Solution X-ray Scattering from Biological Macromolecules

Javier Pérez
Beamline SWING, Synchrotron SOLEIL, Saint-Aubin, France

General Outline

- Introduction
- SAXS basics
- Biophysical information
- A few experimental considerations
- Modelling
- Conclusions

INTRODUCTION

Principles of Small Angle X-ray Scattering in solution

I_0, λ

Sample
C > 0.1 mg/ml
V > 6 µl

Detected : $I(\vec{Q})$

I_T

2D detector

k_f

2θ

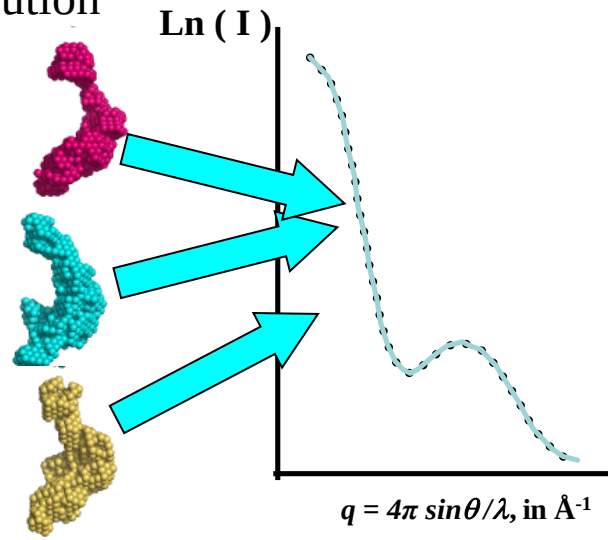
$\vec{q} = \vec{k}_f - \vec{k}_i$

$\vec{k}_i = \frac{2\pi}{\lambda}\vec{e}_i$

Radial average
(isotropic sample)

SAXS provides structural information about macromolecules in solution

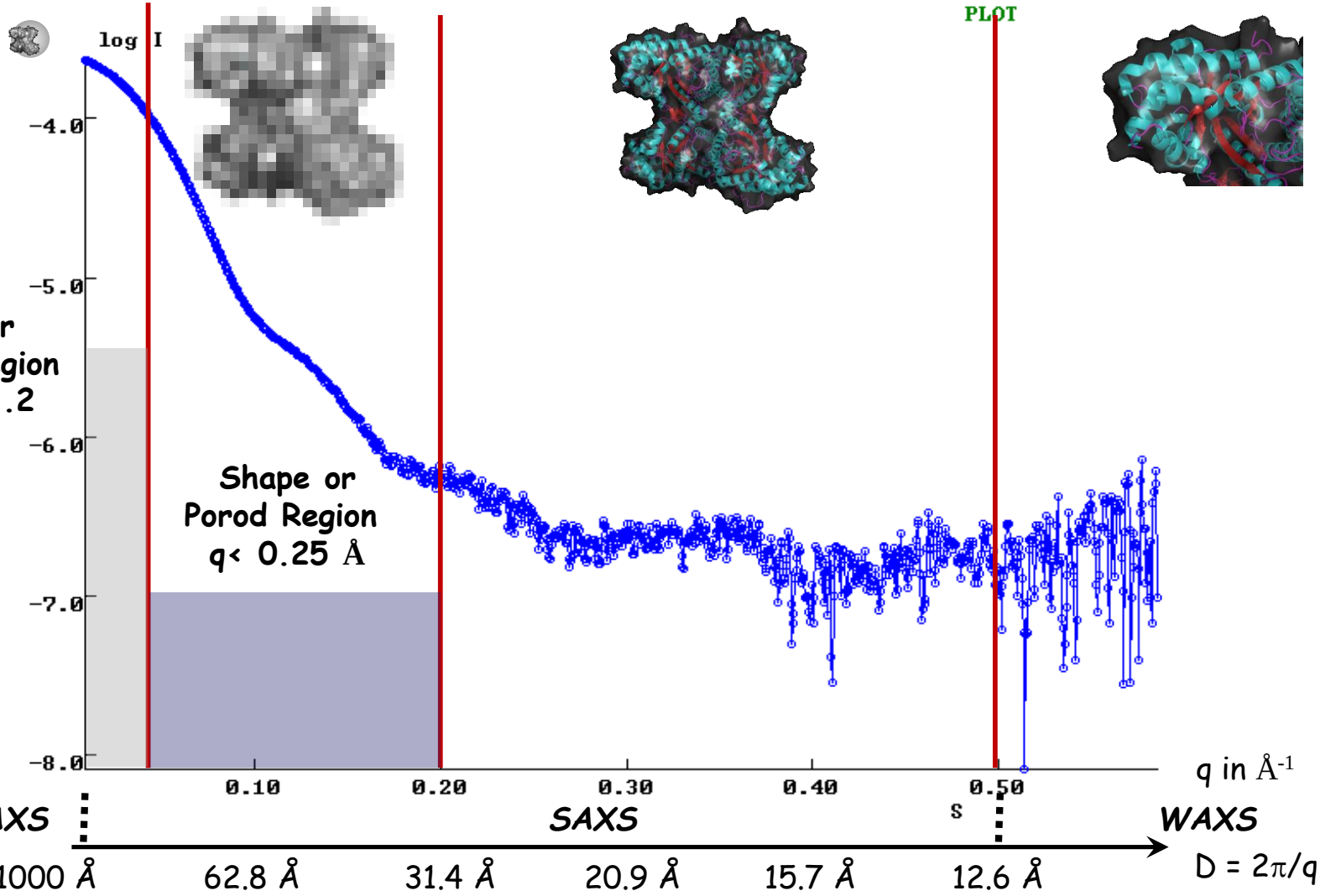
- **Limits**
  - spherically averaged information → low resolution
  - non unicity of the solution
  - does not distinguish elements in a mixture
- **Advantages**
  - solution (no crystal) → kinetics, titration, T°, P
  - relatively easy to carry experiments
  - can be checked against atomic models



SAXS is at its best when complementary (structural) information is available

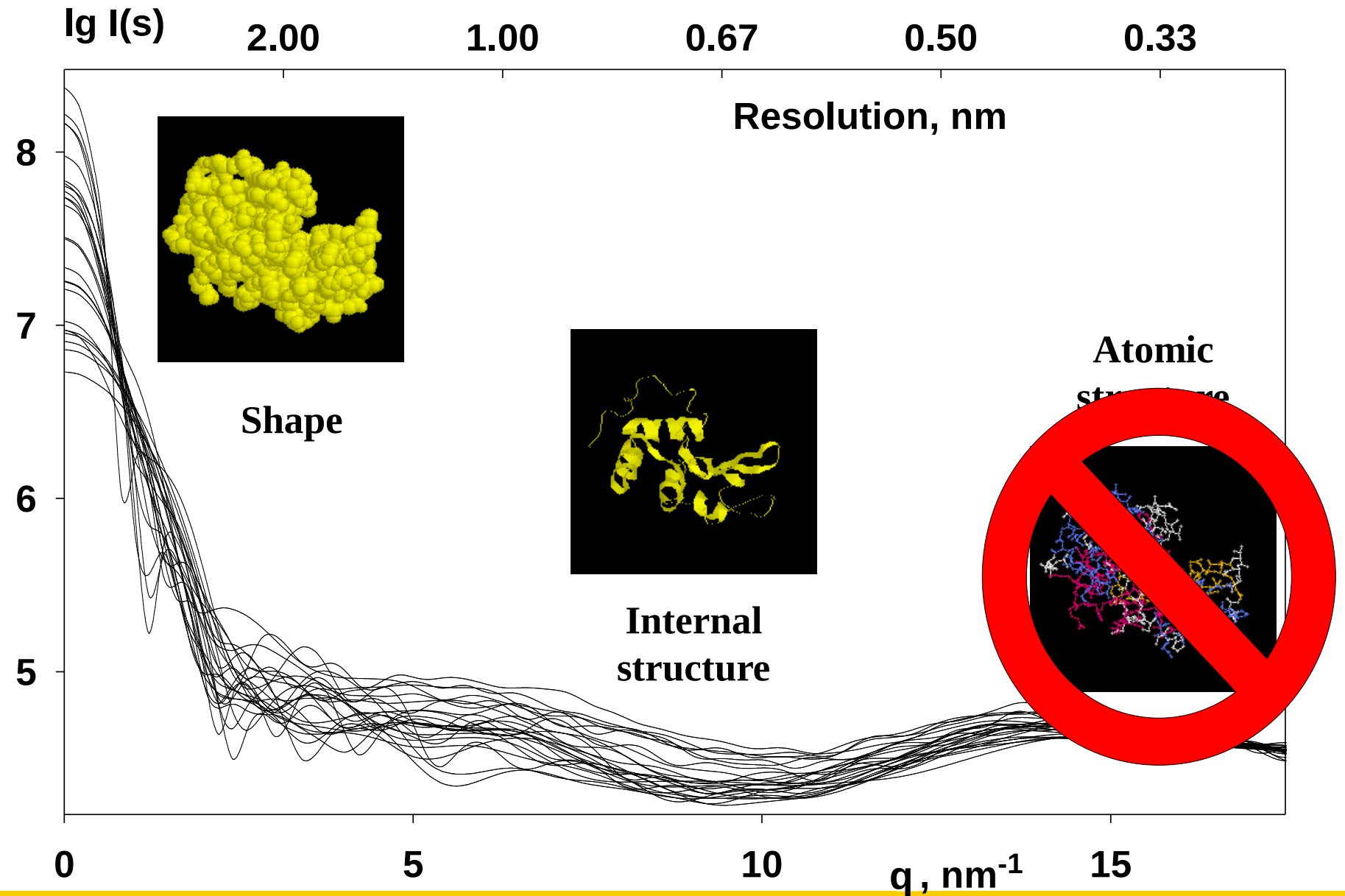
A typical BioSAXS curve from a protein

Global dimension → Global shape → Folding + domains → Secondary structures



Slide from Dmitri Svergun, EMBL Hamburg

What may solution scattering yield?



BioSAXS data output

Data analysis

Guinier range analysis
- Rg (size) and I(0) (mass and oligomeric state)

Pair distribution function p(r) :
- Dmax evaluation
- Rg (size) and I(0) compatibility with Guinier approximation
- Global shape of the object

Kratky plot
- type of structure (globular, elongated or unfolded)

Porod law
- molecular volume if globular protein

Molecular modeling

Cristallographic , NMR structures or complete molecular modeling, ***AlphaFold models***
- theorical curves calculation and data comparison

Nothing is known
- low resolution shape

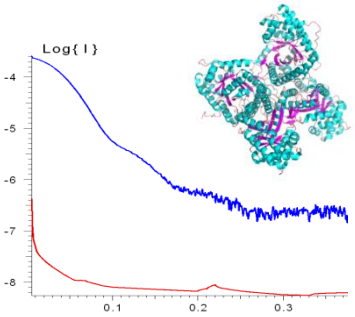
Structures of subunits available
- molecular modeling rigid body against SAXS data

Structures with missing loop or flexible parts
- molecular modeling of missing parts against SAXS data

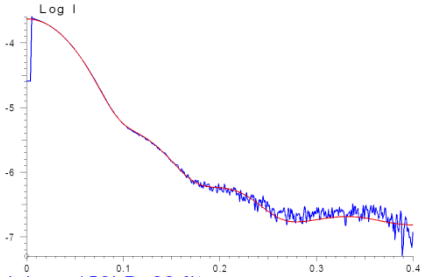
→ « data compatible » models: NOT unique, NOT electronic density map

BioSAXS data modeling, possible programs

1) Theorical model or complete atomic structure available

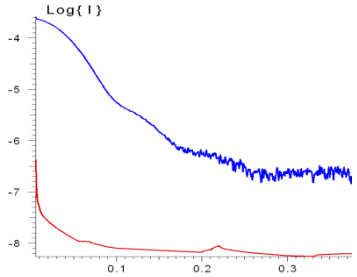


Validation/identification in solution

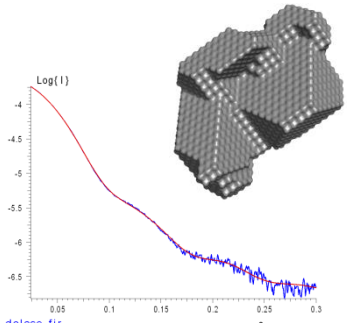


CRYSOL
FOXS
PepsiSAXS
WAXSiS

2) Nothing known (except the curve)

