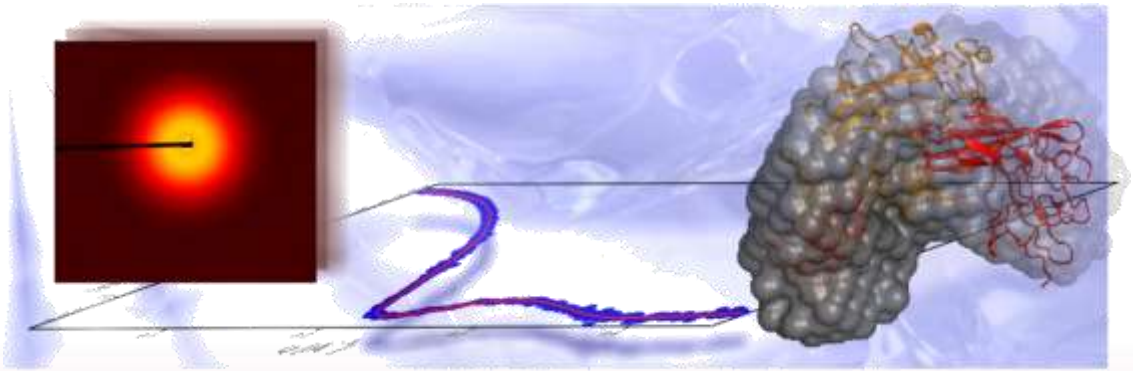
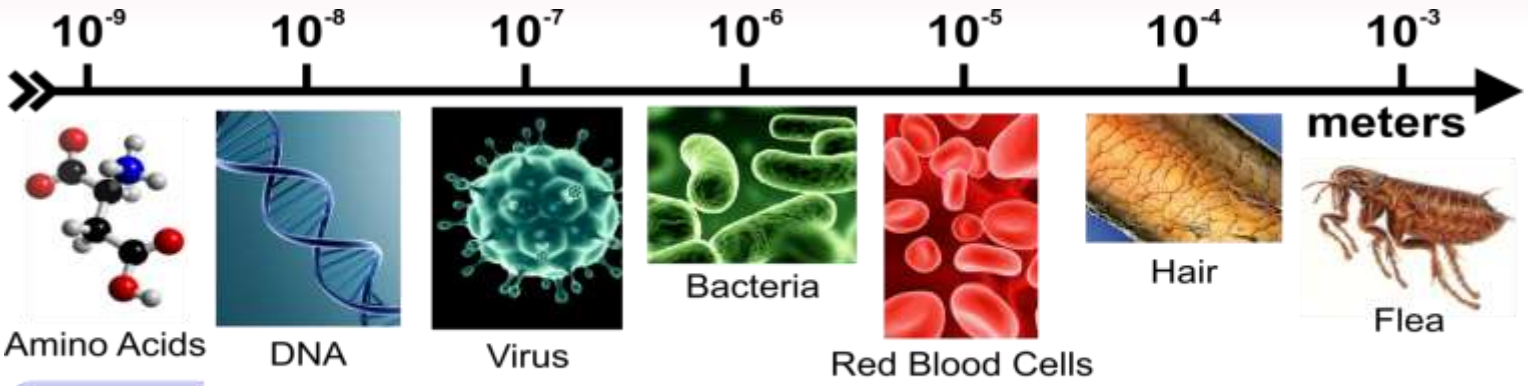
Introduction to Small-Angle X-ray Scattering



