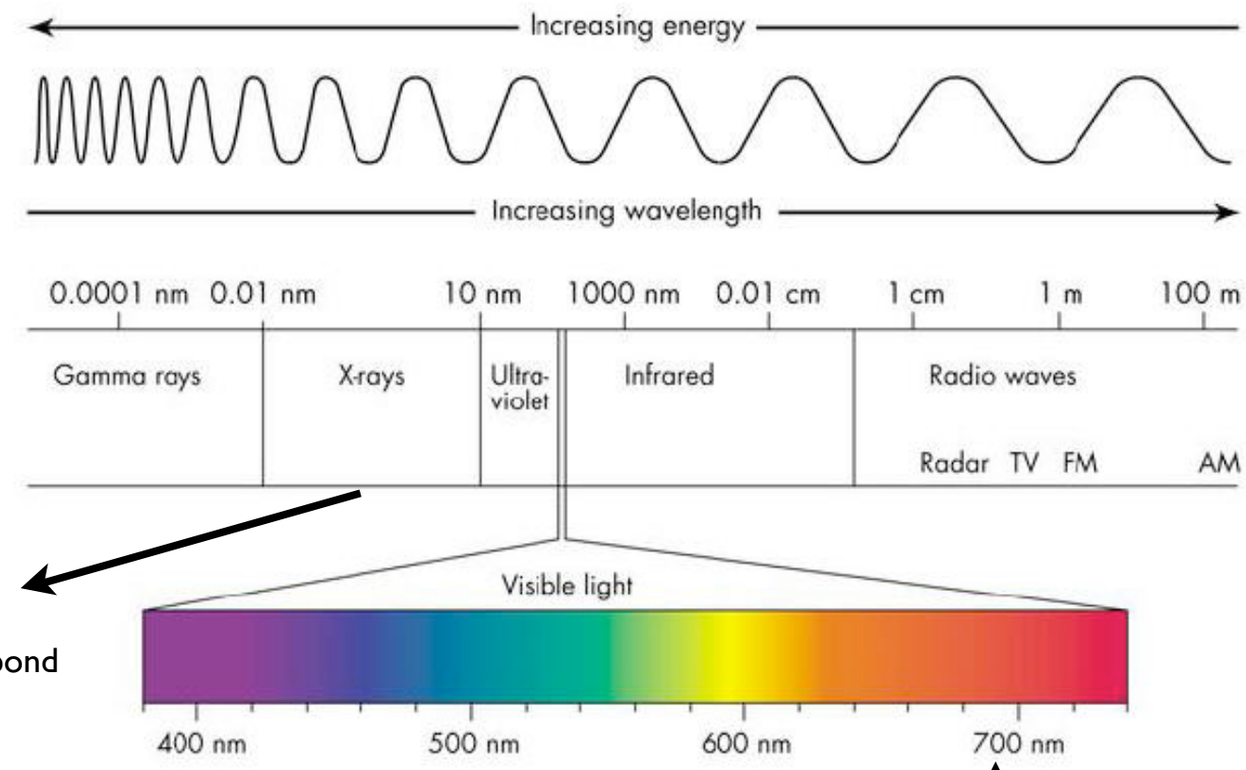
SAXS
Basics for BioSAXS

Robert P. Rambo
Diamond Light Source
B21

Scattering and Size



SAXS 0.11 nm (1.1 Å)

- smaller than C-C bond

690 nm (6900 Å) Light Scattering

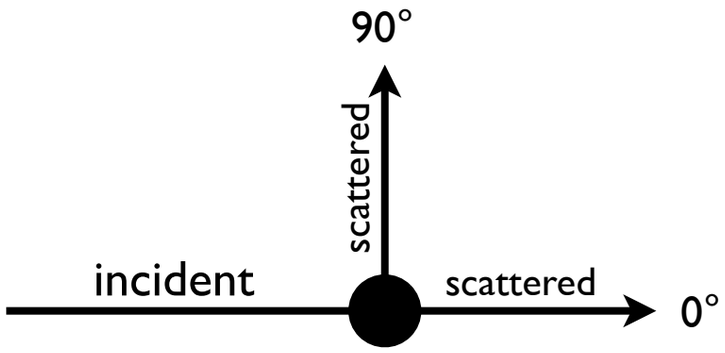
- 3.5x larger than *E.coli*
- *10x larger than a ribosome*

6000x range

Why is the sky blue, why is the sunset red?

Scattering

Effects of size on angular dependence



if R_g << wavelength (or scattering angle)

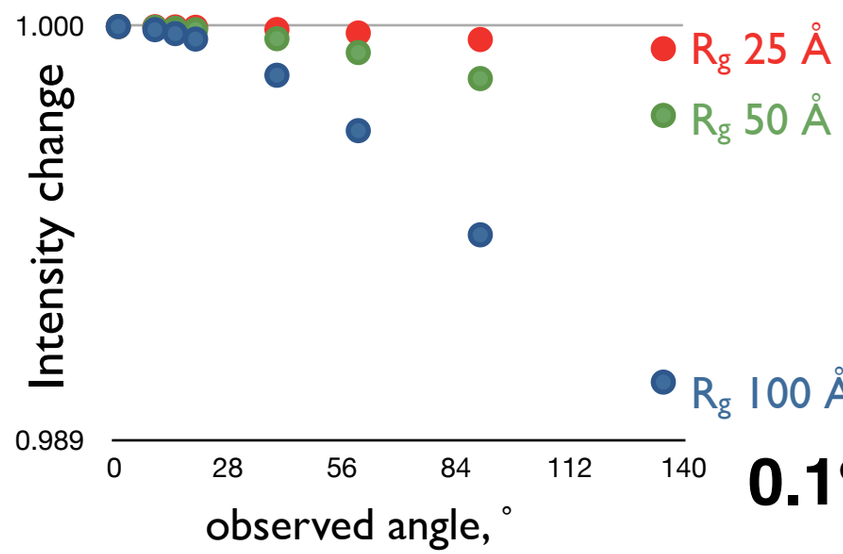
isotropic scattering ⇒ angular independence

Use this type of scattering for:

- determining molecular mass
- assessing monodispersity

	Rg	mass, kDa
ribosome	80-ish	2,000
glucose isomerase	30-ish	173
xylanase	15-ish	22
P4-P6 domain	30-ish	52

λ: 690 nm

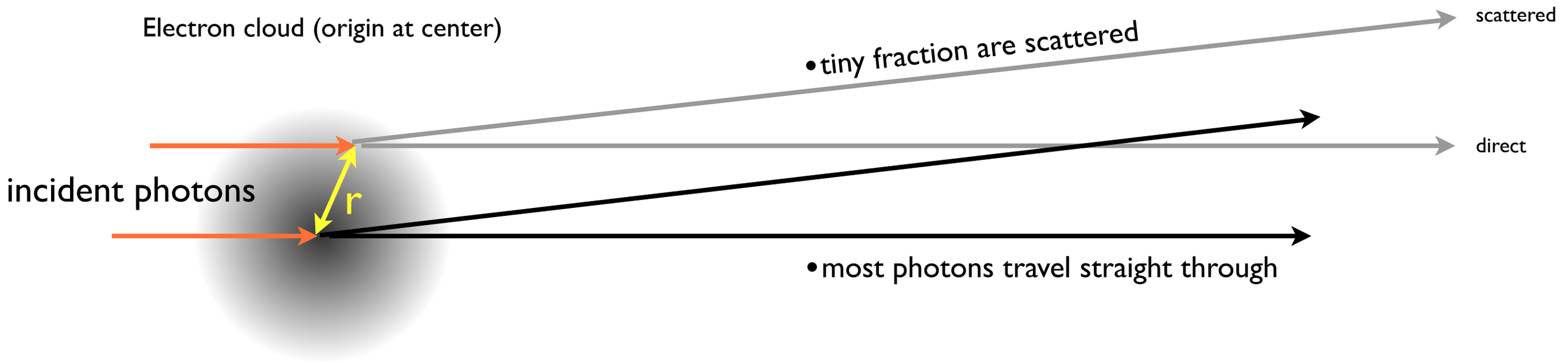


0.1% change

Small change in intensities at 690 nm

Use X-rays to get a bigger angular dependence

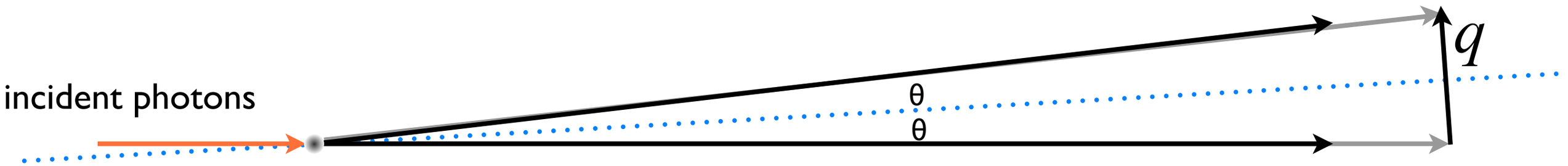
Scattering from a Single Atom



hv: 8 to 12 keV (1.55 to 1.03 Å)
flux: 10^8 (home source) to 10^{12} photons/sec (synchrotron: B21 10^{11} photons/sec)

- Photons are scattering with no change in wavelength (Thompson/Debye/Rayleigh elastic scattering)
- ~ 1% incident photons are scattered

The Scattering Angle

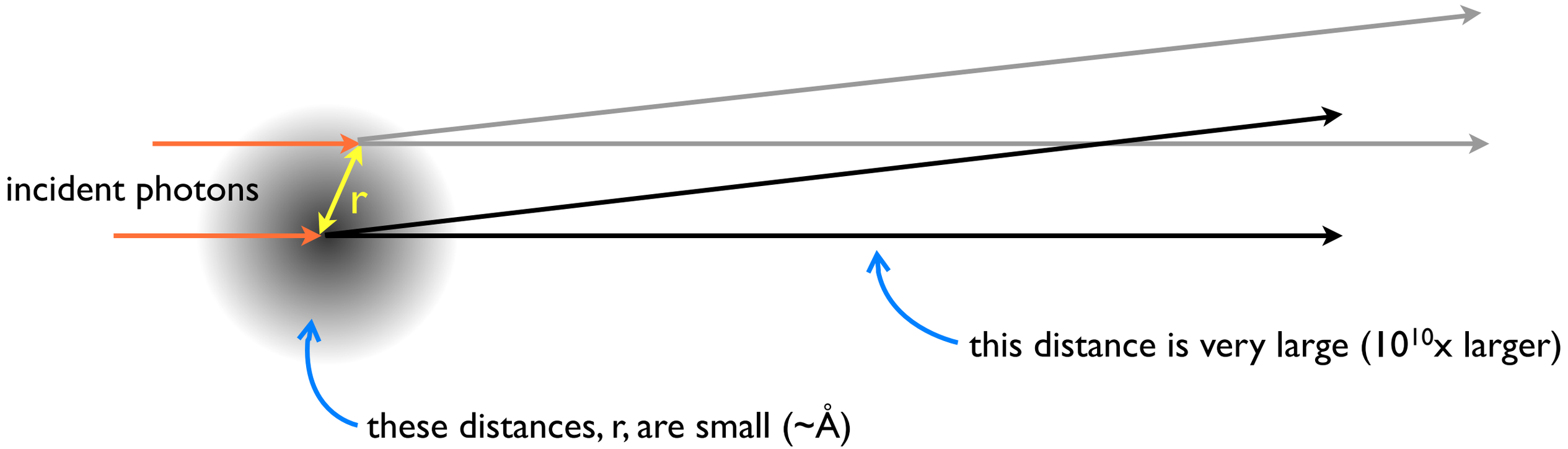


$$q = \frac{4\pi \cdot sin(\theta)}{\lambda}$$

- ***q*** is a vector (*momentum transfer vector*)
- independent of distance to detector and wavelength (λ)
- units are $Å^{-1}$
- defines scattering curve in *reciprocal space*

if *q* was defined only by θ , then scattering angle would be λ-dependent

Scattering from an Object



Scattering is described by a form factor: describes the amplitude of a scattered wave as a Fourier transform of an object's spatial distribution

Molecular Form Factor

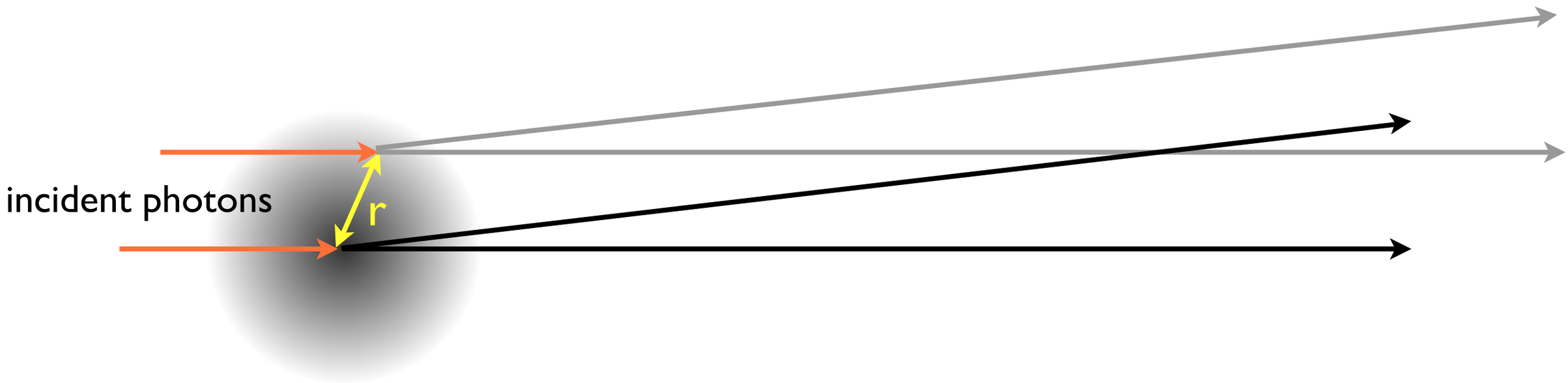
$$A(q) = \int \rho(r) \times e^{iqr} dV$$

spatial distribution

Fourier term

Note: r is in fixed orientation

Scattering from an Object



Molecular Form Factor $A(q) = \int \rho(r) \times e^{iqr} dV$

Scattered Intensity, I(q), is the complex norm of the form factor

I(q) will be intensity of an oriented particle

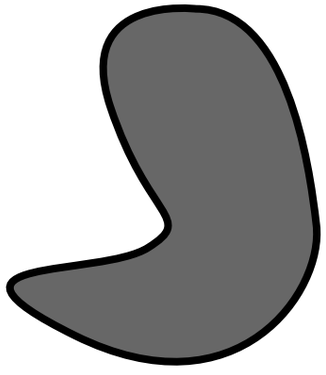
$$I(q) = A(q) \times A^*(q) = \int\int \rho(r) \times \rho(r^*) \times e^{-iq(r-r^*)} dV \; dV^*$$

Complex Norm

Correlation function

Fourier term is now the difference between 2 position vectors. Internal coordinate system becomes internal

Pair-distance Distribution Function



$p(r)$ = 0 when r > d_{max}

- defined in real space
- no negative values ($\Re^+$)
- zero except for defined distances
- expected to be smooth as r → d_{max}

Set of all distances measured within the particle.

$$\sum_{internal\ distances} = \sum r_{ij} = \int \rho(r)\, dr$$

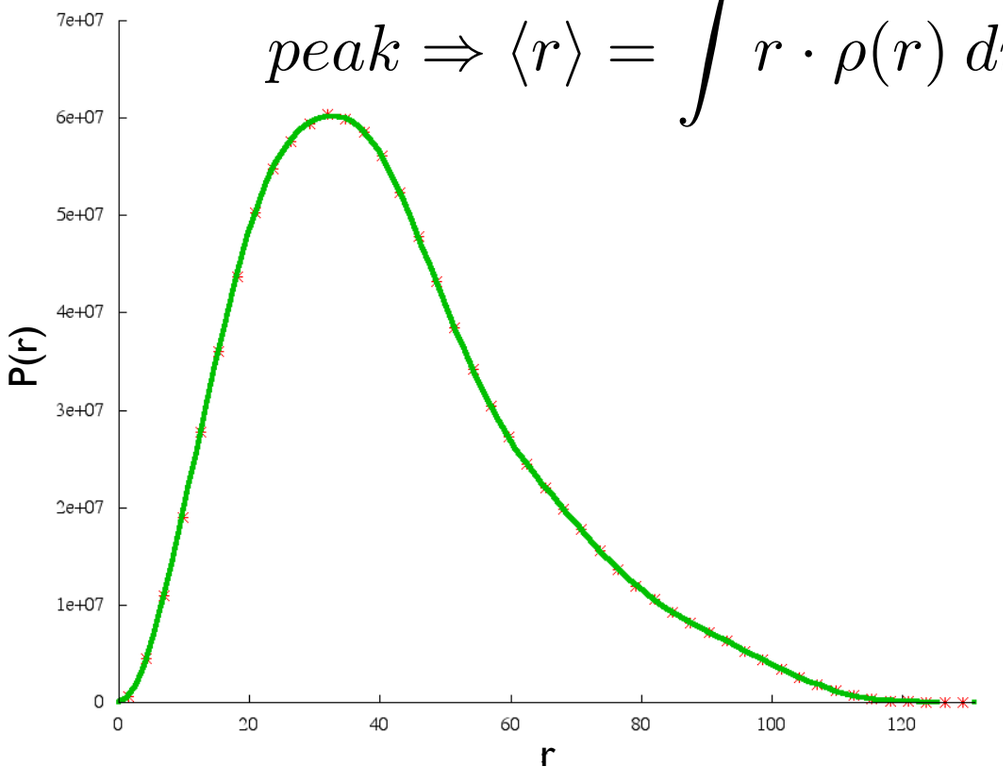
logically this must be $V_{particle}$

$$V_{particle} = \int_0^{d_{max}} \rho(r)\, dr$$

implies $p(r)$ has units of Å^2

$$\rho(r) = r^2\gamma(r)$$ γ(r): correlation function

$$peak \Rightarrow \langle r \rangle = \int r \cdot \rho(r)\, dr$$

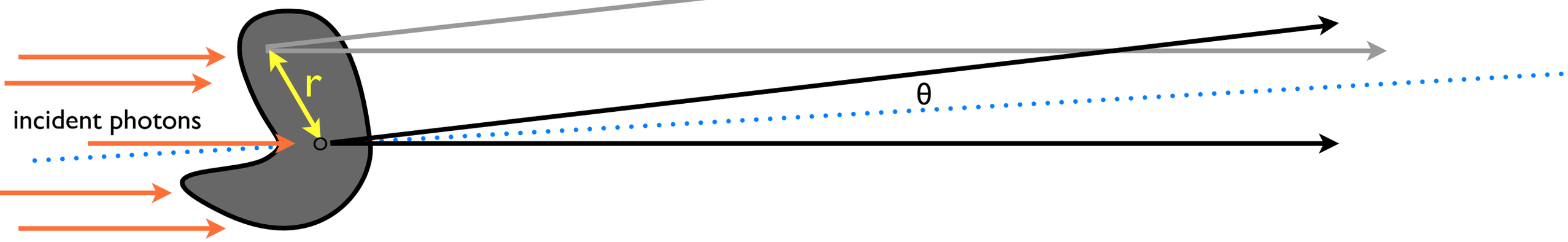


Knowing P(r):