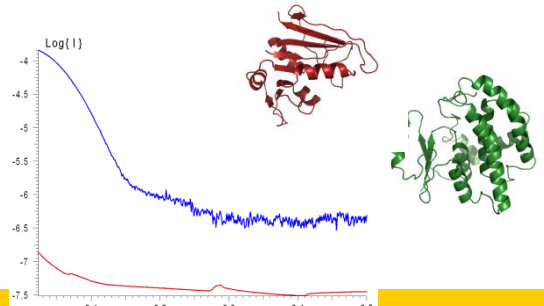


Low resolution model



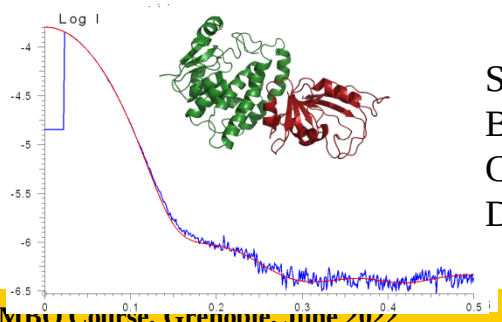
DAMMIN
DAMMIF
GASBOR
MONSA
DENFERT

3) Structures of subunits available



Rigid body modeling of the complex and

molecular modeling of the missing part



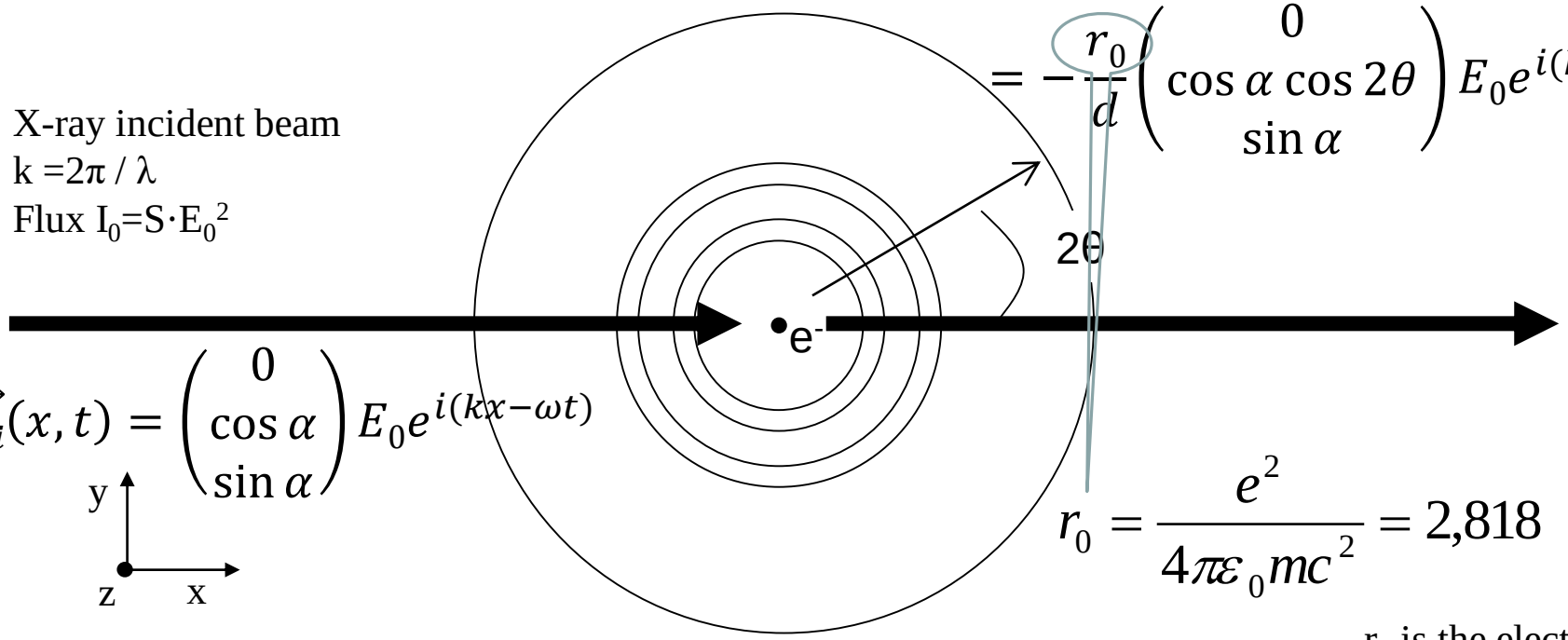
SASREF
BUNCH
CORAL
DADIMODO

SAXS BASICS

Elastic Thompson scattering by an electron

- **What scatter X-rays are <u>the electrons</u>**

X-ray incident beam
k =2π / λ
Flux $I_0=S\cdot E_0^2$



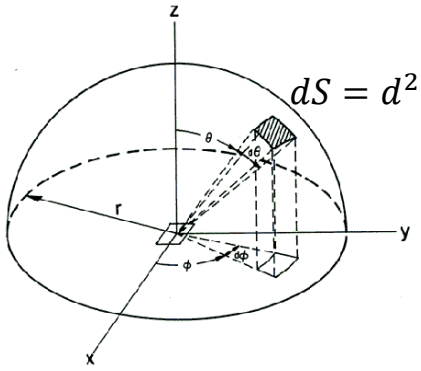
$$\vec{E_s}(d,t) = -\frac{r_0}{d}\begin{pmatrix}0\\ \cos\alpha\cos 2\theta\\ \sin\alpha\end{pmatrix}E_0 e^{i(kd-\omega t)}$$

$$\vec{E_i}(x,t) = \begin{pmatrix}0\\ \cos\alpha\\ \sin\alpha\end{pmatrix}E_0 e^{i(kx-\omega t)}$$

$$r_0 = \frac{e^2}{4\pi\varepsilon_0 mc^2} = 2{,}818\ \ 10^{-15}\,m$$

r_0 is the electron classical radius

δI : flux scattered by one e- through a solid angle dΩ



$$dS = d^2\,\delta\Omega$$

$$\delta I_s = d^2\delta\Omega\left\|\vec{E_s}(r,t)\right\|^2 = \delta\Omega\, b_0 {E_0}^2 * P(\alpha)$$

$$b_0 = \frac{1}{{E_0}^2}\frac{\delta I_s}{\delta\Omega} = {r_0}^2\left(\frac{1+cos^2(2\theta)}{2}\right)$$

b_0 is the electron ***differential scattering cross section***

Scattering amplitude by a particle

Coherent scattering : summing up amplitudes

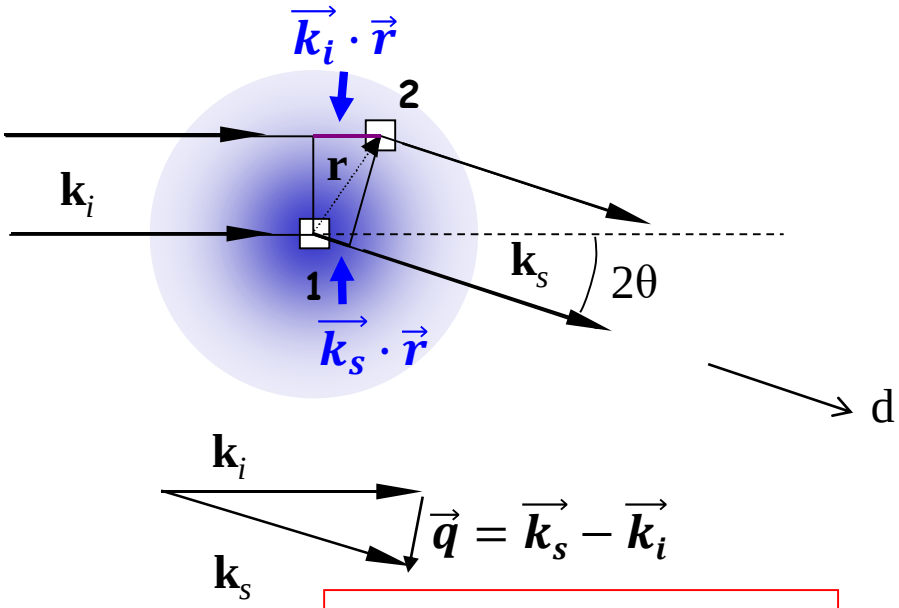
- Waves scattered by two electrons

Electron 1

$$\overrightarrow{\boldsymbol{E_{s1}}}(d,t) = -\frac{r_0}{d}\overrightarrow{\boldsymbol{E_0}}e^{i(kd-\omega t)}$$

Electron 2

$$\overrightarrow{\boldsymbol{E_{s2}}}(d,t) = -\frac{r_0}{d}\overrightarrow{\boldsymbol{E_0}}e^{i(kd-\omega t+\overrightarrow{k_i}\cdot\vec{r}-\overrightarrow{k_s}\cdot\vec{r})}$$



Path shift between waves 1 and 2 :

$$\overrightarrow{\boldsymbol{k_i}}\cdot\vec{\boldsymbol{r}} - \overrightarrow{\boldsymbol{k_s}}\cdot\vec{\boldsymbol{r}} = \left(\overrightarrow{\boldsymbol{k_i}} - \overrightarrow{\boldsymbol{k_s}}\right)\cdot\vec{\boldsymbol{r}} = -\vec{\boldsymbol{q}}\cdot\vec{\boldsymbol{r}}$$

$$\overrightarrow{\boldsymbol{E_{s2}}}(d,t) = -r_0 e^{-i\vec{q}\cdot\vec{r}}\frac{1}{d}\overrightarrow{\boldsymbol{E_0}}e^{i(kd-\omega t)}$$

$\vec{q}$ = Momentum transfer

$$q = \left\|\vec{q}\right\| = \frac{4\pi\sin(\theta)}{\lambda}$$

- The scattered wave by a particle is simply the coherent sum of the spherical waves scattered by all the electrons of the particle.

Particle scattering « amplitude » (length)

$$A(\vec{q}) = -r_0\int_V \rho_e(\vec{r})e^{-i\vec{q}\cdot\vec{r}}d^3\mathbf{r}$$

Intensity scattered by a sample – Auto-correlation function

Scattering *amplitude* $$A(\vec{q}) = -r_0 \int_V \rho_e(\vec{r}) e^{-i\vec{q}\cdot\vec{r}} d^3\mathbf{r}$$

Scattering *intensity* per unit volume : I(**Q**), usual unit: cm^{-1}.

$$I(\vec{q}) = \frac{1}{V} A.A^*(\vec{q}) = \frac{{r_0}^2}{V} \int_V \int_V \rho_e(\vec{r}_1) e^{-i\vec{q}\cdot\vec{r}_1} \rho_e(\vec{r}_2) e^{+i\vec{q}\cdot\vec{r}_2} d^3\mathbf{r}_1 d^3\mathbf{r}_2$$

$$I(\vec{q}) = \frac{{r_0}^2}{V} \int_V \int_V \rho_e(\vec{r}_1)\rho_e(\vec{r}_2) e^{-i\vec{q}\cdot(\vec{r}_1-\vec{r}_2)} d^3\mathbf{r}_1 d^3\mathbf{r}_2$$

Auto-correlation function $\gamma_e(\mathbf{r})$:

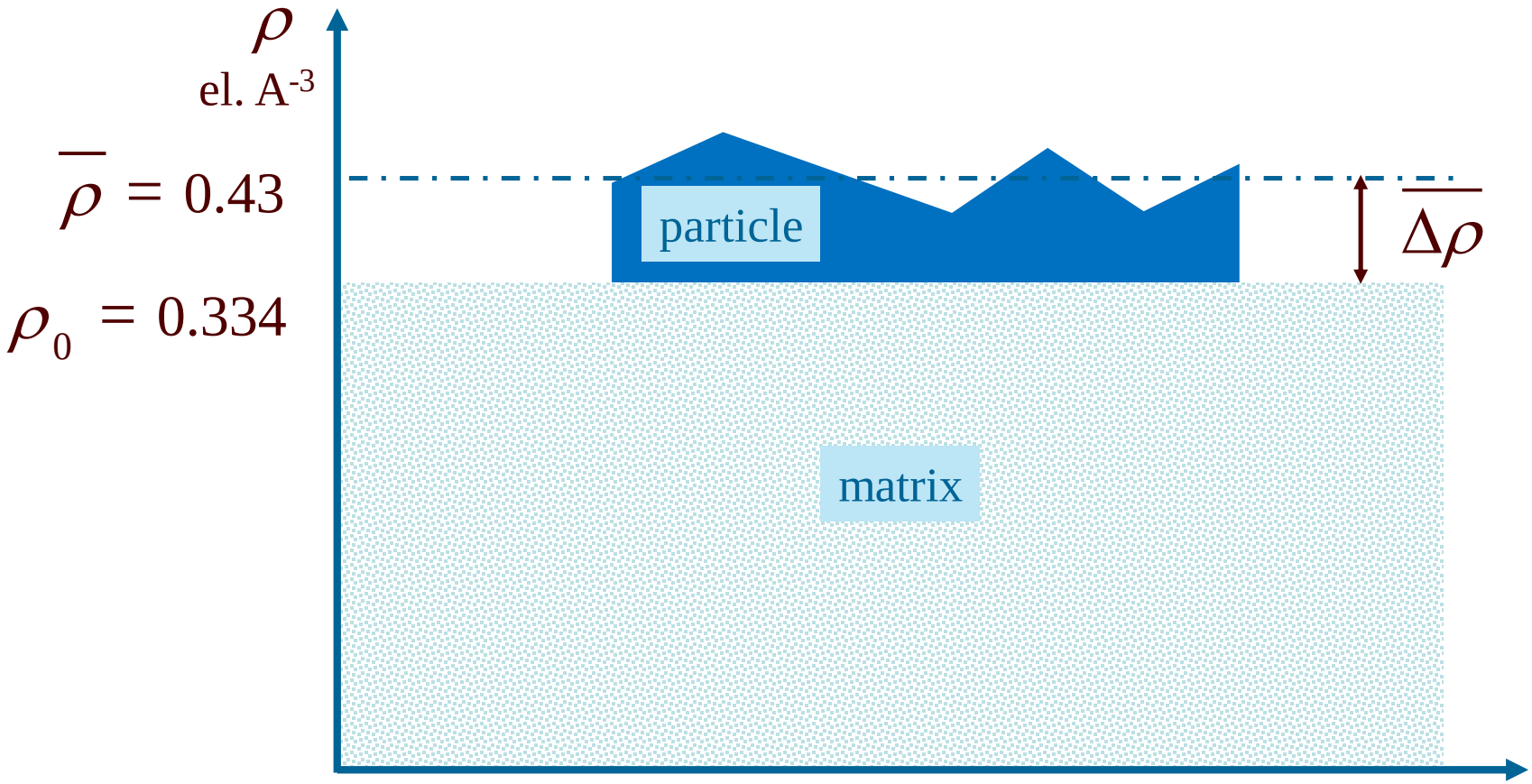
$$\gamma_e(\vec{r}) = \frac{1}{V} \int_V \rho_e(\vec{r}')\rho_e(\vec{r}+\vec{r}') d^3\mathbf{r}'$$

$$I(\vec{q}) = {r_0}^2 \int_V \gamma_e(\vec{r}) e^{-i\vec{q}\cdot\vec{r}} d^3\mathbf{r}$$

The scattered intensity is the Fourier Transform of the electronic density auto-correlation function

Particles in a homogeneous matrix (or buffer)

- A particle is described by the associated electron density distribution $\rho_p(\mathbf{r})$.
- In a matrix, what contributes to scattering is the ***contrast*** of electron density between the particle and the matrix $\Delta\rho(\mathbf{r}) = \rho_p(\mathbf{r}) - \rho_0$ that may be **very small** for biological samples.