Thomas M. Weiss
Stanford University, SSRL/SLAC,
BioSAXS beamline BL 4-2

BioSAXS Workshop, March 28-30, 2016

Sizes and Techniques



EXAFS

Crystallography

Electron Diffraction

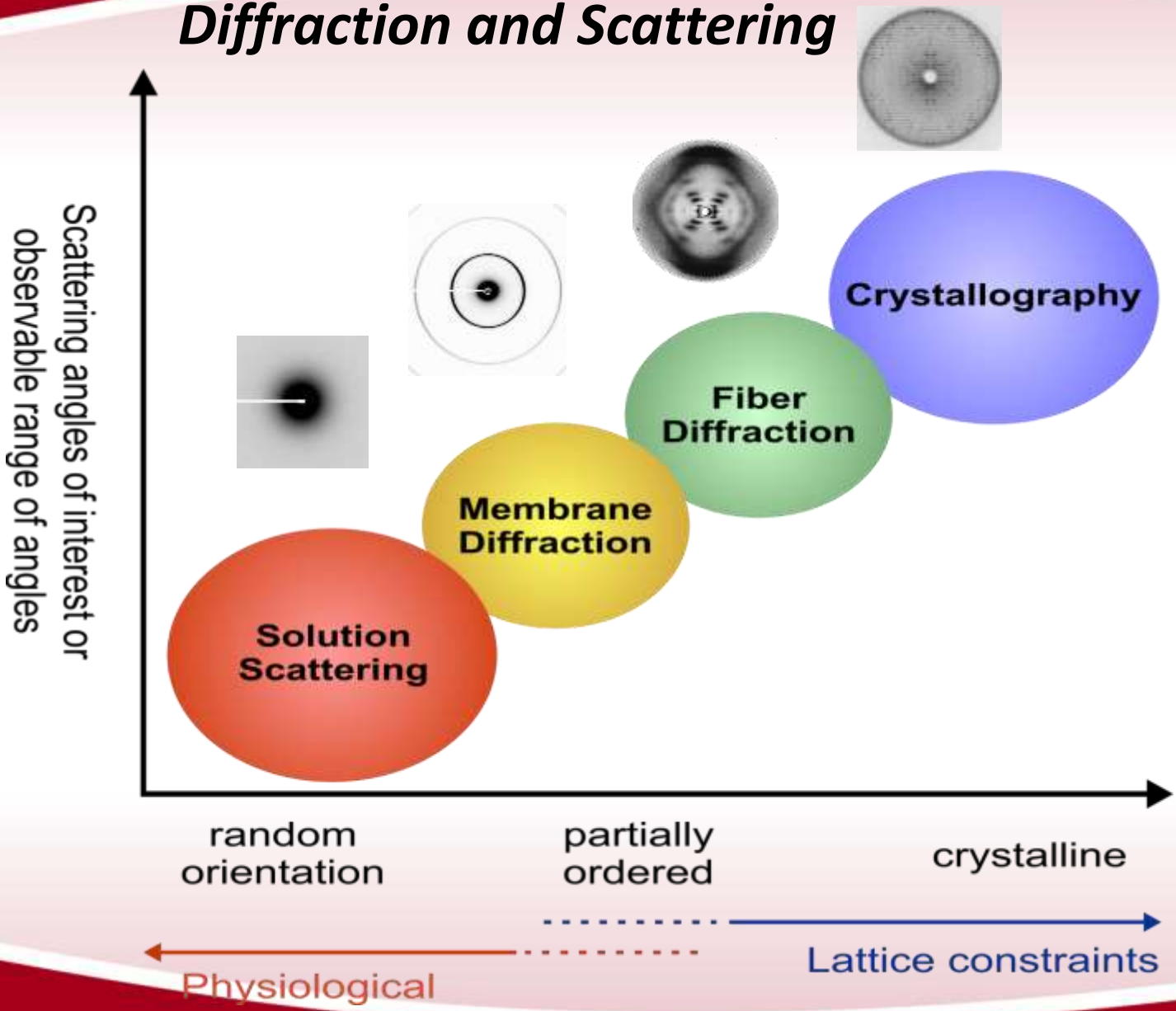
SAXS/SANS

Electron Microscopy

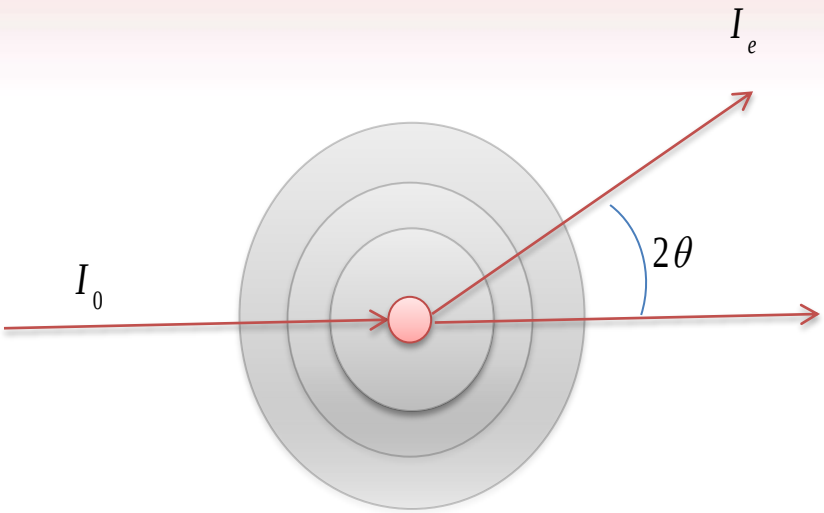
Atomic Force Microscopy

Optical Microscopy

Diffraction and Scattering



Scattering of X-rays from a single electron



I_0 : Intensity of incoming X - rays

2θ : angle of observatio n

I_e : Intensity of scattered X - rays

Thomson formula for the scattered intensity from a single electron

$$I_e = r_0^2 \frac{1+\cos(2\theta)}{2} \frac{1}{r^2} I_0$$

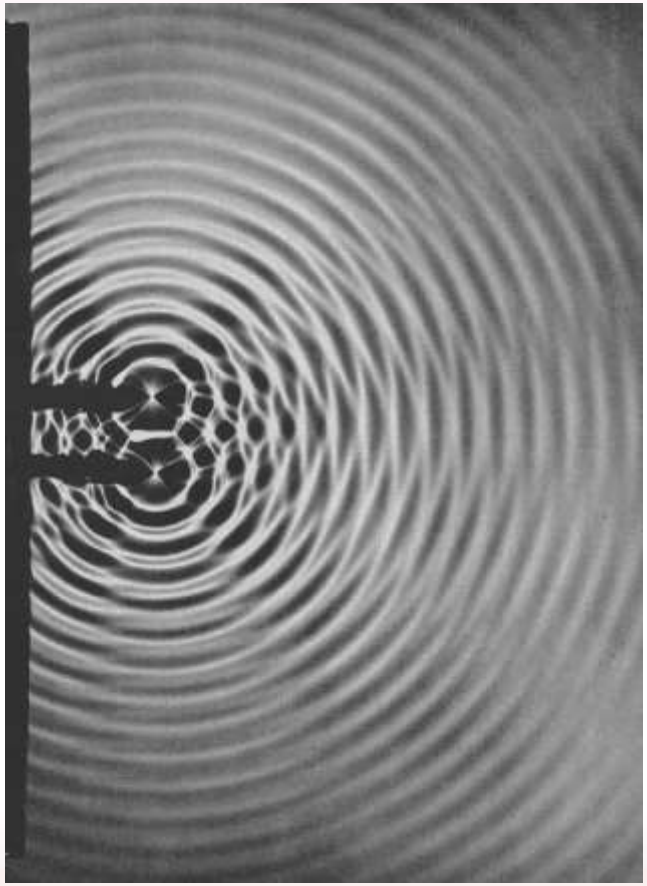
$$r_0 = \frac{e^2}{mc^2} = 2.817\ 10^{-15}\ m$$

Classical electron radius

The Thomson formula plays a central role for all scattering calculations involving absolute intensities. Typically calculated intensities of a given sample will be expressed in terms of the scattering of an isolated electron substituted for the sample.

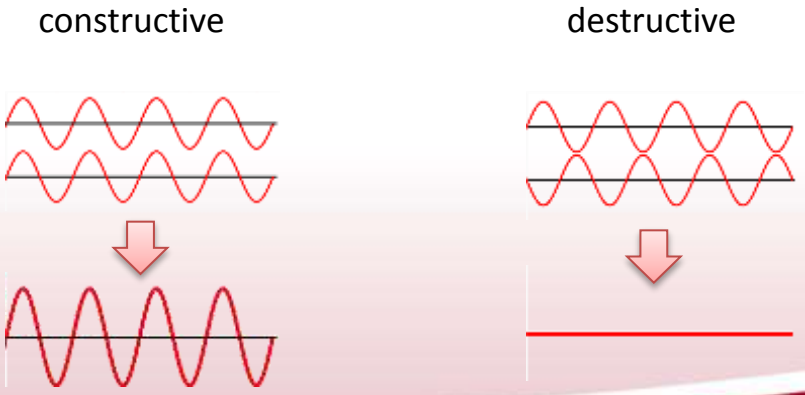
In small angle scattering the slight angle dependence (the so-called polarization factor) in the Thomson formula can be neglected.

Interference of waves



- waves have and amplitude and phase
- interference leads to fringe pattern (e.g. water waves)
- the fringe pattern contains the information on the position of the sources (i.e. structure)
- in X-ray diffraction the intensities (not the amplitudes) of the fringes are measured

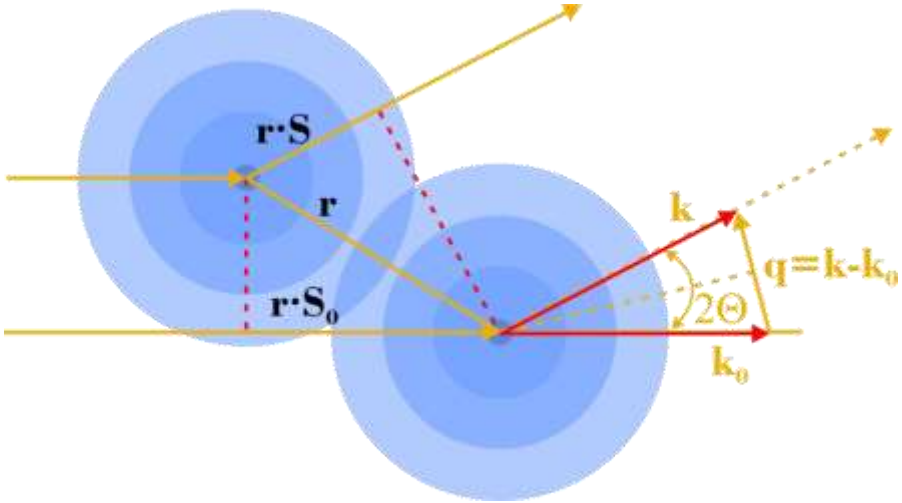
→ ***"phase problem"***



Scattering from two (and more) electrons

scattering vector $\mathbf{q} = \mathbf{k} - \mathbf{k}_0$

with $\mathbf{k} = \frac{2\pi}{\lambda}\mathbf{s}$ and $q = |\mathbf{q}| = \frac{4\pi \sin\theta}{\lambda}$



two electrons

$$F(q) = \sum_{i=1}^{2} f_e \exp(i\mathbf{q}\cdot\mathbf{r}_i) = f_e(1 + \exp(i\mathbf{q}\cdot\mathbf{r}))$$

… generalized to N electrons

$$F(q) = \sum_{i=1}^{N} f_e \exp(i\mathbf{q}\cdot\mathbf{r}_i)$$

Note:
F(q) is the Fourier Transform of the spatial distribution of the electrons

… averaged over all orientations

$$\langle F(q) \rangle = \sum_{i=1}^{N} f_e \frac{\sin(qr)}{qr}$$

using the continuous (radial) distribution $\rho(r)$ of the electron cloud in an atom

$$F(q) \equiv f(q) = \int_0^\infty dr\ \rho(r) r^2 \frac{\sin(qr)}{qr} \qquad \text{with} \qquad f(0) = Z$$

atomic scattering factor

Scattering from Molecules

The scattering amplitude or form factor, F(q), of an isolated molecule with N atoms can be determined in an analogous manner:

$$F(\mathbf{q}) = \sum_{i=1}^{N} f_i(\mathbf{q}) \exp(\, i\mathbf{q} \cdot \mathbf{r}_i)$$

i.e. the Fourier Transform of the atomic distribution

The scattered intensity from the isolated molecule is then

$$I(\mathbf{q}) = \left|F(\mathbf{q})\right|^2 = \sum_{i=1}^{N} \sum_{j=1}^{N} f_i(q) f_j(q) \exp(\, i\mathbf{q} \cdot (\mathbf{r}_i - \mathbf{r}_j))$$

In solution:
average over all orientations

$$\left\langle \exp(\, i\mathbf{q} \cdot (\mathbf{r}_i - \mathbf{r}_j)) \right\rangle = \frac{\sin(\, qr_{ij})}{qr_{ij}}$$

$$I(q) = \sum_{i=1}^{N} \sum_{j=1}^{N} f_i(q) f_j(q) \frac{\sin(\, qr_{ij})}{qr_{ij}}$$

