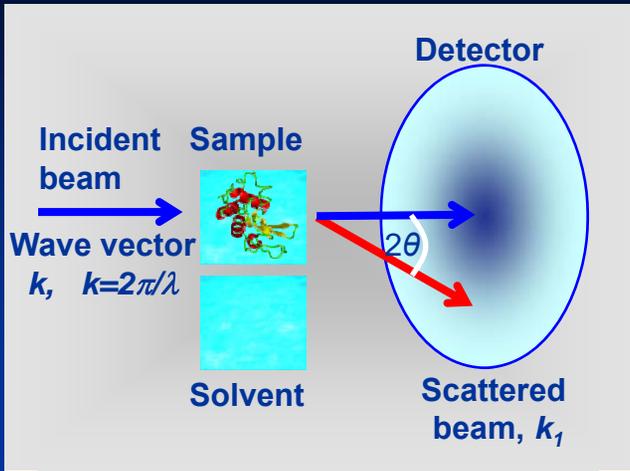# Basics of X-ray scattering by solutions
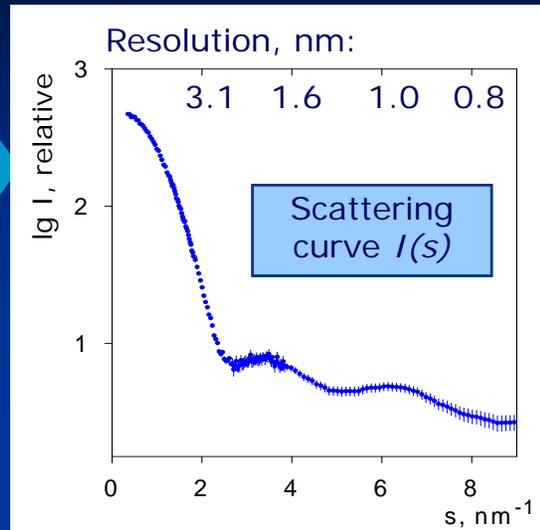
D.Svergun, EMBL-Hamburg

EMBO

EMBL

Small-angle scattering in structural biology

Data analysis

Detector

Incident beam

Sample

Wave vector *k, k=2π/λ*

2θ

Solvent

Scattered beam, k_1

Radiation sources:
X-ray tube   *(λ = 0.1 - 0.2 nm)*
Synchrotron   *(λ = 0.05 - 0.5 nm)*
Thermal neutrons *(λ = 0.1 -   1 nm)*

Resolution, nm:
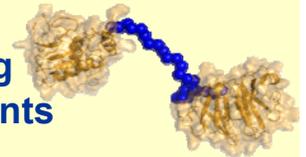3.1   1.6   1.0   0.8

lg I, relative

Scattering curve *I(s)*

s, nm^{-1}

Shape determination

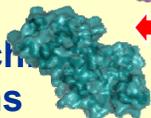**Rigid body modelling**

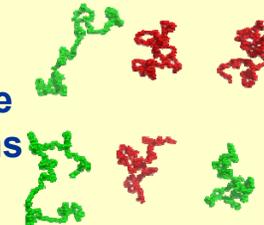**Missing fragments**

Oligomeric mixtures
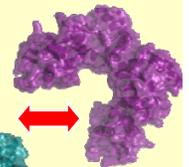
Hierarch... systems

Flexible systems

Complementary techniques

MS

EM

Crystallography

NMR

Bioinformatics

Biochemistry

AUC

FRET

EPR

Additional information

Homology models

Atomic models

Distances

Orientations

Interfaces

General princples of solution SAXS

Small-angle scattering: experiment

Detector

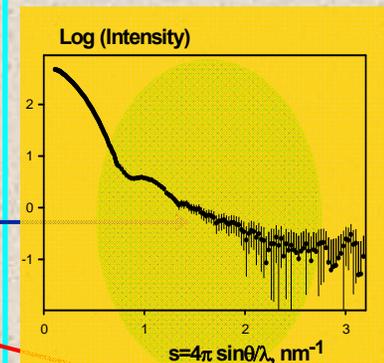**Sample**

Monochromatic beam

Wave vector $\boldsymbol{k}$, $k=2\pi/\lambda$

2θ

Log (Intensity)

$s=4\pi \sin\theta/\lambda$, nm^{-1}

$\boldsymbol{k_1}$

Radiation sources:

X-ray generator (λ = 0.1 - 0.2 nm)
Synchrotron (λ = 0.03 - 0.35 nm)
Thermal neutrons (λ = 0.2 - 1 nm)

Scattering vector $\boldsymbol{s=k_1-k}$,

Scattering by matter

- **X-rays** are scattered mostly by electrons

- **Thermal neutrons** are scattered mostly by nuclei

- Scattering amplitude from an ensemble of atoms $A(s)$ is the Fourier transform of the scattering length density distribution in the sample $\rho(r)$

- Experimentally, scattering intensity $I(s) = [A(s)]^2$ is measured.

Notations

The momentum transfer (i.e. the modulus of the scattering vector is denoted here as $s = 4\pi \sin(\theta)/\lambda$
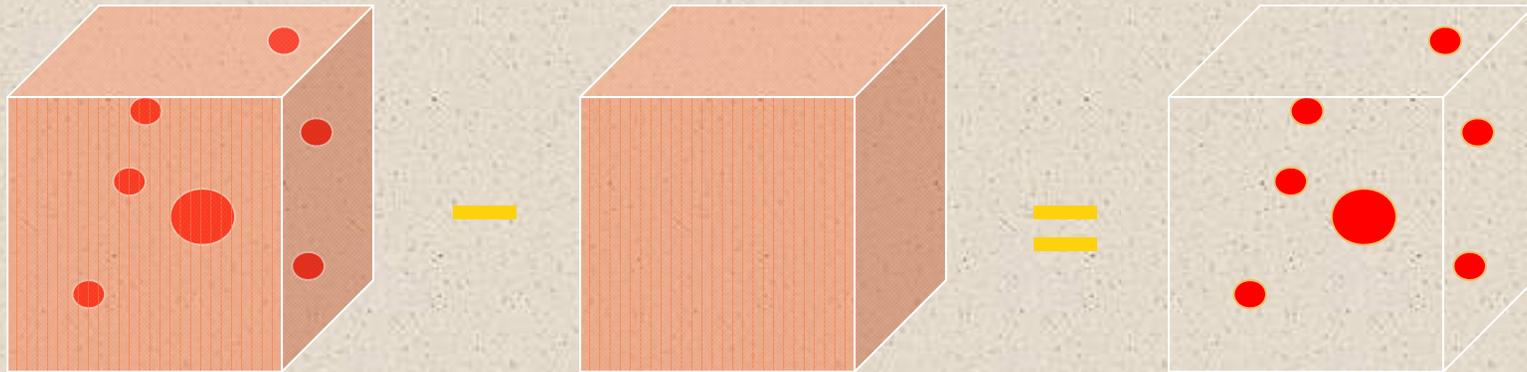
There are also different letters used, like

$$Q = q = s = h = k = 4\pi \sin(\theta)/\lambda$$

Sometimes, the variable $S = 2\sin\theta/\lambda = 2\pi s$ is used, and to add to the confusion, also denoted as "s", or µ or yet another letter. Always check the definition for the momentum transfer in a paper

Small-angle scattering: contrast

$$I_{sample}(s) \quad - \quad I_{matrix}(s) \quad = \quad I_{particle}(s)$$

- ♦ To obtain scattering from the particles, matrix scattering must be subtracted, which also permits to significantly reduce contribution from parasitic background (slits, sample holder etc)

- ♦ **Contrast** $\Delta\rho$ **= <ρ(r) - ρ_s>**, where ρ_s is the scattering density of the matrix, may be very small for biological samples

X-rays versus neutrons

- **X-rays:** scattering factor increases with atomic number, no difference between H and D

- **Neutrons:** scattering factor is irregular, may be negative, huge difference between H and D

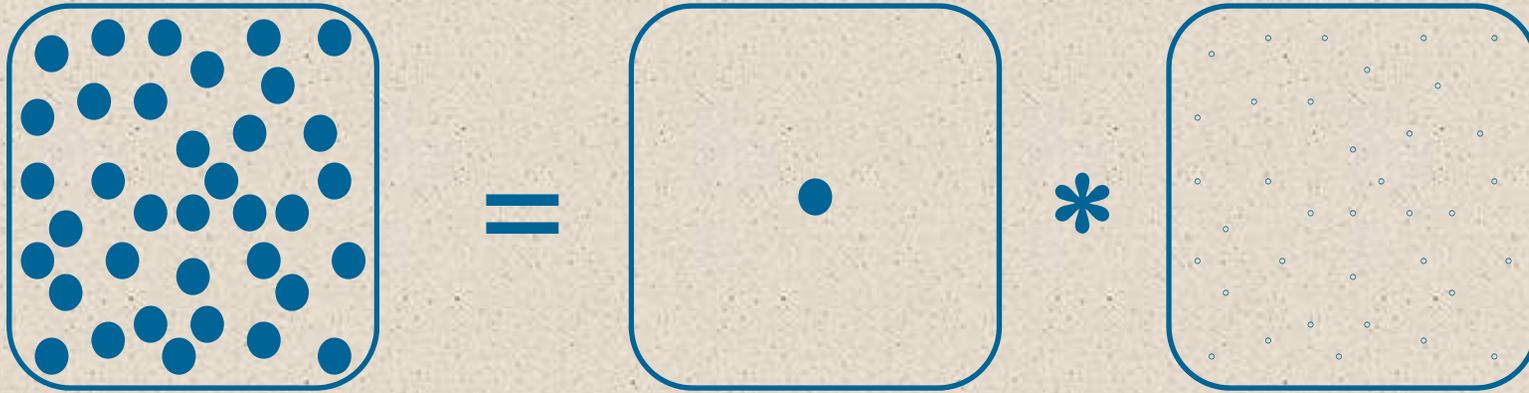| Element	H	D	C	N	O	P	S	Au
At. Weight	1	2	12	14	16	30	32	197
N electrons	1	1	6	7	8	15	16	79
b_X, 10^{-12} cm	0.282	0.282	1.69	1.97	2.16	3.23	4.51	22.3
b_N, 10^{-12} cm	-0.374	0.667	0.665	0.940	0.580	0.510	0.280	0.760