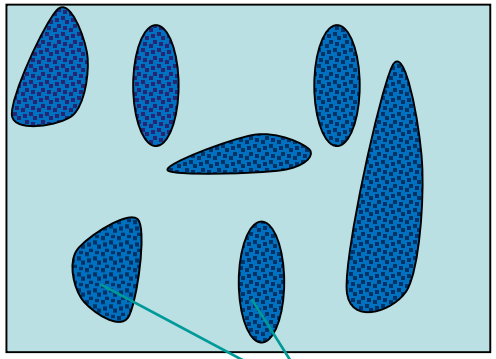
Particles inserted in a "homogeneous matrix"

• Scattering amplitude

$$f(\vec{q}) = -r_0 \int_{V_1} \Delta\rho(\vec{r}) e^{-i\vec{q}\cdot\vec{r}} d^3\mathbf{r}, \vec{q} \neq 0$$

Particles volume

Electronic Density Contrast



Particles

• $\Delta\rho(\vec{r})$ is the contrast of electronic density and describes the scattering object

• $f(\vec{q})$ is the Scattering Amplitude of the ensemble of the particles

• Scattering intensity per unit volume

$$I(\vec{q}) = \frac{1}{V} f(\vec{q}) f^*(\vec{q})$$

Irradiated volume

• $I(\vec{q})$ is expressed in cm^{-1} and is directly related to the measured intensity

Particles in solution

Particles in solution have random orientation, both in time (thermal motion) and in space (no long range correlations). The sample as a whole is therefore **isotropic**. As a result, the scattering intensity only depends on the **modulus** of $\vec{q}$, $q = 4\pi \sin(\theta) / \lambda$.

Scattering from a single particle in solution, averaged over time:

$$I_1(q) = \frac{1}{V}\overline{\left\langle f_1\left(\vec{q}\right) f_1^{*}\left(\vec{q}\right)\right\rangle_{\Omega}}$$

Modulus

Vector

The form factor P(q) is the normalized signature in q–space of a particle in solution.

$$P(q) = \frac{\overline{\left\langle f_1\left(\vec{q}\right) f_1^{*}\left(\vec{q}\right)\right\rangle_{\Omega}}}{r_0^{\,2} V_{obj}^{\,2} \left\langle \Delta\rho \right\rangle^2}$$

$P(0) = 1$

Average Electronic Density contrast

Particle volume

Monodispersity and ideality

- **Monodispersity**
  - Yes ← Identical particles
  - No ← Size and Shape polydispersity

- **Ideality**
  - Yes ← No correlations between particles positions

    (No short-range or long-range interactions)
  - No ← Correlations between particles positions

    (Existence of short-range or long-range interactions)

Ideal and monodisperse solutions

- **Ideal**

$$I(q) = \sum_{i=1,N} I_i(q) = \frac{1}{V} \sum_{i=1,N} \overline{\left\langle f_i\left(\vec{q}\right) f_i^{*}\left(\vec{q}\right)\right\rangle_{\Omega}}$$

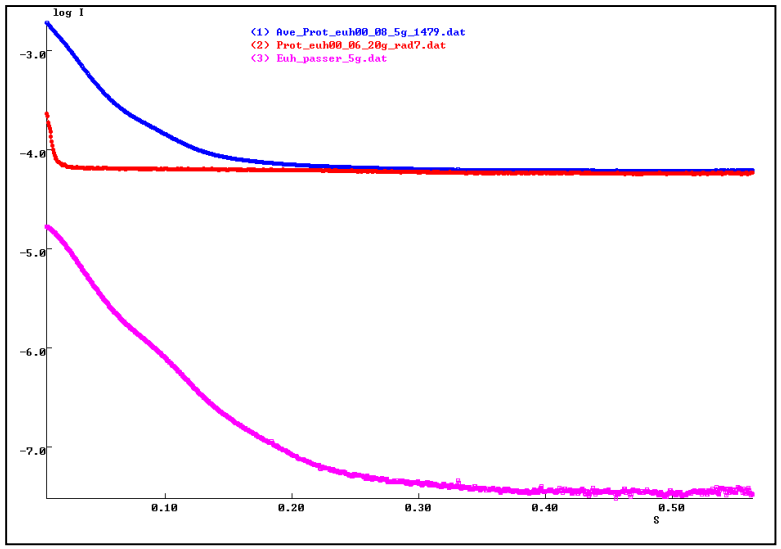
- **Monodisperse**

$$I_i(q) = I_1(q) \qquad \forall \mathrm{i}$$

- **Ideal and monodisperse**

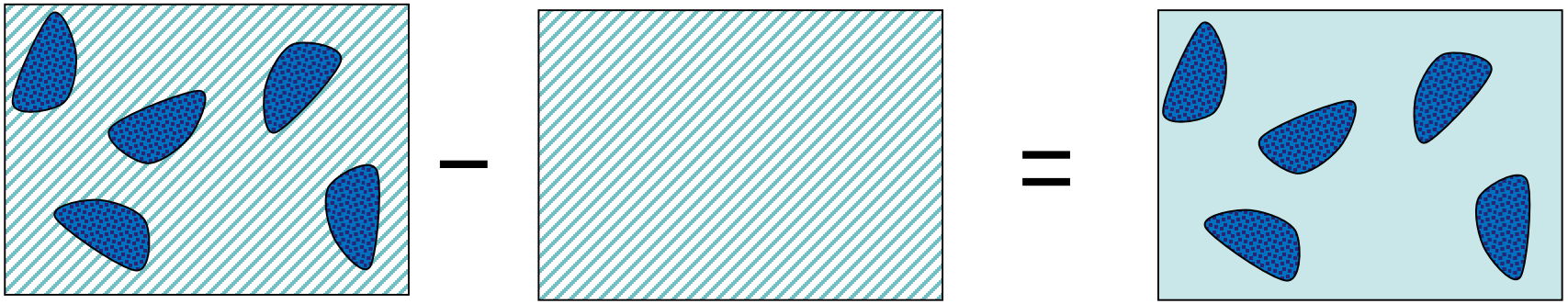
$$I(q) = \mathrm{N}\, I_1(q) = \frac{\mathrm{N}}{V}\left\langle f_1\left(\vec{q}\right) f_1^{*}\left(\vec{q}\right)\right\rangle_{\Omega}$$

Solvent scattering and contrast



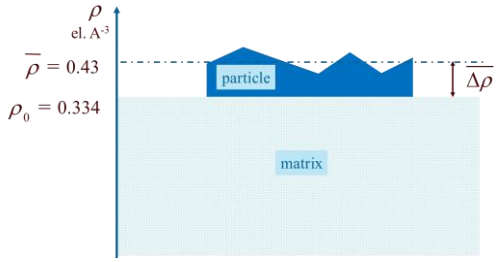
To obtain scattering solely from the contrasting particles, intrinsic solvent scattering must be measured **very accurately** and subtracted, which also permits to subtract contribution from parasitic background (slits, sample holder etc) which should be reduced to a minimum.

$$I_{solution}(q) - I_{buffer}(q) = I_{particles}(q)$$



Do not confuse !

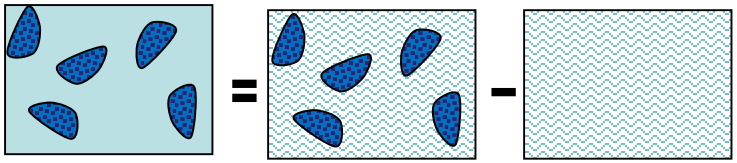
contrast effect



$$\Delta\rho(\vec{r}) = \rho(\vec{r}) - \rho_0$$

$$f(\vec{q}) = -r_0 \int_V \Delta\rho(\vec{r}) e^{-i\vec{q}\cdot\vec{r}} d^3\mathbf{r}, \vec{q} \neq 0 \quad \rightarrow \quad I(q) = \frac{1}{V}\overline{\langle f(\vec{q}) f^*(\vec{q}) \rangle_\Omega}$$

buffer subtraction



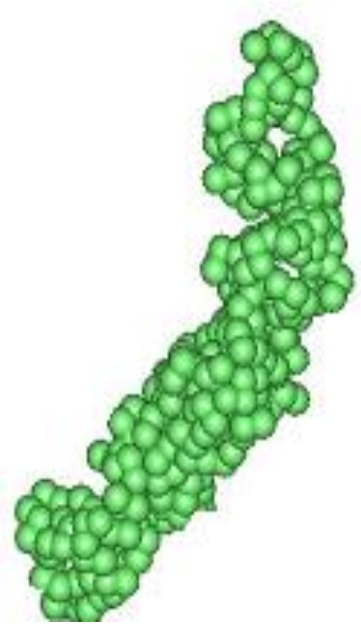
$$I_{particles,\exp}(q) = I_{solution,\exp}(q) - I_{buffer,\exp}(q)$$

$i_1(q)$

Ideality

Monodispersity

$I(q)$



One must check that both assumptions are valid for the sample under study.

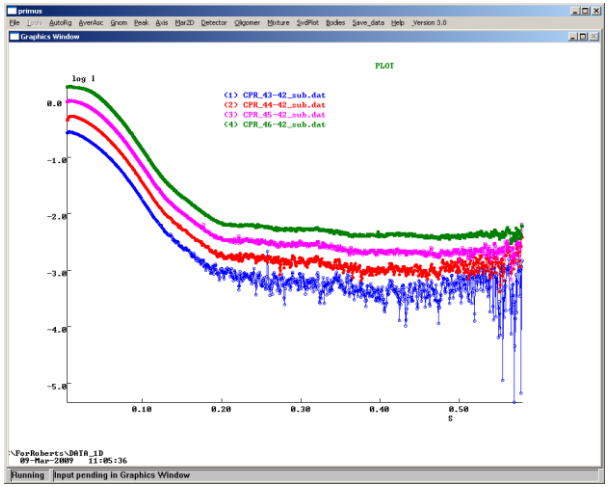
molecule

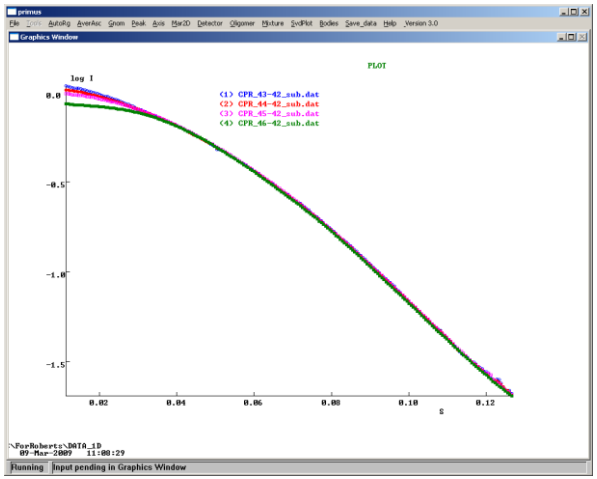
experimental

Checking the validity of both assumptions for the sample under study is crucial for non erroneous data interpretation

- **Size Monodispersity must be checked independently**
  → **Purification protocol :SEC, DLS, AUC, MALS, etc.**

- **Ideality : reached by working in buffers with screened interactions or at high dilution**
  → **In practice : measurements at decreasing concentrations and checks whether the scattering pattern is independent of concentration.**

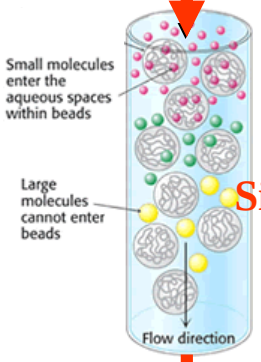




SE-HPLC / Solution Sampler

Flow rate 250 µl/min

- Monodisperse solution
- Aggregation is eliminated
- Oligomeric conformations can be distinguished
- Equilibrium states can be transiently separated
- Perfect background subtraction
- Automatic concentration series

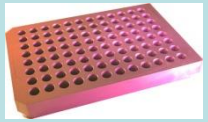


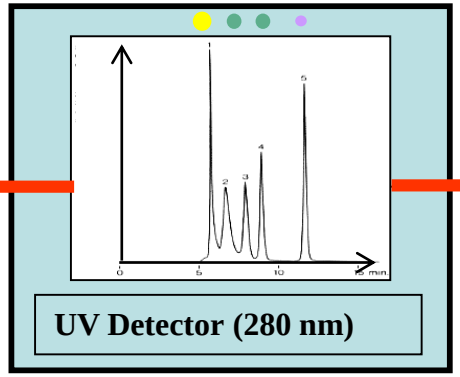
Size Exclusion

Incident X-ray

Pump

Injection-mixing





UV-Vis in

SAXS Cell

UV-Vis out

Flow rate 5-40 µl/min

Pure sample

- Small volumes (~ 10 µl)
- No dilution
- High rate (a few minutes)

BIOPHYSICAL INFORMATION

Biophysical information

- Guinier Analysis
- Kratky plot : why is it so interesting ?
- « Real-space SAXS » : Distance correlation function P(r)

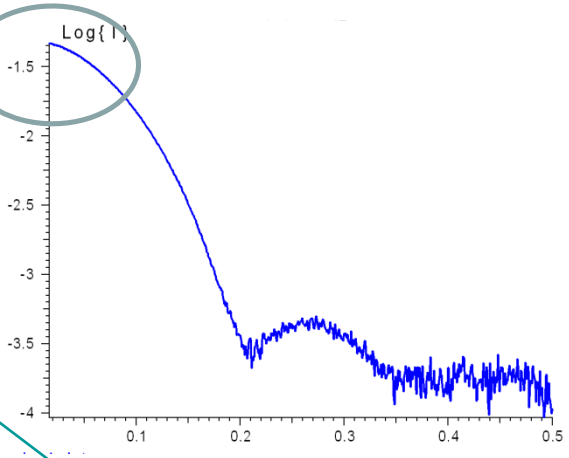
Biophysical information

- Guinier Analysis
- Kratky plot : why is it so interesting ?
- « Real-space SAXS » : Distance correlation function P(r)

Asymptotic behaviour at small angles : Guinier law

Close to q=0, the scattering intensity of a particle can be described by a Gaussian curve.