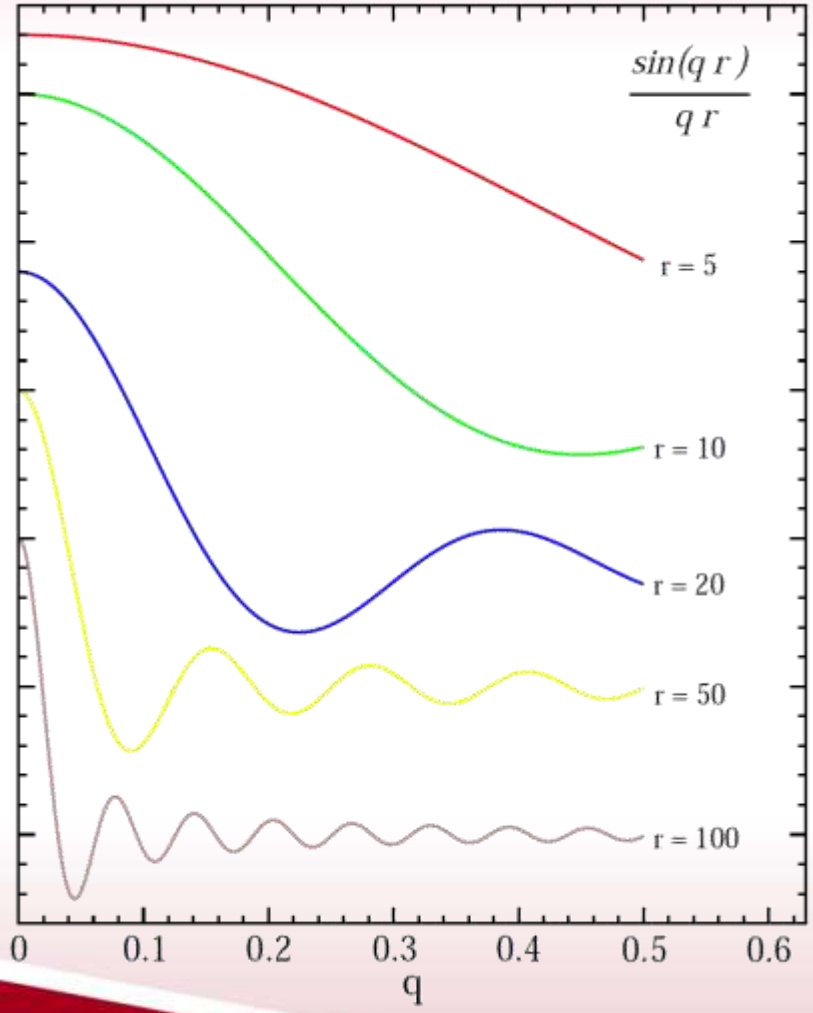
Debye formula

due to solution average only interatomic distances are measured, not atomic coordinates

SSRL

SLAC

Scattering from Molecules



$$I(q) = \sum_{i=1}^{N} \sum_{j=1}^{N} f_i(q) f_j(q) \frac{\sin(qr_{ij})}{qr_{ij}}$$

Each atomic distance r_{ij} in the molecule adds a sinx/x like term to the scattering intensity

- small distance
  low frequency in sinx/x
  dominate signal at **high q**

- large distance
  high frequencies in sinx/x
  dominate the signal at **low q**

Scattering Intensity

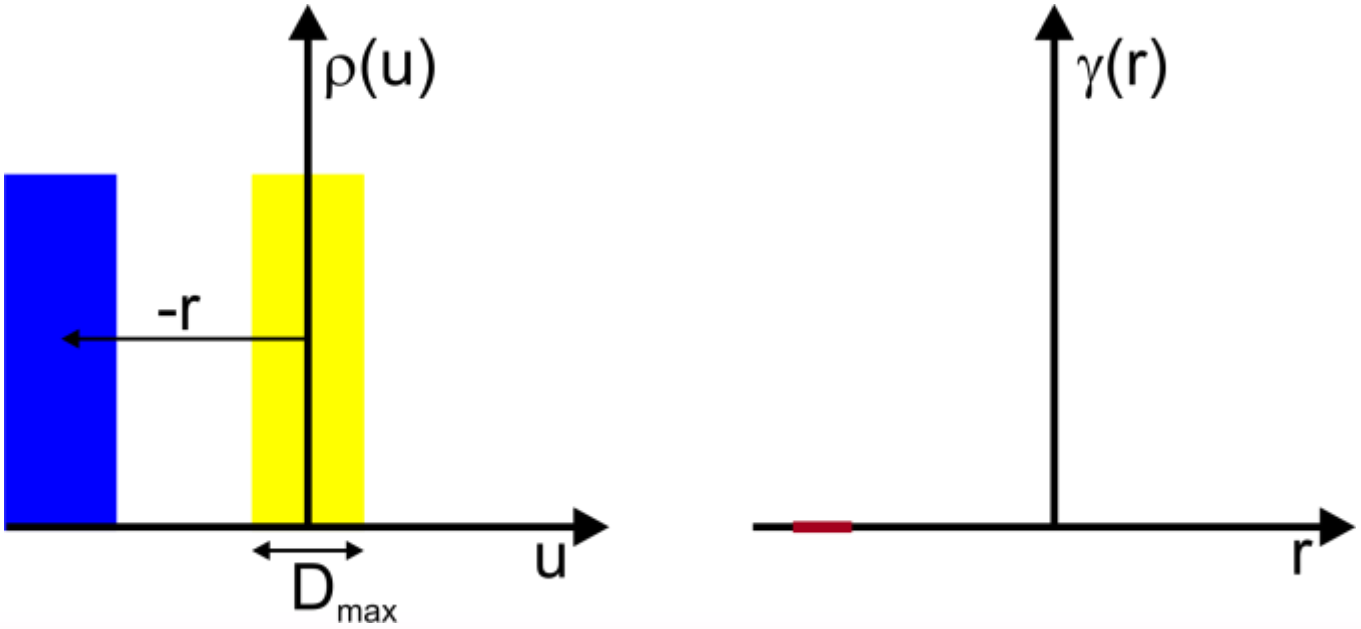
$$I(q) = \left\langle F(\mathbf{q}) F(\mathbf{q})^{*} \right\rangle = \left\langle FT\,[\rho(\mathbf{r})]\, FT\,[\rho(-\mathbf{r})] \right\rangle = \left\langle FT\,[\underbrace{\rho(\mathbf{r}) * \rho(-\mathbf{r})}_{\gamma(\mathbf{r})}] \right\rangle$$

Autocorrelation function

The measured scattering intensity is the spherically averaged **Fourier transform of the auocorrelation** of the electron density of the particle

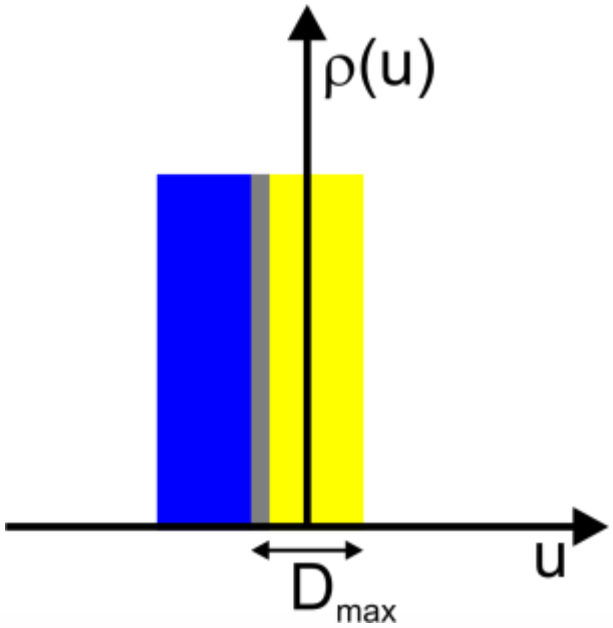
Autocorrelation

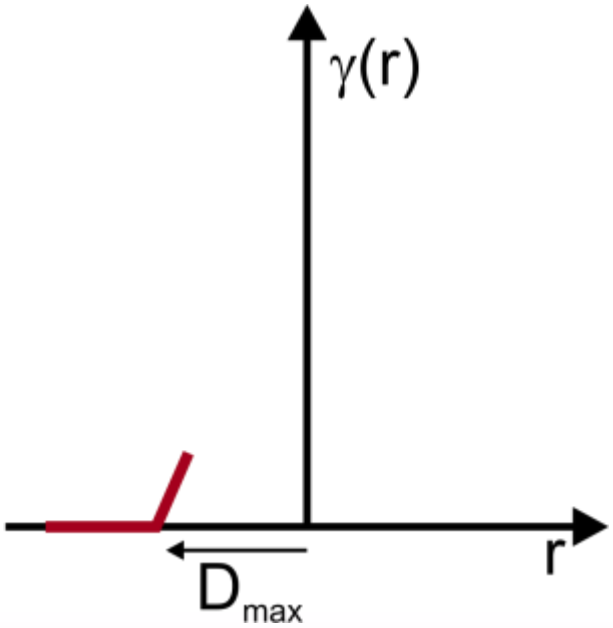
$$\gamma(\mathbf{r}) = \rho(\mathbf{r}) * \rho(-\mathbf{r}) = \int_{V_u} \rho(\mathbf{r}+\mathbf{u})\rho(\mathbf{r})\,dV_u$$