Neutron contrast variation

Contrast of electron density
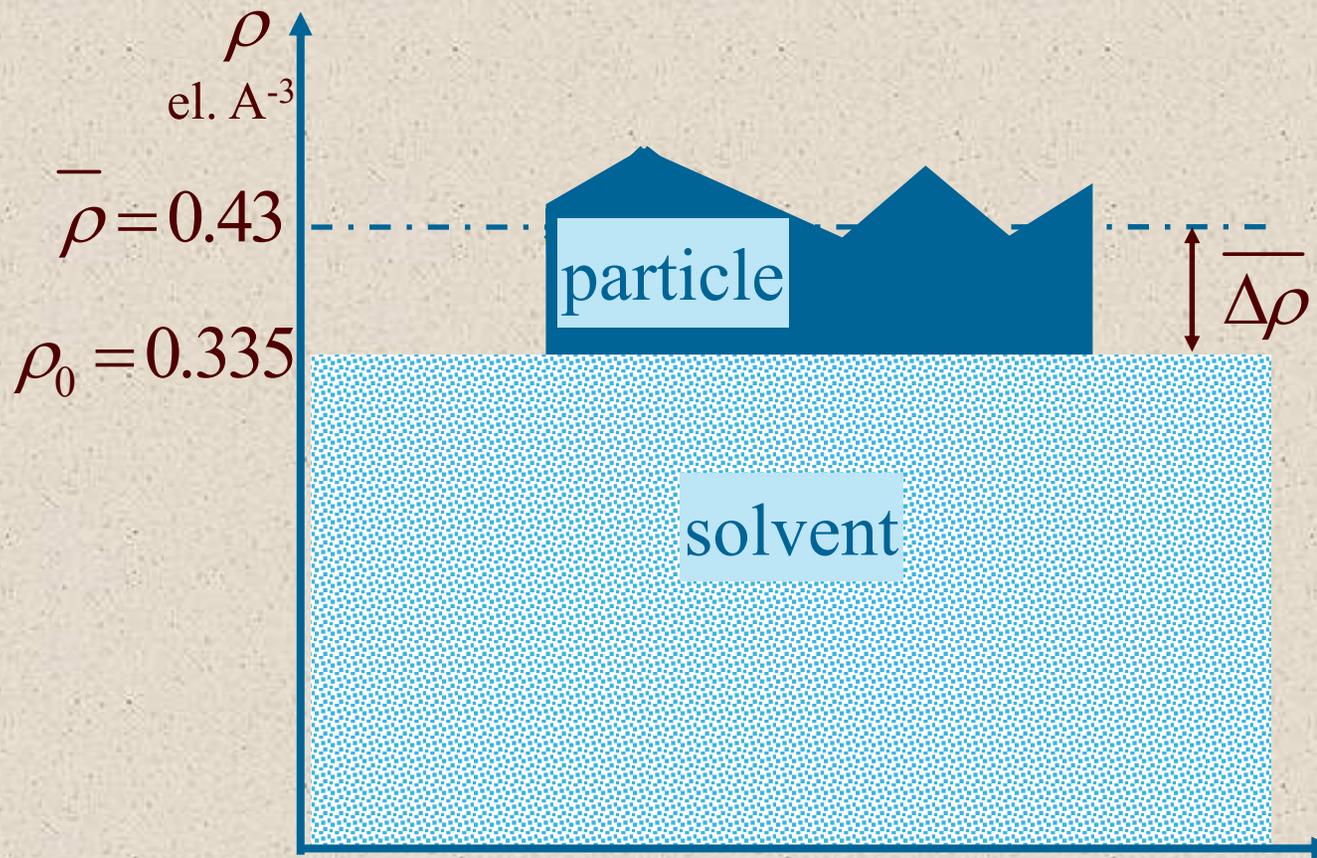
In the equations below we shall always assume that the solvent scattering has already been subtracted

Solution of particles

Solution $\Delta\rho(\mathbf{r})$ $F(c,s)$ = Motif (protein) $\Delta\rho_p(\mathbf{r})$ $F(0,s)$ * . Lattice $d(\mathbf{r})$ $\delta(c,s)$

Solution of particles

For spherically symmetrical particles

$$I(c,s) = I(0,s) \times S(c,s)$$

form factor
of the **particle**

structure factor
of the **solution**

Still valid for globular particles though over a restricted s-range

Solution of particles

- 1 – *monodispersity*: identical particles

- 2 – size and shape polydispersity

- 3 – *ideality* : no intermolecular interactions

- 4 – non ideality : existence of interactions between particles

For most of the following derivations of the structural parameters, we shall make the double assumption 1 and 3

Ideal and monodisperse solution

$$A(\mathbf{s}) = \mathfrak{I}[\varDelta\rho(\mathbf{r})] = \int_V \varDelta\rho(\mathbf{r})\exp(i\mathbf{s}\mathbf{r})d\mathbf{r}$$

Particles in **solution** => thermal motion => particles have random orientations to X-ray beam. The sample is *isotropic.* Therefore, only the *spherical average* of the scattered intensity is experimentally accessible.

Ideality and monodispersity $\mathrm{I}(s) = \mathbf{N}\mathrm{i}_1(s)$

Crystal *versus* **solution**

$$I(c,s) = I(0,s) \times S(c,s)$$

For an ideal crystal, I(**s**) is the three-dimensional scattering intensity from the unit cell

S(**s**) is a sum of δ-functions along the directions of the reciprocal space lattice

$$\mathbf{s} = (h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^*)$$

For an ideal dilute solution, I(**s**)= I(s) is the orientationally averaged intensity of the single particle
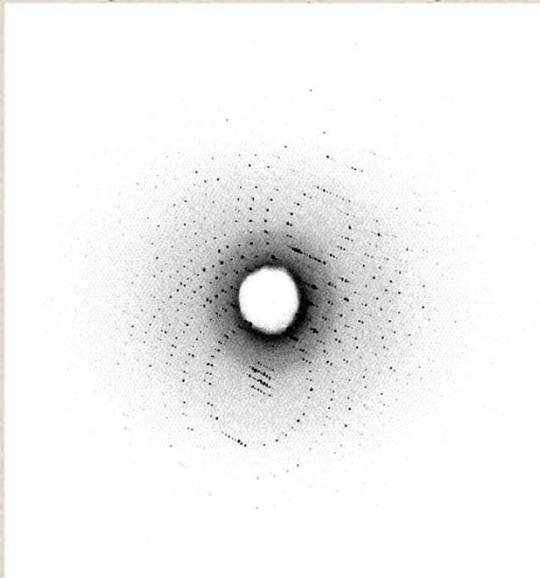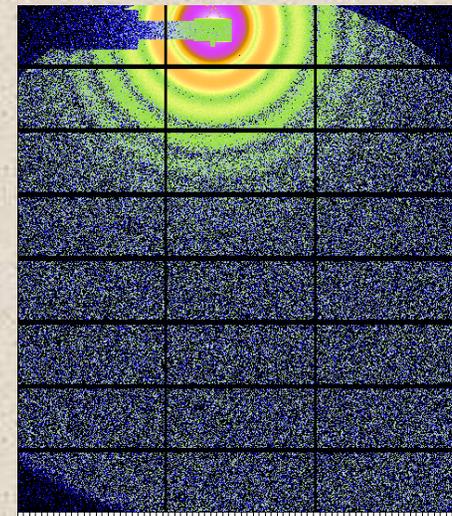
S(s) is equal to unity

Crystal *versus* solution

For an ideal crystal, measured signal is amplified into specific directions allowing measurements to high resolution (d≈λ)

For an ideal dilute solution, I (**s**) is isotropic and concentrates around the primary beam (this is where the name "small-angle scattering" comes from): low resolution (d>>λ).

Main equations and overall parameters

Relation between real and reciprocal space

Using the overall expression for the Fourier transformation one obtains for the spherically averaged single particle intensity

$$I(s) = \left\langle A(\mathbf{s})A^*(\mathbf{s}) \right\rangle_\Omega = \left\langle \int_V \int_V \Delta\rho(\mathbf{r})\Delta\rho(\mathbf{r}')\exp\{i\mathbf{s}(\mathbf{r}-\mathbf{r}')\}d\mathbf{r}d\mathbf{r}' \right\rangle_\Omega$$

or, taking into account that $<exp(isr)>_\Omega = sin(sr)/sr$ and integrating in spherical coordinates,

$$I(s) = 4\pi \int_0^{D_{max}} r^2\gamma(r)\frac{\sin sr}{sr}dr$$

where

$$\gamma(r) = \left\langle \int \Delta\rho(\mathbf{u})\Delta\rho(\mathbf{u}+\mathbf{r})d\mathbf{u} \right\rangle_\omega$$
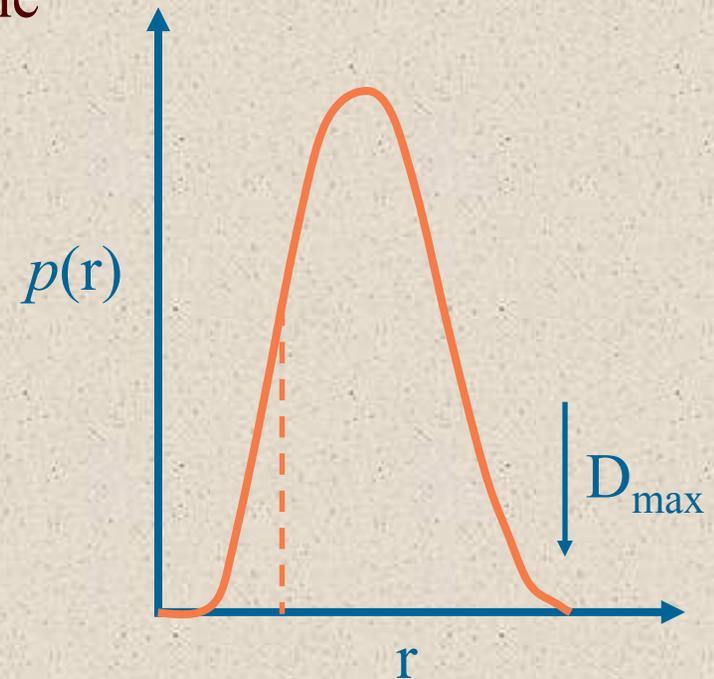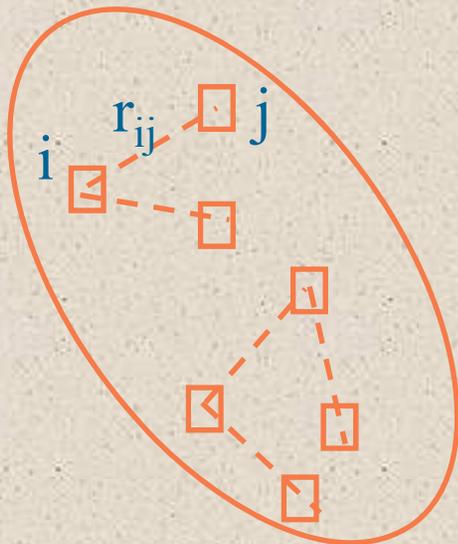
Distance distribution function