The validity domain actually depends on the shape of the particle and is around $q < 1.3 / Rg$ for a globular shape.



Prof. André Guinier
1911-2000
Orsay, France

$$I(q) = I(0)\exp\left(\frac{-q^2 Rg^2}{3}\right)$$

Extrapolated intensity at origin

Radius of gyration

Guinier law, in Log scale :

$$Ln[I(q)] = Ln[I(0)] - \frac{q^2 Rg^2}{3}$$

The Guinier law is equivalent of a linear variation of Ln(I(q)) vs q^2 (Guinier plot). Linear regression on the experimental Guinier plot directly provides Rg and I(0).

Mass retrieval from Guinier analysis

$$I(q) = I(0)\exp\left(\frac{-q^2 Rg^2}{3}\right)$$

Absolute Unit : cm^{-1}

Classical electron radius

$$I(0) = \frac{c \cdot M \cdot r_0^2}{N_A} \cdot \left[\nu_p\left(\rho_{prot} - \rho_{buf}\right)\right]^2$$

$$Rg^2 = \frac{\int_V r^2\left(\rho_{prot}(\vec{r}) - \rho_{buf}\right)d\vec{r}}{\int_V \left(\rho_{prot}(\vec{r}) - \rho_{buf}\right)d\vec{r}}$$

Mass concentration

Electronic density contrast

Protein specific volume

I(0) gives an independent estimation of the molar mass of the protein **(only if the mass concentration and specific volume are precisely known …)**

Typically :

M (kDa) = (1200 ~ 1600) * I0 (cm^{-1}) / C (mg/ml)

Rg depends on the volume AND on the shape of the particle

For globular proteins : $R_g\,(Å) \approx 6.5 * M^{\frac{1}{3}}, M\ in\ kDa$

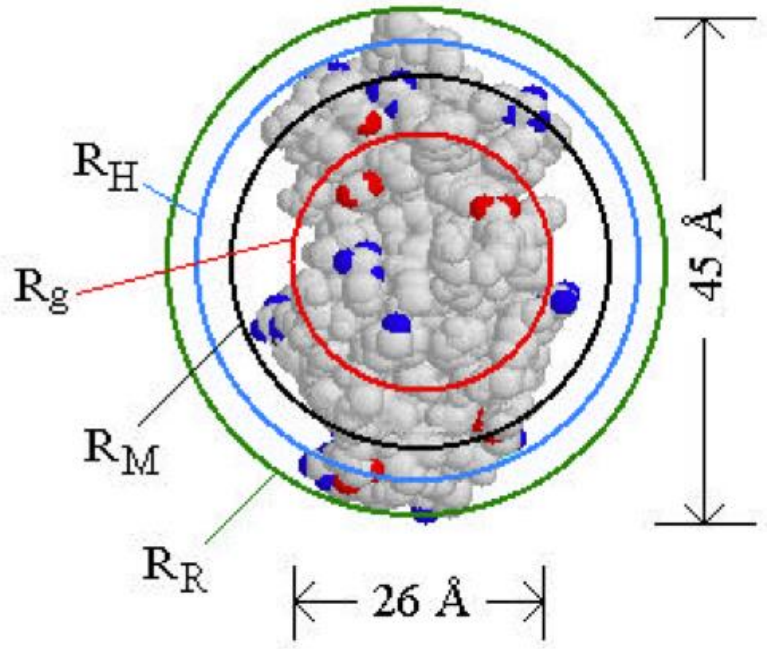
For unfolded proteins : $R_g\,(Å) \approx 8.05 * M^{0.522}$

Bernado et al. (2009), Biophys. J., 97 (10), 2839-2845.

Radius of gyration

$$Rg_{\exp}^{\,2} = \frac{\int_V r^2 \Delta\rho(\vec{r})\, d\vec{r}}{\int_V \Delta\rho(\vec{r})\, d\vec{r}}$$

Lysozyme



graphic: www.silver-colloids.com/Papers/hydrodynamic-radius.pdf

Useful definitions of R_g

$$R_g^2 = \frac{1}{N}\sum \left\| \vec{r}_i - \overrightarrow{r_{COM}} \right\|^2 \quad \text{by atoms}$$

$$R_g^2 = \int r(r) r^2\, dr \Big/ \int r(r)\, dr \quad \text{by electron density}$$

$$R_g^2 = \frac{1}{2N(N-1)} \sum_i \sum_j \left\| \vec{r}_i - \vec{r}_j \right\|^2 \quad \text{by atom pairs}$$

$$R_g^2 = \tfrac{1}{2} \int r^2 p(r)\, dr \Big/ \int p(r)\, dr \quad \text{by pair distribution}$$

R_g radius of gyration
R_H hydrodynamic radius (not always > Rg!)
R_R maximum hard sphere radius
R_M radius of mass-equivalent sphere

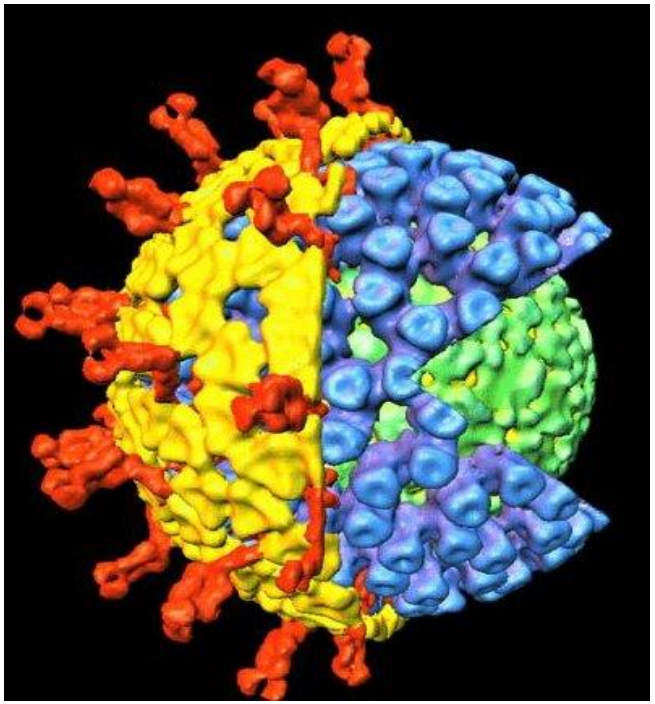
* center of mass of the *electron* density

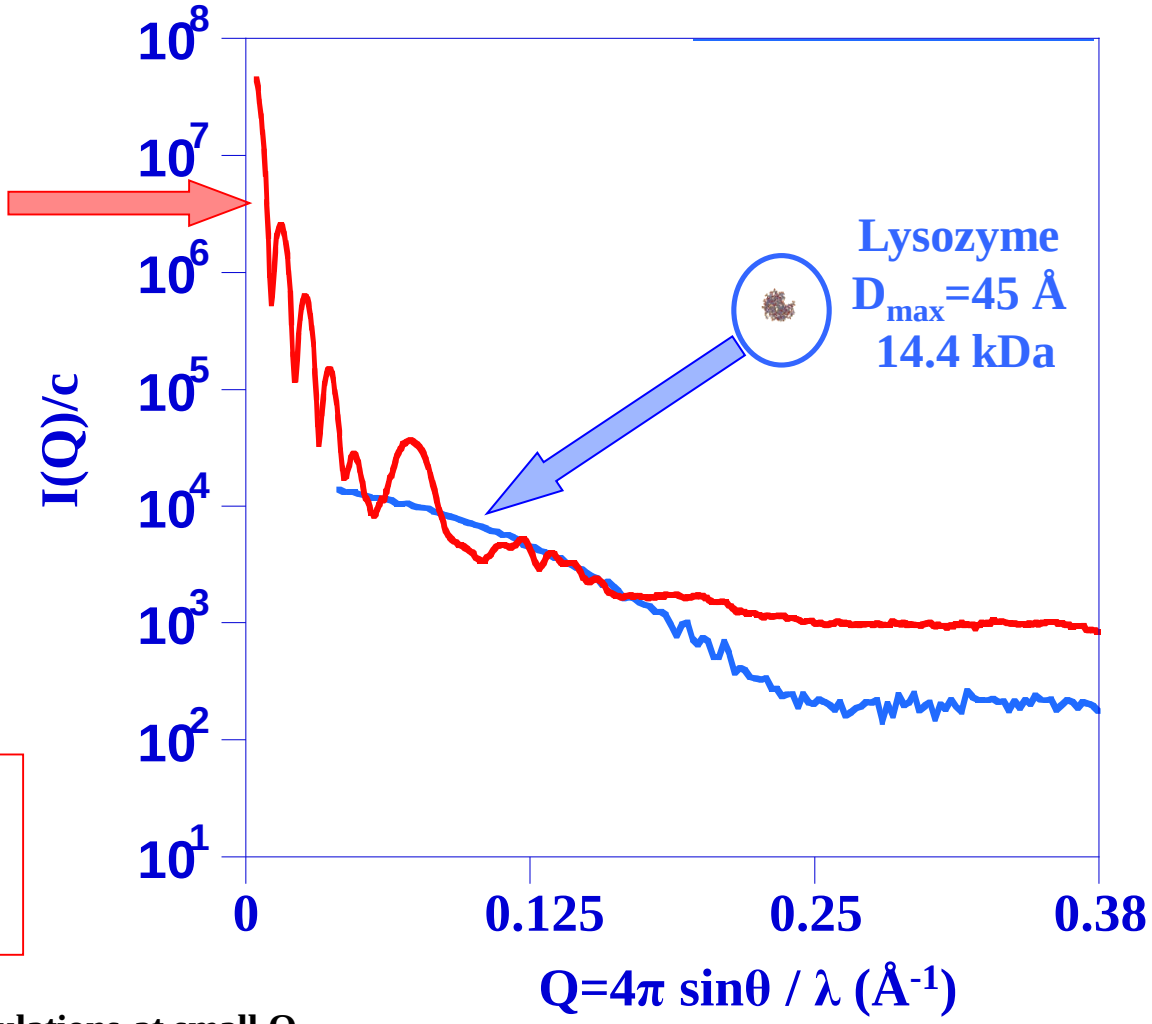
Sphere	$R_g = \sqrt{\frac{3}{5}} R$
Thin rod	$R_g = \sqrt{\frac{1}{12}} L$
Thin disk	$R_g = \sqrt{\frac{1}{2}} R_{disk}$

Courtesy: Richard Gillilan, Cornell U., USA

Basic law of reciprocity in scattering

Rotavirus VLP : diameter = 750 Å, 44 MDa





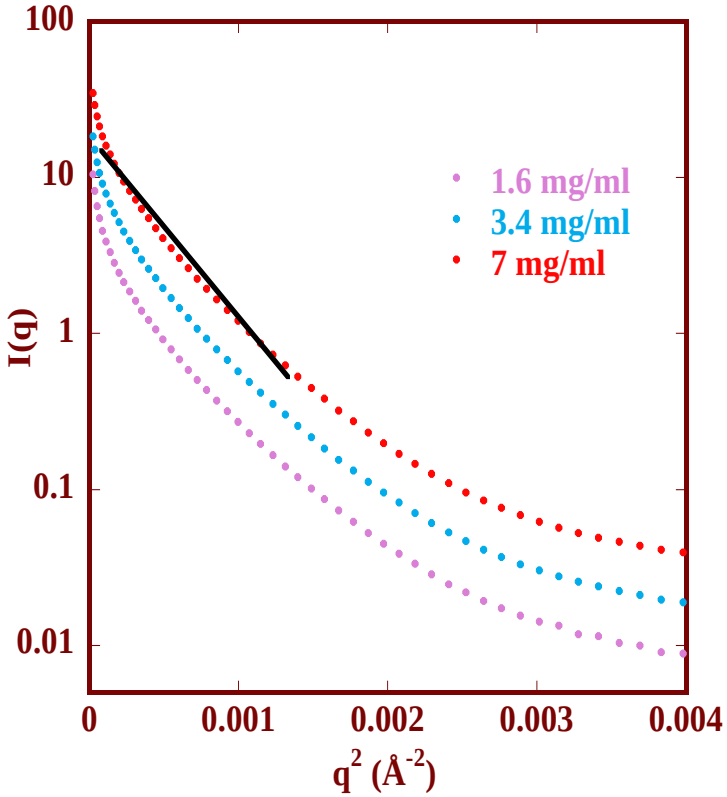
Typical range on beamline SWING
$2\ 10^{-3} < q < 7\ 10^{-1}$ Å^{-1}
750 Å > D_{max}
3100 Å > d_{Bragg} > 9 Å

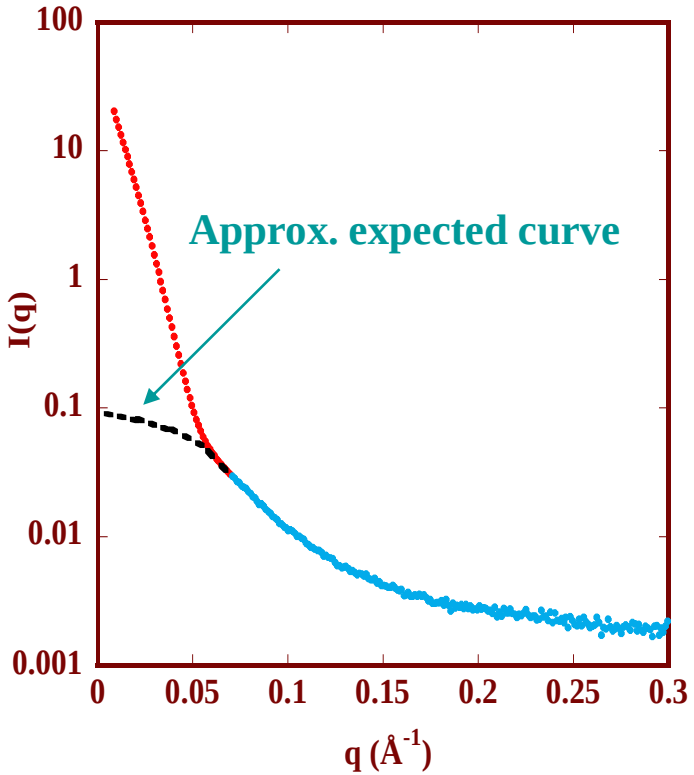
Long distance correlations ⟺ modulations at small Q