Autocorrelation

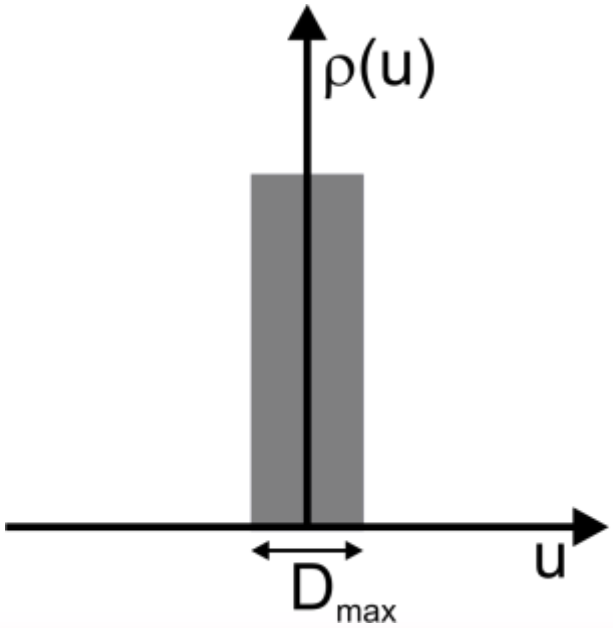
$$\gamma(\mathbf{r}) = \rho(\mathbf{r}) * \rho(-\mathbf{r}) = \int_{V_u} \rho(\mathbf{r}+\mathbf{u})\rho(\mathbf{r})\,dV_u$$

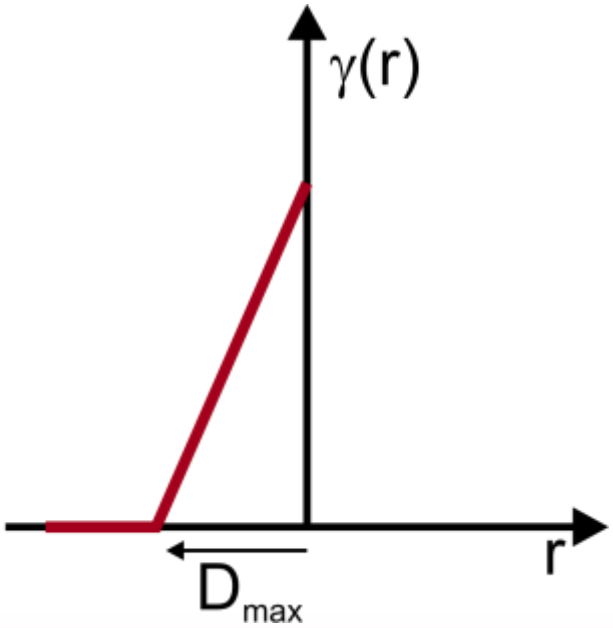




Autocorrelation

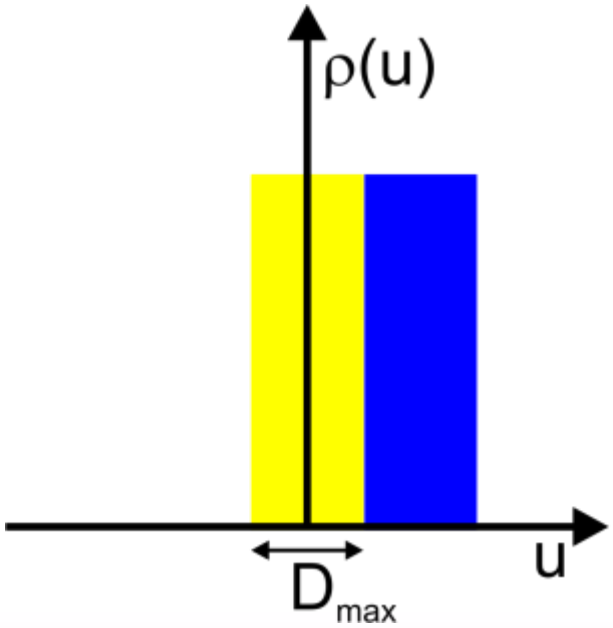
$$\gamma(\mathbf{r}) = \rho(\mathbf{r}) * \rho(-\mathbf{r}) = \int_{V_u} \rho(\mathbf{r}+\mathbf{u})\rho(\mathbf{r})\,dV_u$$

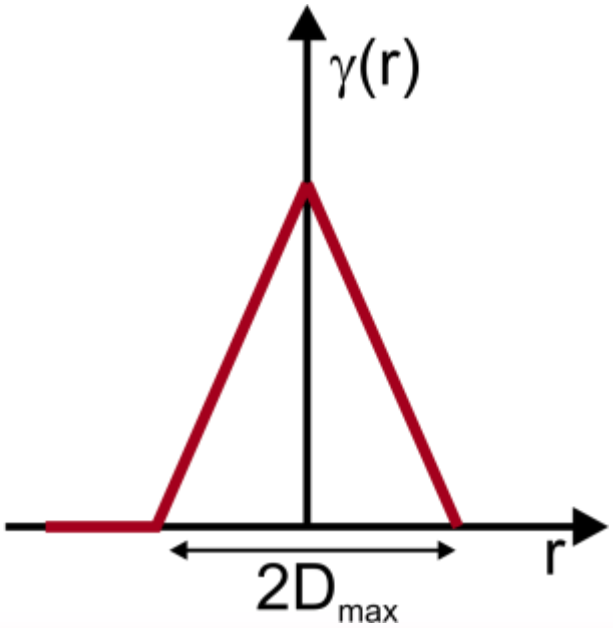




Autocorrelation

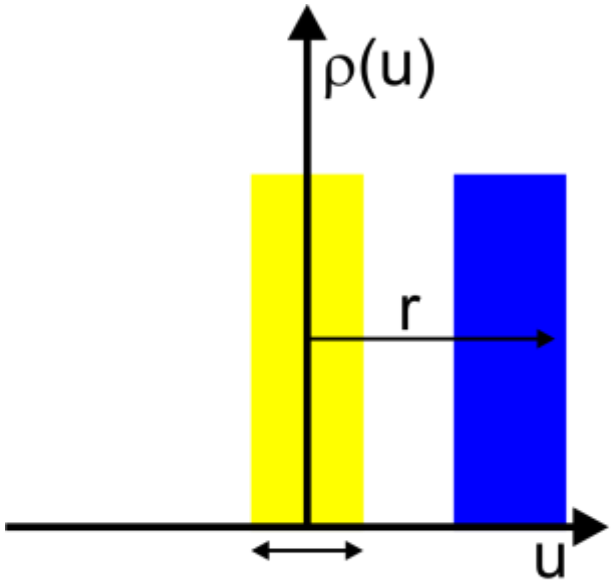
$$\gamma(\mathbf{r}) = \rho(\mathbf{r}) * \rho(-\mathbf{r}) = \int_{V_u} \rho(\mathbf{r}+\mathbf{u})\rho(\mathbf{r})\,dV_u$$

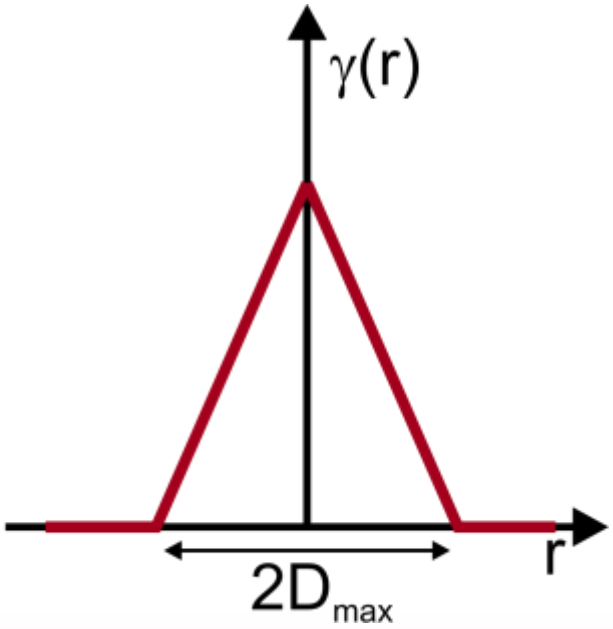




Autocorrelation

$$\gamma(\mathbf{r}) = \rho(\mathbf{r}) * \rho(-\mathbf{r}) = \int_{V_u} \rho(\mathbf{r}+\mathbf{u})\rho(\mathbf{r})\,dV_u$$





Autocorrelation

$$\gamma(\mathbf{r}) = \rho(\mathbf{r}) * \rho(-\mathbf{r}) = \int_{V_u} \rho(\mathbf{r}+\mathbf{u})\rho(\mathbf{r})\,dV_u$$

For a homogeneous particle

$$\rho(\mathbf{r}) = \rho \quad \mathbf{r} \in V$$
$$= 0 \quad \mathbf{r} \notin V$$