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

$\gamma_0(r)$: **probability** of finding a point at r from a given point

number of el. vol. i $\propto$ V - number of el. vol. j $\propto 4\pi r^2$

number of pairs (i,j) separated by the

distance r $\propto 4\pi r^2 V \gamma_0(r) = (4\pi/\rho^2)p(r)$

Debye formula

If the particle is described as a discrete sum of elementary scatterers,(e.g. atoms) with the atomic scattering factors $f_i(s)$ the spherically averaged intensity is
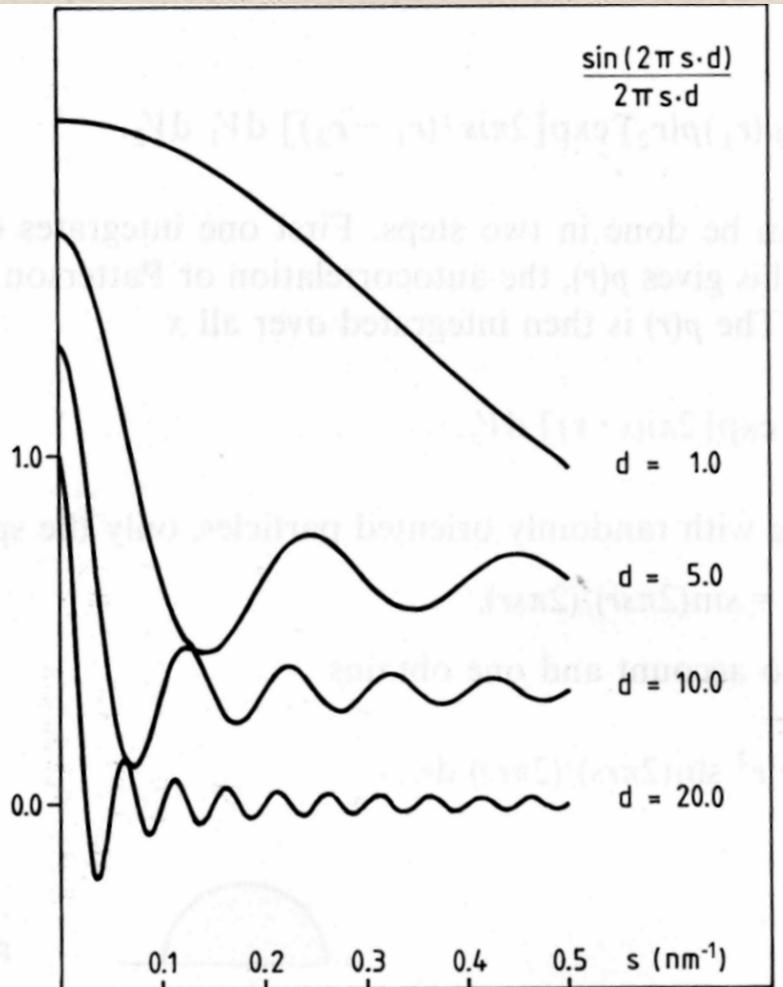
$$I(s) = \sum_{i=1}^{K} \sum_{j=1}^{K} f_i(s) f_j(s) \frac{\sin(sr_{ij})}{sr_{ij}}$$

where $r_{ij} = \left| \mathbf{r}_i - \mathbf{r}_j \right|$

The Debye (1915) formula is widely employed for model calculations

Contribution of distances to the scattering pattern

$$\frac{\sin(2\pi s \cdot d)}{2\pi s \cdot d}$$

d = 1.0

d = 5.0

d = 10.0

d = 20.0

1.0

0.0

0.1 0.2 0.3 0.4 0.5 s (nm⁻¹)

In isotropic systems, each distance $d = r_{ij}$ contributes a *sin(x)/x* –like term to the intensity.
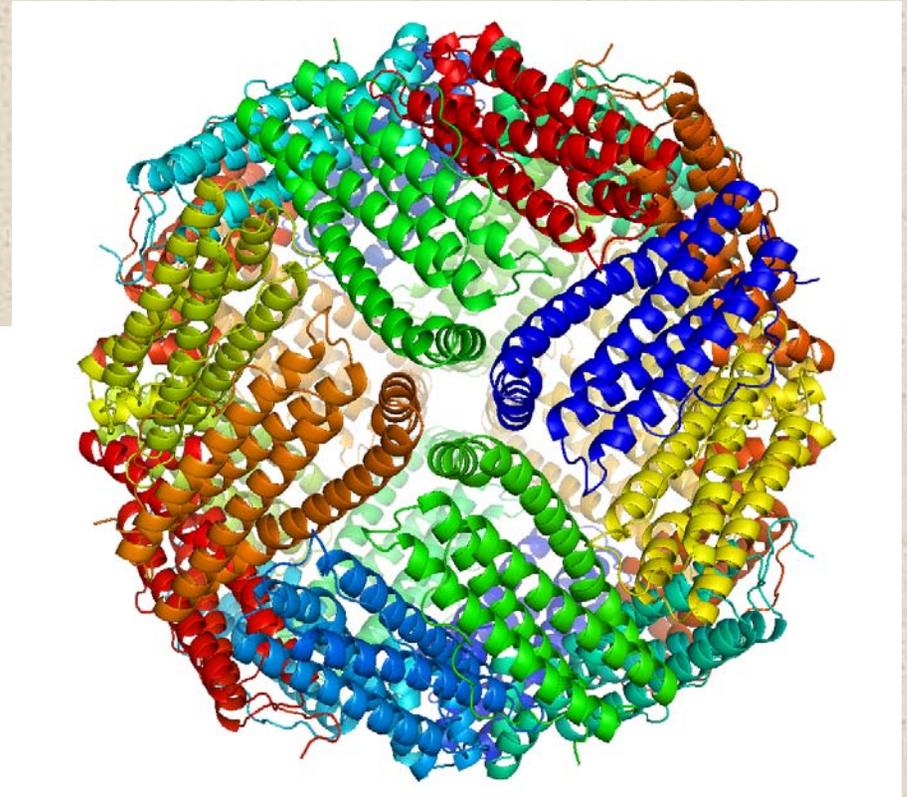
Large distances correspond to high frequencies and only contribute at **low angles** (i.e. at low resolution, where particle shape is seen)

Short distances correspond to low frequencies and contribute over a large angular range.
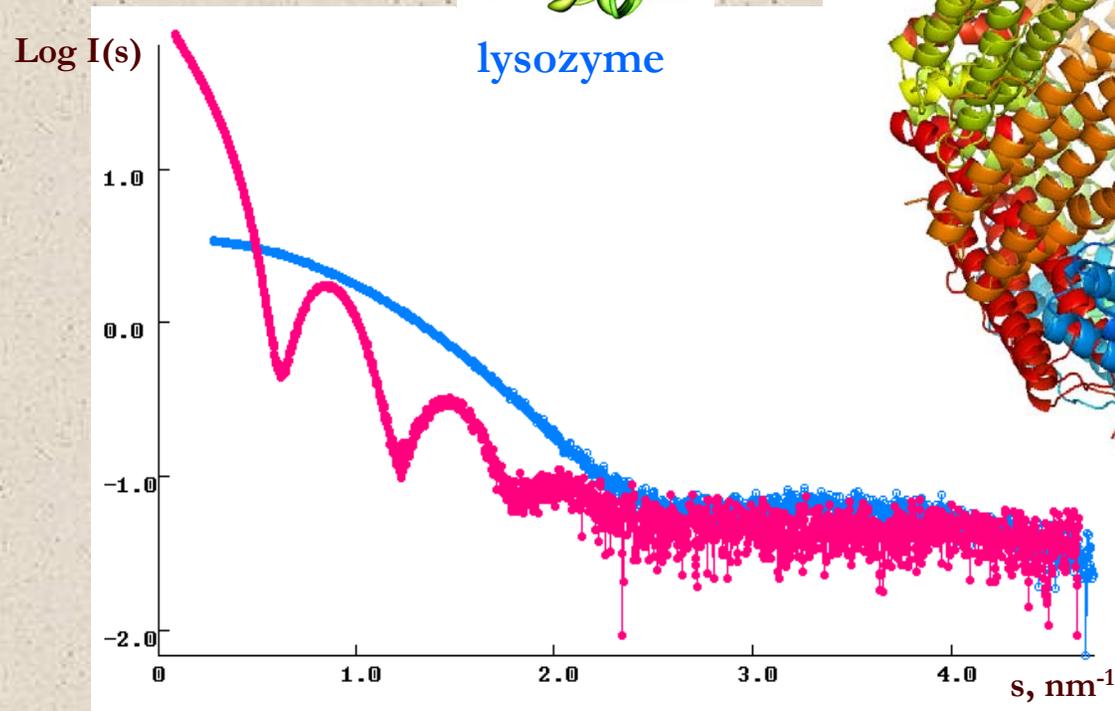Clearly at **high angles** their contribution dominates the scattering pattern.

Small and large proteins: comparison
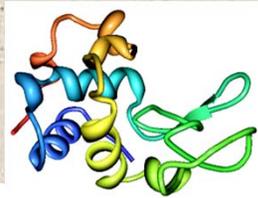
lysozyme

apoferritin

Guinier law

Near *s=0* one can insert the McLaurin expansion
sin(sr)/sr≈1-(sr)²/3+... into the equation for I(s) yielding

$$I(s) = I(0)\left[1 - \frac{1}{3}R_g^2 s^2 + O(s^4)\right] \cong I(0)\exp\left(-\frac{1}{3}R_g^2 s^2\right)$$

This is a classical formula derived by Andre Guinier (1938) in his first SAXS application (to defects in metals). The formula has two parameters, forward scattering and the radius of gyration

$$I(0) = \int_V \int_V \Delta\rho(\mathbf{r})\Delta\rho(\mathbf{r}')d\mathbf{r}d\mathbf{r}' = 4\pi \int_0^{D_{max}} p(r)dr = (\Delta\rho)^2 V^2$$

$$R_g = \int_0^{D_{max}} r^2 p(r)dr \left[2\int_0^{D_{max}} p(r)dr\right]^{-1}$$

ideal
monodisperse

Intensity at the origin

$$i_1(0) = \int_{V_r} \int_{V_{r'}} \Delta\rho(\mathbf{r})\Delta\rho(\mathbf{r'})dV_{\mathbf{r}}dV_{\mathbf{r'}}$$

$$i_1(0) = \Delta m^2 = (m - m_0)^2 = \left[\frac{M}{N_A}\overline{v_P}(\rho - \rho_0)\right]^2$$

$$c = \frac{NM}{N_A V}$$ is the concentration (w/v), e.g. in mg.ml^{-1}

$$I(0) = \frac{cMV}{N_A}\left[\overline{v_P}(\rho - \rho_0)\right]^2$$

ideal
monodisperse

Intensity at the origin

If :   the concentration c (w/v), —
        the partial specific volume v_P,
        the intensity on an absolute scale,
        i.e. the number of incident photons
are known,
Then, the **molecular mass** of the particle can be
determined from the value of the intensity at the origin

In practice, MM can be determined from the data on
relative scale by comparison with I(0) of a reference protein
(e.g. BSA, lysozyme or cytochrom C)

ideal
monodisperse

Radius of gyration

Radius of gyration : $$R_g^2 = \frac{\int_{V_r} \Delta\rho(\mathbf{r})r^2 dV_{\mathbf{r}}}{\int_{V_r} \Delta\rho(\mathbf{r})dV_{\mathbf{r}}}$$

R_g is the quadratic mean of distances to the center of mass weighted by the contrast of electron density.

R_g is an *index of non sphericity*.