Evaluation of the solution properties

Irreversible aggregation

→ Useless data: the whole curve is affected





I(0): > 150 fold the expected value for the given MM

Swing – Domaine 1-242 de RRP44 – 07/08

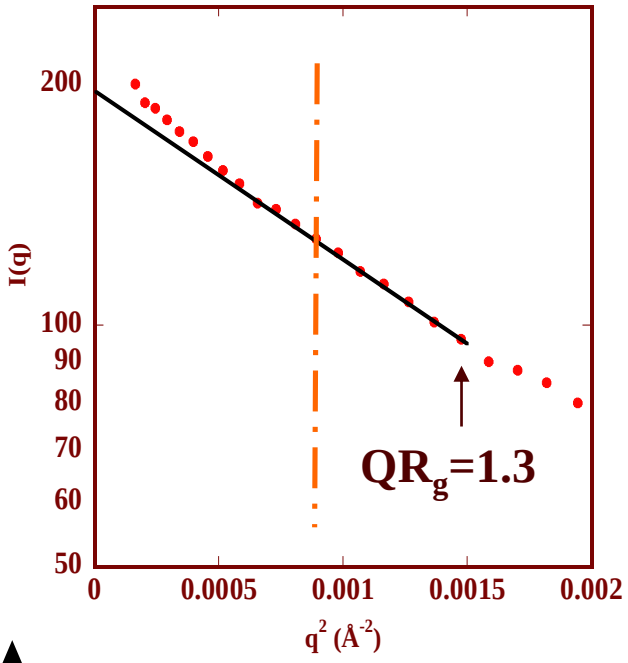
(Courtesy D. Durand, IBBMC, Orsay)

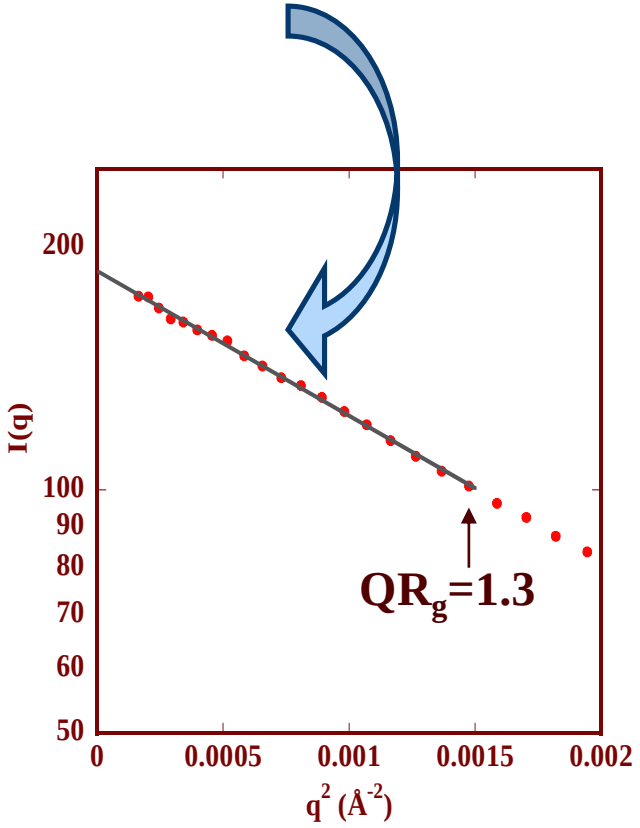
Evaluation of the solution properties

Weak aggregation → **possible improvement**

centrifugation, buffer change

Nanostar –PR65 protein





R_g ~ 38 Å – too high!!

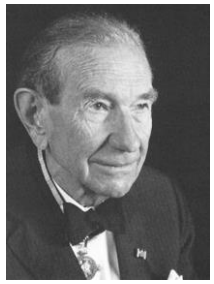
R_g ~ 36 Å

(Courtesy D. Durand, IBBMC, Orsay)

Biophysical information

- Guinier Analysis
- Kratky plot : why is it so interesting ?
- « Real-space SAXS » : Distance correlation function P(r)

Kratky Plot

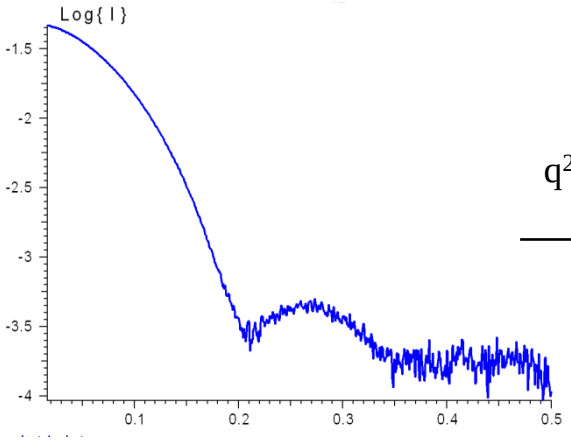


Prof. Otto Kratky
1902-1995
Graz, Austria

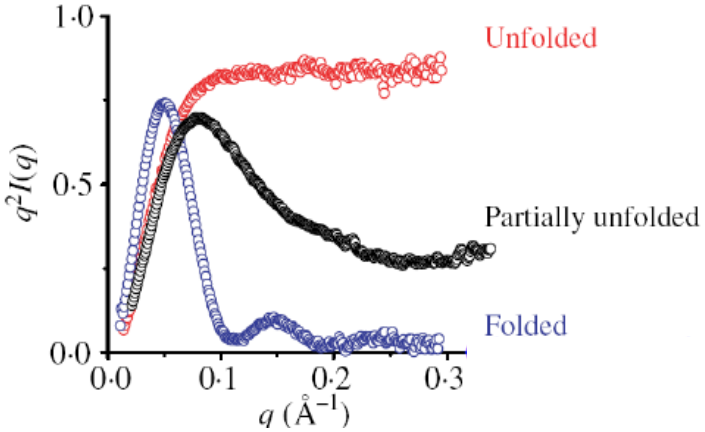
SAXS provides a sensitive means to ***evaluate the degree of compactness*** of a protein:

- To determine whether a protein is globular, extended or unfolded
- To monitor the folding or unfolding transition of a protein

This is most conveniently represented using the so-called Kratky plot:



q^2 I(q) *versus* q



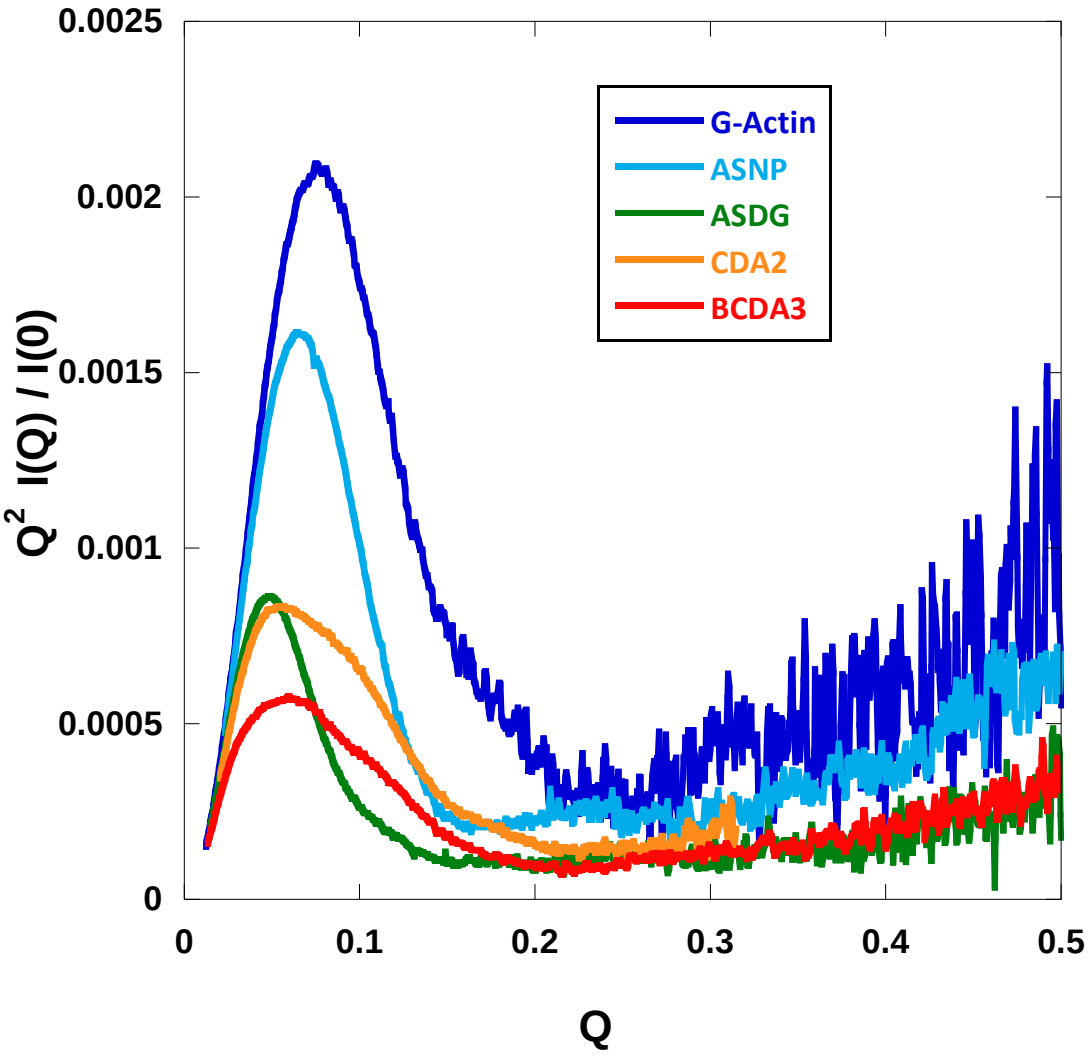
Putnam, D., et al. (2007) Quart. Rev. Biophys. 40, 191-285.

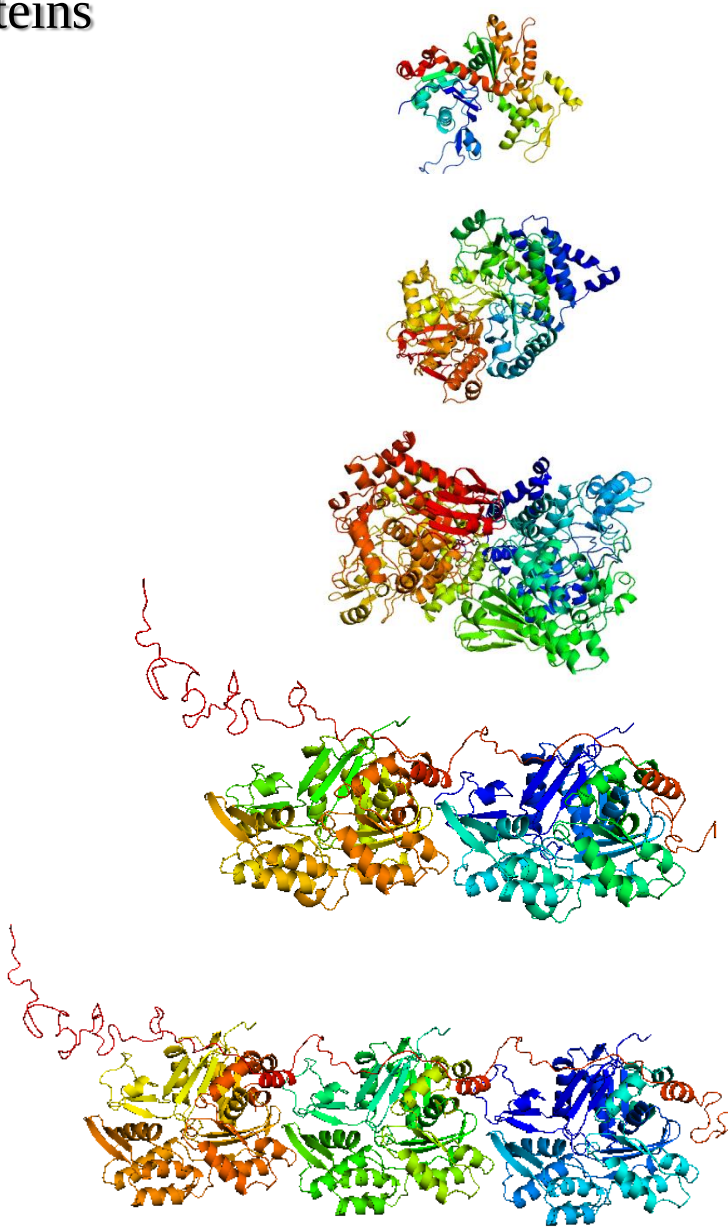
Folded particle : ***bell-shaped curve*** (asymptotic behaviour $I(q) \sim q^{-4}$)

Random polymer chain : ***plateau*** at large q-values (asymptotic behaviour in $I(q) \sim q^{-2}$)

Extended polymer chain : ***increase*** at large q-values (asymptotic behaviour in $I(q) \sim q^{-1.x}$)

