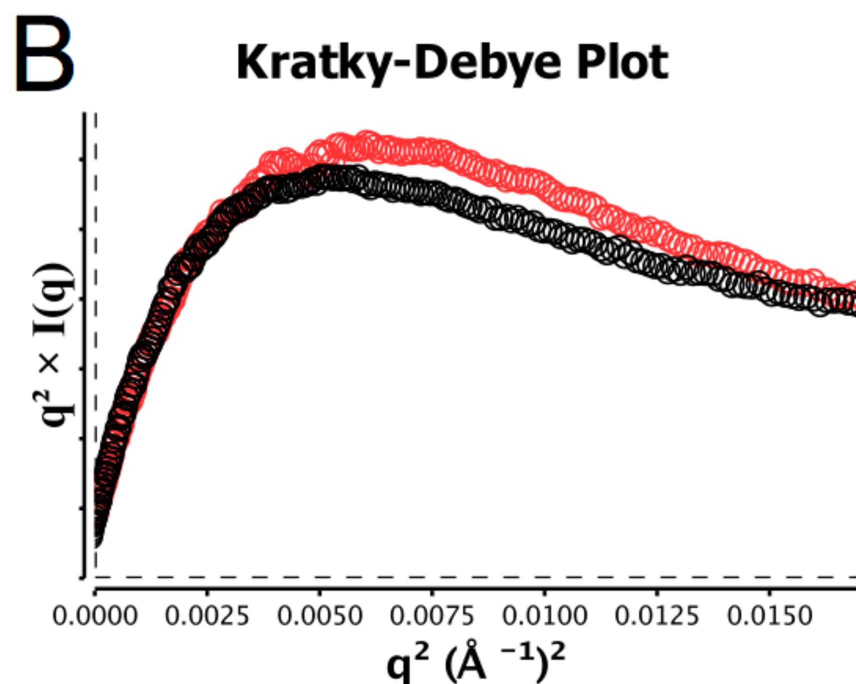
Decrease in Porod exponent (3.4 to 1.9) suggests: RNA becomes flexible in absence of Mg^{2+}

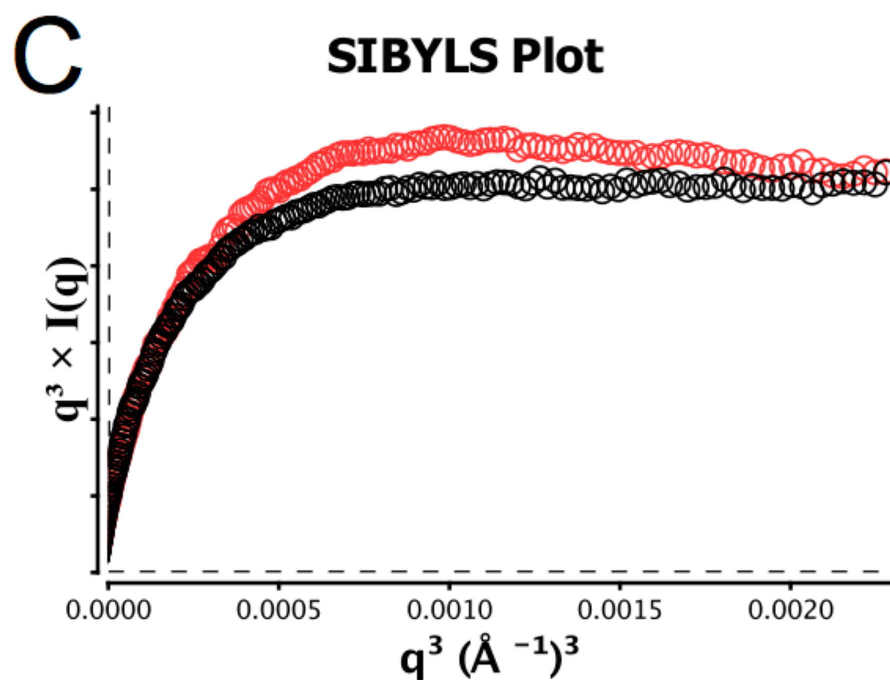
Mechanistically, this is akin to an 'induced fit"