For a given volume the smallest R_g is that of a sphere : $$R_g = \sqrt{\frac{3}{5}}R$$
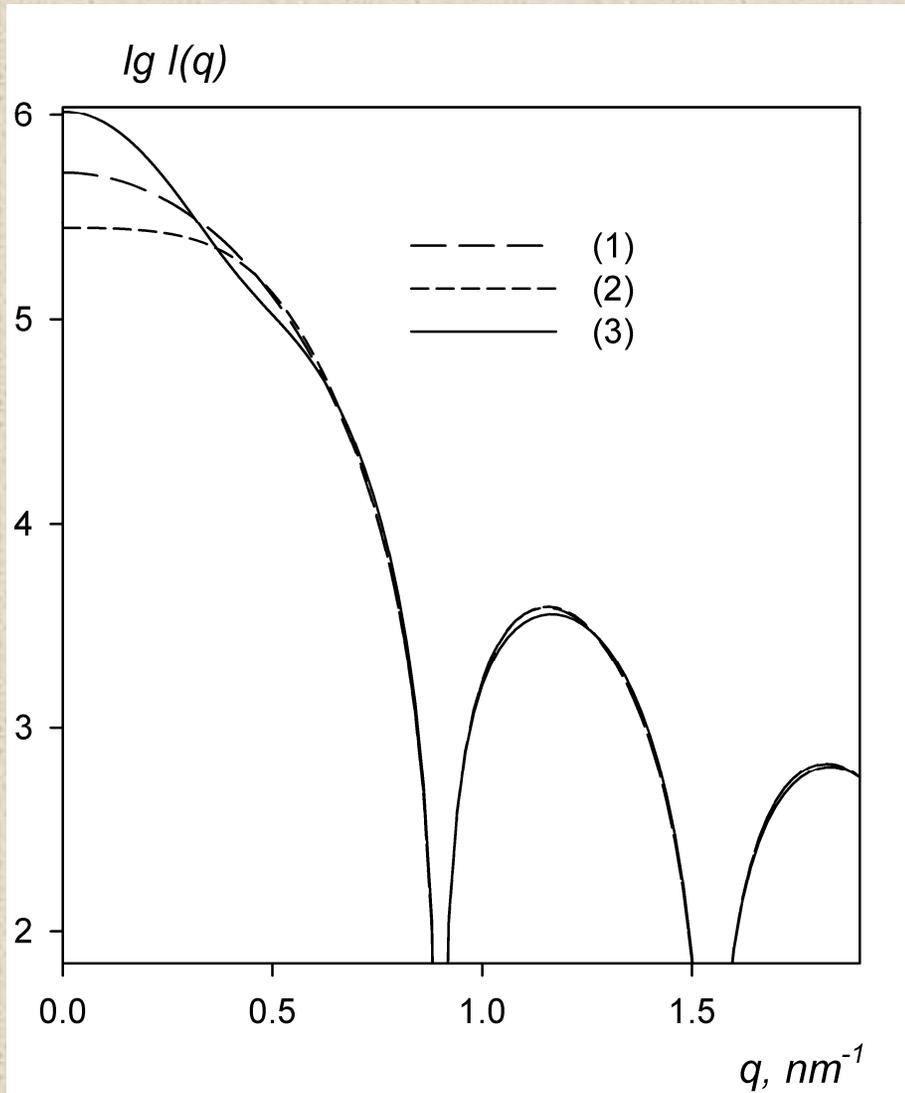
Ellipsoid of revolution (a, b)

$$R_g = \sqrt{\frac{2a^2 + b^2}{5}}$$

Cylinder (D, H)

$$R_g = \sqrt{\frac{D^2}{8} + \frac{H^2}{12}}$$

ideal monodisperse

Repulsion versus attraction

Computed scattering:
(1), from a solid sphere with R = 5 nm,
(2), from a solution of non-interacting hard spheres with $v = 0.2$,
(3), from a dumbbell (divided by a factor of two)

At higher angles, interactions are not visible; at small angles, repulsion diminishes the scattering, attraction increases the scattering

Structure factor

The structure factor at a given concentration c can be obtained from the ratio of the experimental intensity at this concentration to that obtained by extrapolation to infinite dilution or measured at a sufficiently low concentration c_0 were all correlations between particles have vanished:

$$S(q,c) = \frac{c_0 I_{\exp}(q,c)}{c I(q,c_0)}$$

Analytical expression

For non-interacting hard spheres model, based on the statistical-mechanical treatment of hard-sphere fluids (Percus and Yevick 1958), *S(s)* was expressed as an analytical function of two parameters, the sphere radius *R* and the volume fraction *f*, *S(s, R, v)= (1- C(s, R, f))⁻¹* (Ashcroft and Leckner 1966). Here, *C(s, R, f)* is the Fourier transformation of the interparticle correlation function

$$C(s, R, \phi) = -\frac{24\phi}{x^6}\left\{\alpha x^3\left[\sin x - x\cos x\right] + \beta x^2\left[2x\sin x - (x^2 - 2)\cos x - 2\right] + \right.$$

$$\left. \gamma\left[(4x^3 - 24x)\sin x - (x^4 - 12x^2 + 24)\cos x + 24\right]\right\}$$

where *x=2sR*, *α=(1+2f)²/(1-f)⁴*, *β=-6f(1+f/2)²/(1-f)⁴* and *γ=f(1+2f)²/[2(1-f)⁴]*.

Structure factor
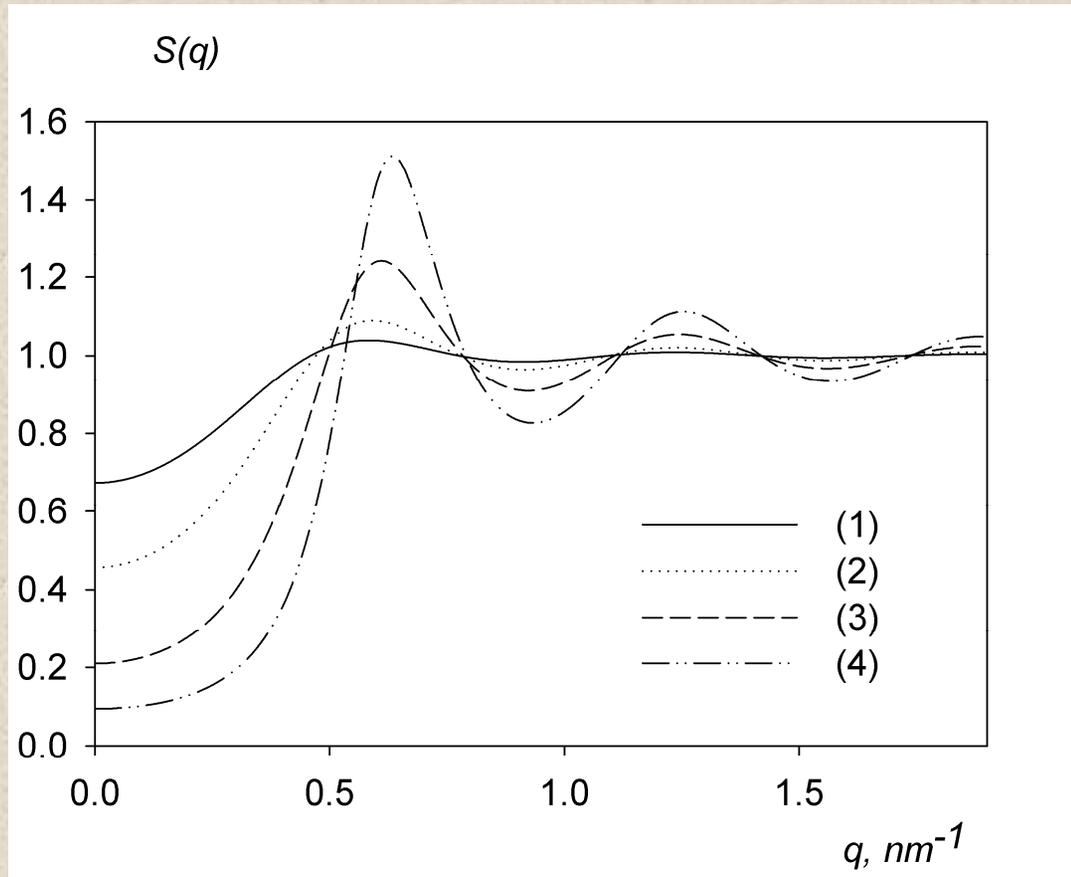
Structure factors computed for systems of hard spheres with the radius 5 nm (curves 1-4 correspond to the volume fractions $v = 0.05, 0.1, 0.2, 0.3$ respectively).

Virial coefficient

In the case of moderate interactions, the intensity at the origin varies with concentration according to :

$$I(0,c) = \frac{I(0)_{ideal}}{1 + 2A_2Mc + ...}$$

Where A_2 is the second virial coefficient which represents pair interactions and $I(0)_{ideal}$ is proportional to c.
A_2 is evaluated by performing experiments at various concentrations c.
A_2 is proportional to the slope of $c/I(0,c)$ vs c.

To obtain $I(0, s)$, this extrapolation to infinite dilution is performed for different angles.

Guinier plot example

The law is generally used under its log form :

$$\ln[I(s)] = \ln[I(0)] - [sR_g]^2 / 3$$

A linear regression yields two parameters : I(0) (y-intercept)
                                      R_g from the slope

Validity range :
0 <sR_g<1.3

ideal monodisperse
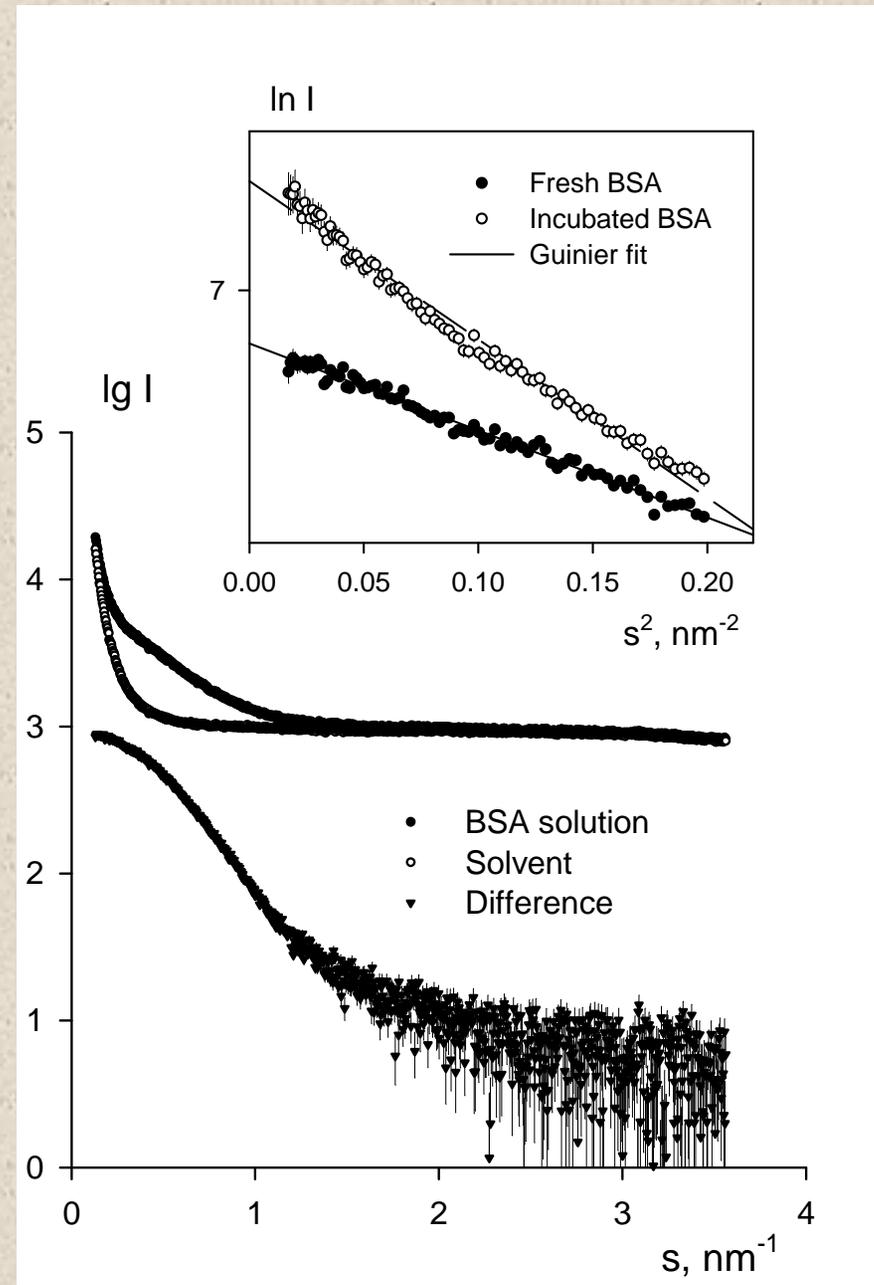
Rods and platelets

In the case of very elongated particles, the radius of gyration of the cross-section can be derived using a similar representation, plotting this time sI(s) *vs* s²