Kratky Plots of folded proteins





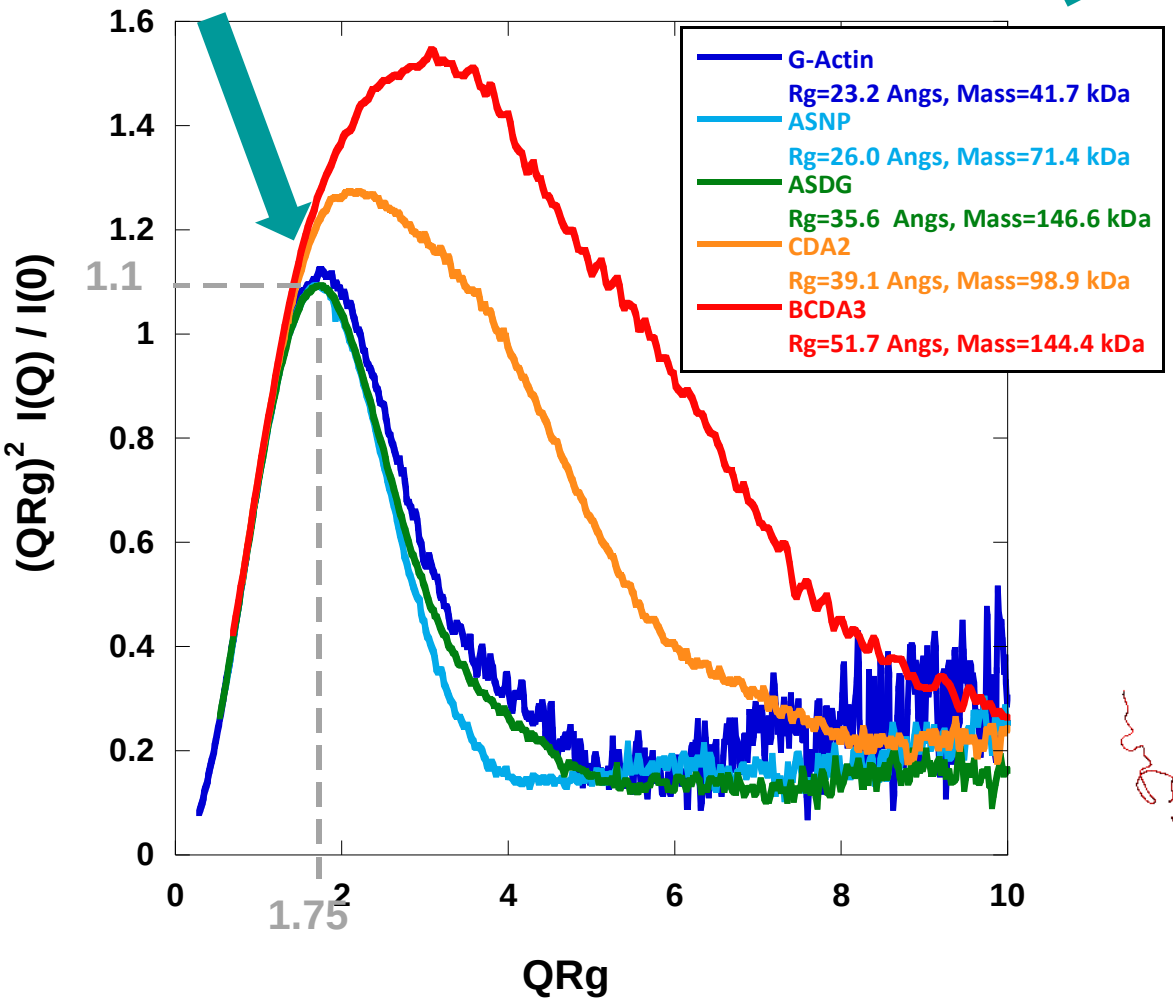
Folded proteins display a bell shape. Can we go further?

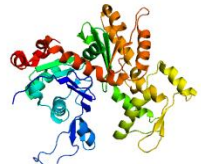
Dimensionless Kratky Plots of folded proteins

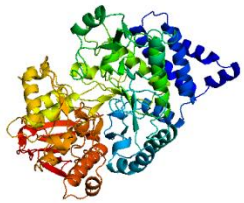
Introduced for biology in Durand et al. (2010), J. Struct. Biol. **169**, 45-53.

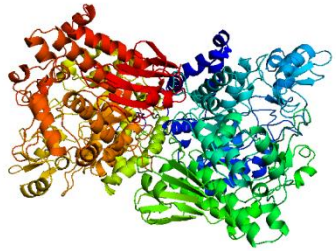
For globular structures, DLKPs fold into the same maximum

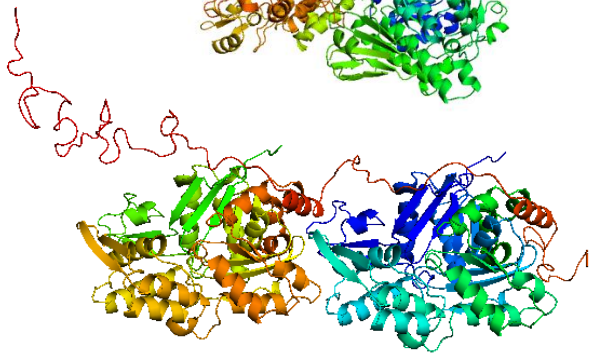
The relation $M_{Rg}(kDa) \approx (Rg / 6.5)^3$ only works for the globular structures, not the elongated

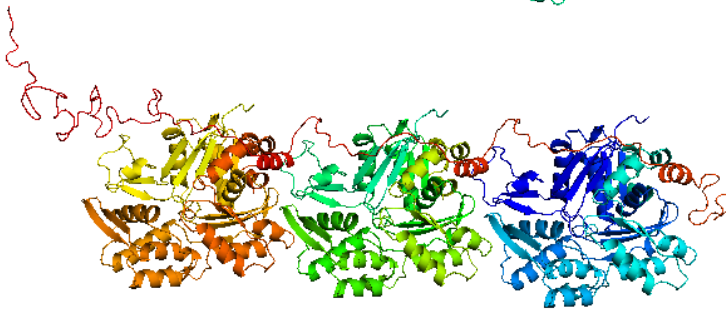








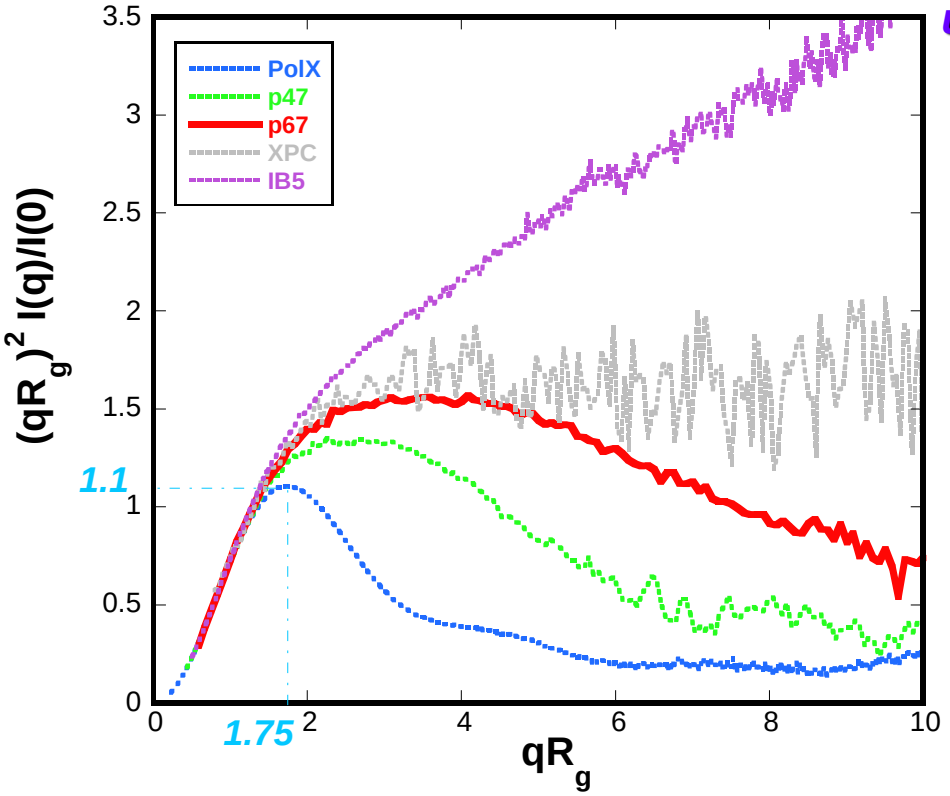




The maximum value on the dimensionless bell shape tells if the protein is globular.

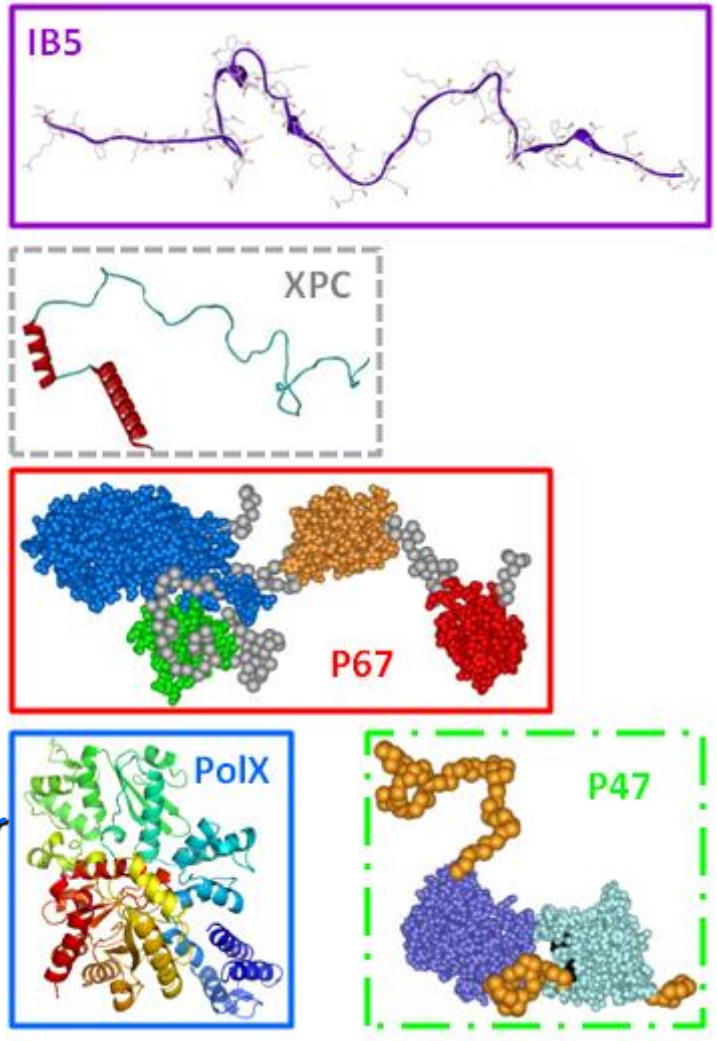
Dimensionless Kratky Plots of (partially) unfolded proteins

Receveur-Bréchot V. and Durand D (2012), Curr. Protein Pept. Sci., 13:55-75.



unfolded

globular



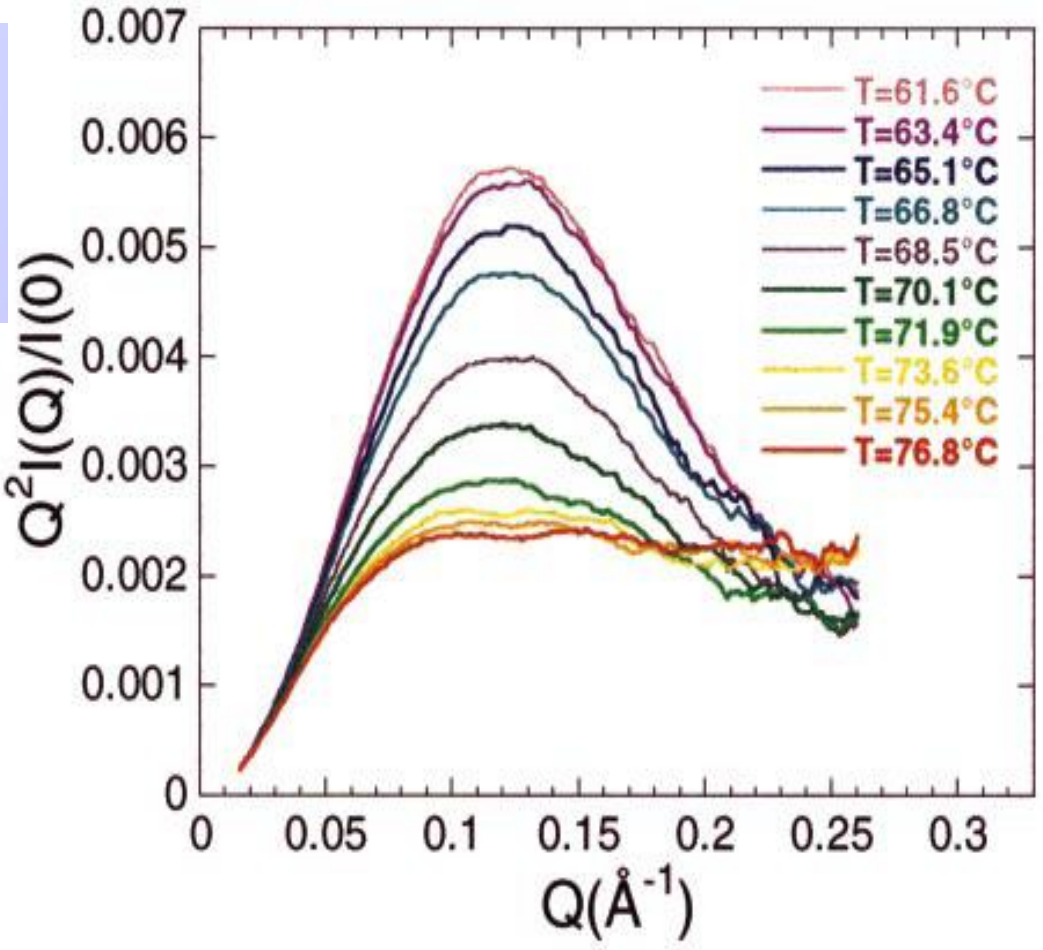
The bell shape vanishes as folded domains disappear and flexibility increases.

The curve increases at large Q as the structure extends.

Kratky Plot : NCS heat unfolding

In practice, thin Gaussian chains do not exist.