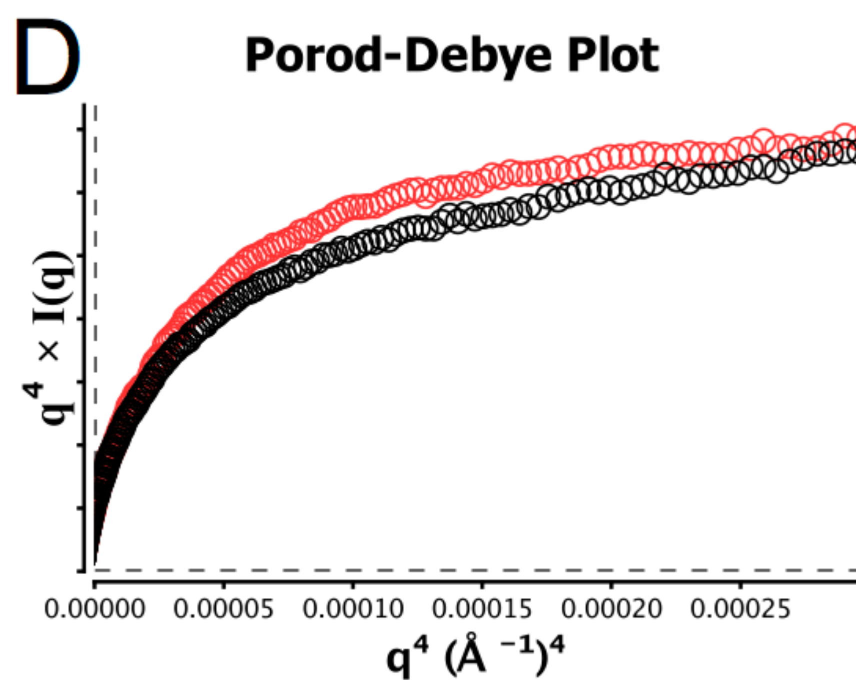
SAXS can inform on binding mechanism

Flexibility plots









SAM riboswitch (+/-) ligand
Zoom in on data where Porod-Debye approximation is true (can have more than one Porod region)
Look for the plot with a plateau, nearly flat with a positive slope (C and D)
Red(+) is more compact than black (-)

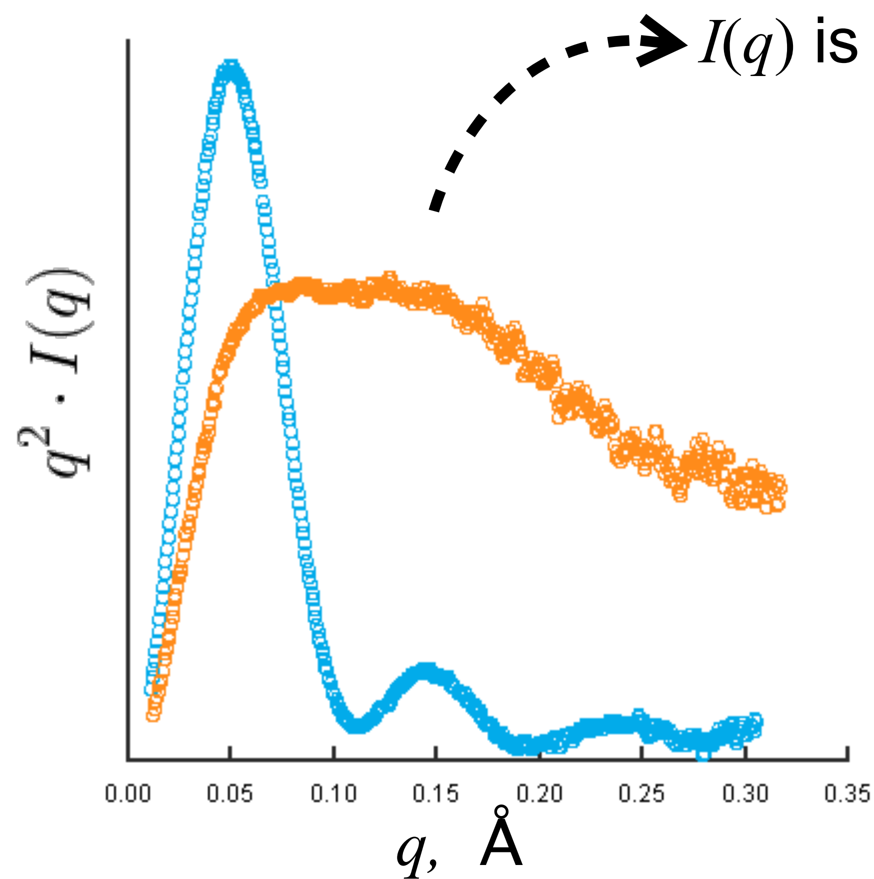
Dimensionless Kratky

scale free analysis

Receveur-Brechot V, Durand D. How random are intrinsically disordered proteins? A small angle scattering perspective. Curr Protein Pept Sci. 2012 Feb;13(1):55-75.