Spherical average

$$\gamma(r) = \langle \gamma(\mathbf{r}) \rangle$$

Characteristic Function

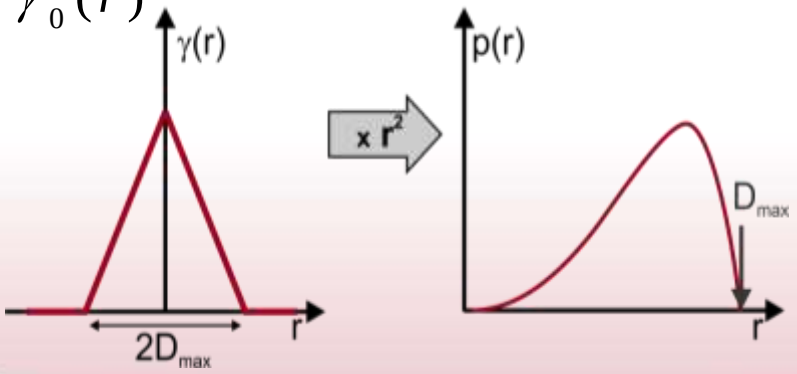
$$\gamma_0(r) = \gamma(r)/\gamma(0)$$

with $\gamma(0) = \rho^2 V$

"probability of finding a point within the particle at a distance r from a given point"

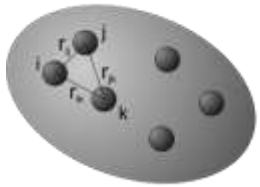
Pair distance distribution function:

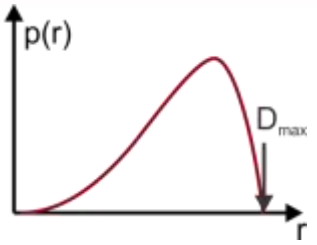
$$p(r) = r^2\gamma(r) = \rho^2 V r^2 \gamma_0(r)$$



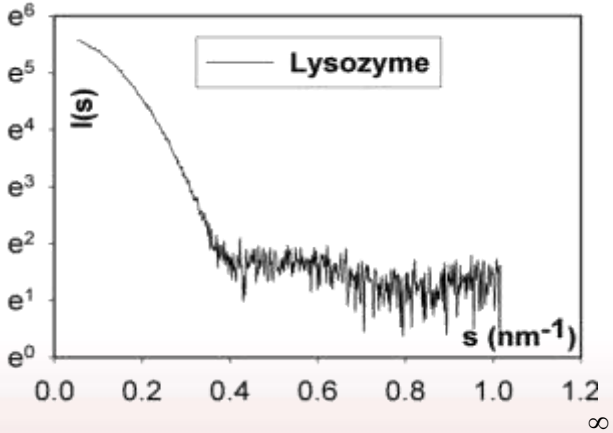
Pair distance distribution function p(r)

The p(r) function represents the histogram of distances between pairs of points within the particle. **Dmax** is the maximum diameter in the particle.

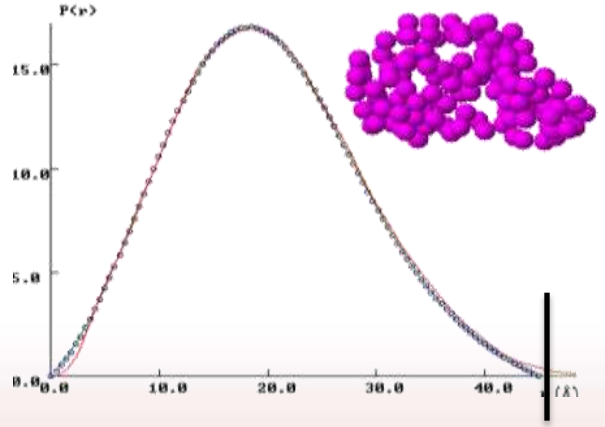




Measured scattering intensity

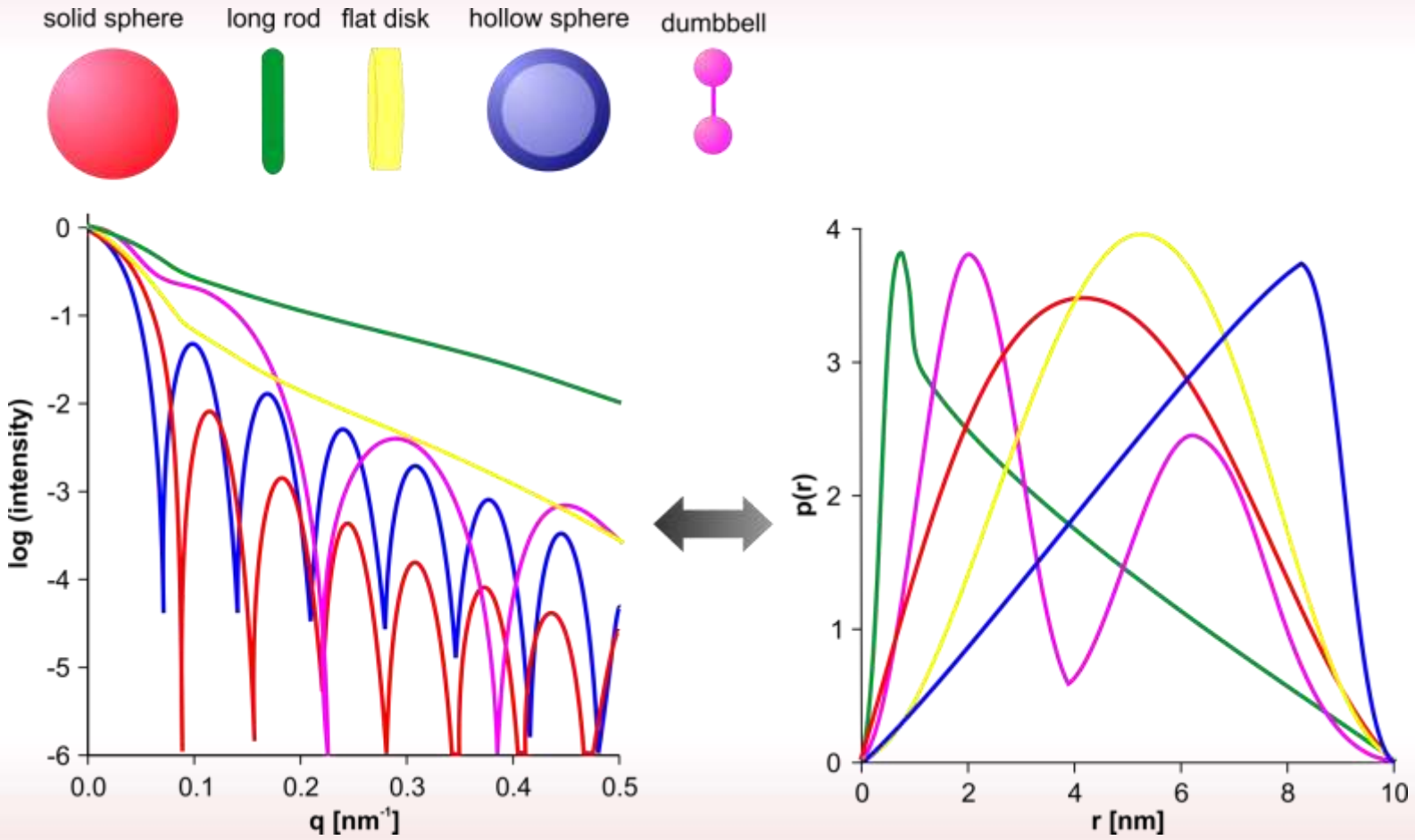


Pair distance distribution



$$p(r) = 4\pi \int_0^\infty I(q)\, qr\, \sin(qr)\, dq$$

Scattering from model structures



Adopted from Svergun & Koch, “SAS studies of biological macromolecules in solution”, Rep. Prog. Phys. 66 (2003) 1735-1782, Fig. 5 (c)I

Particles in Solution

For solution scattering we typically require the following characteristics:

- **Monodisperse**, i.e. identical particles

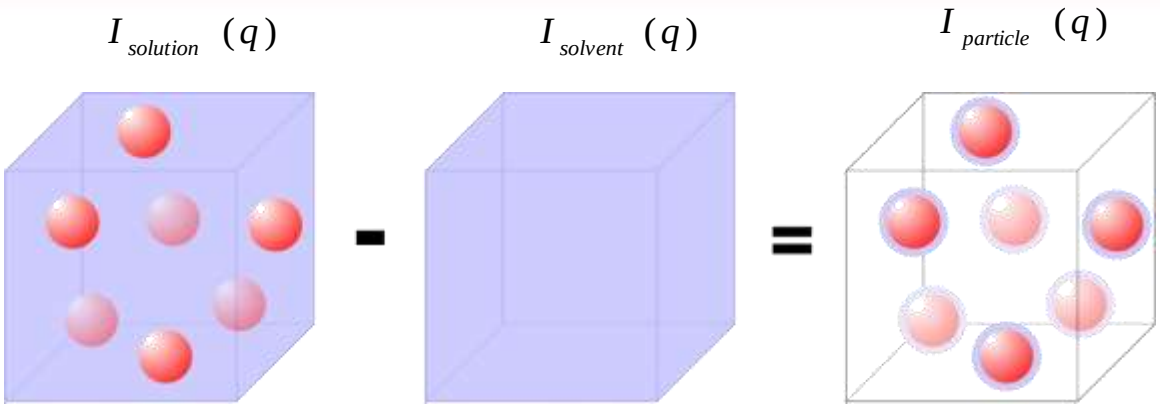
$$i_j(q) = i_1(q) \qquad \forall j$$

- **Uncorrelated**, i.e. no inter-molecular interactions present

$$I(q) = \sum_j n_j i_j(q)$$

$$I(q) = N i_1(q)$$

Background Scattering and X-ray Contrast

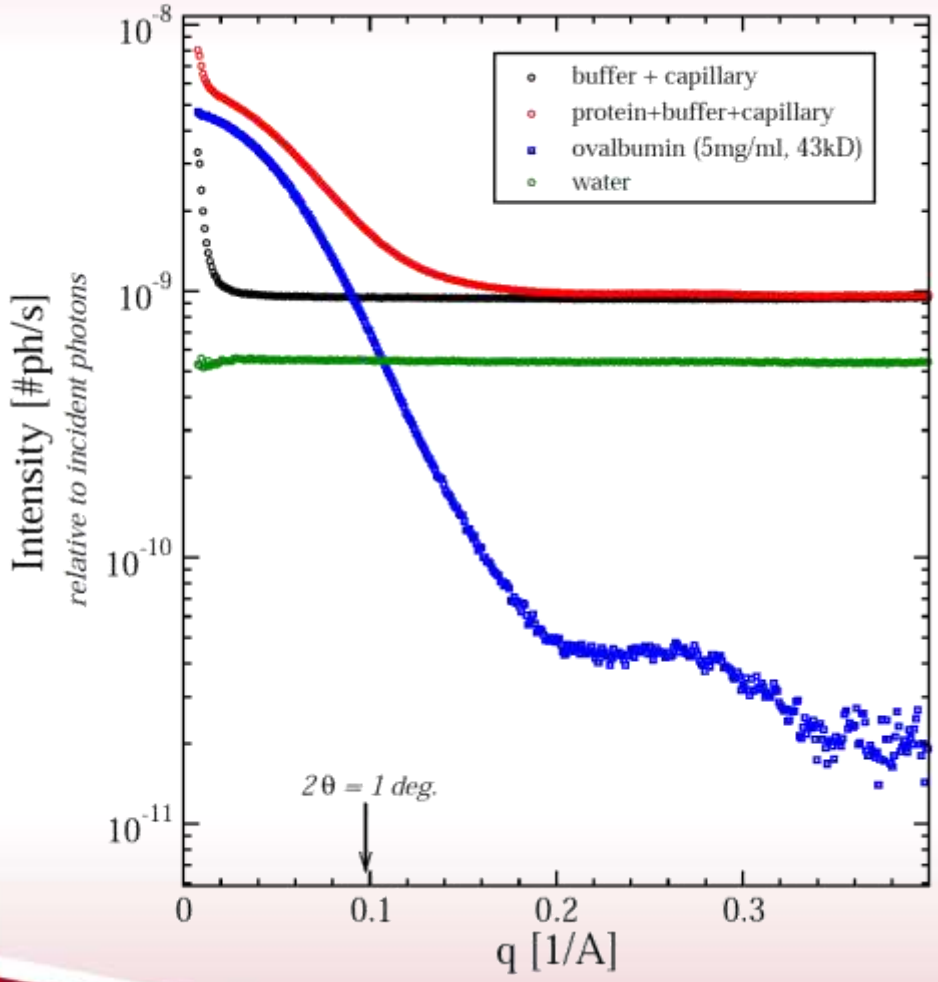


• The solvent scattering background must be properly subtracted to obtain the signal from the particles

• the contrast, that makes the particles “visible” for X-rays, is the difference in electron density of the particle versus the solvent

$$(\rho(r) - \rho_s)^2$$

Protein solution scattering data



- weak level of scattering at small angles
- drops off quickly for higher angles
- due to low contrast scattering level of background and sample is very similar except for the lowest angles
- background and sample scattering need to be measured with high accuracy

- a 1mg/ml solution of a globular protein of the size of lysozyme (14kD) scatters on the order of:

1 out of 10^6 incident photons

"…. one in a million!"

SSRL

SLAC

X-ray Contrast and Contrast Variation

Substance	Average Contrast ($x10^{10}$ cm^{-2})
Protein	2.5
Nucleic Acid	6.7
Fatty Acid	-1.1
Carbohydrates	4.5

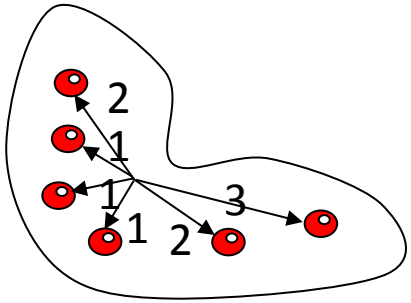
- change contrast by adding salts (e.g. CsBr), sucrose or glycerol to the solvent
- but that changes the chemical environment for the particles
- other possibility to change contrast is anomalous scattering

Note:
Contrast variation is widely used in neutron scattering, due to the large scattering length difference of hydrogen and deuterium

Introducing the Radius of Gyration

R_g^2 is the average electron density weighted squared distance of the scatters from the centre of the object

$$R_g^2 = \frac{\int \mathbf{r}^2 \rho(\mathbf{r}) d\mathbf{r}}{\int \rho(\mathbf{r}) d\mathbf{r}}$$



$R_g^2 = (1^2 + 1^2 + 1^2 + 2^2 + 2^2 + 3^2)/6 = 20/6$

$R_g = \sqrt{3.333} = 1.82$

- Solid sphere radius R:

  $R_g = \sqrt{(3/5)}\ R$

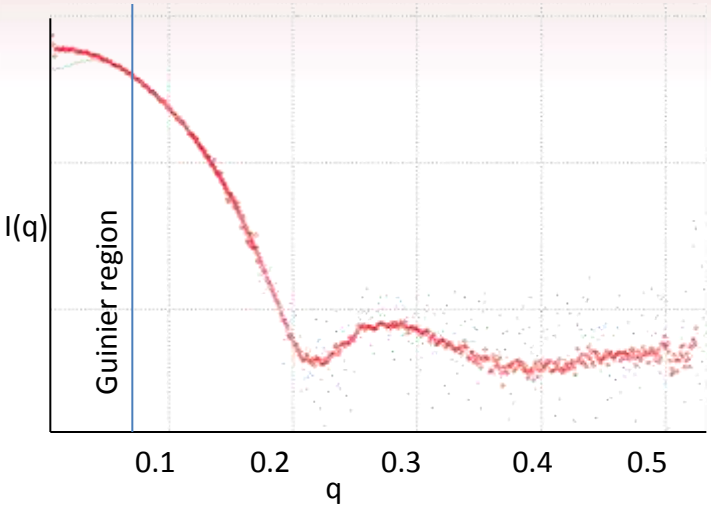
- Thin rod length L

  $R_g = \sqrt{(1/12)}\ L$

- Thin disk radius R:

  $R_g = \sqrt{(1/2)}\ R$

The Guinier approximation



The low-q region of the scattering curve is characteristic for the overall dimension of the particle.

$$\lim_{q \to 0} I = I_0 \exp\left(-\frac{1}{3} q^2 R_g^2\right)$$

I_0 is proportional to M_w

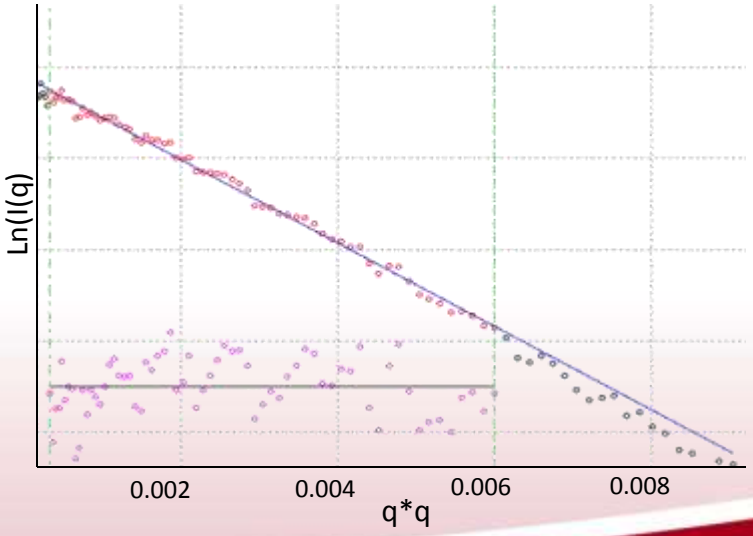
Radius of gyration: size of the particle