In spite of the plateau at T=76°C, NCS is not a Gaussian chain when unfolded, but a thick chain with persistence length

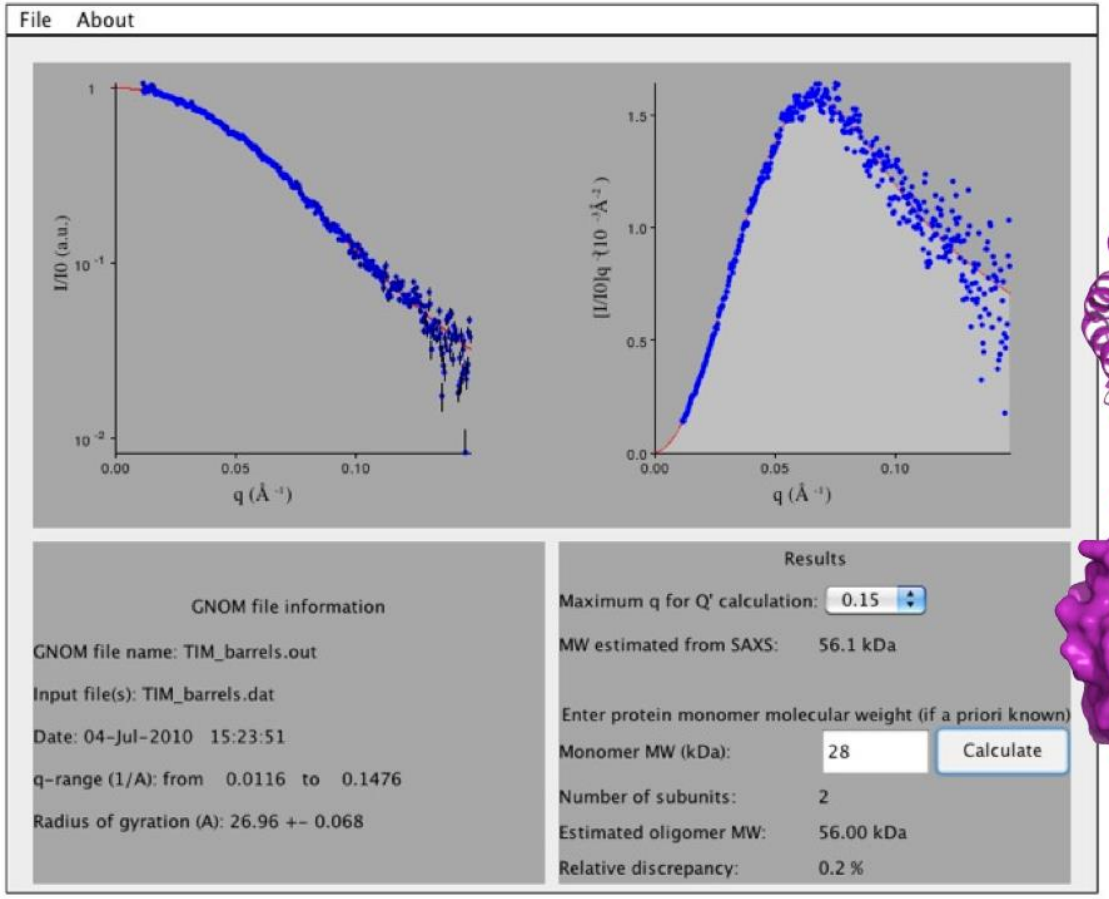


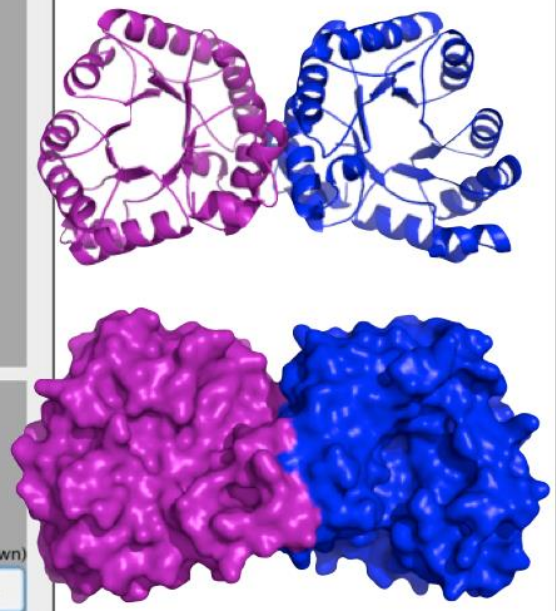
Pérez et al., *J. Mol. Biol.*(2001), 308, 721-74

Molecular Weight estimation based on Porod invariant

http://www.ifsc.usp.br/~saxs/saxsmow.html

- does not require knowledge of concentration
- relies on Porod Volume theory **+ structural database**
- does **not** work for proteins with unfolded domains





Other methods for MW estimation based on similar though different grounds were developed

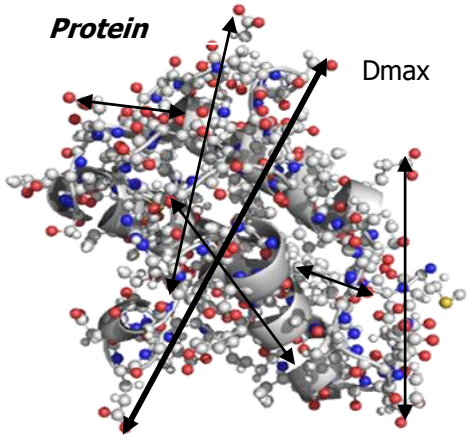
Rambo R. And Tainer J. (2013), Nature, **496**, 477-481.

Biophysical information

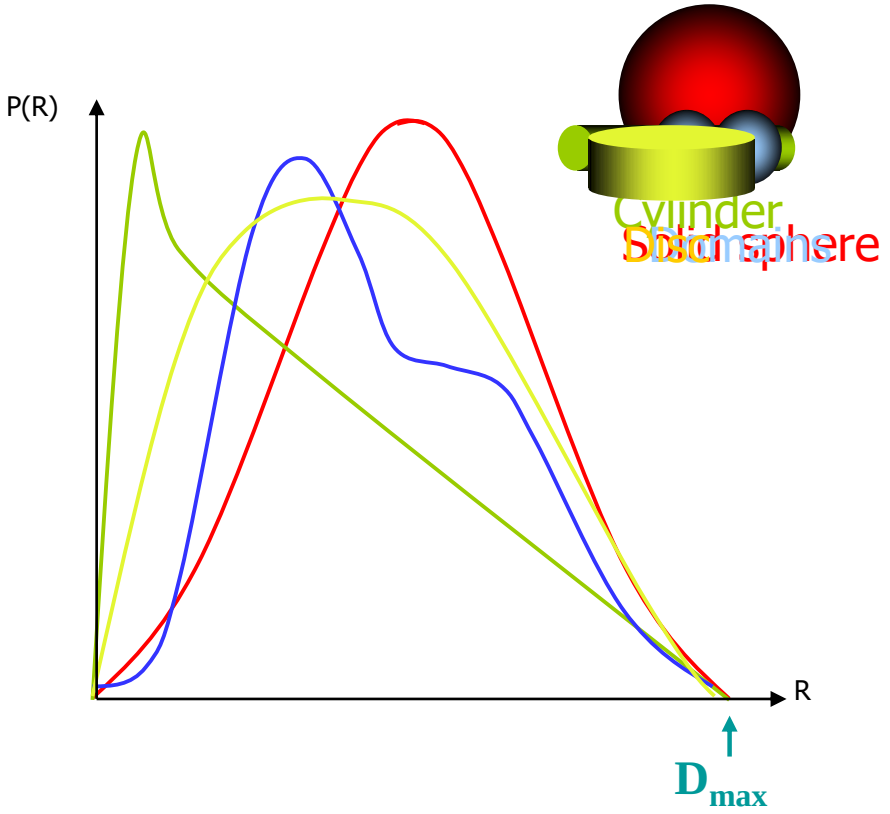
- Guinier Analysis
- Kratky plot : why is it so interesting ?
- « Real-space SAXS » : Pair distribution function P(r)

Pair Distribution Function p(r)

The distance distribution function p(r) is proportional to the average number of atoms at a given distance, r, from any given atom within the macromolecule.



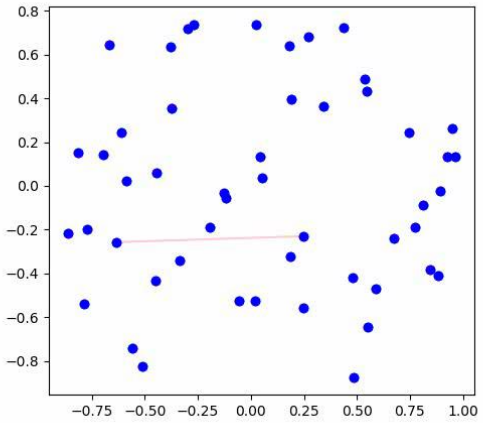
p(r) vanishes at r = D_{max}

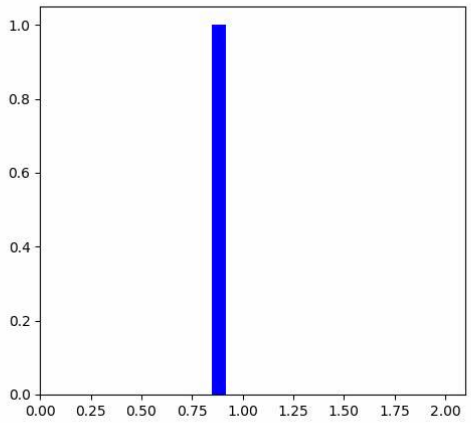


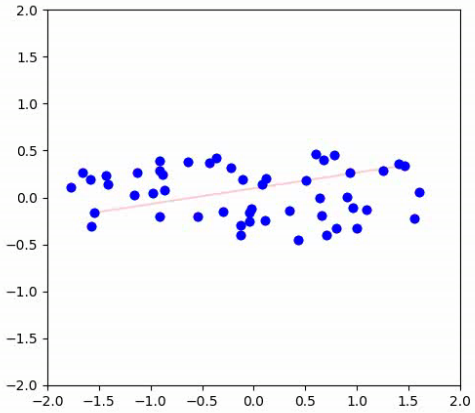
The distance distribution function characterises the shape of the particle **in real space**

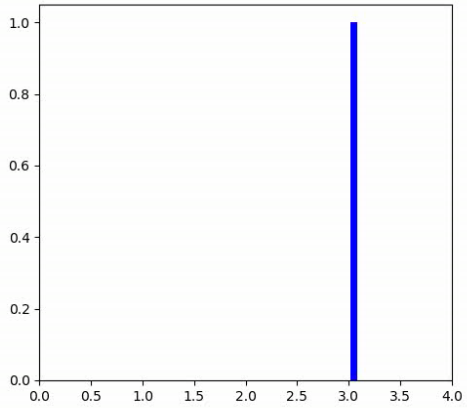
Pair Distribution Function p(r)

Marc-André Delsuc:
thanks for the python script









Relation between p(r) and I(Q)

Intensity is the Fourier Transform of self-correlation function $\gamma_{obj}(r)$:

$$I(q) = 4\pi r_e^2 \varphi \int_{V_{obj}} \gamma_{obj}(r) r^2 \frac{\sin(qr)}{qr} dr$$

Fourier Transform for isotropic samples

And :

$$p(r) = \gamma_{obj}(r) r^2$$

Then :

$$I(q) = 4\pi r_e^2 \varphi \int_0^D p(r) \frac{\sin(qr)}{qr} dr$$

And :

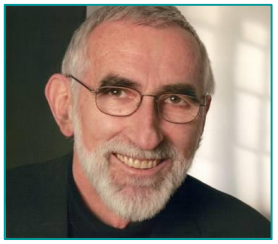
$$p(r) = \frac{r^2}{2\pi^2 \varphi r_e^2} \int_0^\infty q^2 I(q) \frac{\sin(qr)}{qr} dq$$

p(r) could be directly derived from I(q). Both curves contain the same information.

However, direct calculation of p(r) from I(q) is made difficult and risky by $[q_{min}, q_{max}]$ truncation and data noise effects.

Back-calculation of the Distance Distribution Function

Glatter, O. *J. Appl. Cryst.* (1977) **10**, 415-421.



Prof. Otto Glatter
Guinier Prize 2012
Graz, Austria

Main hypothesis : the particle has a « finite » size, characterised by D_{max}.