Durand D, et al. J Struct Biol. 2010 Jan;169(1):45-53.

Multiply $I(q)$ by $(q \bullet R_g)^2$ and divide by $I(0)$



$I(q)$ is scaled by c and V

$$I_{macromolecules}(q) = I_{macromolecule}(q) \cdot c \cdot k$$

Divide by I(0)

$$I_{particles}(0) = c \cdot I_{particle}(0) = c \cdot (\Delta\rho)^2 \cdot V^2$$

$I(q)$ is independent of concentration and normalized to V

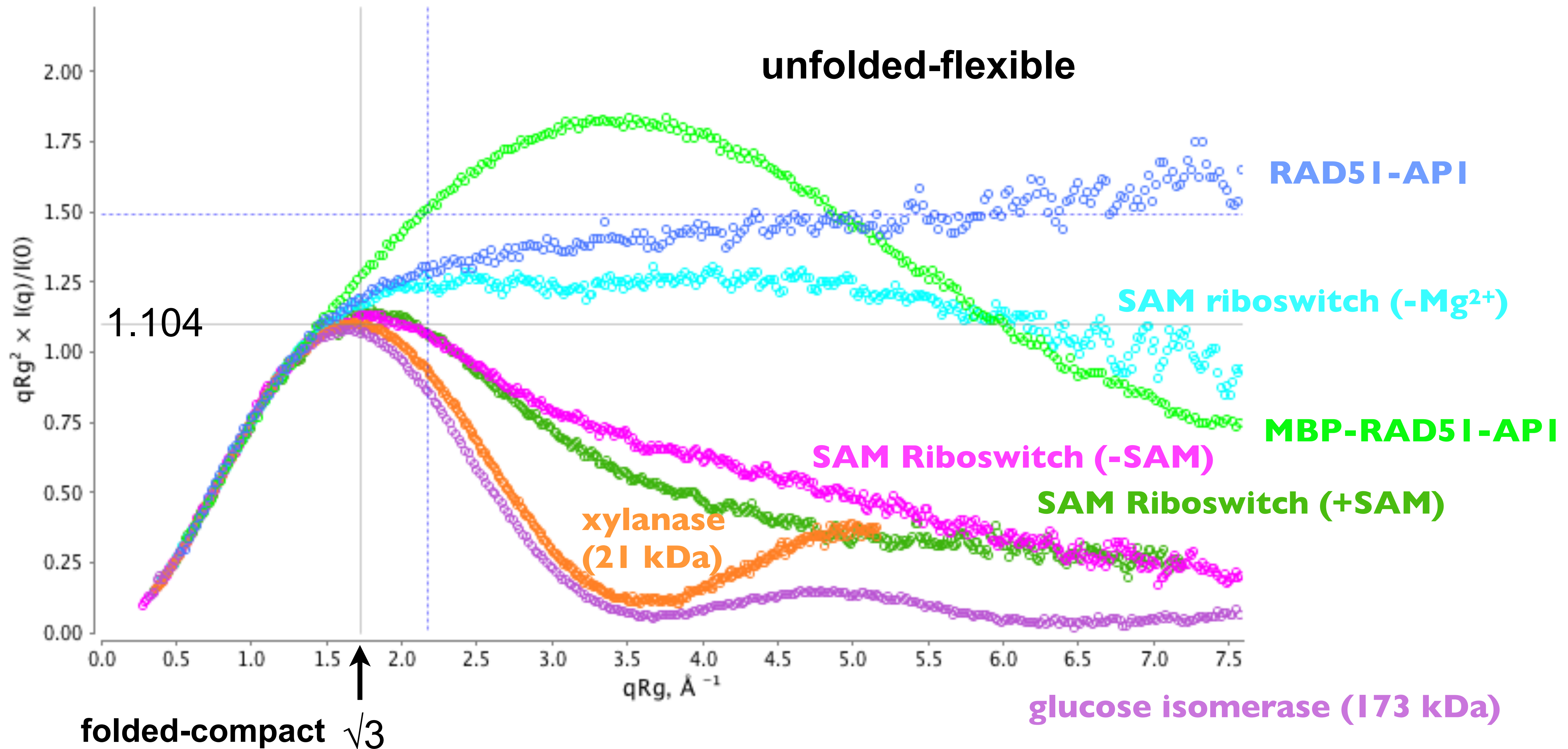
Still have units of Å^{-2}, multiply by R_g^2

What does it all mean?