1. Determine $V_{particle}$
2. R_g (real space)
3. Correlation function

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

Scattering from a Particle



In SAXS:

- particles are not centered at origin
- particles are sampled in all orientations

Debye factor

$$exp\{-2\pi i(\mathbf{q}\cdot\mathbf{r})\} \Rightarrow (exp\{-2\pi i(\mathbf{q}\cdot\mathbf{r})\})_{average} = \frac{sin(\mathbf{q}\cdot\mathbf{r})}{\mathbf{q}\cdot\mathbf{r}}$$

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

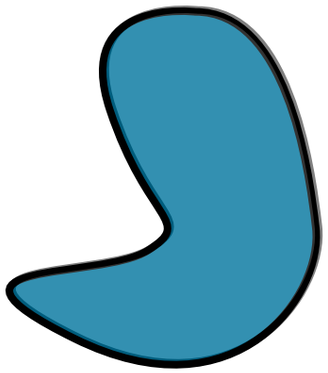
Pair-distance distribution function

$p(r)$ = 0 when r > d_{max}

- defined in real space
- no negative values ($\Re^+$)
- zero except for defined distances

Scattering by an Inhomogeneous Solid, Debye P and Bueche AM. 1949 Journal of Applied Physics

Correlation Function

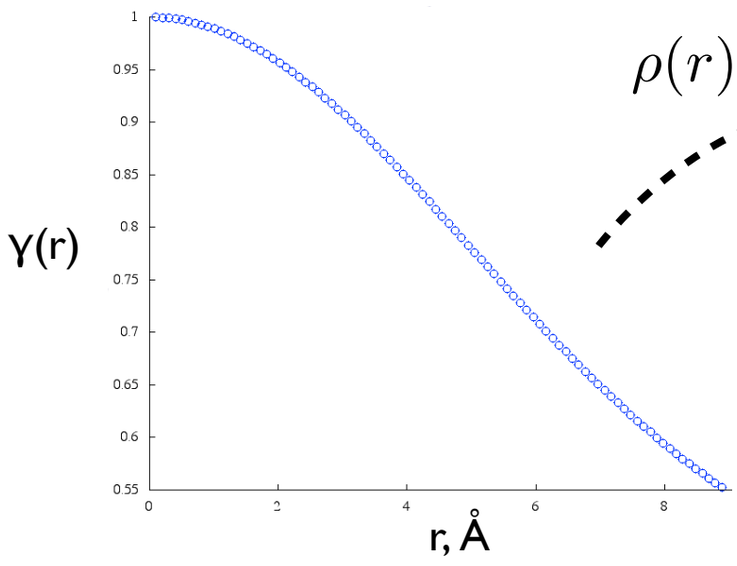


recalculate

- displace particle by a small distance, Δr
- calculate how well the particle correlates with its previous self

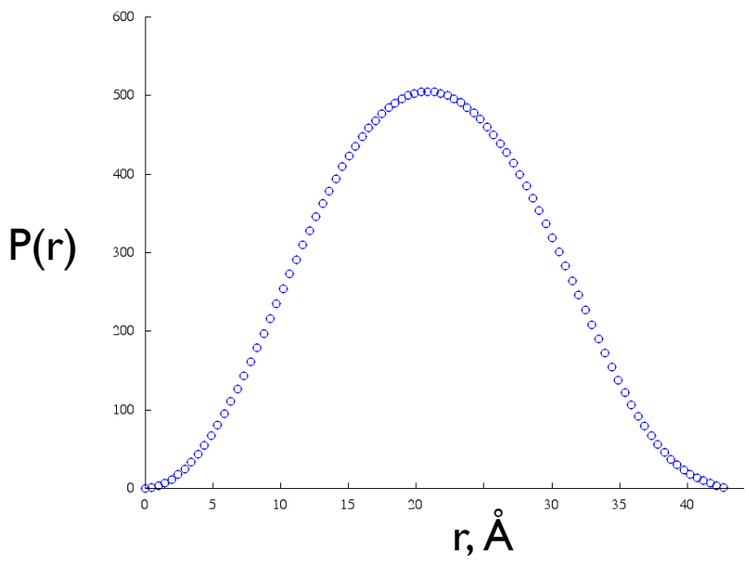
increase Δr

Correlation Function



$$\rho(r) = r^2\gamma(r)$$

Pair-distance Function



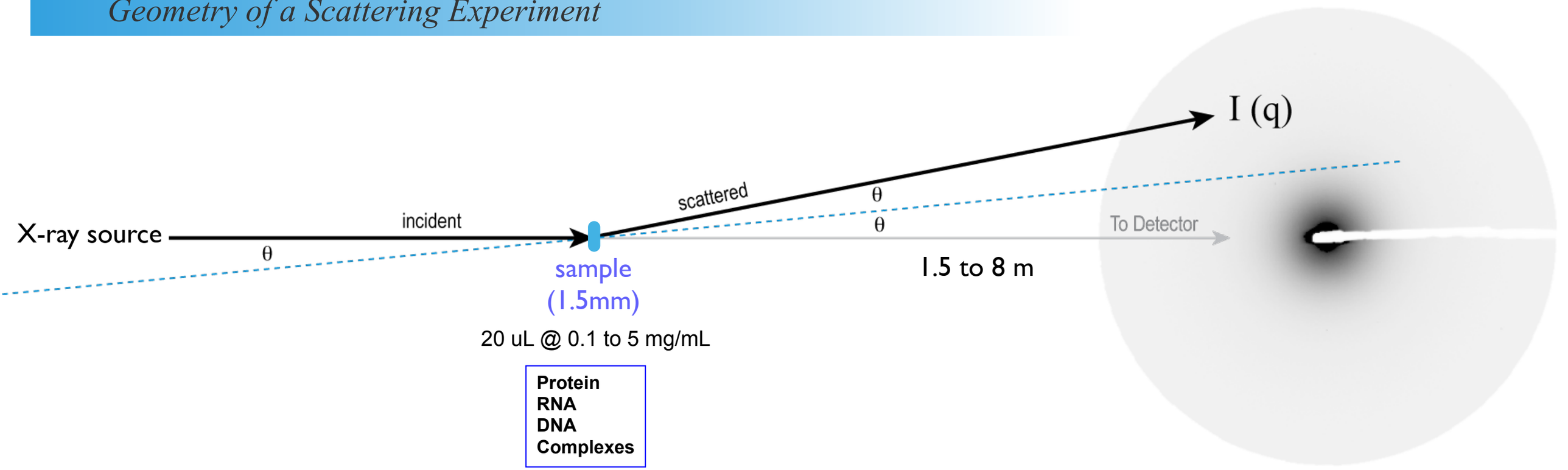
Maximum self-correlation occurs at r = 0
Correlation decays to 0 at r > d_{max}

l_c, mean width of γ(r)

$$l_c = \frac{l_{ave}^2}{l_{ave}}$$

SAXS

Geometry of a Scattering Experiment



In a real experiment, scattering contributions from:

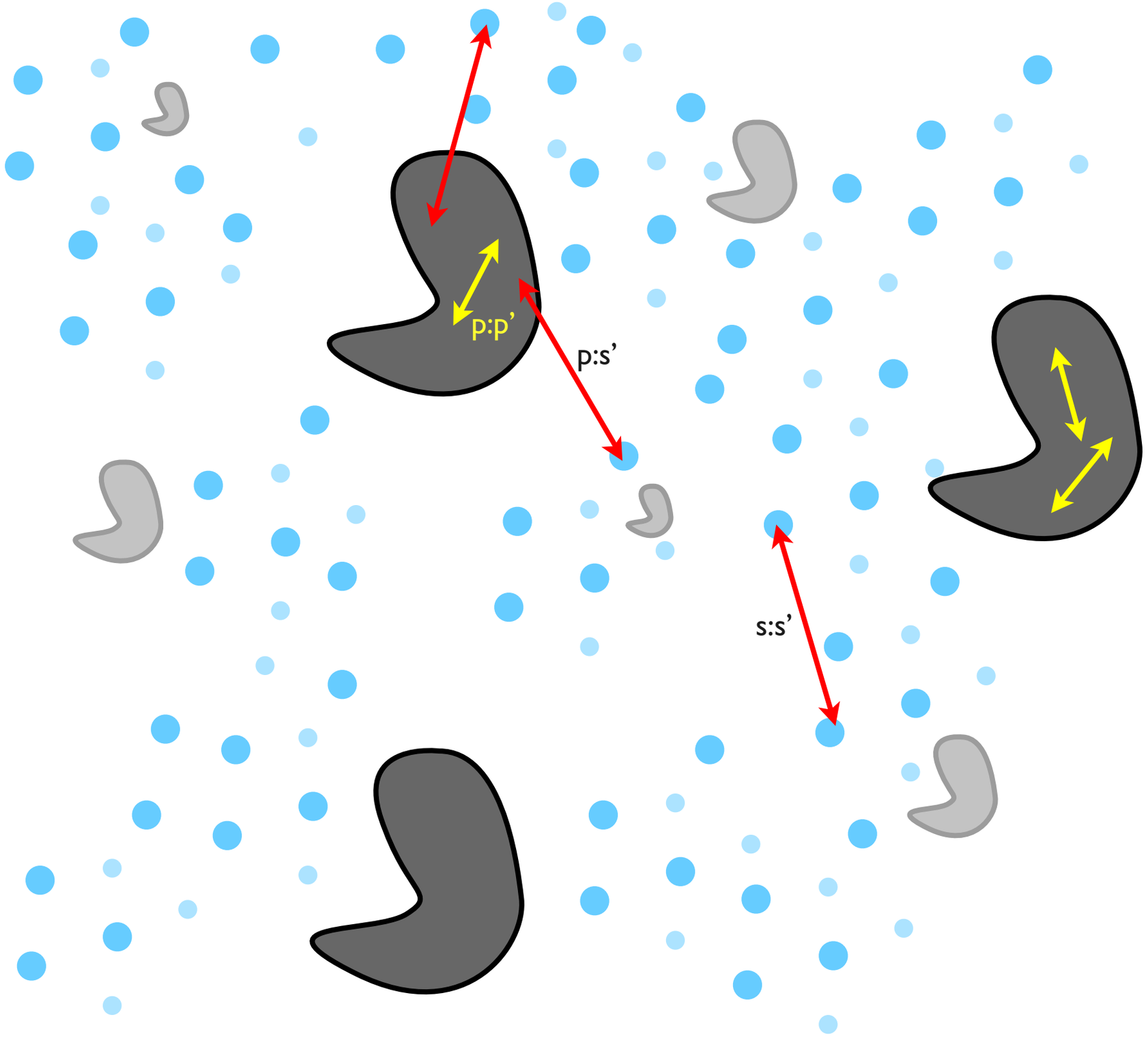
1. **sample cell**
2. **solvent**
3. **air**

Correct for above by buffer subtraction:

- **need at least 1 SAXS measurement of “buffer” sample**
- **subtract from SAXS data sample**

Debye Method

Buffer Subtraction



Using a distance vector:

$$r = \frac{2\pi}{q}$$

Randomly through **r** in and make note of its ends and count.

Several end-to-end pairs:

1. $n_p : n_s$
2. $n_s : n_{s'}$ (intra)
3. $n_p : n_{p'}$ (intra)
4. $n_p^i : n_p^i$ (inter)

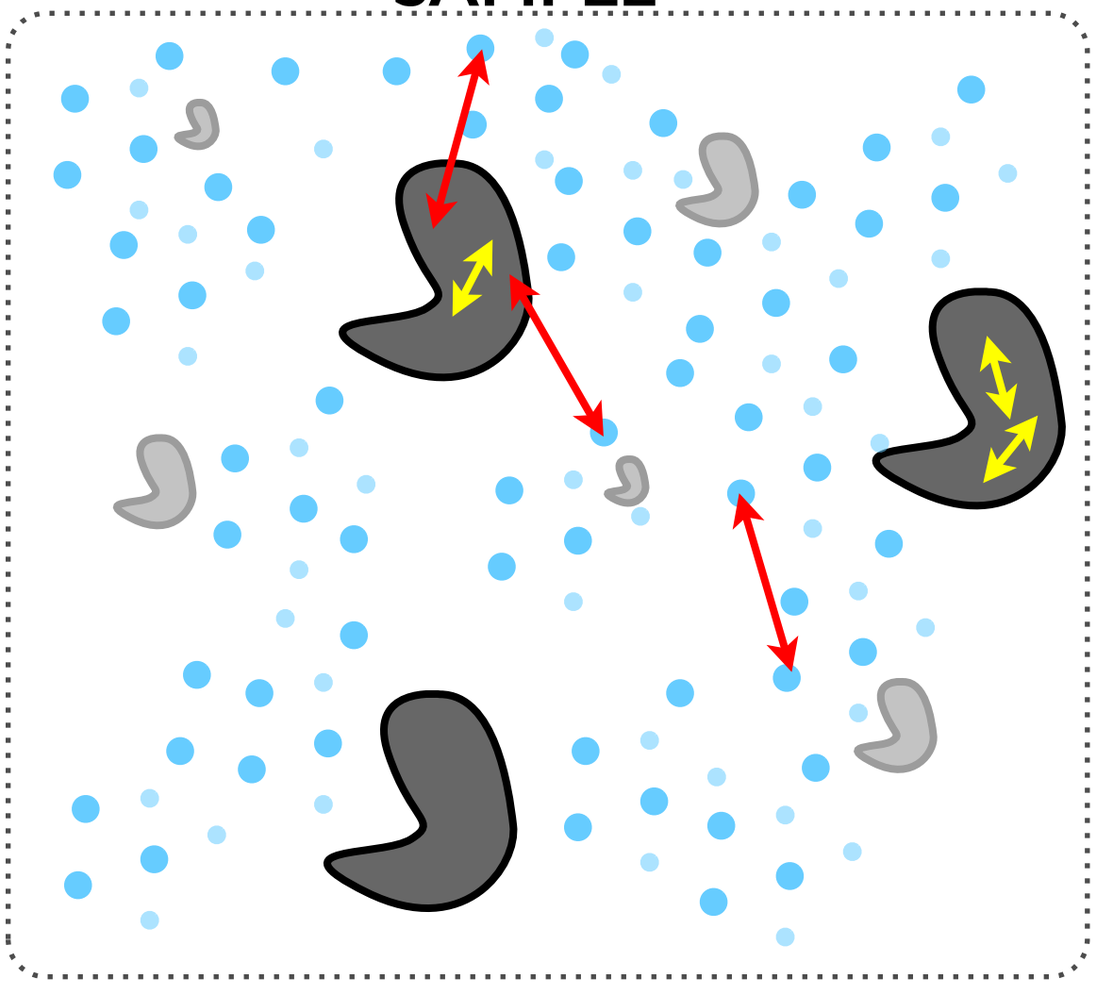
$$\sum_{n_s:n_{s'}} f_s \cdot f_{s'} \frac{sin(qr)}{qr}$$