$$sI(s) \cong I_C(0)\exp\left(-\frac{1}{2}R_c^2 s^2\right)$$

In the case of a platelet, a thickness parameter is derived from a plot of s²I(s) *vs* s² :

$$s^2 I(s) \cong I_T(0)\exp\left(-R_t^2 s^2\right)$$

with $R_t = T/\sqrt{12}$ T : thickness

ideal
monodisperse

Distance distribution function

$$p(r) = \frac{r^2}{2\pi^2} \int_0^\infty s^2 I(s) \frac{\sin sr}{sr} \, dr$$

In theory, calculation of p(r) from I(s) is simple.

Problem : I(s) - is only known over [s_{min}, s_{max}] : truncation

- is affected by experimental errors and possible instrumental distorsions due to the beam-size and the bandwidth $\Delta\lambda/\lambda$ (neutrons)

$\Rightarrow$ Fourier transform of ***incomplete and noisy data*** is an ***ill-posed problem***.

Solution : Indirect Fourier Transform (suggested by O. Glatter, 1977).

p(r) is parameterized on [0, D_{max}] by a linear combination of orthogonal functions, where D_{max} is the particle diameter.

Implemented in several programs, including GNOM (part of ATSAS)

ideal monodisperse

Distance distribution function

The radius of gyration and the intensity at the origin are derived from p(r) using the following expressions :

$$R_g^2 = \frac{\int_0^{D_{max}} r^2 p(r)dr}{2\int_0^{D_{max}} p(r)dr} \quad \text{and} \quad I(0) = 4\pi \int_0^{D_{max}} p(r)dr$$

This alternative estimate of R_g makes use of the whole scattering curve, and is much less sensitive to interactions or to the presence of a small fraction of oligomers.

Comparison of both estimates : useful cross-check

ideal
monodisperse

Porod invariant and volume

Following the Parseval theorem for Fourier transformations

$$Q = \int_0^\infty s^2 I(s)\,ds = 2\pi^2 \int_V (\Delta\rho(\mathbf{r}))^2\,d\mathbf{r}$$

Q is called the Porod invariant, which is computed from the intensity but provides the mean square electron density contrast.

For homogeneous particles, $Q=2\pi^2(\Delta\rho)^2 V$, and, taking into account that $I(0)=(\Delta\rho)^2 V^2$, the excluded (Porod) volume of hydrated particle in solution (Porod, 1952) is

ideal
monodisperse

$$\mathbf{V=2\pi^2 I(0)/Q} \ .$$

The asymptotic regime : Porod law

Integrating the Fourier transformation for I(s) by parts and using that for particles with a *sharp interface $\gamma'(D_{max})$ = 0*, one has

$$I(s) \cong 8\pi s^{-4}\gamma'(0) + O_1 s^{-3} + O_2 s^{-4} + o(s^{-5})$$

where O_1, O_2 are oscillating trigonometric terms of the form *sin (sD_{max})*. The main term responsible for the intensity decay at high angles is therefore proportional to s^{-4}, and this is known as Porod's law (1949). For homogeneous particles, $\gamma'(0)$ is equal to $-(\Delta\rho)^2 S/4$, where *S* is the particle surface.

ideal monodisperse

Correlation volume

One can also calculate 'correlation volume' using the (better converging) integral from the intensity weighted with s, not s^2

$$V_c = \frac{I(0)}{\int_0^\infty sI(s)ds} = \frac{V_p}{2\pi l_c}$$

where l_c is the particle 'correlation length'. This quantity has no direct physical meaning (having units of surface). Rambo & Tainer (2013) showed that the ratio $(V_c^2)/R_g$ can be used to estimate molecular weight of particles from an empirical power law dependence.
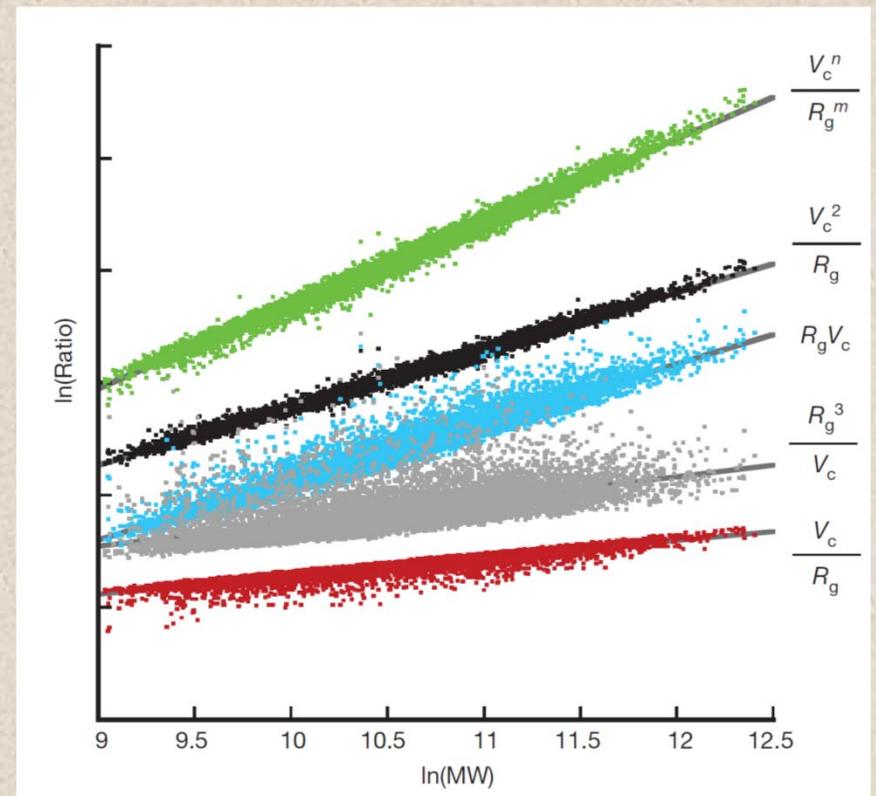
Kratky plot

A plot of $s^2I(s)$ vs s provides a sensitive means of *monitoring the degree of compactness* of a protein.

Globular particle : bell-shaped curve

Unfolded particle: plateau or increase at large s-values
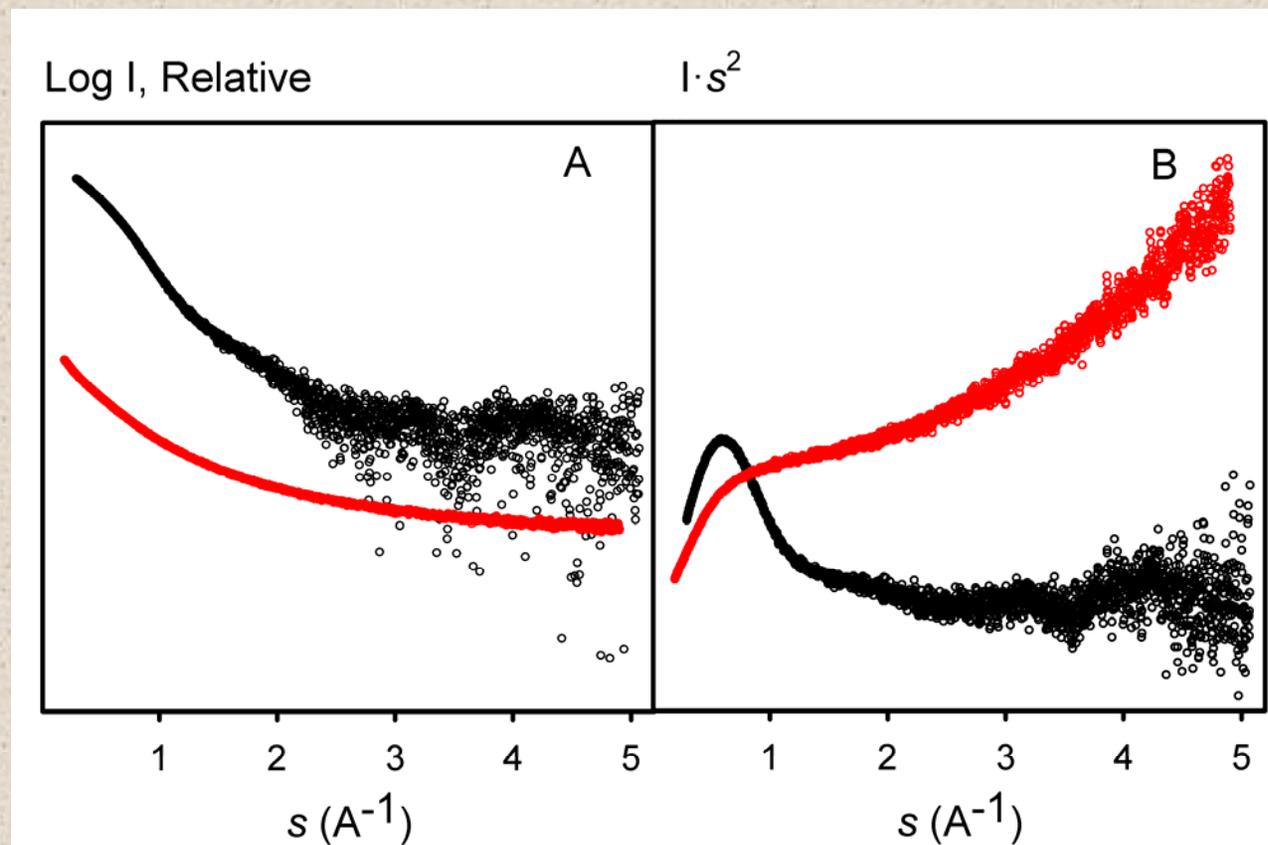
Dimensionless Kratky plot
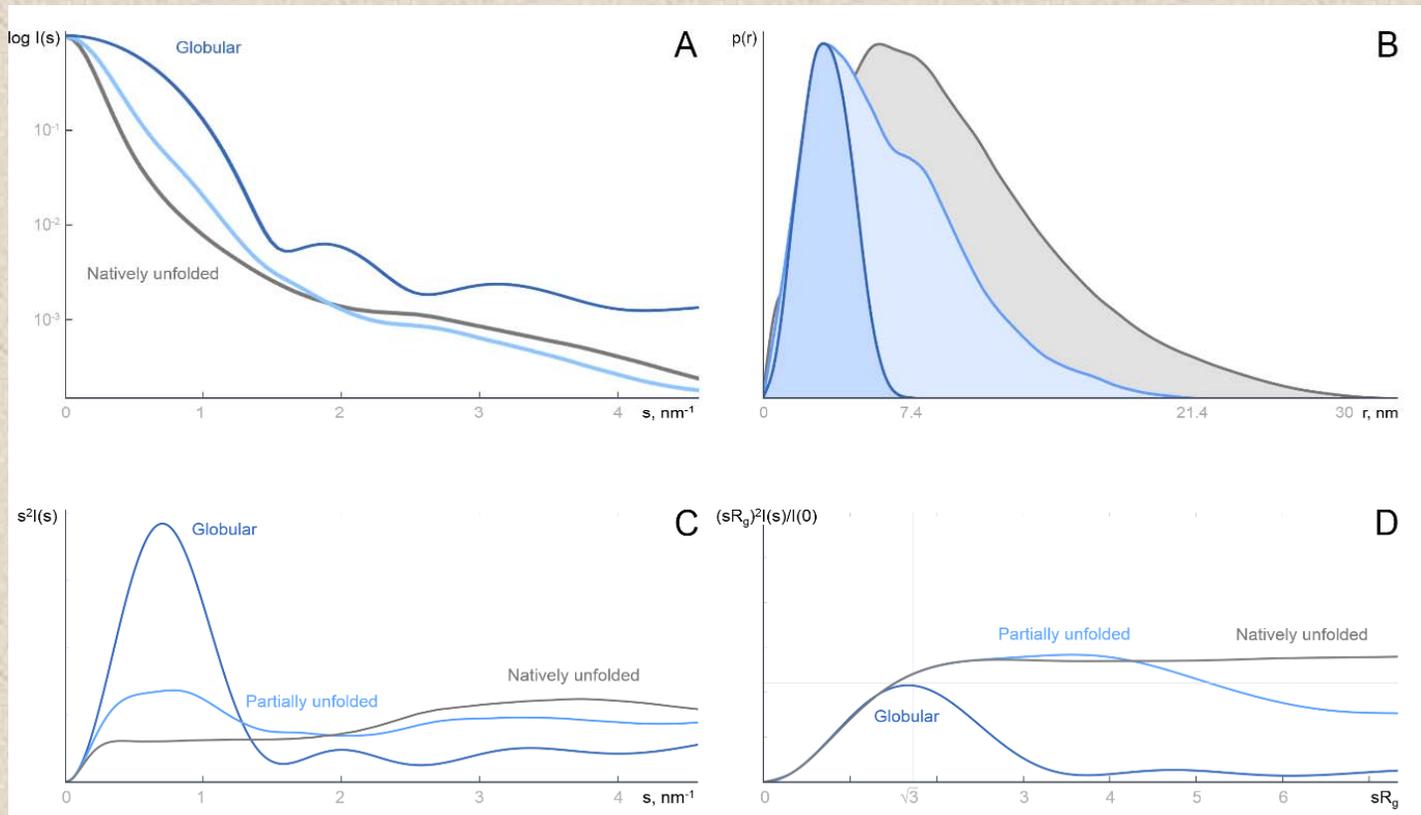
Durand D, Vivès C, Cannella D, Pérez J, Pebay-Peyroula E, Vachette P, Fieschi F. (2010) J Struct Biol. **169**: 45-53.