Dimensionless Kratky

scale free analysis

SCÅTTER ≡ Kratky Plot

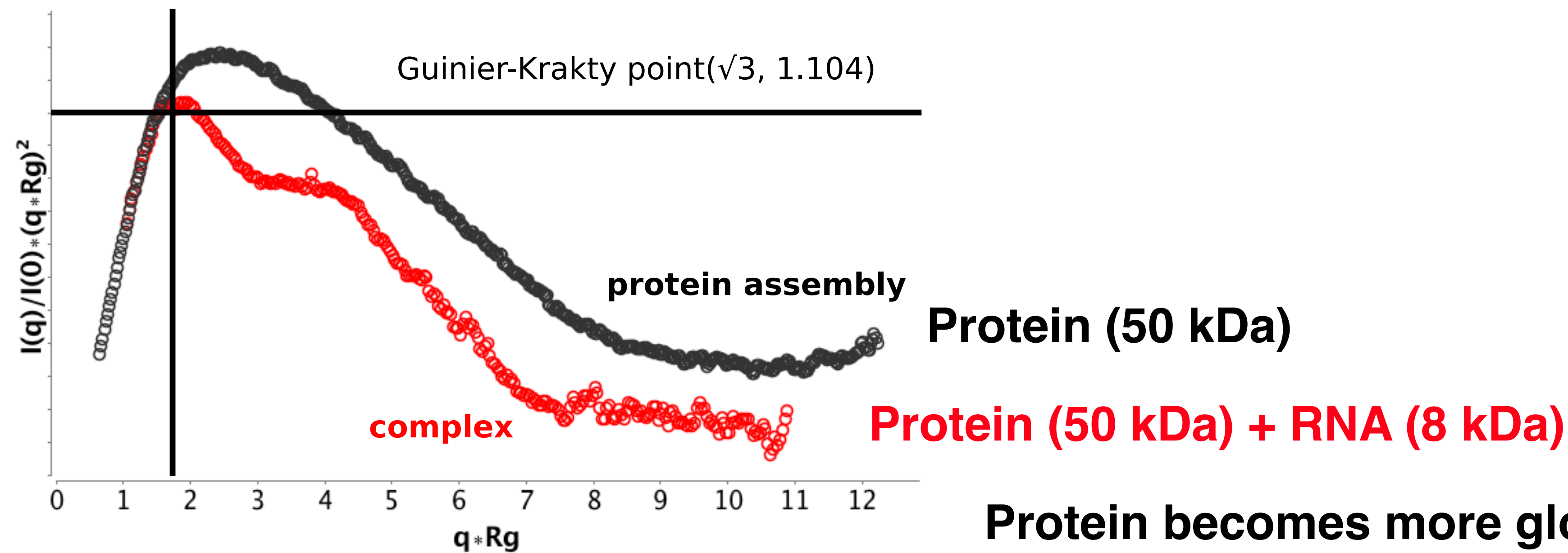


Globular particle peaks at Guinier-Kratky point (√3, 1.104)