$$\sum_{n_p:n_s} f_p \cdot f_s \frac{sin(qr)}{qr}$$

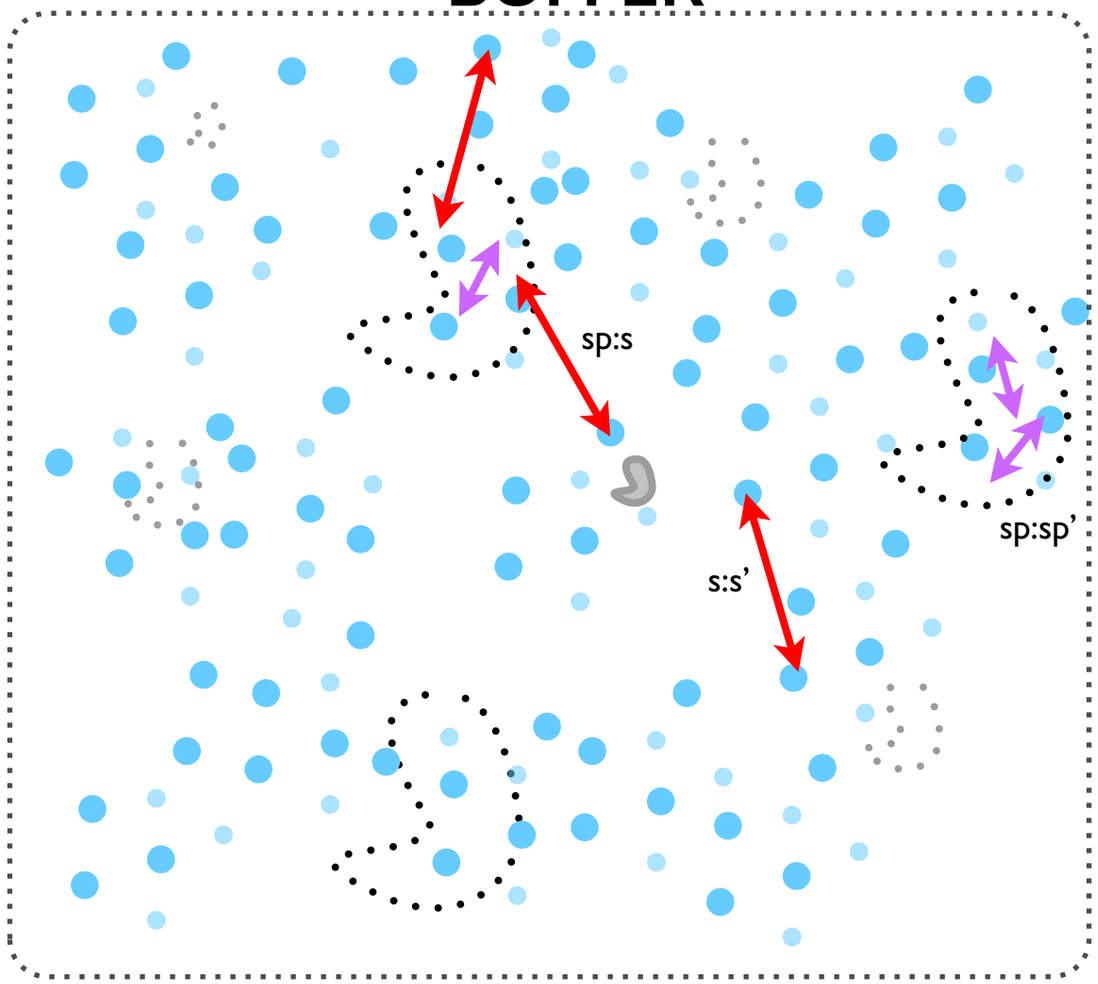
$$\sum_{n_p:n_{p'}} f_p \cdot f_{p'} \frac{sin(qr)}{qr}$$

Debye Method

Buffer Subtraction



—



What's left in the difference?

$$\sum_{n_s:n_{s'}} f_s \cdot f_{s'} \frac{sin(qr)}{qr}$$

$$\sum_{n_p:n_s} f_p \cdot f_s \frac{sin(qr)}{qr}$$

$$\sum_{n_p:n_{p'}} f_p \cdot f_{p'} \frac{sin(qr)}{qr}$$

—

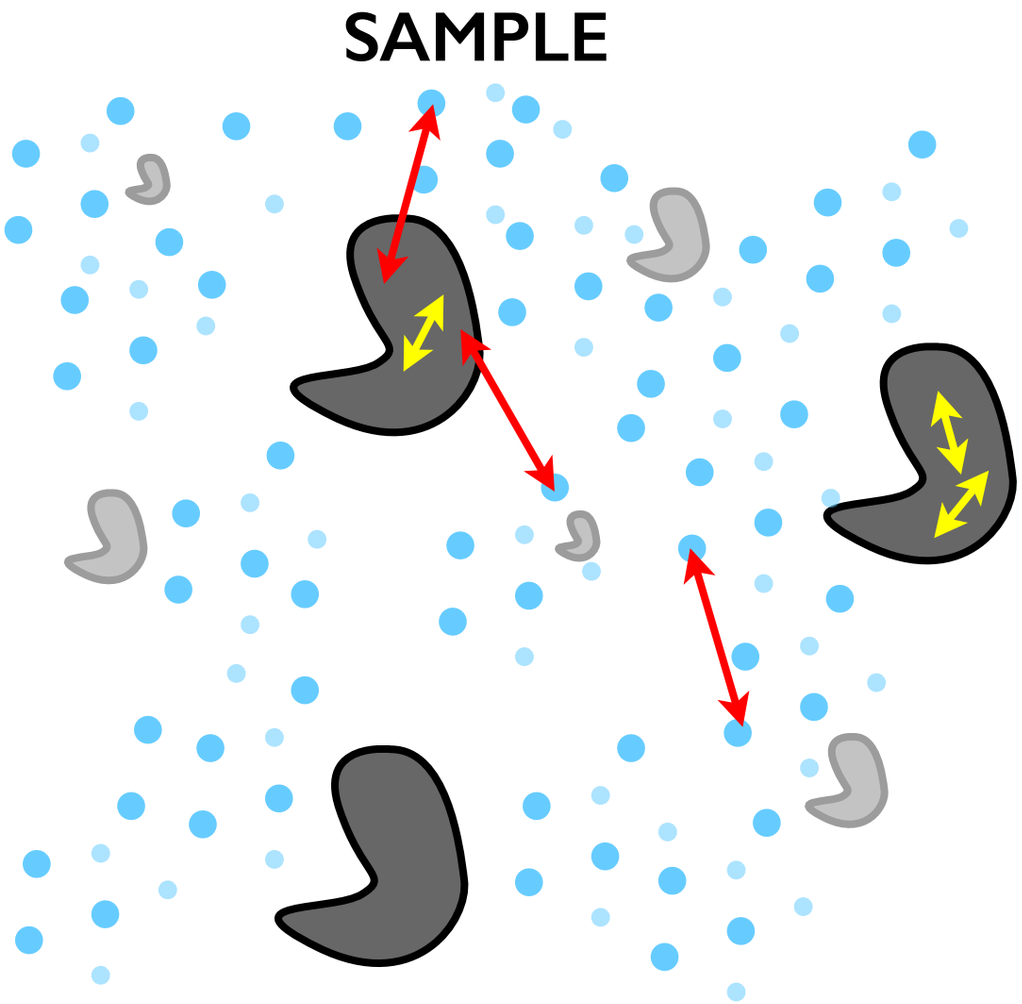
$$\sum_{n_s:n_{s'}} f_s \cdot f_{s'} \frac{sin(qr)}{qr}$$

$$\sum_{n_{sp}:n_s} f_{sp} \cdot f_s \frac{sin(qr)}{qr}$$

$$\sum_{n_{sp}:n_{sp'}} f_{sp} \cdot f_{sp'} \frac{sin(qr)}{qr}$$

Debye Method

Buffer Subtraction



$$\sum_{n_p:n_s} f_p \cdot f_s \frac{sin(qr)}{qr} \quad - \quad \sum_{n_{sp}:n_s} f_{sp} \cdot f_s \frac{sin(qr)}{qr}$$

$$\sum_{n_p:n_{p'}} f_p \cdot f_{p'} \frac{sin(qr)}{qr} \quad \sum_{n_{sp}:n_{sp'}} f_{sp} \cdot f_{sp'} \frac{sin(qr)}{qr}$$

Actual measured SAXS curve contains artifacts:

$$I_{obs}(q) = I_{sample}(q) - I_{buffer}(q) = I_{ps}(q) + I_{pp'}(q) - I_{sp:s}(q) - I_{sp:sp'}(q)$$

particle scattering — internal scattering — excluded volume — internal scattering

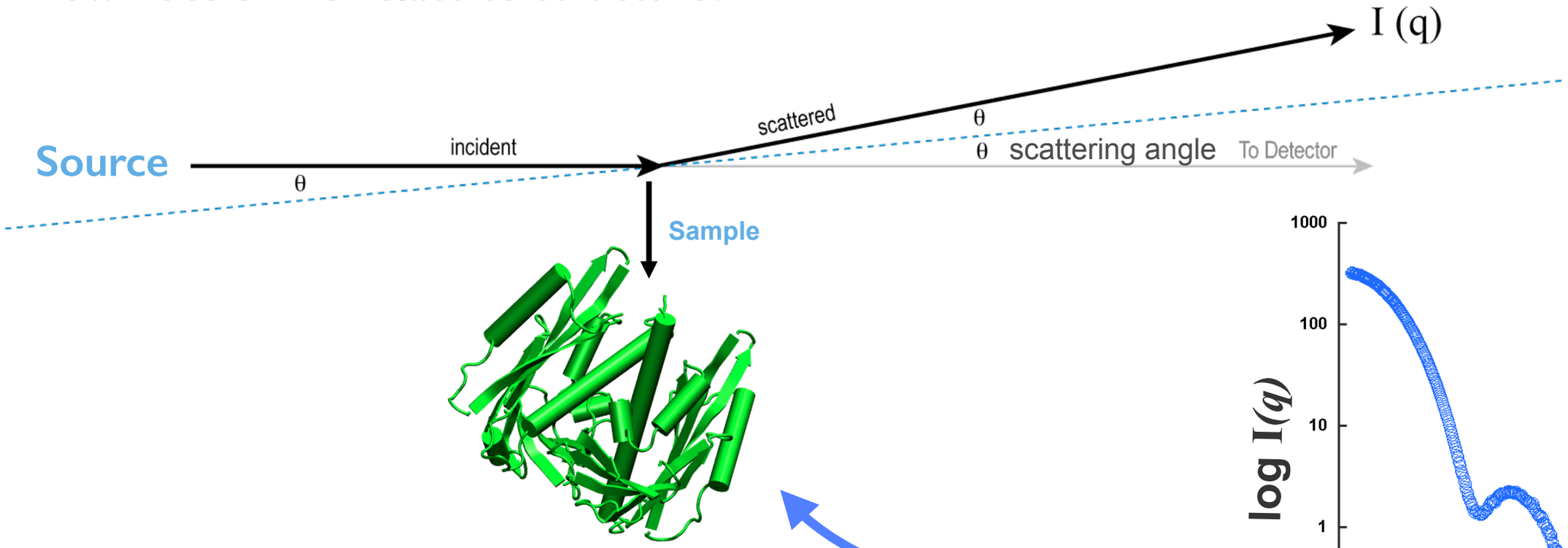
$$I_{ps}(q) > I_{sp:s}(q) >> I_{pp'}(q), I_{sp:sp'}(q)$$ (for q < 0.18-ish)

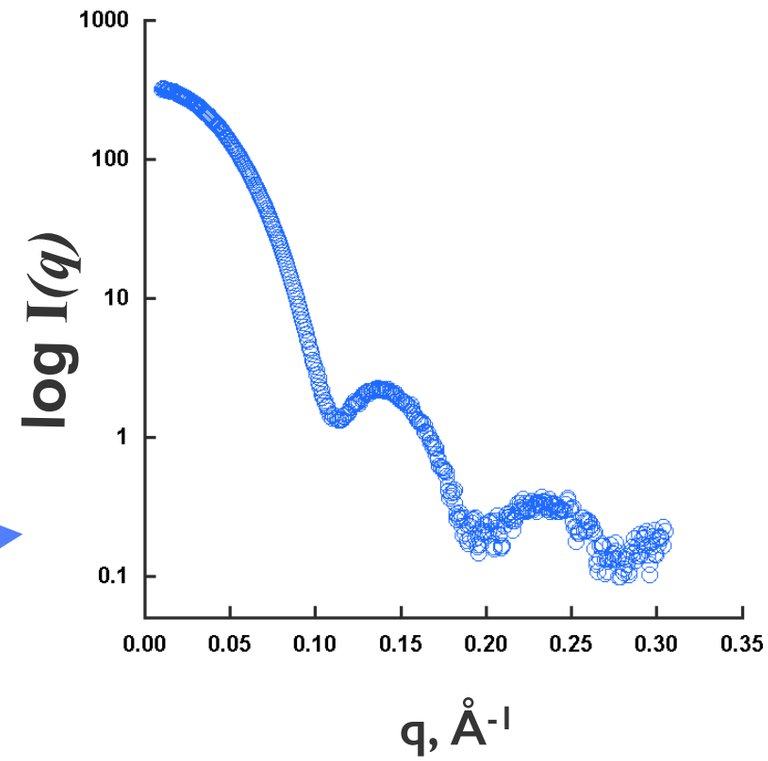
$$I_{ps}(q), I_{sp:s}(q) \approx I_{pp'}(q), I_{sp:sp'}(q)$$ (for q > 0.18-ish)

SAXS

Small Angle X-ray Scattering

How does SAXS relate to structure?



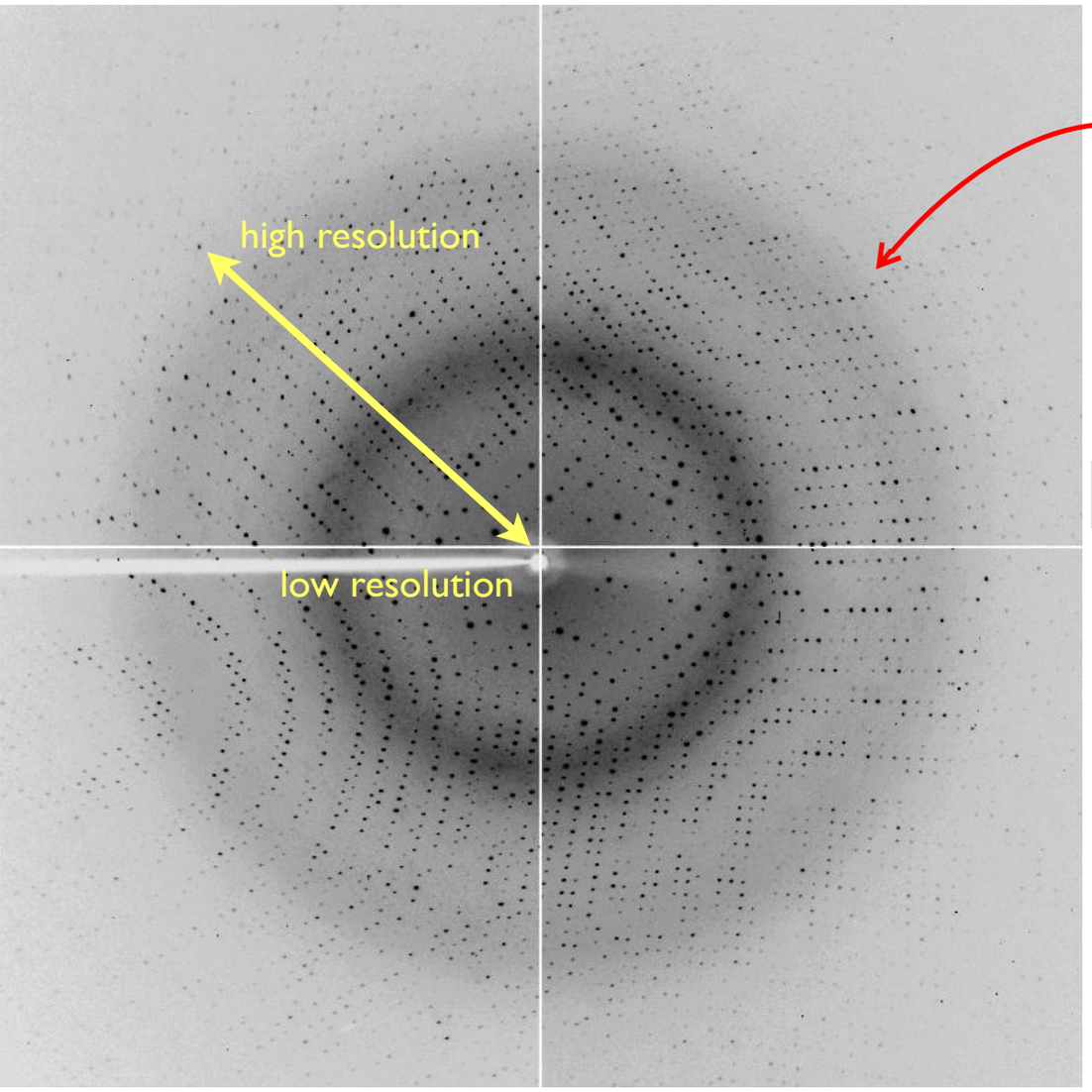


$$I_{molecule}(q) = \int p(r) \cdot \frac{\sin(q \cdot r)}{q \cdot r}\, dr$$

Set of all distance vectors within the particle

P(r) ~ Pair Distance Distribution Function

Diffraction vs SAXS



Diffraction is a consequence of an ordered array of scatters.

- each "spot" is a reflection, $I_{obs}(h\ k\ l)$
- no reflection ⇒ no data!

Resolution is measured as the increase in reflections away from center.

An increase in diffraction resolution ⇒ an increase in observed reflections, $I_{obs}(h\ k\ l)$

$$\rho(x\,y\,z)_{e_n density} = \frac{1}{V}\sum_h\sum_k\sum_l \left|F(h\,k\,l)\right|^2 \cdot e^{\left|2\pi\, i(hx+ky+lz)+i\,\alpha(hkl)\right|}$$

$F_{h\,k\,l}$ observed as $I_{h\,k\,l}$

In crystallography, I_{obs} can be related back to a structure in real space (e_n^- density).

Not true for SAXS!