Scattering from three 60 kDa proteins: globular (dark blue), half- (light blue) and fully disordered (gray). A, B: scattering intensity I(s) and p(r) functions in (inverse) nanometres. C: Kratky plot $s^2 I(s)$ *vs.* s. D: Normalized (or "dimensionless") Kratky plot $(s R_g)^2 I(s)/I(0)$ *vs.* $s R_g$.

Summary of model-independent information

$I(0)/c$, i.e. molecular mass *(from Guinier plot or p(r) function)*

Radius of gyration R_g *(from Guinier plot or p(r) function)*

R_g of thickness or cross-section *(anisometrc particles)*

Second virial coefficient A_2 *(extrapolation to infinite dilution)*

Maximum particle size D_{max} *(from p(r) function)*

Particle volume V *(from I(0) and Porod invariant)*

Molecular mass estimate from correlation volume V_c *(empirical)*

Specific surface S/V *(from I(0), Porod invariant and asymptotics)*

Globularity/ unfodedness *(from (dimensionless) Kratky plot)*

Small-angle scattering: experiment
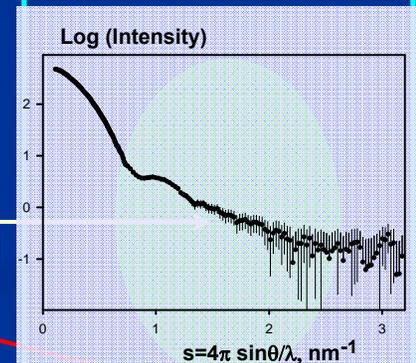
Detector

Sample

Monochromatic beam

Wave vector k, $k = 2\pi/\lambda$

Log (Intensity)

$s = 4\pi \sin\theta/\lambda$, nm^{-1}

2θ

k_1

Radiation sources:

X-ray generator (λ = 0.1 - 0.2 nm)
Synchrotron (λ = 0.03 - 0.35 nm)
Thermal neutrons (λ = 0.2 - 1 nm)

Scattering vector $s = k_1 - k$,
$s = 4\pi \sin\theta/\lambda$

Crystal versus solution

Crystal versus solution

- Thousands of reflections
- 3D, high resolution
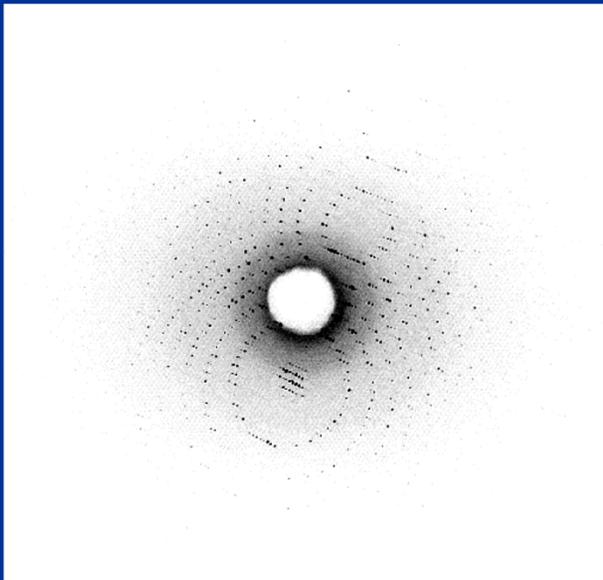
- A few Shannon channels
- 1D, low resolution

- Data undersampled,
  $\Delta s = 2\pi / D$
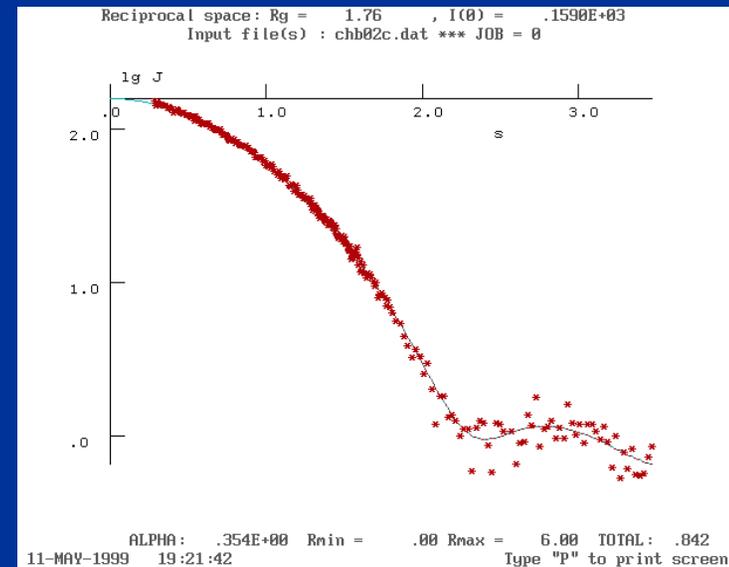
- Data oversampled,
  $\Delta s << \pi / D$

Crystal *versus* solution

- For SAXS solution studies, one does not need to grow crystals
- SAXS is not limited by molecular mass and is applicable under nearly physiological conditions
- Using solution SAXS, one can more easily observe responses to changes in conditions
- SAXS permits for quantitative analysis of complex systems and processes

- In solution, no crystallographic packing forces are present

Scattering from dilute macromolecular solutions (monodisperse systems)

$$I(s) = 4\pi \int_0^D p(r)\frac{\sin sr}{sr}dr$$

The scattering is proportional to that of a single particle averaged over all orientations, which allows one to determine size, shape and internal structure of the particle at low (*1-10 nm*) resolution.

Example of strong attractive interactions
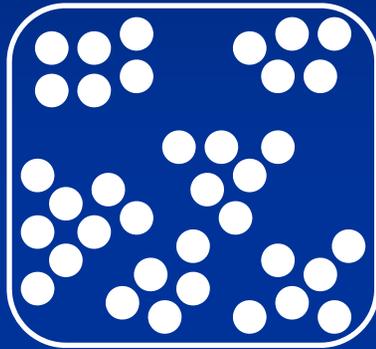
γ-crystallins c=160 mg/ml
in 50mM Phosphate pH 7.0

A. Tardieu et al., LMCP (Paris)
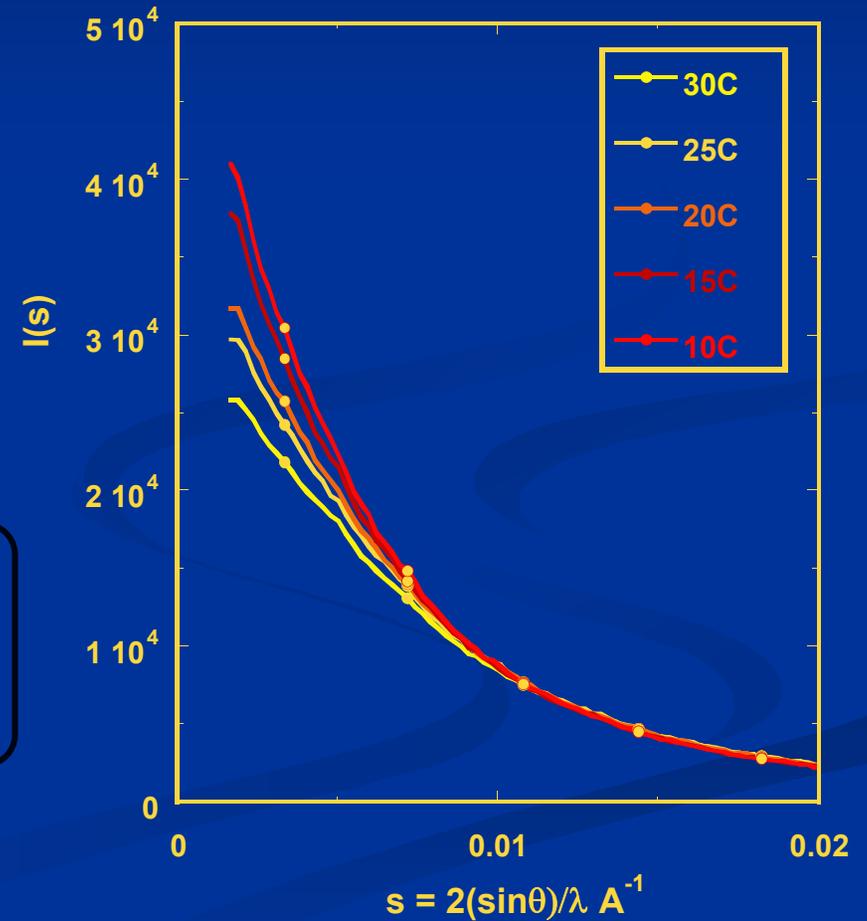
Example of strong repulsive interactions
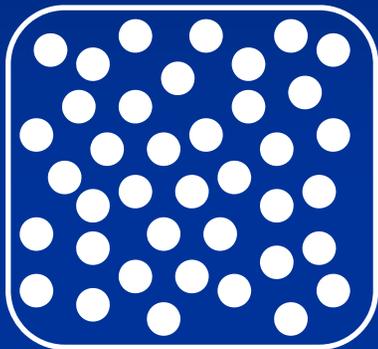
ATCase in
10mM borate buffer pH 8.3