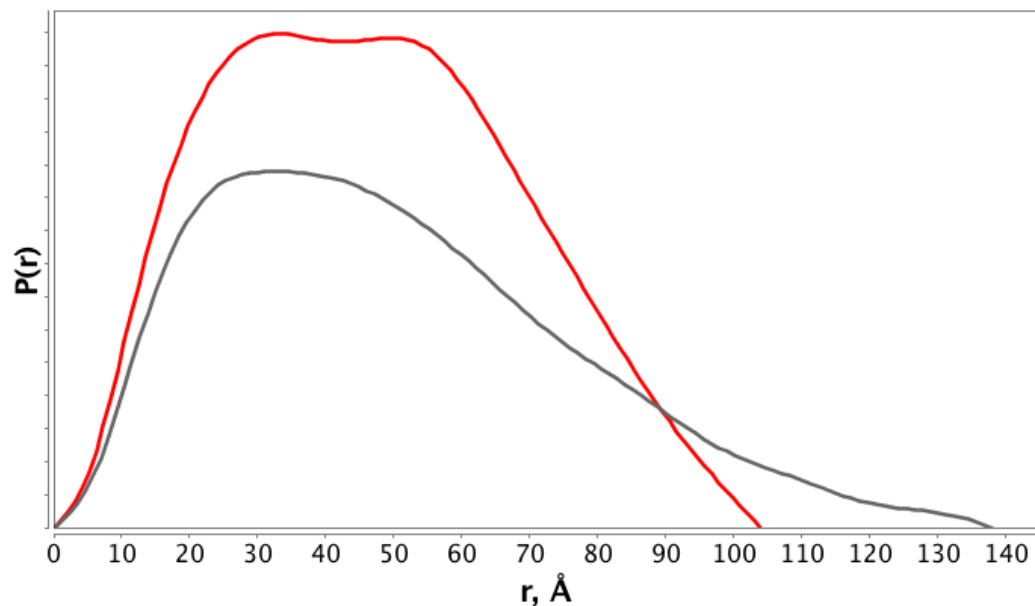
Flexible, unfolded bounded: 1.104 < peak < 2 (Debye equation Gaussian chain)

Characterize Binding or Conformational Changes

Dimensionless Kratky Plot



Protein becomes more globular with RNA present
Kratky plot suggests a large conformational change



P(r)-distribution confirms conformational change

Model Independent Analysis (no dummy atoms)

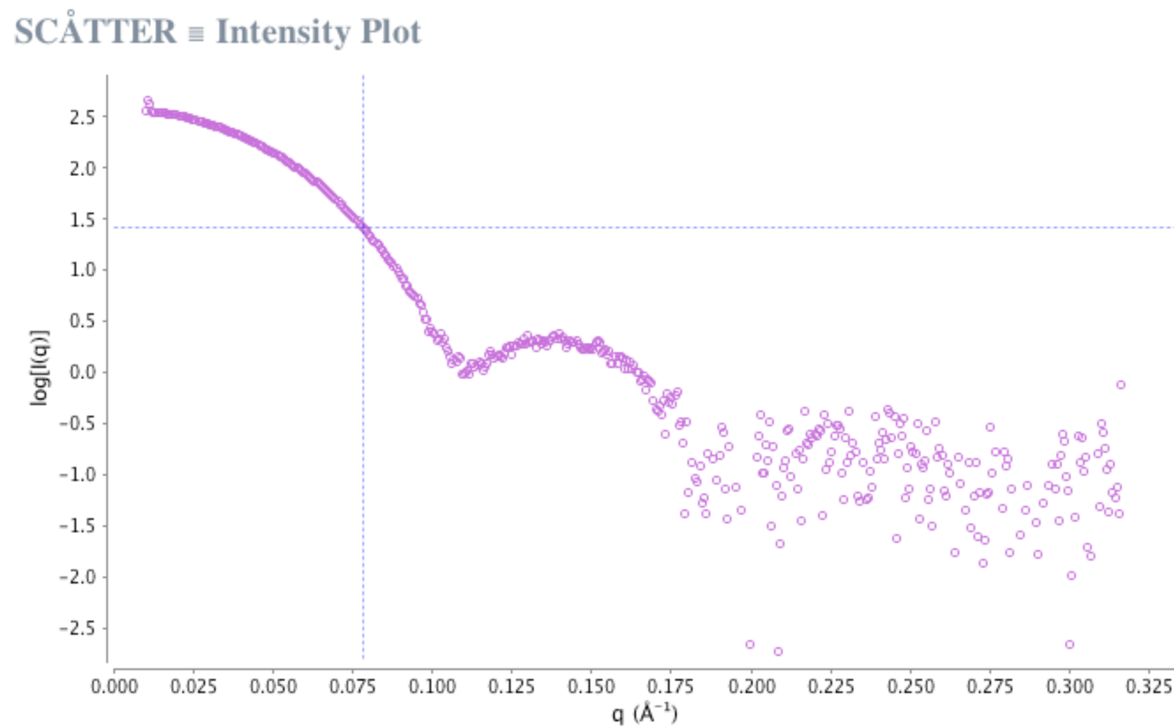
SAXS: A Measurement of TWO

Situation with poor buffer subtraction or low concentrations:
Remember, SAXS consists of two measurements:

1. **buffer**
2. **sample**

If buffer (background) is measured *n* times, perform *n* buffer subtractions and merge the data

Single 1 Second exposure



Overlay of 5