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

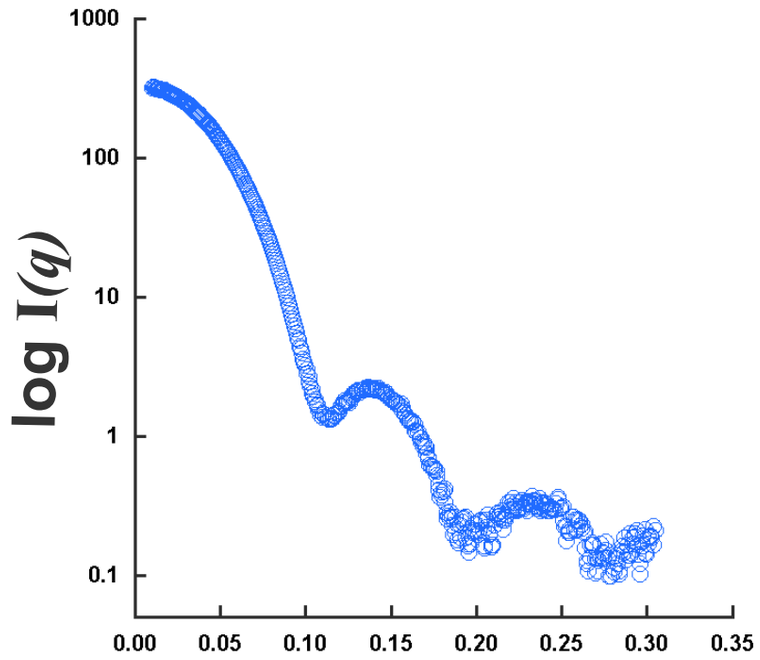
Is there an equivalent formalism in SAXS?

What constitutes a high or low resolution data set in SAXS?

Resolution

d-spacing Vector

Define d-spacing vector (from crystallography): $d = \frac{2\pi}{q}$

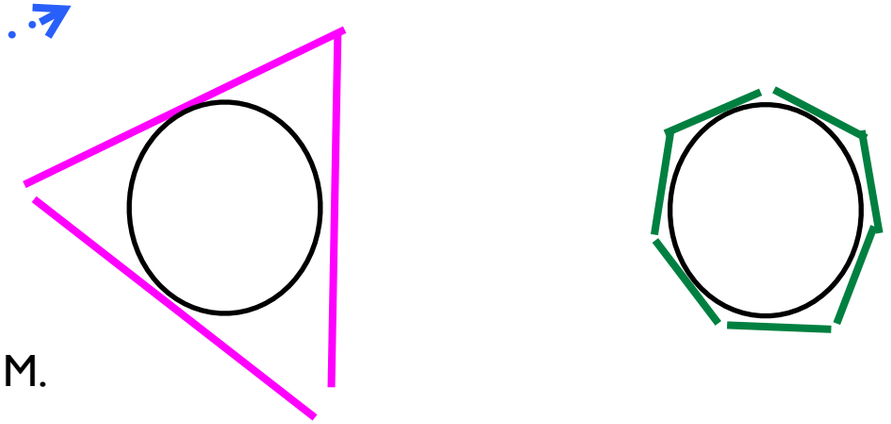


q (Å	d (Å)
0.001	6283
0.01	628
0.05	120
0.1	62
0.15	42
0.2	30
0.3	21

d-spacing vector:

- sets the size of your molecular ruler.
- provides *fractal* dimension to SAXS.

What's the circumference?



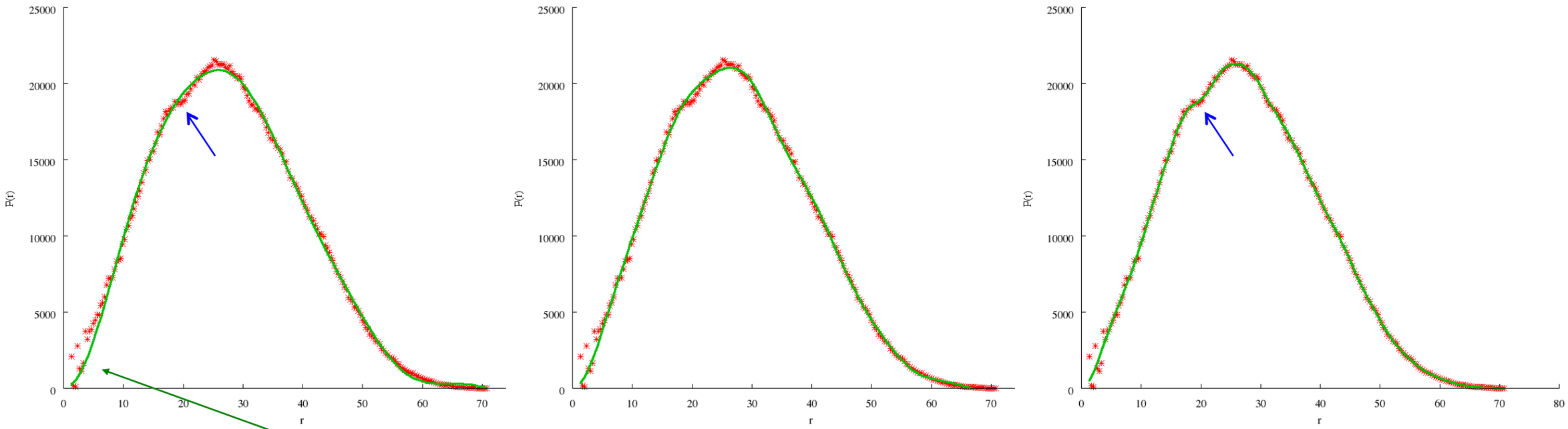
How long is the coast of Great Britain? Science 1967 Mandelbrot M.

Resolution

P(r) function

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

q_{max} = 0.3 Å^{-1}(21Å) q_{max} = 0.4 Å^{-1} q_{max} = 1.3 Å^{-1}(4.8Å)



Fitted P(r) curves at increasing q_{max}.

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

Low resolution, curve (green) is very smooth, increasing resolution adds more bumps to curve.

Increasing q_{max} increases observed information content, must correct for internal scattering.

Radius-of-Gyration

Small Angle X-ray Scattering

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

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

For a given particle, changes in conformation should → change P(r) → change R_g

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

R_g can be approximated using visible or X-ray photons depends on:

- size of the particle
- angular dependence of scattering

Guinier

Small Angle X-ray Scattering

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

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

Taylor Series Expansion of sin(x)

$$\sin(q \cdot r) = \sin(a) + \frac{\cos(a)}{1!}(q \cdot r - a) - \frac{\sin(a)}{2!}(q \cdot r - a)^2 - \frac{\cos(a)}{3!}(q \cdot r - a)^3 + \frac{\sin(a)}{4!}(q \cdot r - a)^4 + \ldots$$

Evaluating the function at a = 0

$$\sin(q \cdot r) = (q \cdot r) - \frac{1}{3!}(q \cdot r)^3 - \frac{1}{5!}(q \cdot r)^5 + \ldots$$ Using data at q near 0 like 0.001, gives a polynomial representation of sin(*q r*)

Polynomial substitution of the sine function

$$I(q) = 4\pi \int_0^{d_{max}} P(r) \cdot \frac{1}{q \cdot r} \left[(q \cdot r) - \frac{1}{3!}(q \cdot r)^3 - \frac{1}{5!}(q \cdot r)^5 + \ldots \right] dr$$

essentially zero, consider (q = 0.003)

Distribute

$$I(q) = 4\pi \int_0^{d_{max}} P(r)\, dr - 4\pi \cdot \frac{1}{3!} \int_0^{d_{max}} P(r) \cdot (q \cdot r)^2 dr - 4\pi \cdot \frac{1}{5!} \int_0^{d_{max}} P(r) \cdot (q \cdot r)^4 dr + \ldots$$

Guinier

Small Angle X-ray Scattering

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

Polynomial substitution of the sine function

$$I(q) = 4\pi \int_0^{d_{max}} P(r) \cdot \frac{1}{q \cdot r} \left[(q \cdot r) - \frac{1}{3!}(q \cdot r)^3 - \frac{1}{5!}(q \cdot r)^5 + \ldots \right] dr$$

Distribute

essentially zero, consider (q = 0.003)

$$I(q) = 4\pi \int_0^{d_{max}} P(r)\, dr - 4\pi \cdot \frac{1}{3!} \int_0^{d_{max}} P(r) \cdot (q \cdot r)^2 dr - 4\pi \cdot \frac{1}{5!} \int_0^{d_{max}} P(r) \cdot (q \cdot r)^4 dr + \ldots$$

Define $I(0)$ and R_g

$$I(q) = \underbrace{4\pi \int_0^{d_{max}} P(r)\, dr}_{I(0)} \left(1 - \left(\frac{q^2}{3!} \right) \cdot \frac{\int_0^{d_{max}} r^2 \cdot P(r)\, dr}{\int_0^{d_{max}} P(r)\, dr} \right)$$

$$R_g^2 = \frac{1}{2} \cdot \frac{\int_0^{d_{max}} r^2 \cdot P(r)\, dr}{\int_0^{d_{max}} P(r)\, dr}$$

Substitute

$$I(q) = I(0) \cdot \left(1 - \frac{q^2 \cdot R_g^2}{3} \right)$$

Guinier method approximates scattering equation using Taylor/McLaurin series expansion

Guinier

Small Angle X-ray Scattering

Guinier method approximates scattering equation using Taylor/McLaurin series expansion

Approximation with q close to zero

$$I(q) = I(0) \cdot \left(1 - \frac{q^2 \cdot R_g^{\,2}}{3} \right)$$

Simply a Taylor Series expansion of e^x

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots \quad \text{where} \quad x = -\frac{q^2 \cdot R_g^{\,2}}{3}$$

Substitute

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

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

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

thus a plot of ln$I(q)$ vs q^2 will have a linear region

How valid is the approximation?

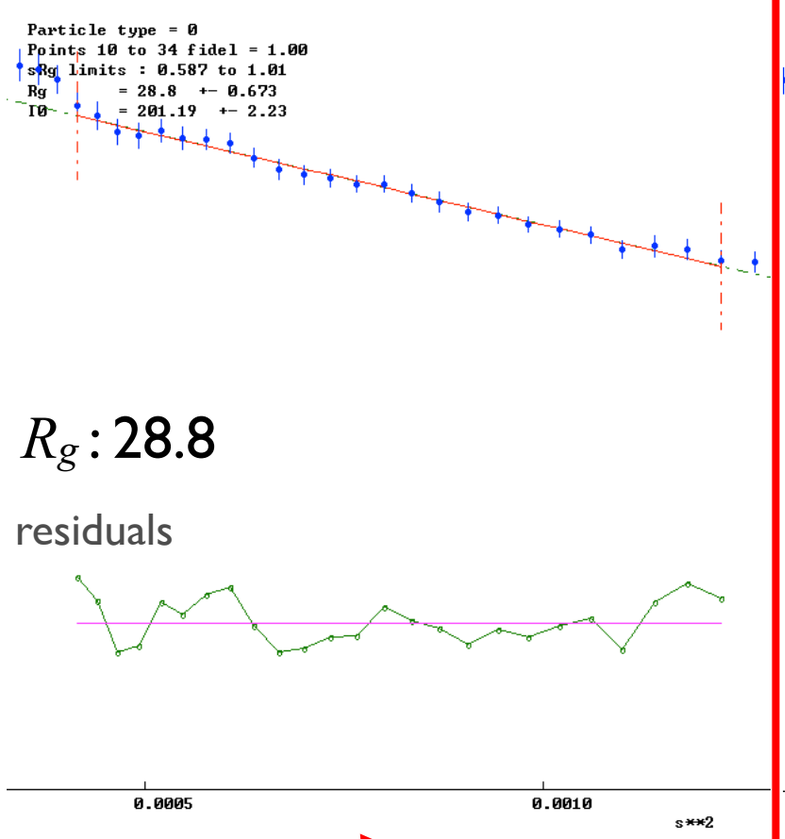
How well does the Guinier R_g approximate $R_g^{\text{real space}}$?

Guinier

Small Angle X-ray Scattering

How valid is the approximation?
How well does the Guinier R_g approximate $R_g^{\text{real space}}$?