P.Vachette et al., CNRS-Universite Paris-sud, Orsay
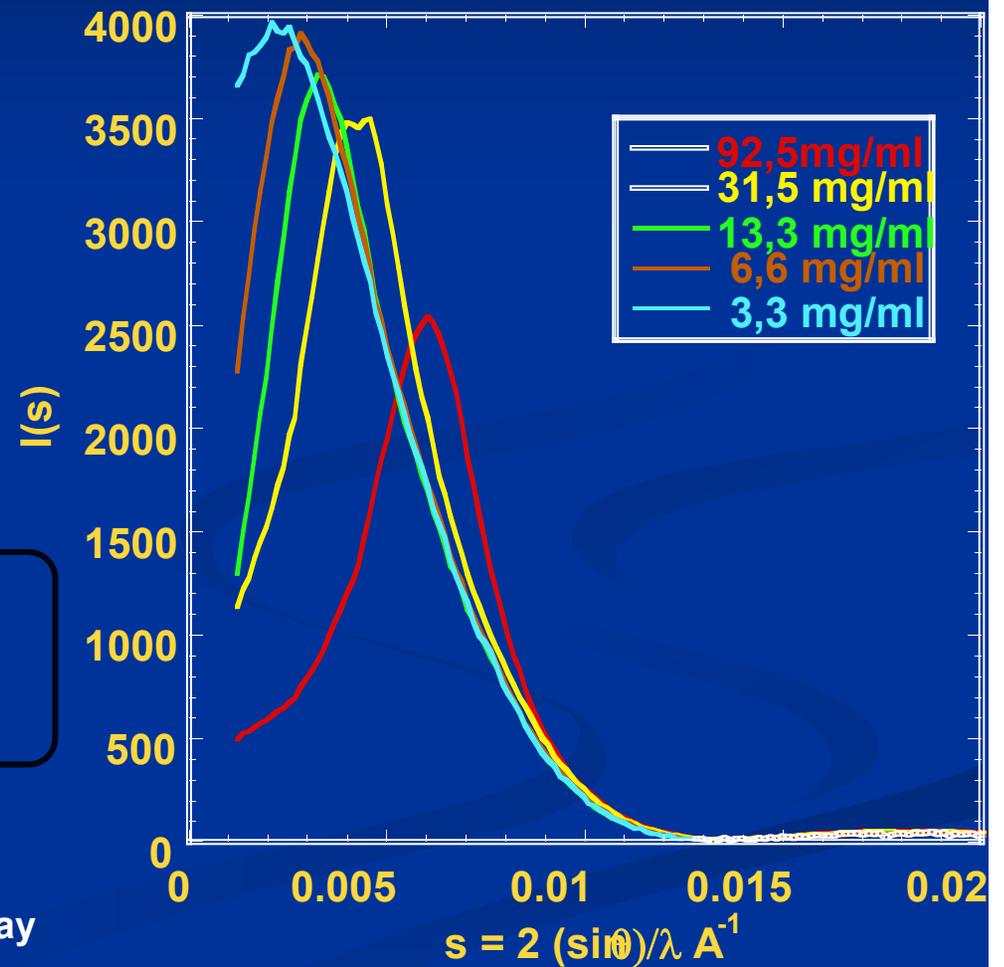
In dilute solutions, scattering is related to the shape (or low resolution structure)
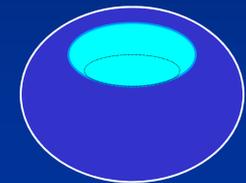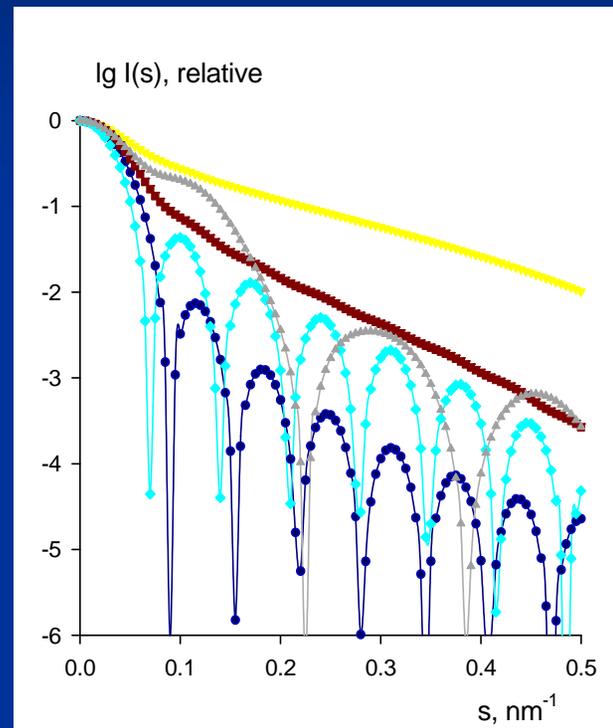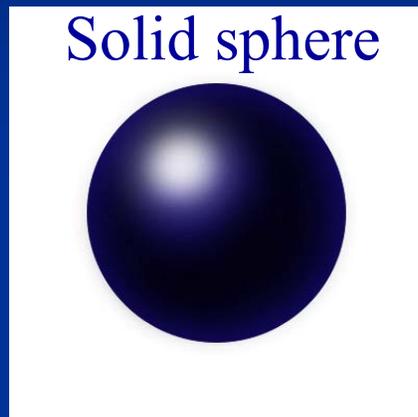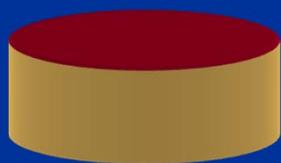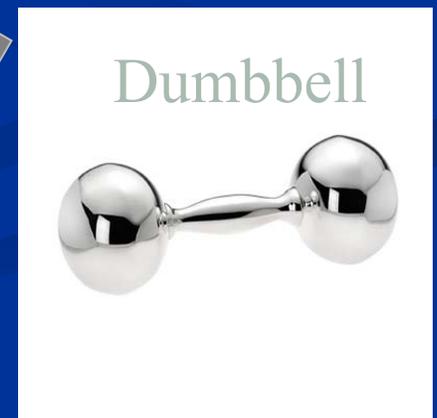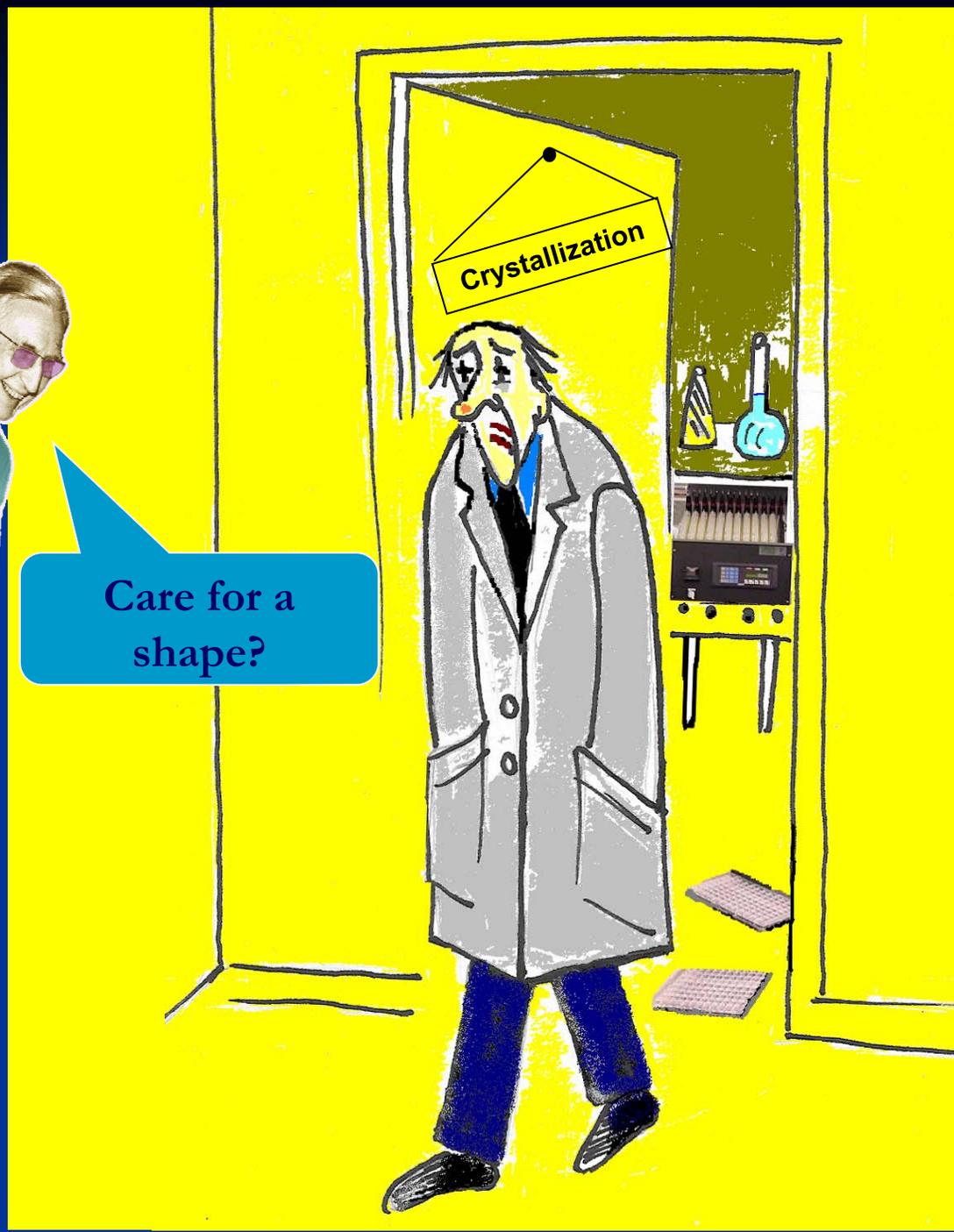
Methods development at EMBL-Hamburg