"The Guinier Plot"

Plot ln I against q^2 →Straight line, slope $-R_g/3$

$$\ln I = \ln I_0 - \frac{1}{3} q^2 R_g^2$$

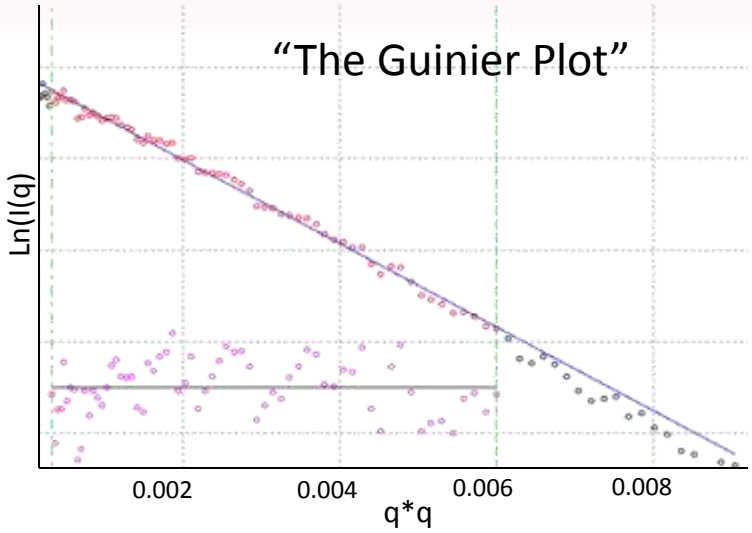
Deviation from the straight line in the Guinier plot indicate intermolecular interaction or aggregation



The Guinier approximation

Plot ln I against q^2 →Straight line

$$\ln I = \ln I_0 - \frac{1}{3} q^2 R_g^2$$



Recall: $$I(q) = \sum_{i=1}^{N} \sum_{j=1}^{N} f_i(q) f_j(q) \frac{\sin(qr_{ij})}{qr_{ij}}$$

thus $$I_0 = \left(\sum_i f_i \right)^2$$

i.e. the number of (excess) electrons in the sample

$$I_0 = \frac{cMV}{N_a} \left[\nu_{partial} (\rho - \rho_0)^2 \right]$$

c: concentration
M: molecular mass
V: Volume

ν: partial specific Vol.
ρ: prot. e-density
ρ_0: solvent e-density

Radius of Gyration

$$R_g^2 = \frac{\int \mathbf{r}^2 \rho(\mathbf{r}) d\mathbf{r}}{\int \rho(\mathbf{r}) d\mathbf{r}}$$

Alternatively to using the Guinier plot to determine the Rg of the protein of can also use the following experession involving the P(r) function:

$$R_g = \frac{\int_0^{D_{\max}} r^2 p(r) dr}{2 \int_0^{D_{\max}} p(r) dr}$$

i.e. Rg equals the second moment of the electron density distribution as well as half the second moment of the distance distribution function

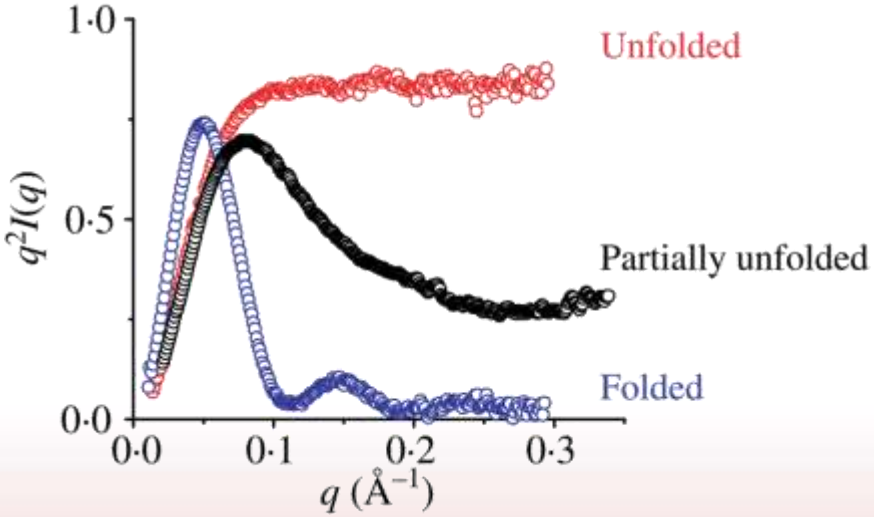
This is often better than using the Guinier plot as it involves the whole scattering curve

Radius of gyration for proteins and viruses

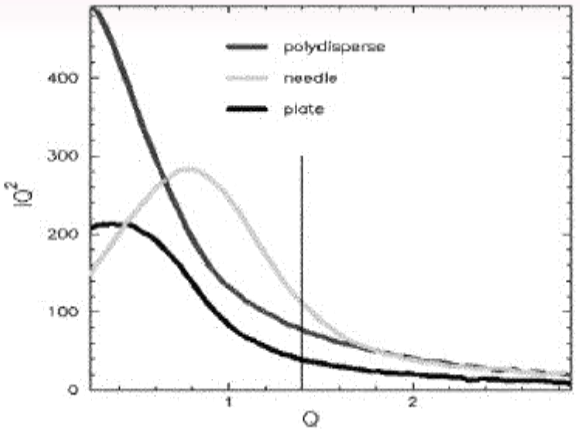
	Molecular Weight	Rg (A)
Ribonuclease	12700	14.8
Lysozyme	14800	14.5
B-lactoglobulin	36700	21.7
BSA	68000	29.5
Myosin	493000	468
Brome Mosaic Virus	$4.6\ 10^6$	134
TMV	$3.9\ 10^6$	924

Kratky analysis

- Kratky plot: $I*q^2$ vs. q
- sensitive to morphology of particle
- sensitive to the compactness of a protein
- unfolded and folded states of proteins are easy to distinguish

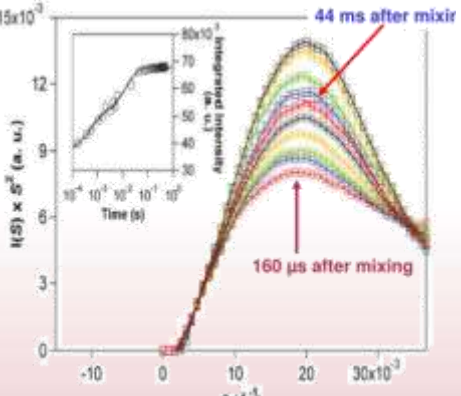


Putnam *et al., Quat.Rev.Bioph 40,3 (2007), Fig24*



Hiller *et al., Biomaterials, 24 (2003), Fig5*

Example: folding of cytochrome C



Akiyama et al., PNAS, 99, (2002)

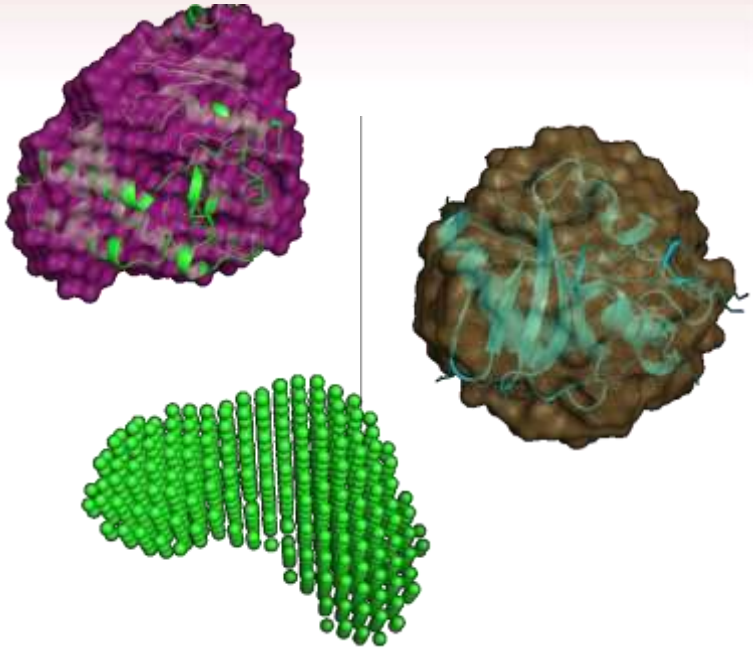
ab-initio *structure determination*

Envelop models

- using spherical harmonics to produce molecular envelopes that fit the experimental scattering data