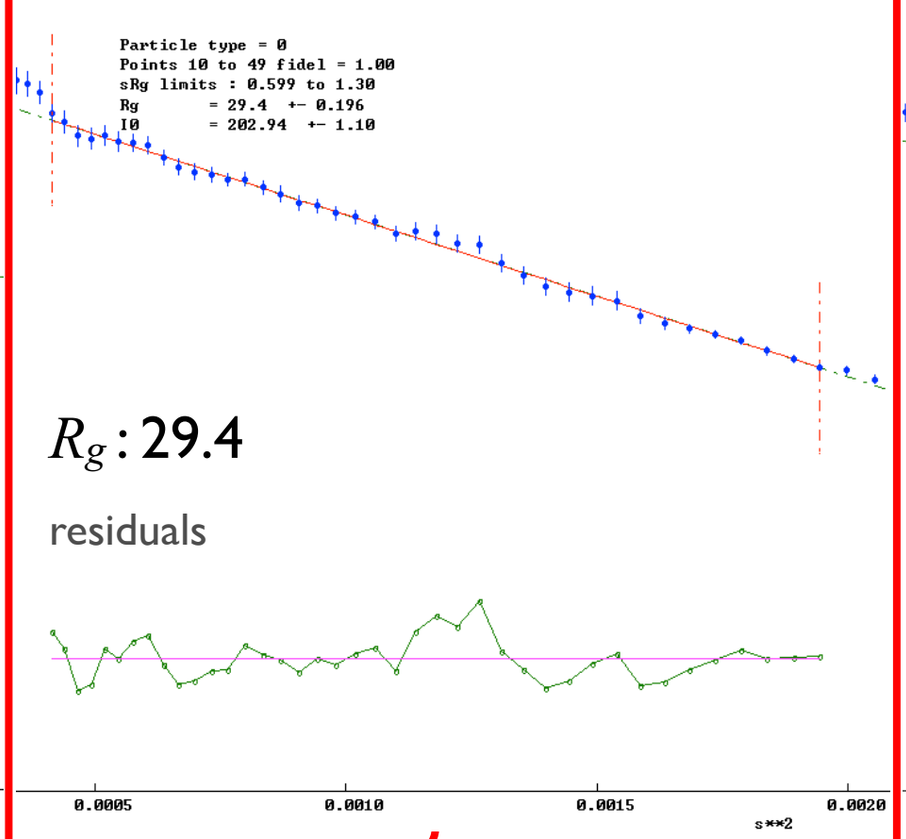
$q \cdot R_g < 1.0$



R_g : 28.8

residuals

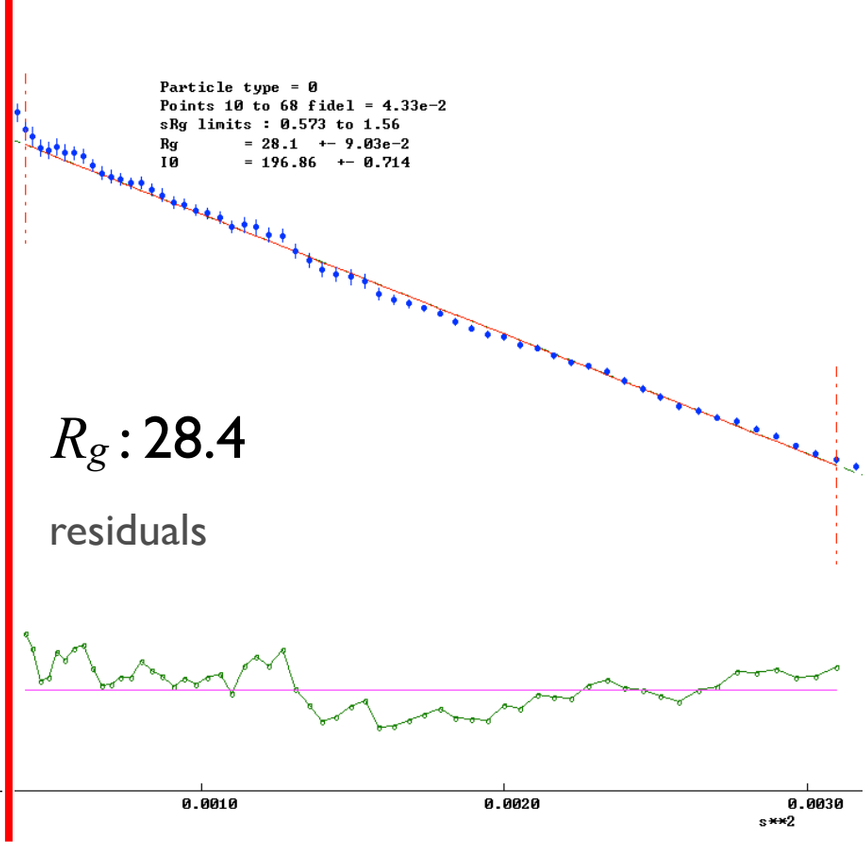
$q \cdot R_g < 1.3$



R_g : 29.4

residuals

$q \cdot R_g < 1.5$



R_g : 28.4

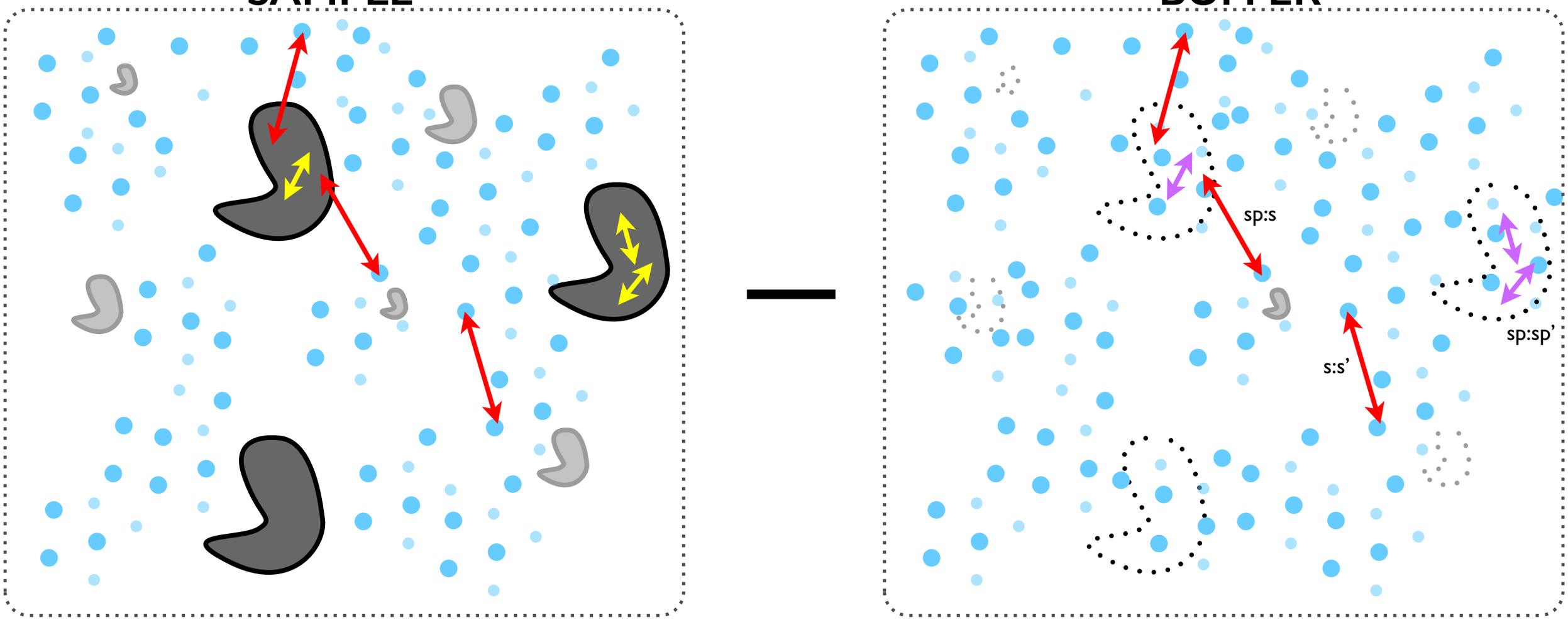
residuals

16 fewer data points

We recommend determining using data where $R_g < 1.3$

Scattering Contrast

Buffer Subtraction



$$\sum_{n_s:n_{s'}} f_s \cdot f_{s'} \frac{sin(qr)}{qr} + \sum_{n_p:n_s} f_p \cdot f_s \frac{sin(qr)}{qr} + \sum_{n_p:n_{p'}} f_p \cdot f_{p'} \frac{sin(qr)}{qr} - \sum_{n_s:n_{s'}} f_s \cdot f_{s'} \frac{sin(qr)}{qr} - \sum_{n_{sp}:n_s} f_{sp} \cdot f_s \frac{sin(qr)}{qr} - \sum_{n_{sp}:n_{sp'}} f_{sp} \cdot f_{sp'} \frac{sin(qr)}{qr}$$

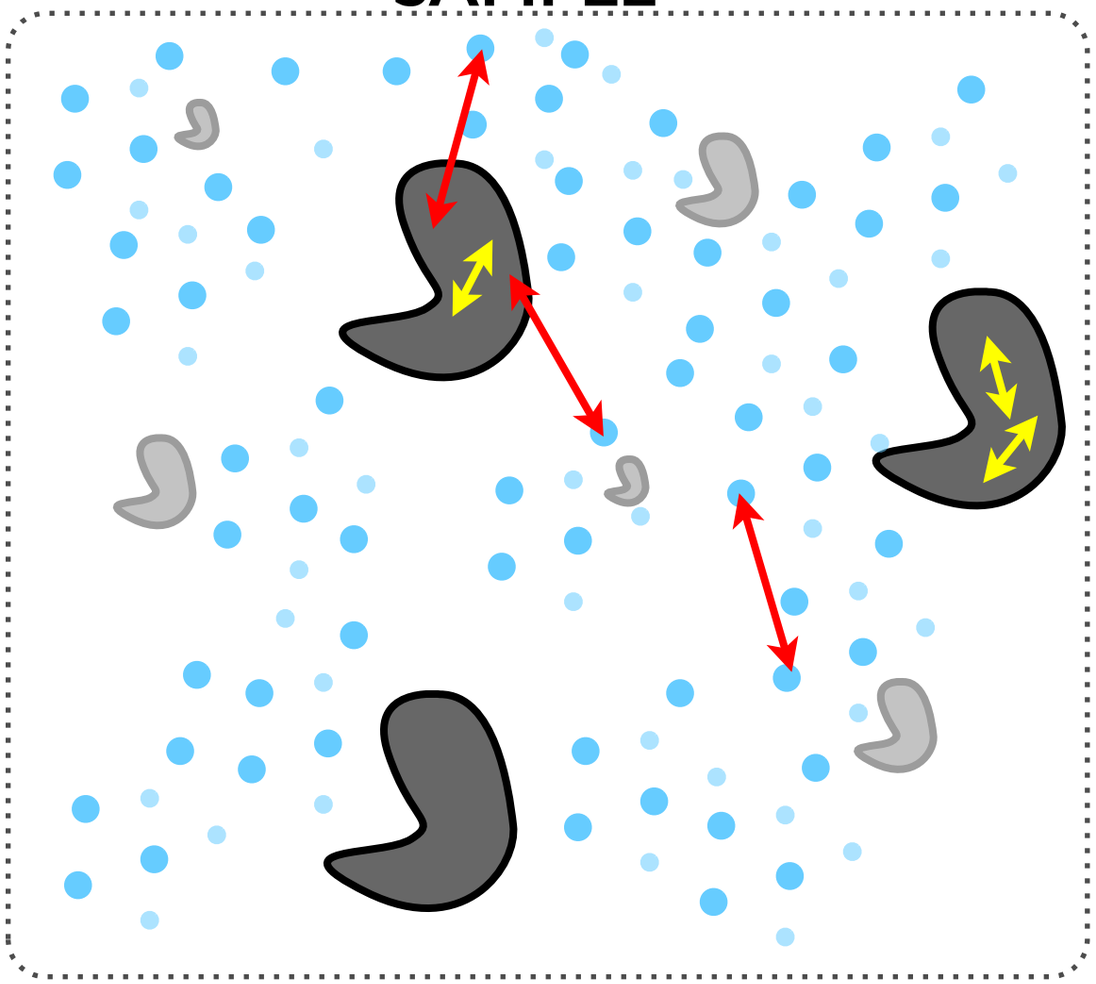
At low resolution:

- solvent e_n^- density is an average from small molecules
- change f_s by adding sucrose, salts, etc.

What happens if $f_s = f_p$?

Scattering Contrast

Buffer Subtraction



In general, e_n^- density(ρ) of $\rho_{protein}$, ρ_{RNA}, $\rho_{DNA} \neq \rho_{solvent}$

$$\rho_{water} = 0.334\ e/Å^3$$

$$\rho_{lysozyme} = 0.414\ e/Å^3$$

$$\rho_{RNA} = 0.621\ e/Å^3$$

$$\sum_{n_s:n_{s'}} f_s \cdot f_{s'} \frac{sin(qr)}{qr}$$

$$\sum_{n_p:n_s} f_p \cdot f_s \frac{sin(qr)}{qr}$$

$$\sum_{n_p:n_{p'}} f_p \cdot f_{p'} \frac{sin(qr)}{qr}$$

Scattering of the particle more correctly written as:

$$I_{particle}(q) = (\Delta\rho)^2 V \cdot \int_0^{d_{max}} P(r) \cdot \frac{sin(q \cdot r)}{q \cdot r} dr$$

consider as q → 0:

$$\lim_{q \to 0} \frac{sin(q \cdot r)}{q \cdot r} = 1$$

$$I_{particle}(0) = (\Delta\rho)^2 V \cdot \int_0^{d_{max}} P(r) \cdot 1\ dr = (\Delta\rho)^2 \cdot V^2$$

I(0) is directly proportional to particle's volume scaled by Δρ

Single to Many Particle Scattering

$$I_{macromolecules}(q) = I_{macromolecule}(q) \cdot c \cdot \frac{I_e \cdot N_l \cdot P \cdot d}{M \cdot a^2}$$

substitute constants for *k*

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

rearrange

$$\frac{1}{k} \cdot \frac{1}{c} \cdot I_{macromolecules}(q) = I_{macromolecule}(q)$$

$$I_{\sin gle\ particle}(q) = k' \cdot \frac{1}{c} \cdot I_{sample}(q)$$

limit as $q \rightarrow 0$

$$I_{\sin gle\ particle}(0) = (\Delta\rho)^2 \cdot V^2 = (\Delta\eta_e)^2$$

c - concentration (gm/cm^3)
I_e – 7.9 x 10^{-26} (cm^2)
N_L- 6.0223 x 10^{23} (mol^{-1})
P – total energy over the irradiated area
d - sample thickness (cm)
M – molecular weight (gm • mol^{-1})
a – distance to detector (cm)

Allows for the determination of M.W. by either:

1. Standard curve with proteins of known M.W.
2. Determination of I(0) on an absolute scale.

Scattering Contrast

Mass Estimation and I(0)

$$I_{particle}(0) = (\Delta\rho)^2 V \cdot \int_0^{d_{max}} P(r) \cdot 1 \, dr = (\Delta\rho)^2 \cdot V^2$$

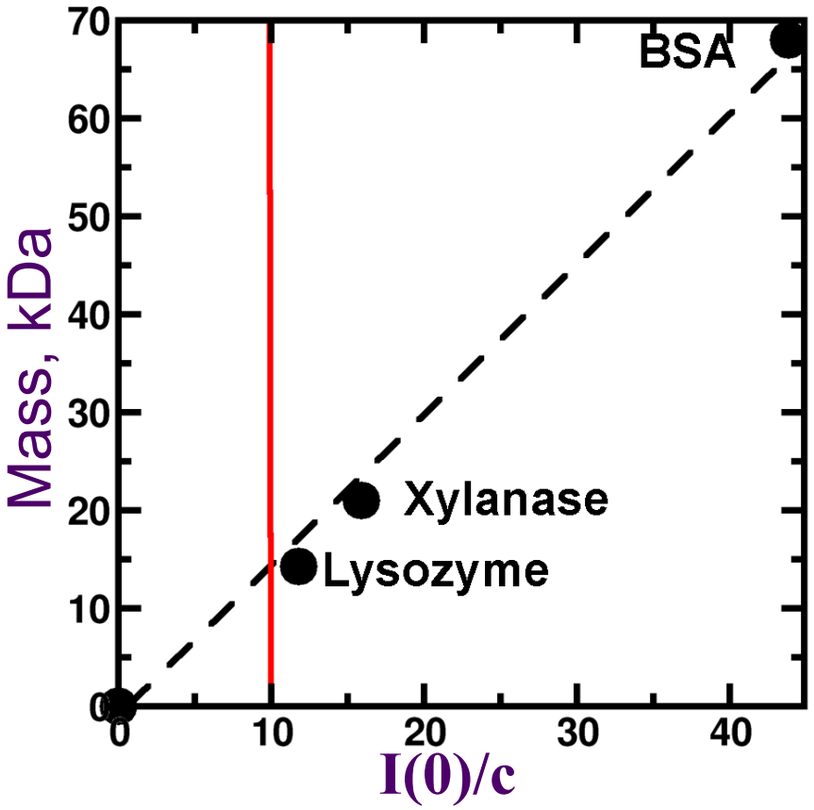
In real experiments, $I_{particle}(0)$ has to be scaled by concentration, c

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

Estimated by Guinier method

divide both sides by [particle]

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



For a given protein, ratio is a constant at a specific $(\Delta\rho)^2$

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

This relationship can be used to make a standard curve to determine:

- mass of either protein, RNA or particles of same composition
- requires accurate knowledge of concentration

Scattering Contrast

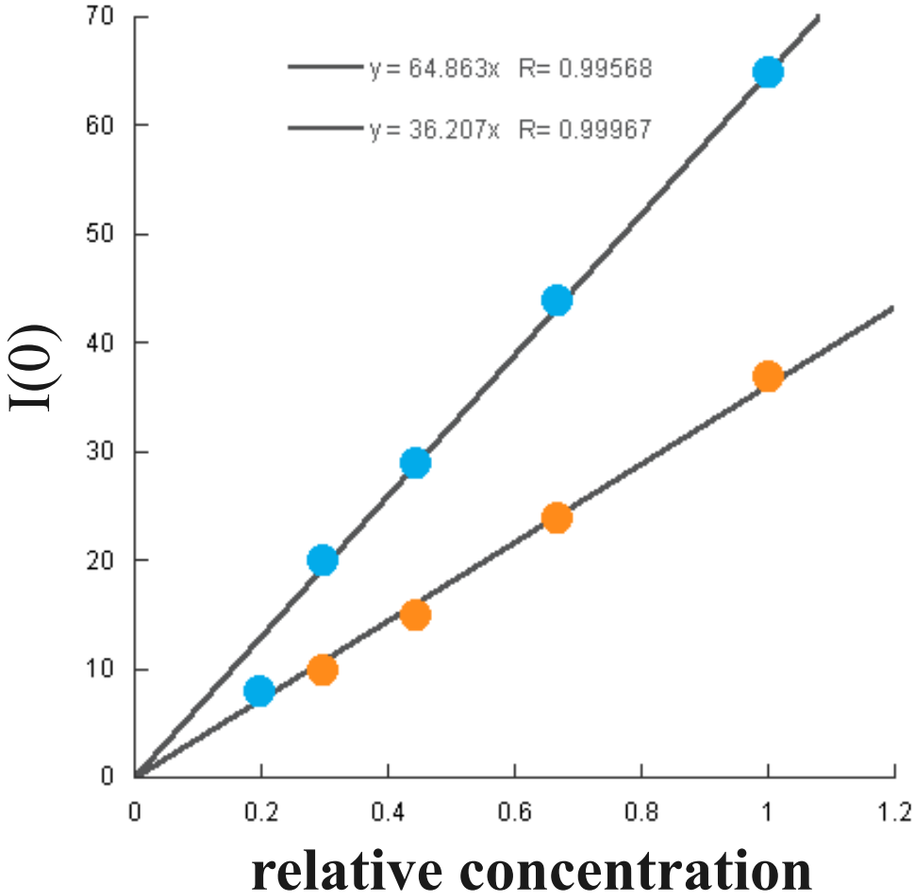
Alternative method for mass determination from I(zero)

- use a single standard (xylanase)
- do a dilution series (e.g., 2/3rds)
- determine slope

Performed as a 2/3rds dilution series:

52 ul sample	34 ul sample	34 ul sample	34 ul sample
	17 ul buffer	17 ul buffer	17 ul buffer

Xylanase in Two Different Buffers



$$\frac{\Delta I_{protein}(0)}{\Delta c} = m_{protein} \propto Mass$$

$$Mass_{unkn} = \frac{m_{unkn}}{m_{std}} \cdot Mass_{std}$$

SAXS data must be collected in same buffer as sample.

Porod Invariant

Assessing flexibility

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

Assumption: defined Δρ between particle and solvent and scatterer has homogenous electron density

Integration of data transformed as $q^2 \bullet I(q)$ 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.

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

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

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

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

l_c, correlation length

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

R_g, radius-of-gyration

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

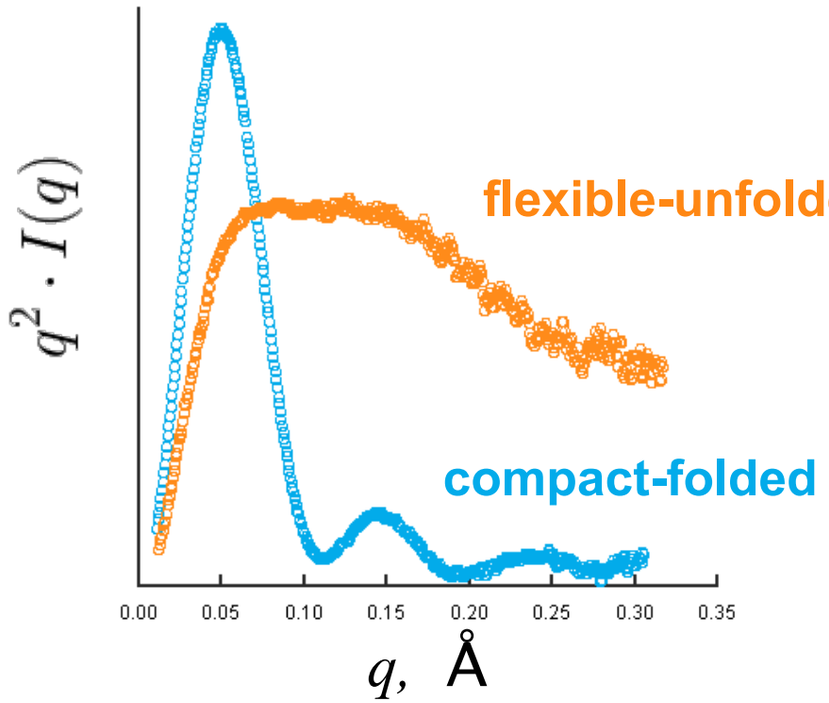
Does not require Q
Concentration independent
Contrast independent (as long as structure does not change)
Essentially normalized to I(0)

Porod 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 q2 • I(q) should show a curve that captures an area
Define area means transformed data converges.
Qualitative assessment of flexibility/unfoldedness

Can do quantitatively!