Data processing and manipulations

Rigid body refinement

Employed by over
14,000 users worldwide

Ab initio modeling suite

Analysis of mixtures
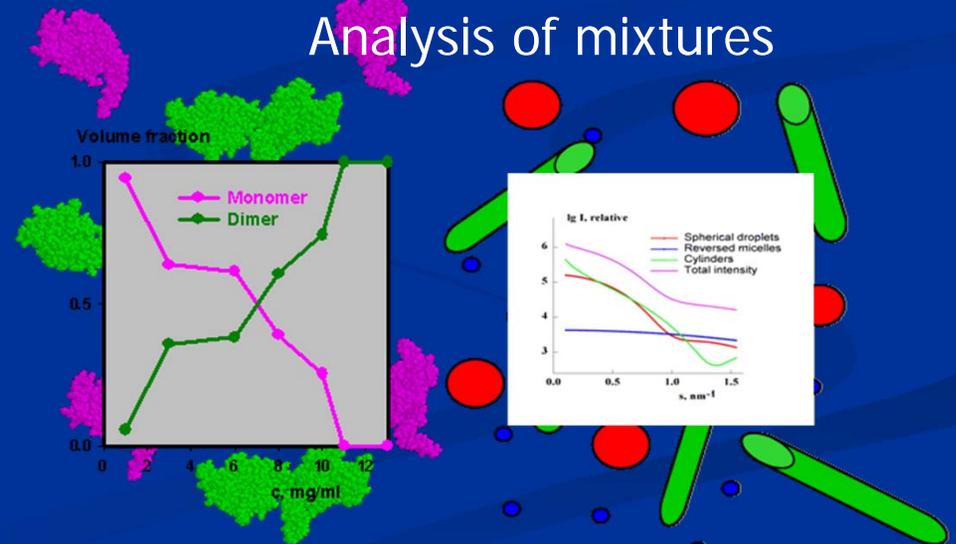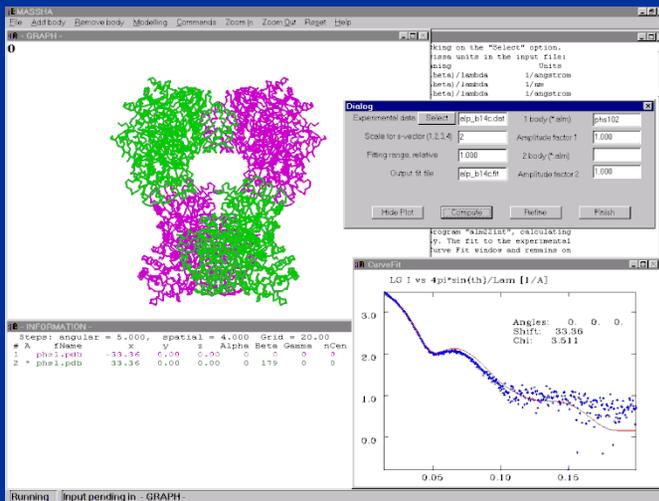
Franke D, Petoukhov MV, Konarev PV, Panjkovich A, Tuukkanen A, Mertens HDT, Kikhney AG, Hajizadeh NR, Franklin JM, Jeffries CM, Svergun DI (2017). *J. Appl. Cryst.* **50**, 1212
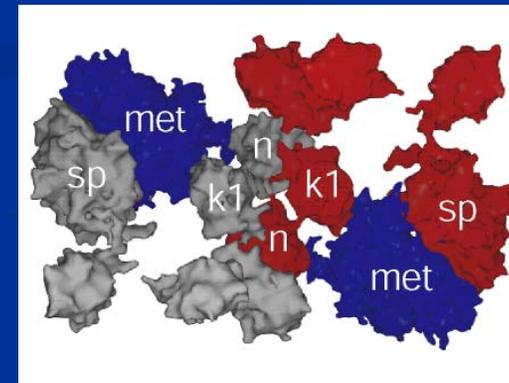
"Simple" monodisperse systems

Shape and conformational changes
of macromolecules and complexes

Validation of high resolution models
and oligomeric organization

Rigid body models of complexes
using high resolution structures

Addition of missing fragments to high
resolution models

Scattering from mixtures

$$I(s) = \sum_k v_k I_k(s)$$

The scattering is proportional to that of a single particle averaged over all orientations, which allows one to determine size, shape and internal structure of the particle at low (*1-10 nm*) resolution.

For equilibrium and non-equilibrium mixtures, solution scattering permits to determine the number of components and, given their scattering intensities $I_k(s)$, also the volume fractions
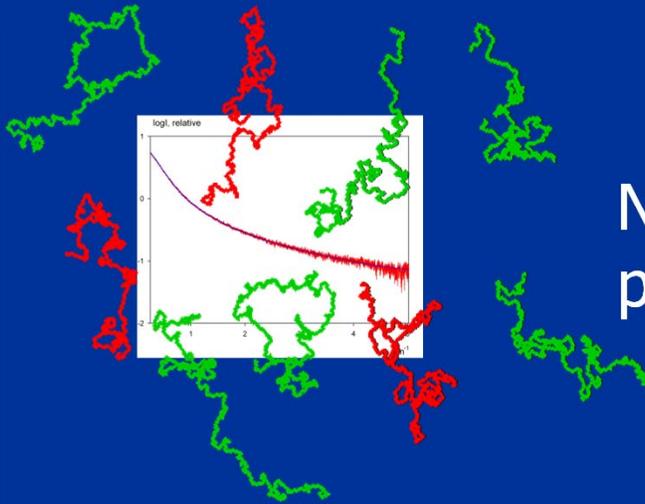
Complicated systems: mixtures and processes

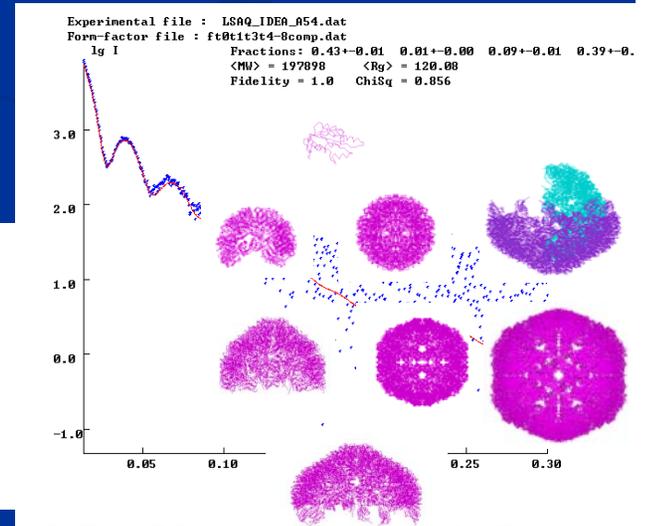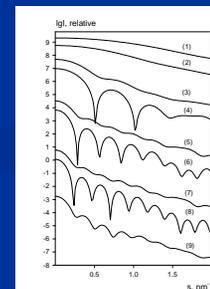Equilibrium oligomeric mixtures

Stoichiometry and complex formation
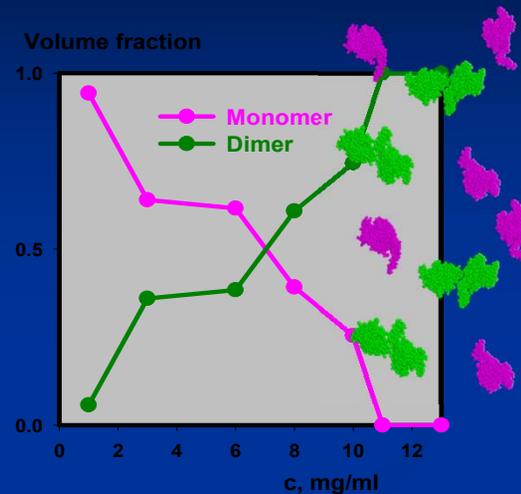
Natively unfolded proteins and multidomain proteins with flexible linkers

Protein folding/unfolding kinetics

Assembly/disassembly processes

A roadmap of biological SAS data analysis
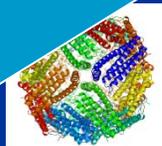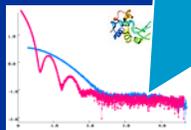
ATSAS 2.8.3 tutorials, practical sessions, SASQuest and remote data collection at the P12 beamline in Hamburg by D.Franke, C.Jeffries, H.Mertens and A.Panjkovich
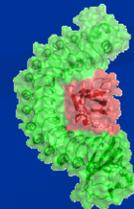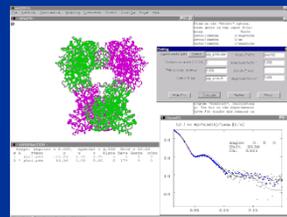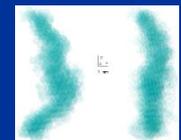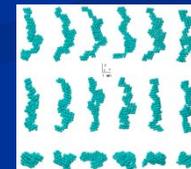
A word of caution

- Sample preparation
- Experiment
- Data processing
- Unambiguous interpretation
- Changing conditions
- Relation to function

Books on SAXS

" The origins" (no recent edition) : Small Angle Scattering of X-rays. A. Guinier and A. Fournet, (1955), in English, ed. Wiley, NY

Small-Angle X-ray Scattering: O. Glatter and O. Kratky (1982), Academic Press. PDF available on the Internet at http://physchem.kfunigraz.ac.at/sm/Software.htm

Structure Analysis by Small Angle X-ray and Neutron Scattering. L.A. Feigin and D.I. Svergun (1987), Plenum Press.
PDF available on the Internet at http://www.embl-hamburg.de/ExternalInfo/Research/Sax/reprints/feigin_svergun_1987.pdf

Small Angle X-Ray and Neutron Scattering from Solutions of Biological Macromolecules. D.I,Svergun, M.H.J. Koch, P.A.Timmins, R.P. May (2013) Oxford University Press, London.

Recent reviews on solution SAS

Tuukkanen A, Spilotros A, Svergun DI (2017) Progress in small-angle scattering from biological solutions at high brilliance synchrotrons. *IUCrJ,* **4**, 518-528.

Mertens HDT, Svergun DI (2017) Combining NMR and Small Angle X-ray Scattering for the study of biomolecular structure and dynamics. *Arch Biochem Biophys*, **628,** 33-41.

Trewhella J (2016) Small-angle scattering and 3D structure interpretation. *Curr Opin Struct Biol.* **40**,1-7.

Vestergaard, B. (2016) Analysis of biostructural changes, dynamics, and interactions – Small-angle X-ray scattering to the rescue. *Arch Biochem Biophys*, **602**, 69–79.

Kikhney AG, Svergun DI (2015) A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. *FEBS Lett.* **589**, 2570-2577

Byron O, Vestergaard B. (2015) Protein-protein interactions: a supra-structural phenomenon demanding trans-disciplinary biophysical approaches. *Curr Opin Struct Biol,* **35**, 76–86