Bead models

- fitting the scattering data using bead as scattering centers
- so-called dummy residues (scattering centers representing the Ca atoms of the residues)



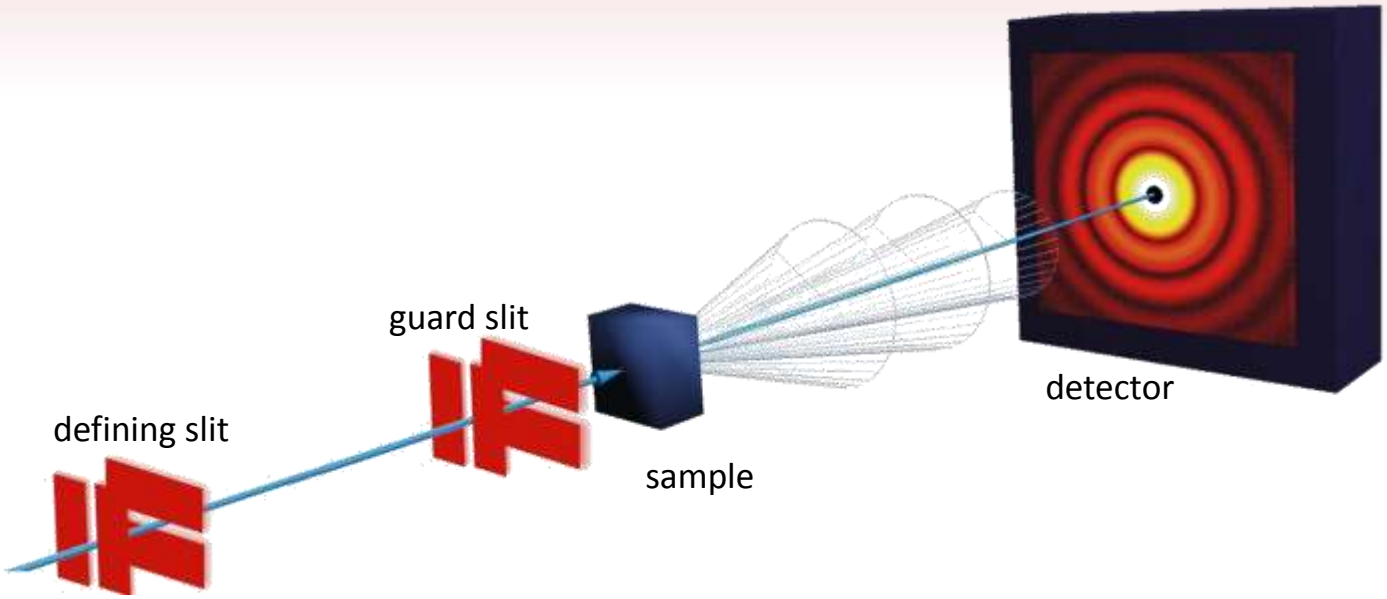
Can be extremely powerful particularly if combined with (partial) crystal structures if available!

but be careful:
you will always get a structure from these programs, but it doesn't mean they make sense

Program packages:

- ATSAS from EMBL Hamburg (Svergun group)
- IMP from UCSF (Sali group)
- SAXS3D from Stanford (Doniach group)
- SASTBX from LBL (Zwart group)
- …

Experimental setup

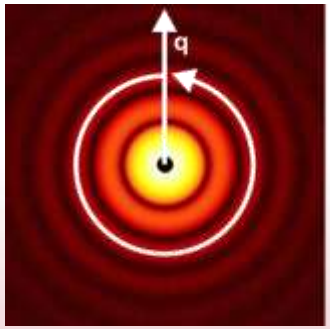


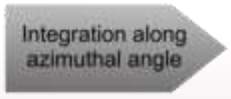
$$I(q) = N\left|F(q)\right|^2 S(q)$$

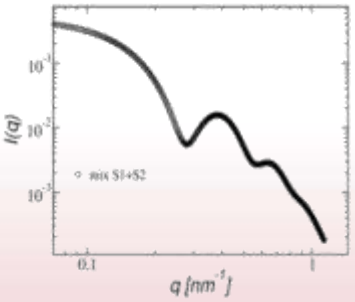
$$q = \frac{4\pi \sin\theta}{\lambda}$$

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

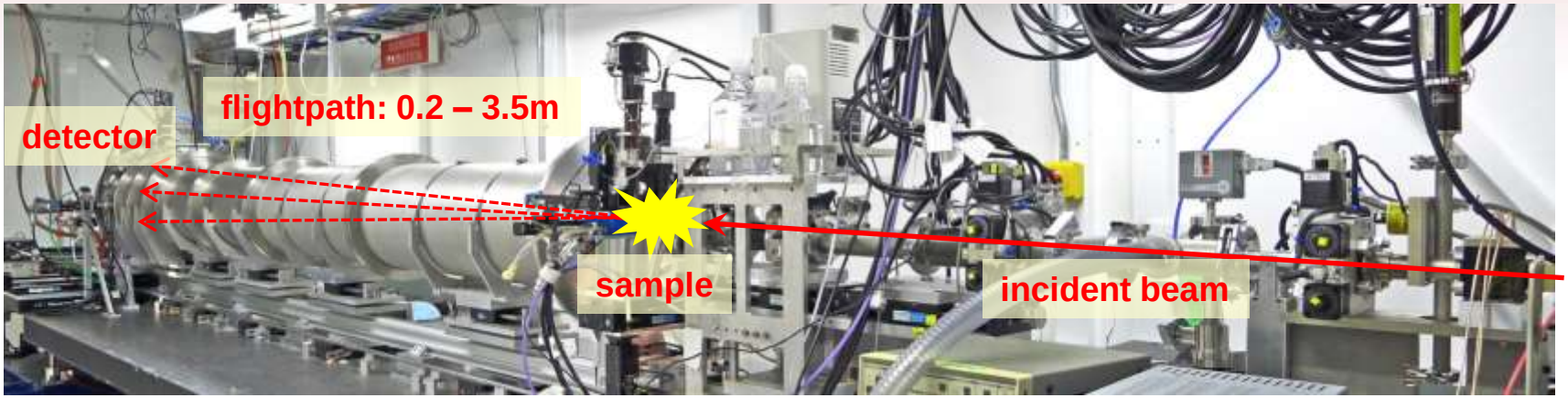
isotropic scattering signal:
the 2D detector image is integrated to yield I(q) vs q





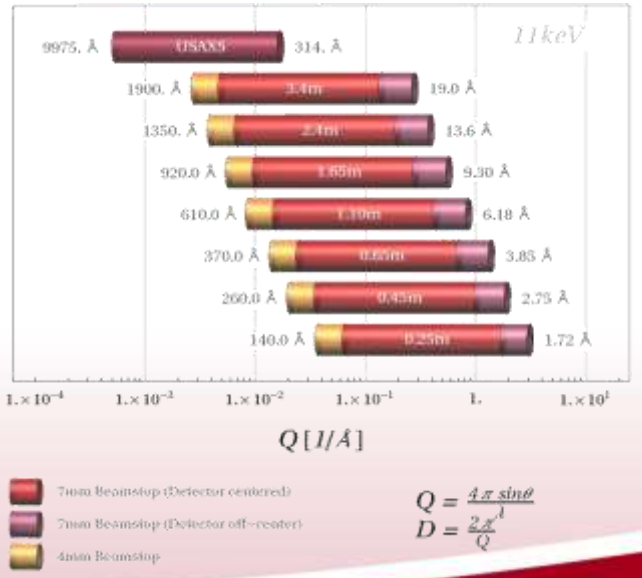


BioSAXS instrument at SSRL BL 4-2



- widely re-configurable instrument for
  - static and time-resolved solution scattering
  - lipid/fiber diffraction
  - grazing incidence scattering
  - anomalous scattering
- variety of advanced sample environments
  - solution scattering robot with attached analysis pipeline
  - in-line size-exclusion chromatography setup
  - stopped-flow mixer with low sample consumption
  - humidity chamber for lipid studies
  - high-throughput LCP screening setup

Q-range Q= 0.003/Å ... 4.2/Å



Why do SAXS?

Structural information obtainable from SAXS

- Radius of gyration (globular, cross-sectional etc.)
- molecular weight (monomer, dimer, multimer ..)
- pair-distance distribution function
- low-resolution envelope of molecule and ab-initio structures (about 1nm resolution)
- unfolded vs folded (Kratky plot)
- interaction potentials

Systems that can be studies by SAXS

- study protein at physiological conditions
- time-resolved studies possible (reaction kinetics)
- large protein complexes (no need for crystals)
- unfolded or partially folded proteins
- complex systems (protein-DNA, protein-lipid …)

Thank you