Detecting Flexibility

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

Scattering by a Gaussian Coil

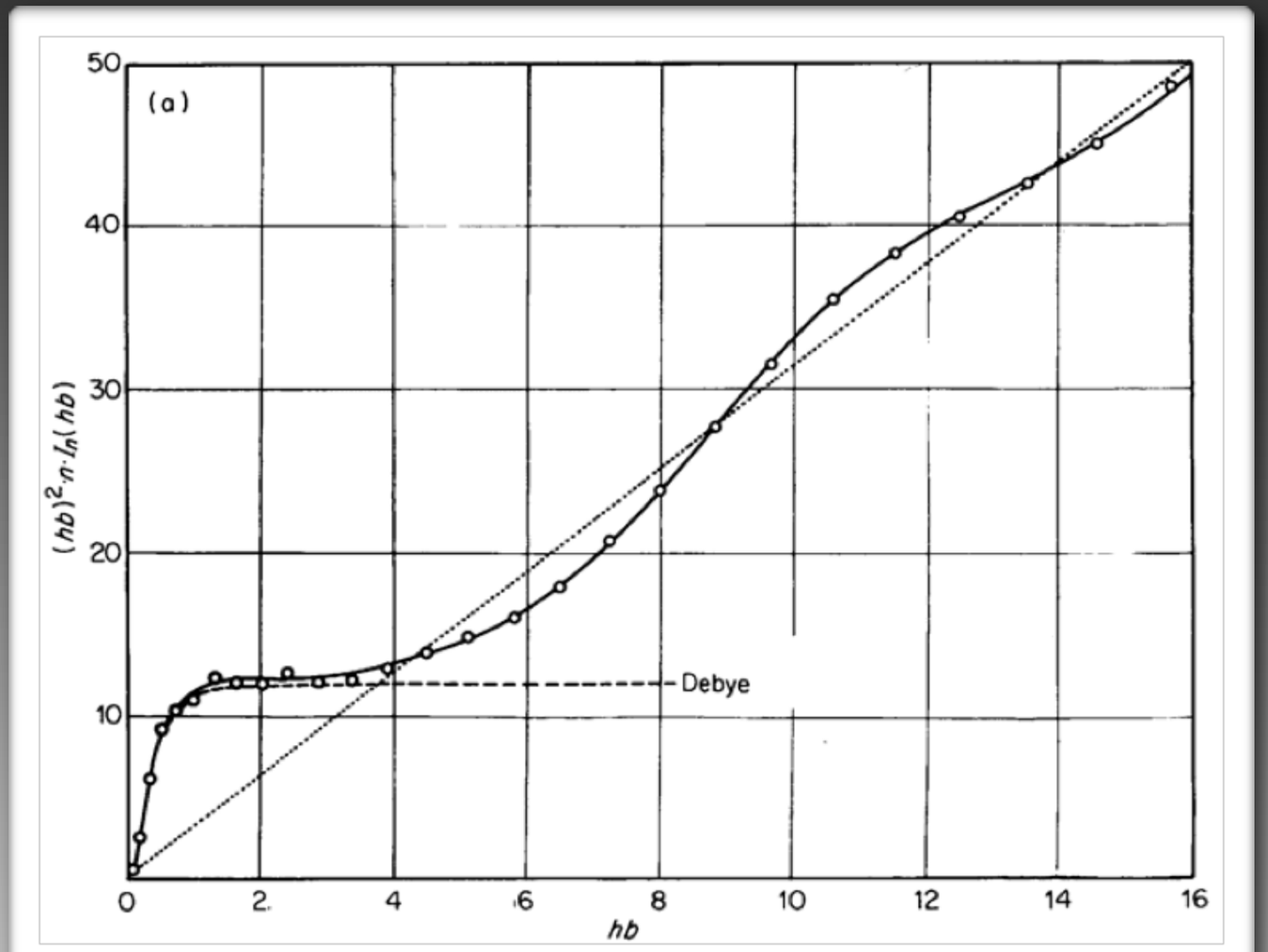
$$I(q) = \frac{2(e^{-R_g^2 \cdot q^2} + R_g^2 \cdot q^2 - 1)}{(R_g^2 \cdot q^2)^2}$$

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 $q^2 \cdot R_g^4 << 1$

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

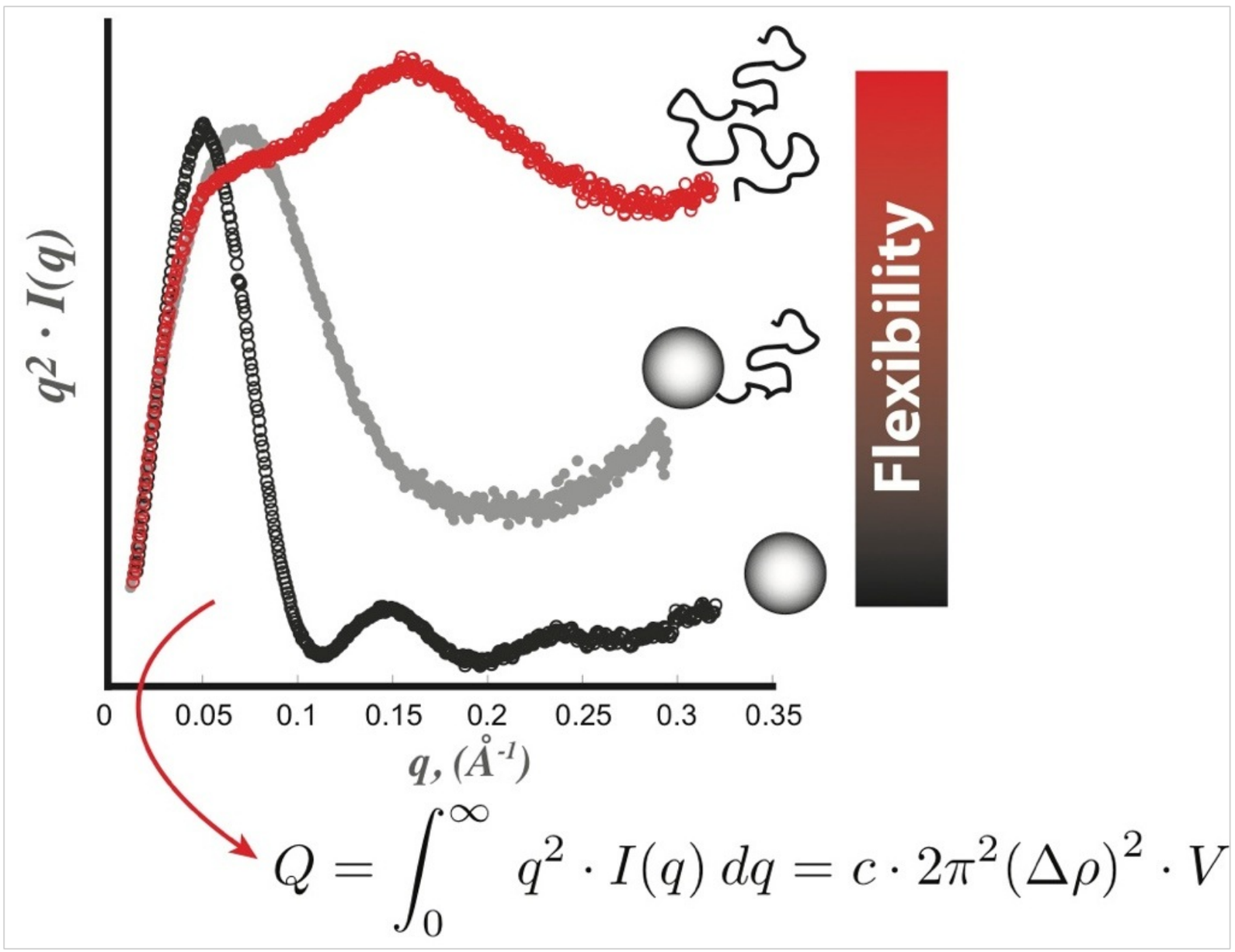


$q^2 \bullet I(q)$ becomes constant at high q
Creates hyberbolic curve
Basis for Kratky Plot ($q^2 \bullet I(q)$ *vs* q)

Kratky Plot

Qualitative Assessment of flexibility

for $q \cdot R_g$ > 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**•** ***I(q)*** **becomes constant at high** ***q***

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

Power Law Relationship

log vs log plot... quantitating flexibility?

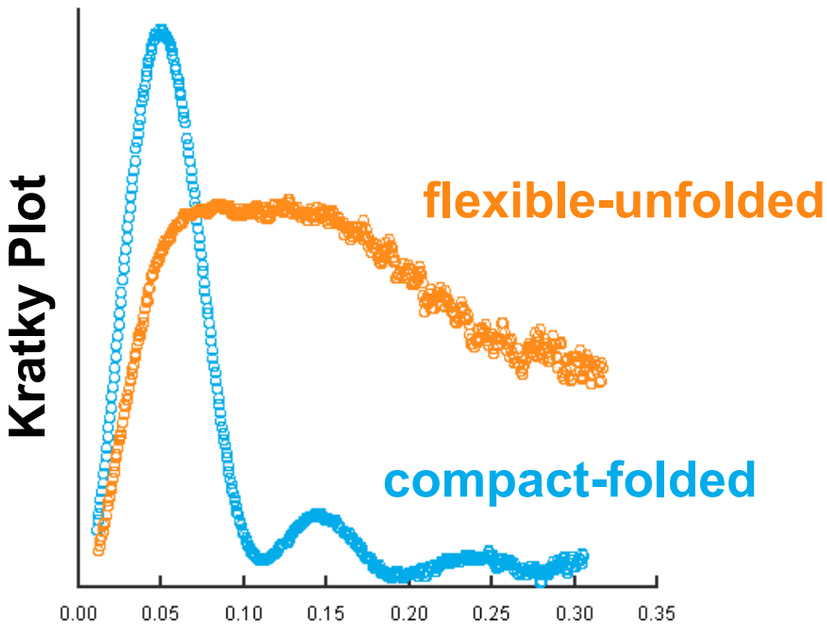
if particle is flexible, should see a plateau in $q^2 \cdot I(q)$ vs. q

Porod ⟷ **Debye**

$$q^4 \cdot I(q) = constant \qquad constant \approx q^2 \cdot I(q)$$

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

Defines a power law relationship!



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

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

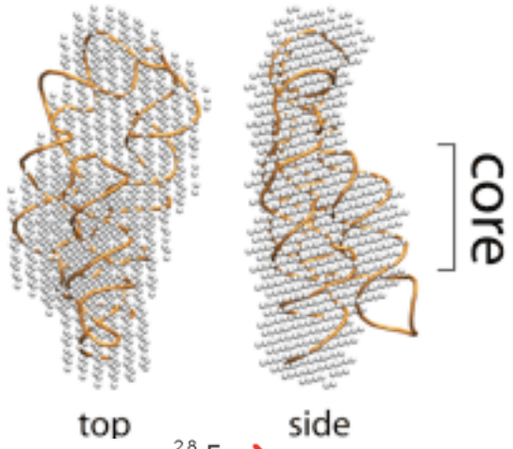
Low-resolution SAXS
Requires small amounts of sample

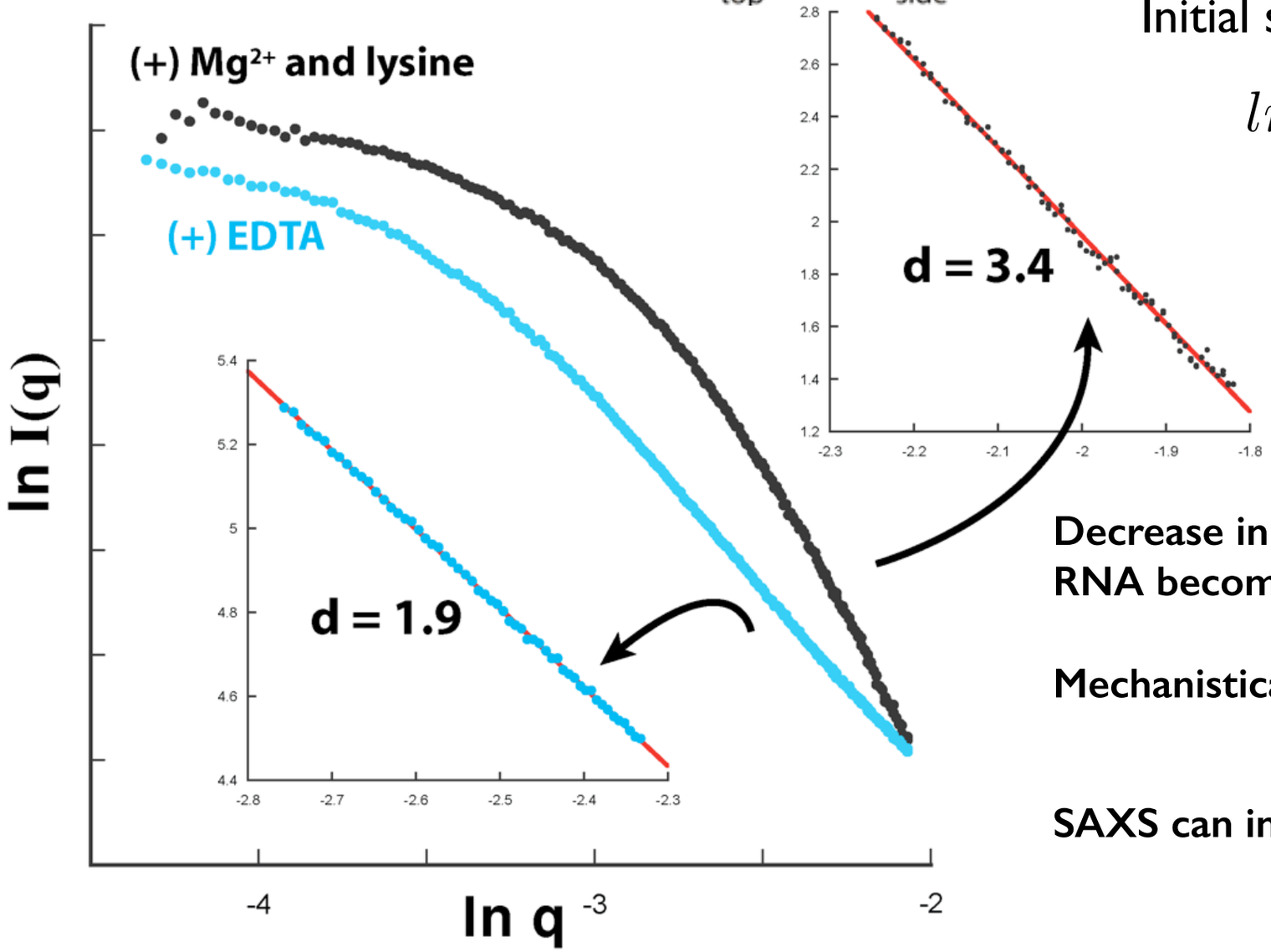
QUANTIFYING FLEXIBILITY

example Lysine Riboswitch

Lysine riboswitch requires:

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





Initial slope defines the Porod-Debye region

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

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

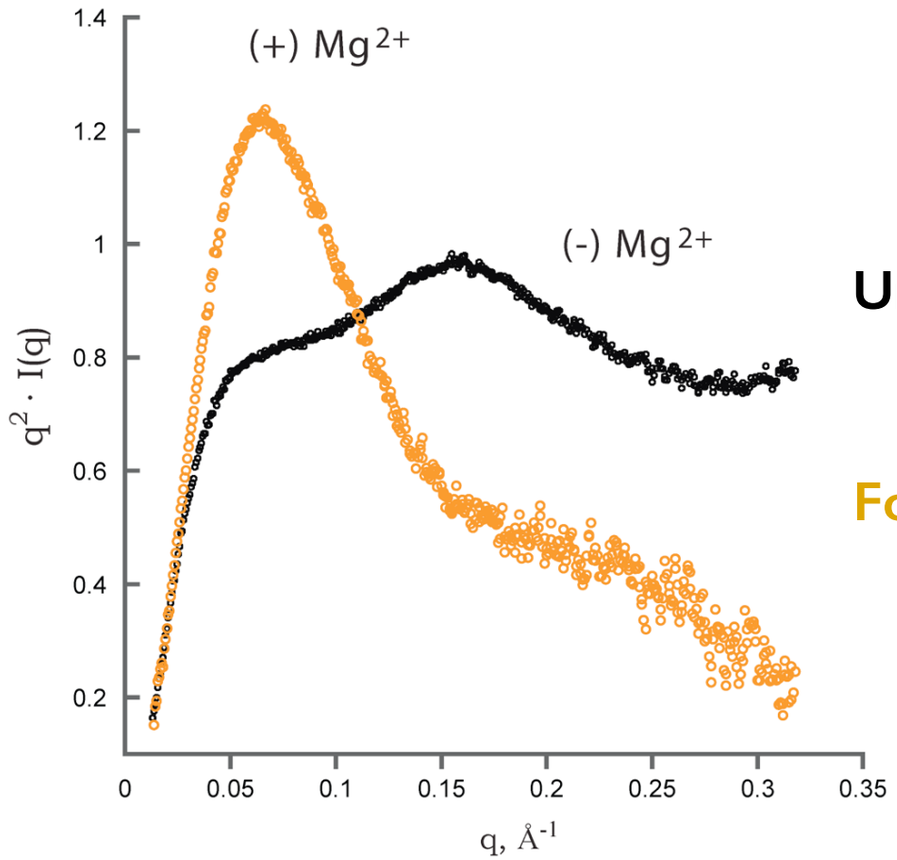
Porod Invariant

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.

Kratky Plot



Unfolded particle diverges, does not capture a defined area.

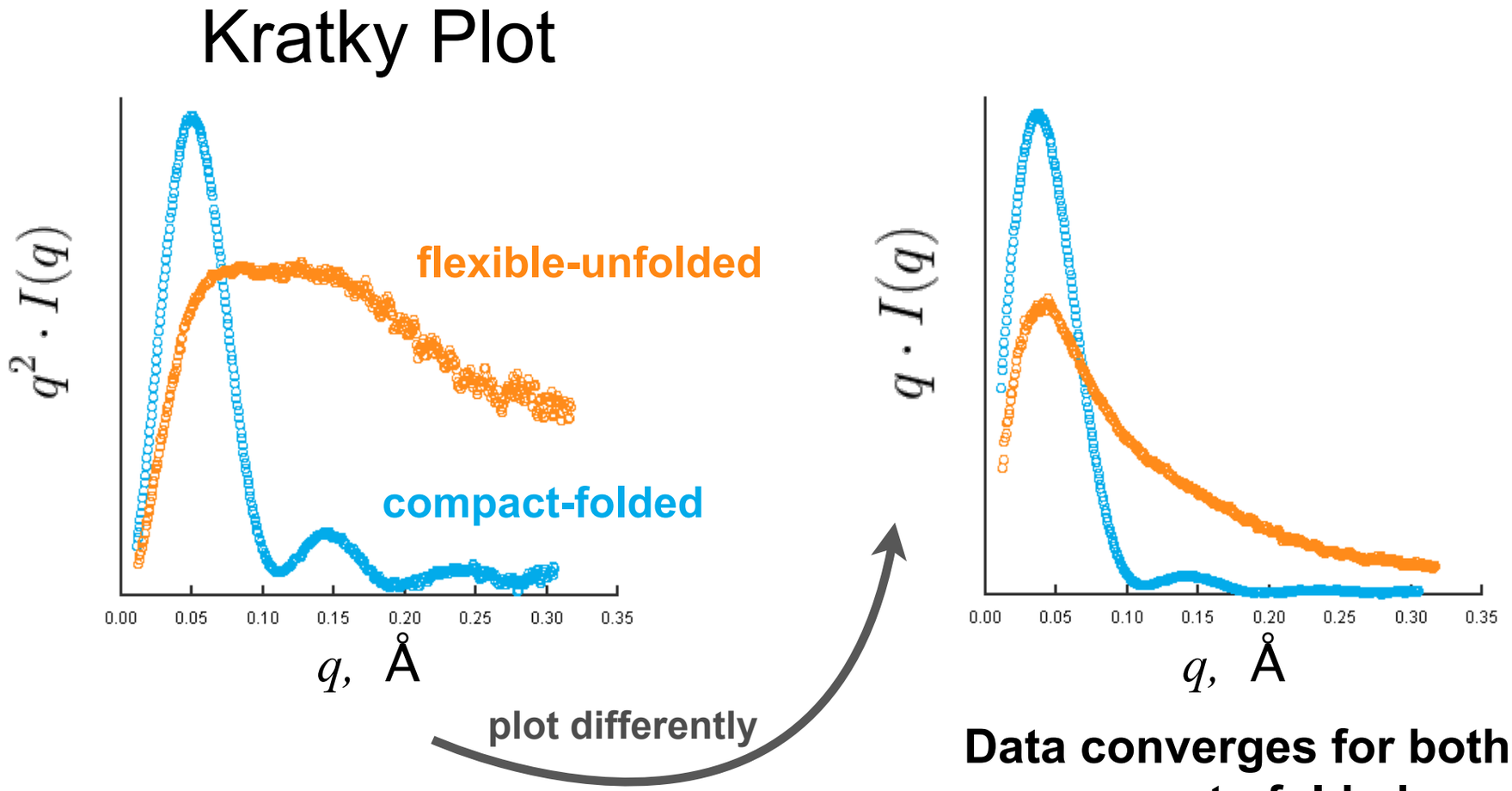
(flexible, unfolded, gaussian chain like)

Folded particle displays convergence towards higher q

(folded, compact particle)

No Q implies, volume and I_c are no longer defined for flexible particles.

Defining a new Invariant



Data converges for both

- **compact - folded**
- **flexible - unfolded**

What does the integrated area mean?

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

1. substitute for *I(q)*

l_c is the expected correlation length

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

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

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

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

5. collect like terms

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

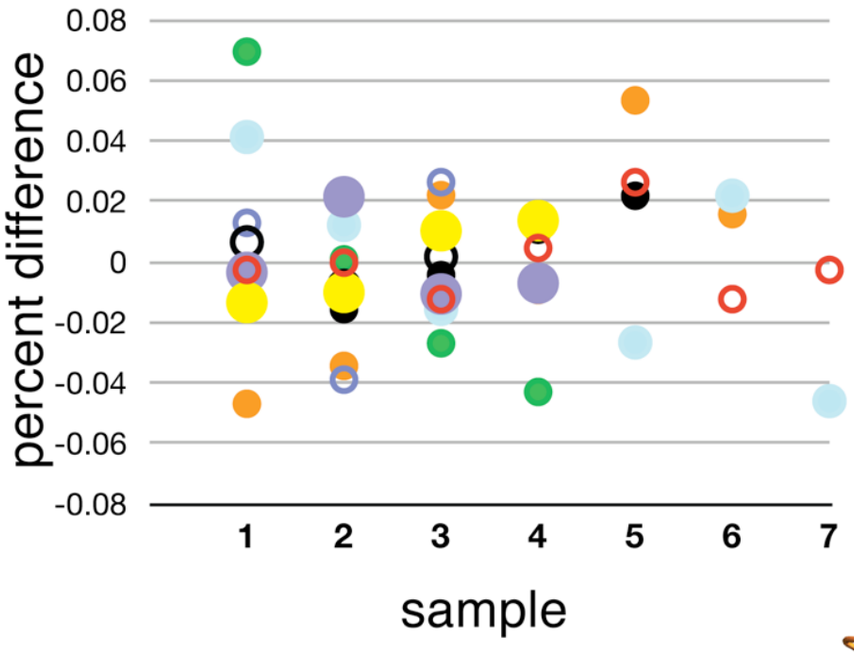
4. substitute $P(r) = 4\pi\, r^2\, \gamma(r)$

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

correlation function

V_c: A Novel SAS Ratio

V_c

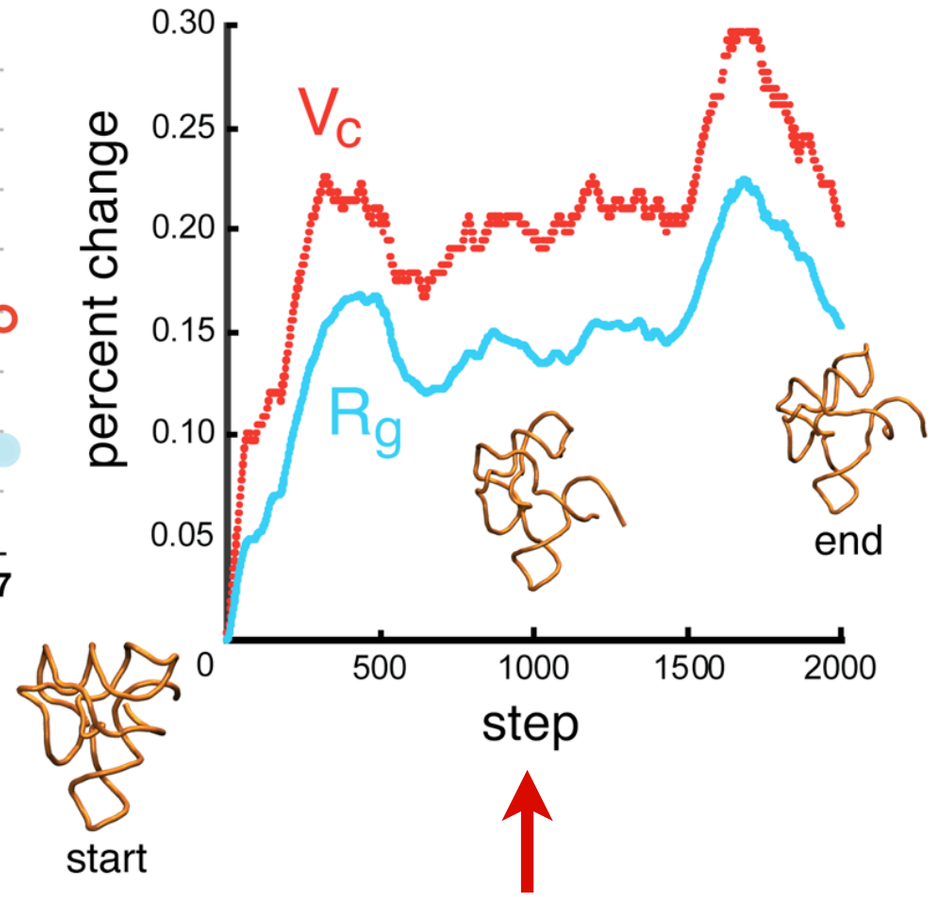


- **8 different protein and RNA samples**
- **4 to 7 different concentrations**

67% variance is contained within 2% mean

CONCENTRATION INDEPENDENCE!

MD simulation of SAM-1



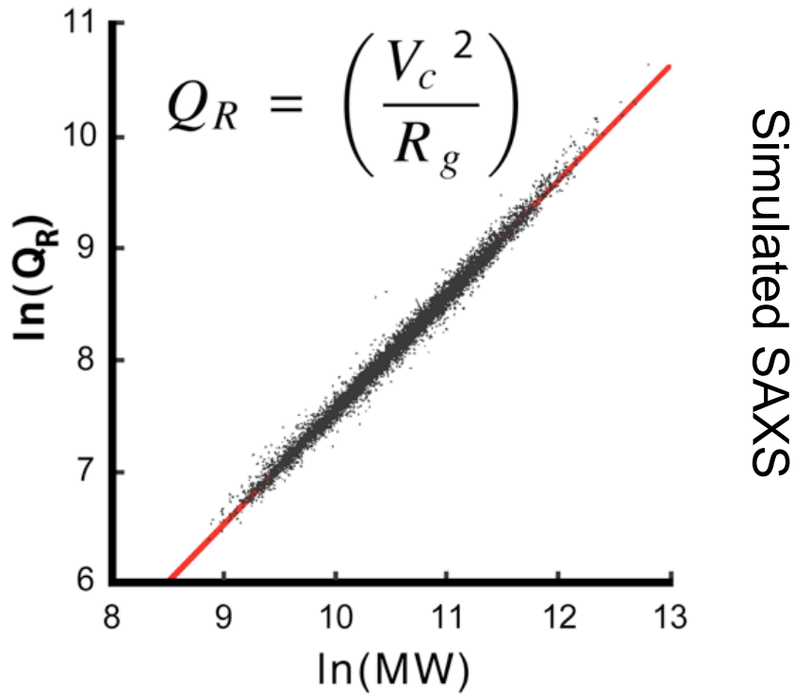
V_c sensitive to conformational state like R_g

Direct Mass Determination

the power law distribution

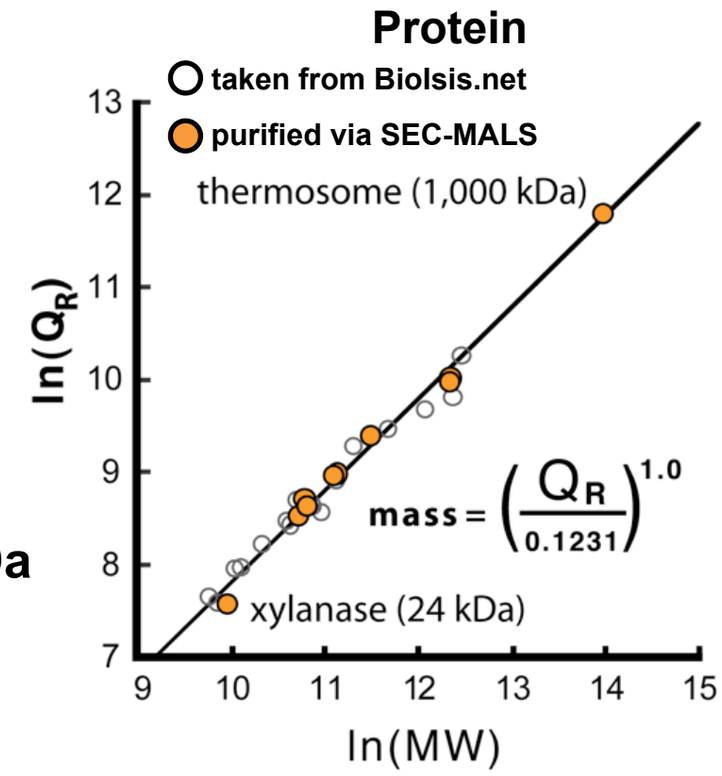
9446 PDB entries range from 8 to 400 kDa (protein only)

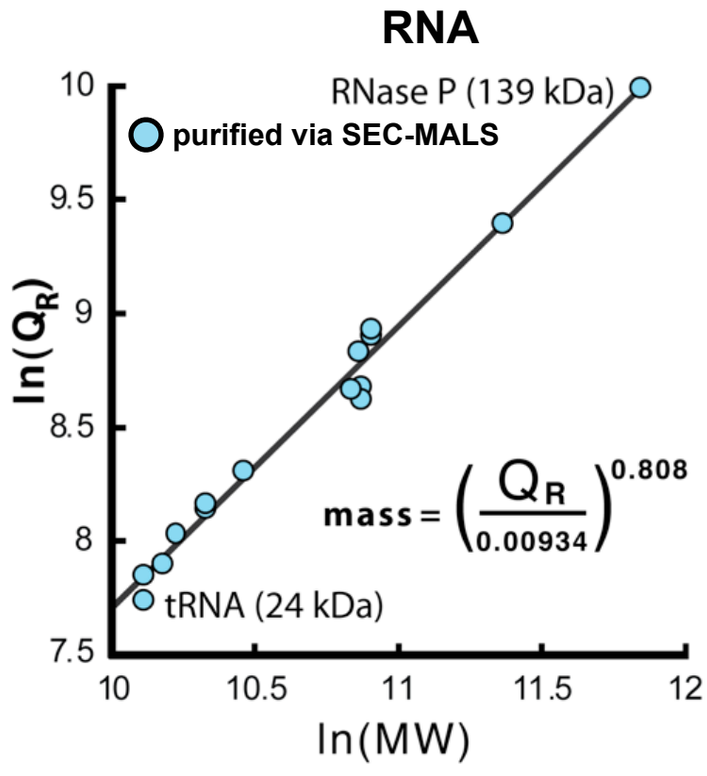
- **Q_R scales with mass, linear via power-law distribution.**
- **Using actual data, 9% mass error with previously frozen samples.**
- **Linear relationship covers a large mass range 20 to 1,000 kDa.**
- **Effective for RNA samples 5% error.**



Simulated SAXS

Use to infer mixtures:
...expect 26 kDa and get 40 kDa
- **monomer ↔ dimer?**

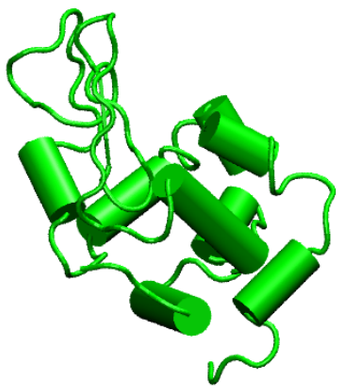




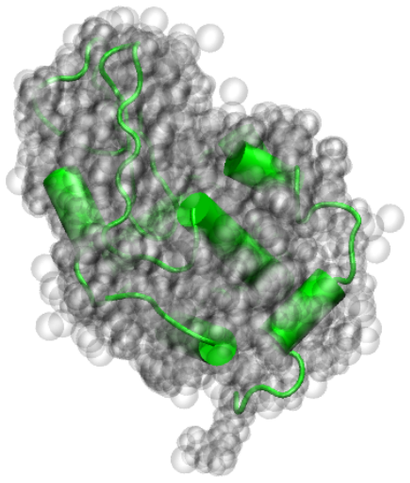
Experimental SAXS

Distance Distribution Function

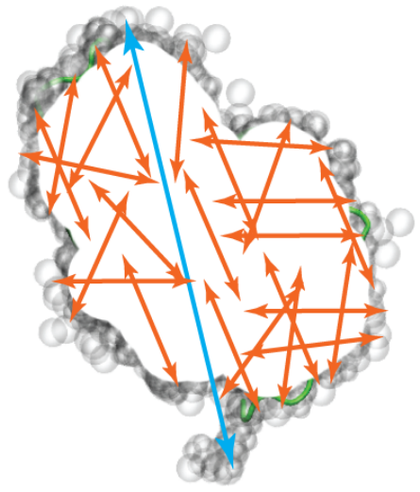
2° Structure



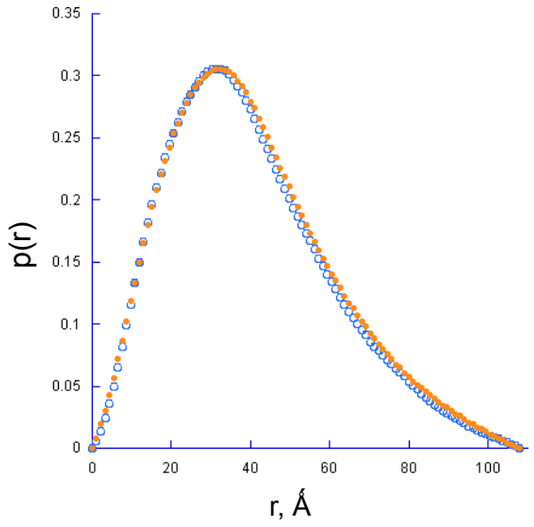
Molecular Envelope



Distance Distribution



p(r) Function



P(r) ~ pair-distribution function

- Not a proper mathematical function
- Counts all the pairwise interatomic distances between e_n^- within the macromolecule.
  31 kDa macromolecule → 2,086 atoms → ~ 2,175,000 distance vectors

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

Properties of P(r)

P(-r) = - P(r) thus P(r) is an "odd" function *i.e.*, $f(x) = x^3$, $sin(x)$,...

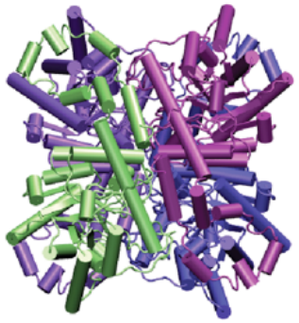
Defined on $0 < r < d_{max}$

How do we calculate P(r) from *I(q)* data?

Indirect Fourier Transform

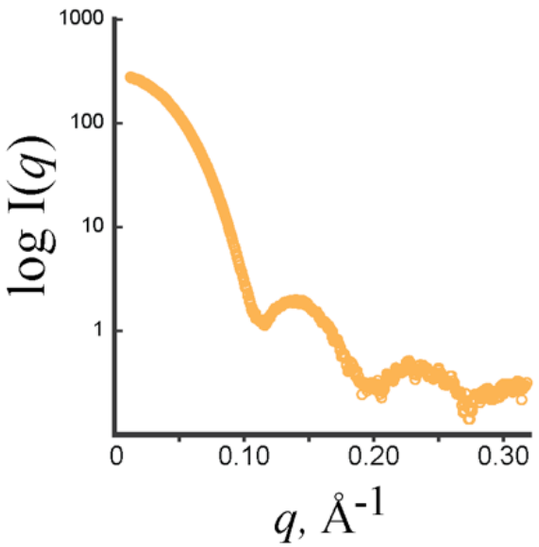
A measured SAXS curve determines a unique P(r)-distribution.
A P(r) distribution (from a model) can be used to determine a scattering curve.

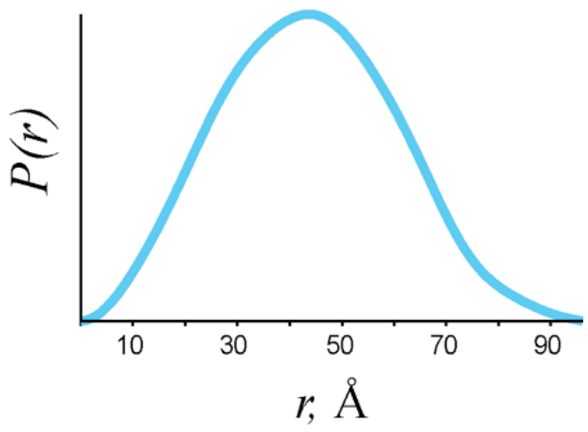
I(q) vs q
(Reciprocal Space)



P(r)-distribution
(Real Space)

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





Gnom (Svergun)
- guess at the distribution

GIFT (Glatter's method)
- use cubic splines

Moore's Method
- use Fourier sine series

Expect a smooth curve
Minimize oscillations
No negative values
Iterative process in determining dmax
Difficulties in finding a P(r) solution suggest poor sample.

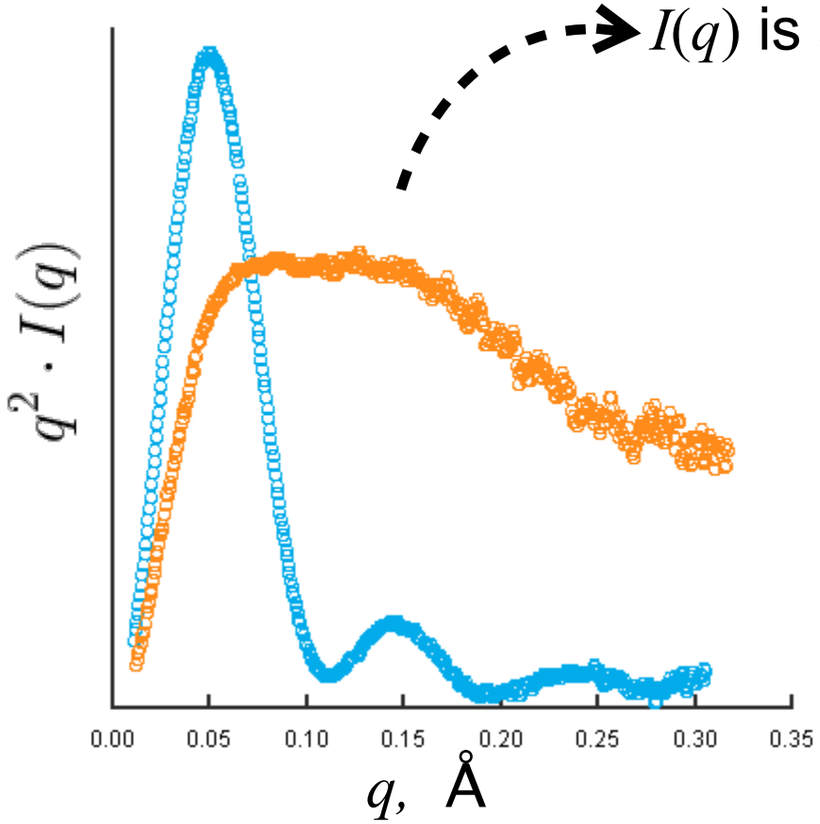
Dimensionless Kratky

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 $(q \cdot R_g)^2$ 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?

Use Guinier approximation to get some insights...

Dimensionless Kratky

Guinier approximation relates scattering to R_g

Derivation shows all particles that can be approximated by Guinier relation should have a peak value occurring at: $q \cdot R_g = 1.732$

→ peak value of $3 \cdot e^{-1} = 1.104$

Really about particles that can be approximated by the same correlation function such as:

$$\gamma(r) = e^{-\frac{r}{a}}$$

Starting with Guinier approximation:

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

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

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

Do a change of variables letting $u = q \cdot R_g$:

$$f(u) = (u)^2 \cdot \frac{I(q)}{I(0)} = (u)^2 \cdot e^{-\frac{(u)^2}{3}}$$

Find the first maxima by taking the derivative and solving for f'(u) = 0:

$$f'(u) = 2u \cdot e^{-\frac{u^2}{3}} - u^2 \cdot \frac{2u}{3} \cdot e^{-\frac{u^2}{3}}$$

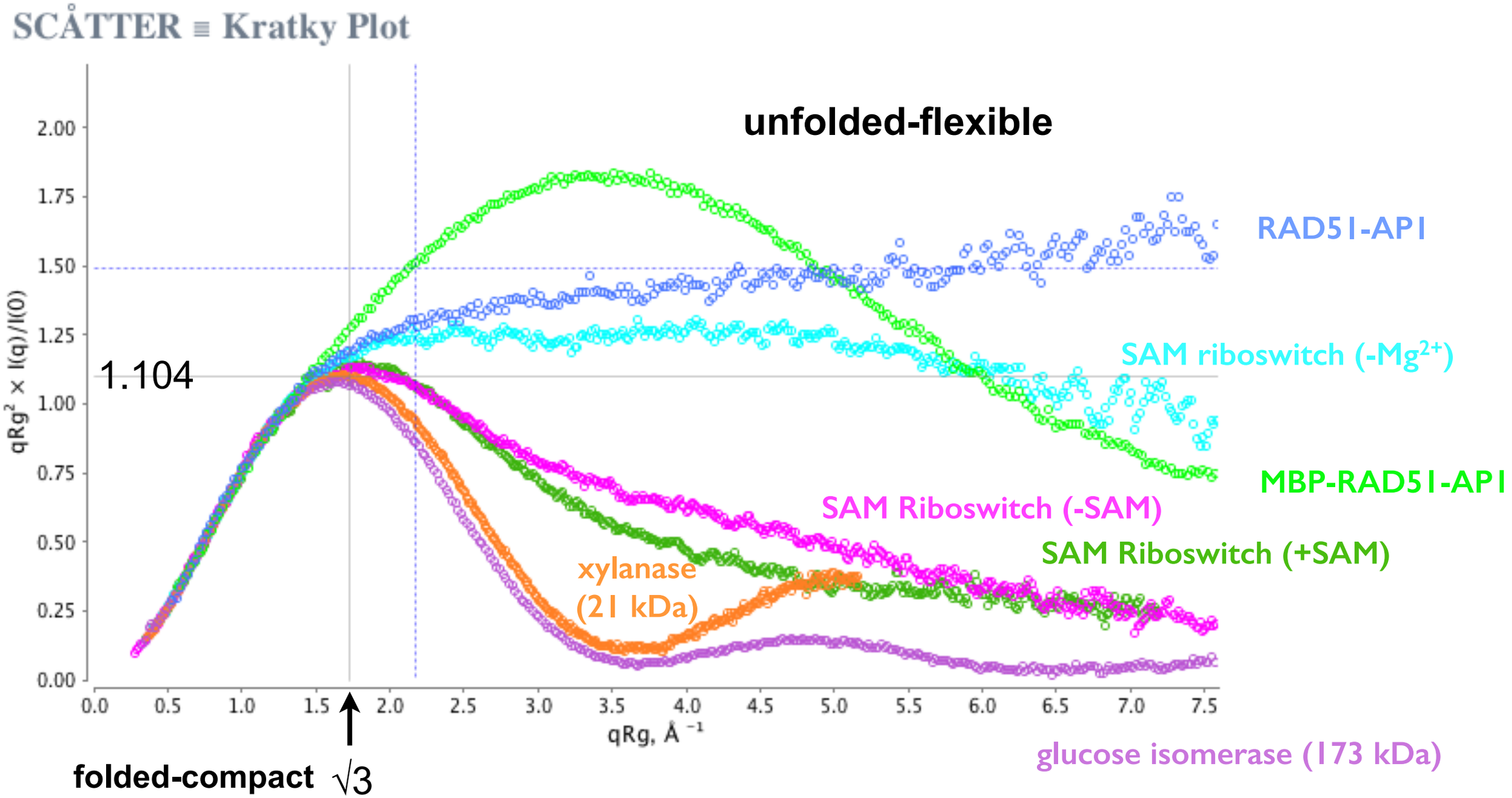
$$2u \cdot e^{-\frac{u^2}{3}} = \frac{2u^3}{3} \cdot e^{-\frac{u^2}{3}}$$

Take the square root of both sides thus solving for u or $(q \cdot R_g) = \sqrt{3}$:

$$3 = u^2$$

Dimensionless Kratky

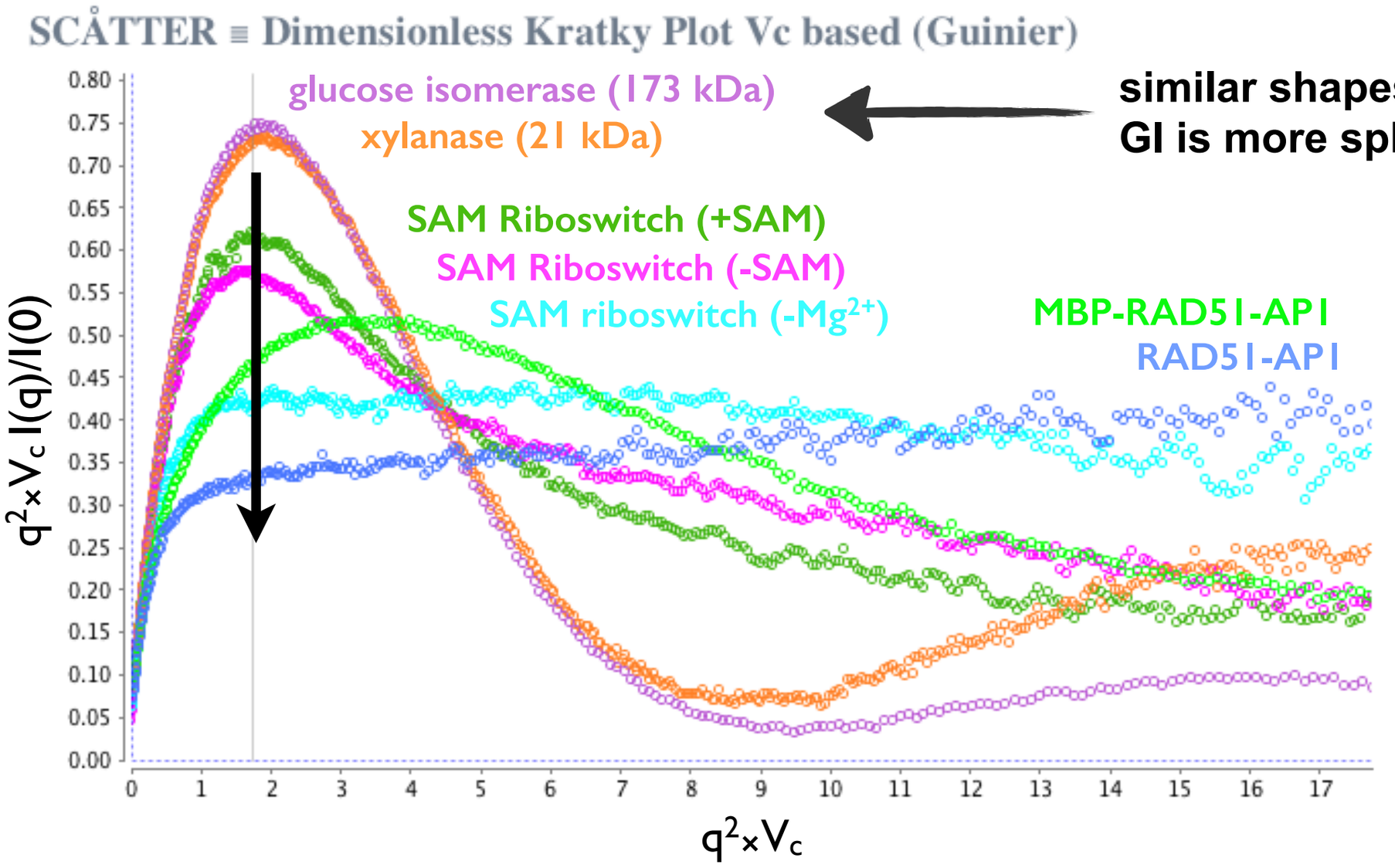
only a button away



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

Dimensionless Kratky

using Volume-of-Correlation



similar shapes but not exactly.
GI is more spherical than XY

Peak is inversely proportional to S-to-V ratio
Max value is 0.82 (sphere)
For a fixed molecule, any decrease suggests increase in surface area
Illustrate differences better than previous (see SAM)
Fully unfolded particle should have largest S-to-V ratio (low on graph)

Information Content

Moore, P.B. J. Appl. Cryst. (1980). 13, 168-175

Using Shannon Sampling theorem, P. Moore determined the number of independent parameters that can be extracted from a single SAXS curve.

$$I(q) = 8\pi \cdot \sum_{n=1}^{N} a_n \cdot \frac{1}{q} \cdot \left[\frac{\pi \cdot n \cdot d_{\max} (-1)^{n+1} \cdot \sin(d_{\max} \cdot q)}{(\pi \cdot n)^2 - (d_{\max} \cdot q)^2} \right] \longleftrightarrow p(r) = 8\pi \cdot r \sum_{n=1}^{N} a_n \cdot \sin\left(\frac{\pi \cdot r \cdot n}{d_{\max}} \right)$$

Represent SAXS data using a Fourier sine series

How large should N be?

Consider the denominator…

$$(\pi \cdot n)^2 - (d_{\max} \cdot q)^2 \quad \Rightarrow \quad (\pi \cdot n)^2 \neq (d_{\max} \cdot q)^2$$

The inequality naturally limits the expansion.

q_{max}	d_{max}	n
0.32	71	7
0.32	240	25

Notice, increasing d_{max} naturally increases n (same for q_{max})

make sense?

- **Logically, a larger macromolecule would require a more "complicated" equation to describe it.**
- **Similar to diffraction… larger object ⇒ larger unit cell ⇒ increase in I_{obs}. (lysozyme vs ribosome)**

What Can SAXS Do?

Assess solution state of biopolymer

Does MX represent solution state (~40% of the time)
Ensemble modeling

Characterize folded state of the biopolymer

Particle dimensions (d_{max}, R_g, R_c, mass, volume)
Assess compactness (Porod Exponent)

Monitor/Detect Conformational Changes

Magnitude of change dictates resolution range
Easy to detect by examining ratio of SAXS curves to reference state
Visualize by P(r) distribution

Volumeteric Modelling

Bead model representation of the scattering particle (DAMMIN/F)

Atomistic Modelling

Refinement of existing PDB structure

- add back missing elements (chains, domains)
- refine homology model (ALLOS-MOD FOXS Server UCSF)
- rigid body modelling

References

“Super-Resolution in Solution X-ray Scattering and Its Applications to Structural Systems Biology”
Annual Review of Biophysics, 2013 Volume 42, Pages 415-441 Rambo, R.P. and Tainer, J.A.

“Small-Angle Scattering for Structural Biology — expanding the frontier while avoiding the pitfalls.”
Protein Sci. 2010 19(4):642-57. Jacques DA, Trewhella J.

"Solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution"
Q Rev Biophys. 2007 Aug;40(3):191-285. Putnam CD, Hammel M, Hura GL, Tainer JA

“Small Angle X-ray Scattering”
Book circa 1982 Glatter O. and Kratky O. (very technical, freely available online)