- D_{max} is proposed by the user
- p(r) is expressed over [0, D_{Max}] by a linear combination of orthogonal functions

$$p_{theoret}(r) = \sum_{1}^{M} c_n \varphi_n(r)$$

- I(q) is calculated by Fourier Transform of $p_{theoret}(r)$

$$I(q) = 4\pi \, r_e^{\,2} \varphi \int_0^{D_{max}} p_{theoret}(r) \frac{\sin(q \cdot r)}{q \cdot r} dr$$

Dr. Dmitri Svergun
Guinier Prize 2018
Hamburg, Germany

Svergun (1988) : program "GNOM"

M ~ 30 - 100 ⇒ ill-posed LSQ ⇒ regularisation method

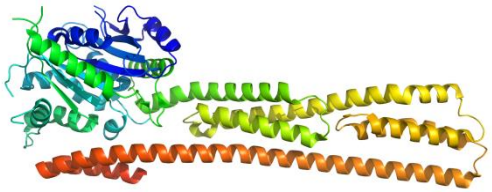
+ "Perceptual criteria" : smoothness, stability, absence of systematic deviations

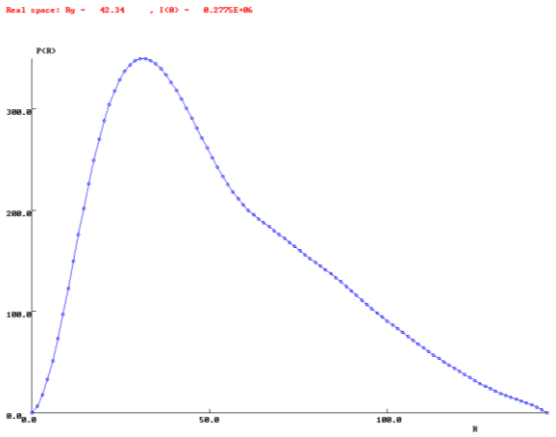
- Each criterium has a predefined weight
- The solution is given a score calculated by comparison with « ideal values »

Distance Distribution Function

Experimental examples

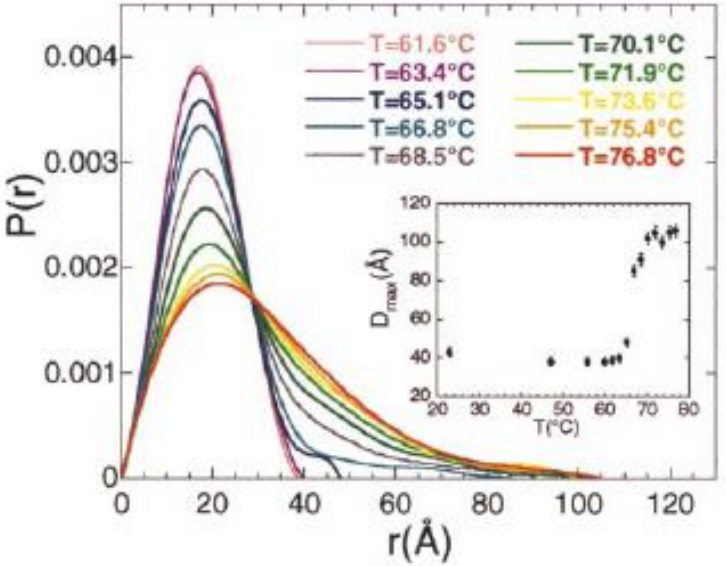
GBP1





Heat denaturation of Neocarzinostatin

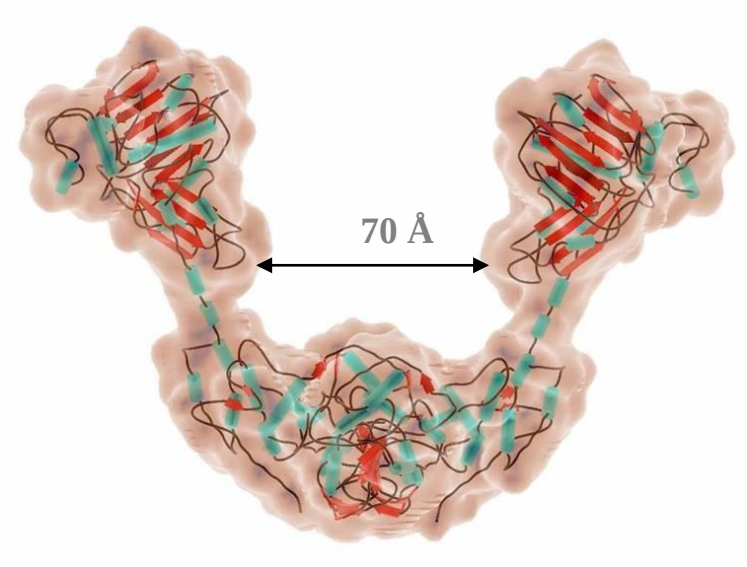
Pérez et al., J. Mol. Biol. (2001) 308, 721-743



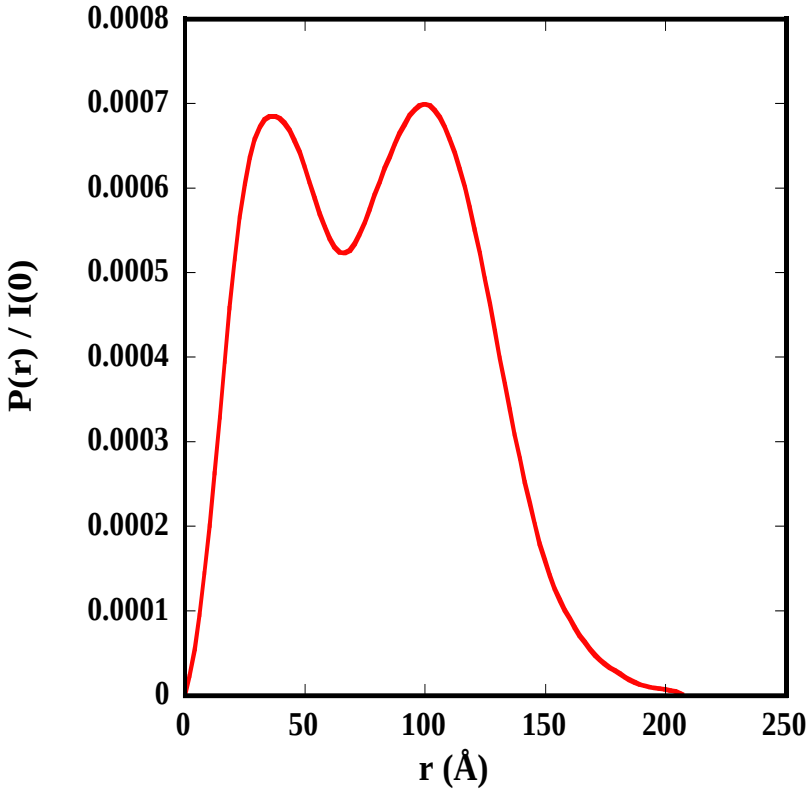
Distance Distribution Function

Experimental examples

Topoisomerase VI



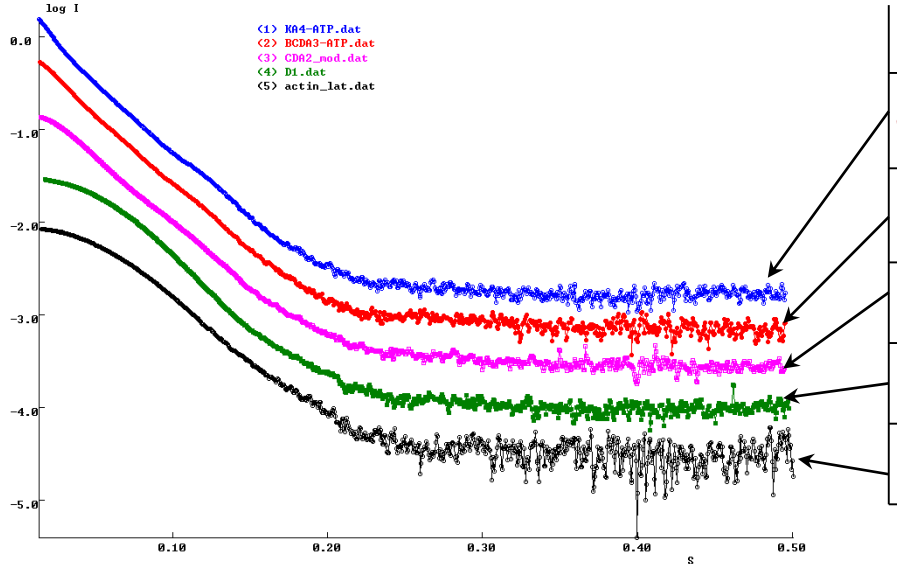
Bimodal distribution

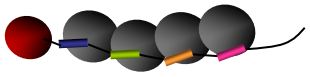


M. Graille et al., Structure (2008), *16*, 360-370.

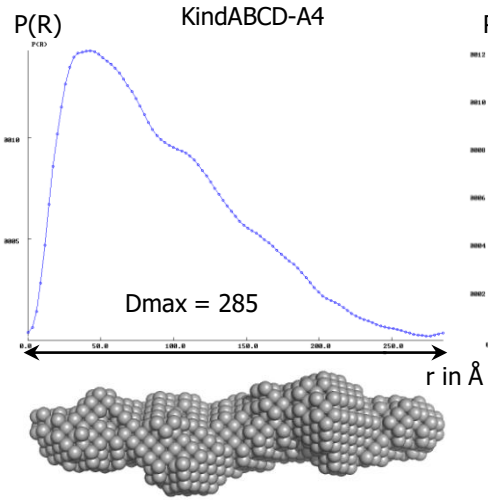
Distance Distribution Function

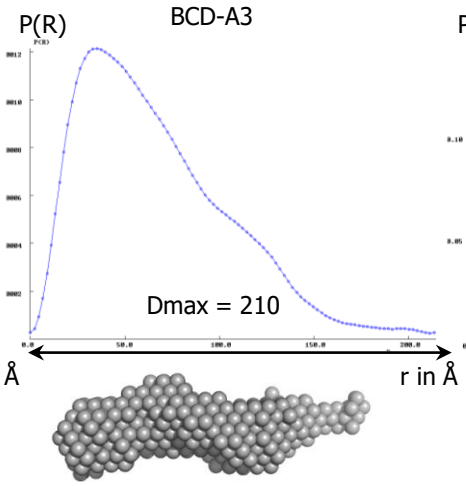
Scattering curves obtained on different complexes Spire-Actin and Actin alone

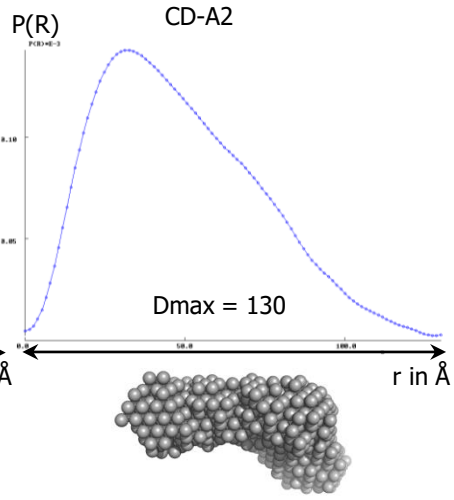


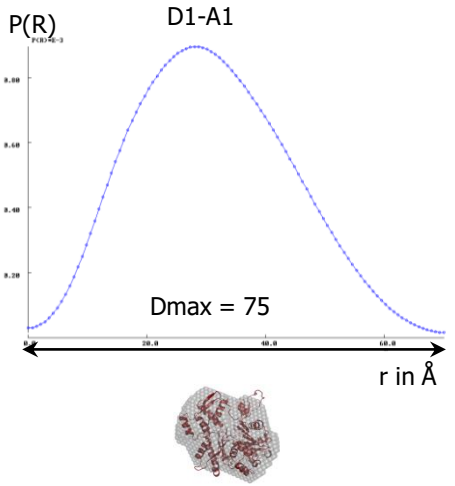
Complexes	Radius of gyration	Maximum diameter
	75.5 Å	285 Å
	55.5 Å	210 Å
	38.9 Å	130 Å
	25 Å	75 Å
	23.1 Å	70 Å

Histogram of intramolecular distances and ab initio molecular enveloppes determined using DAMMIF









Distance Distribution Function

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

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

and

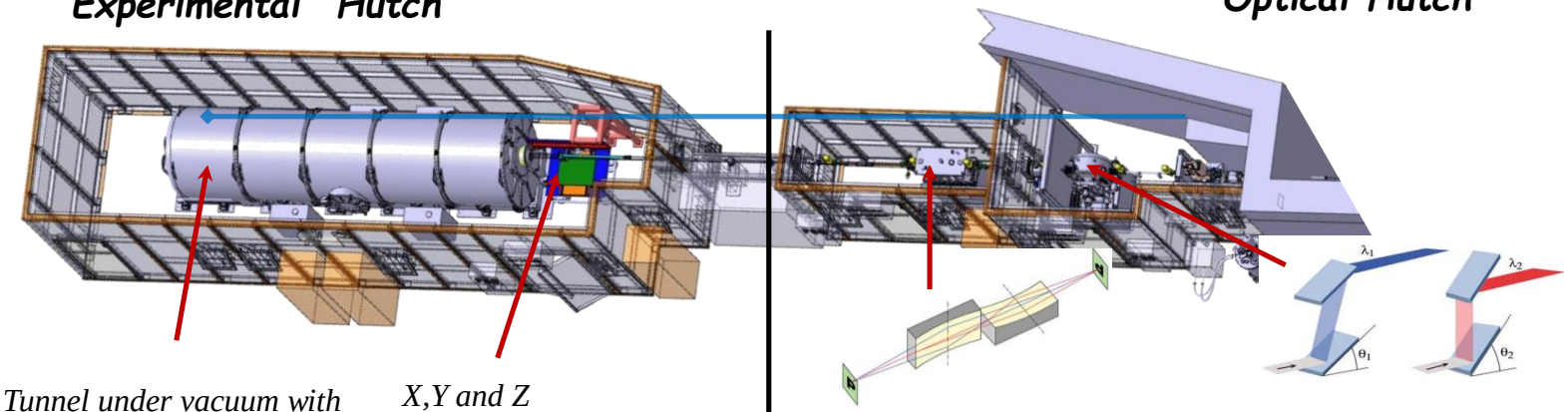
$$I(0) = 4\pi r_e^2 \varphi \int_0^{D} p(r) dr$$

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

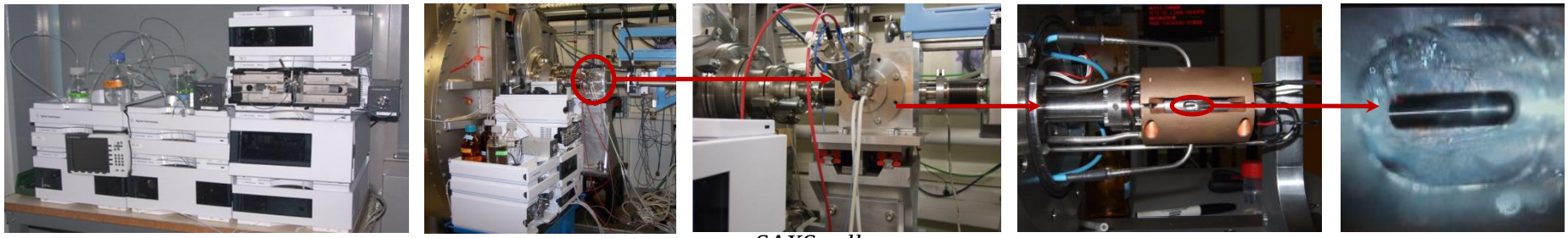
Comparison of estimates from Guinier analysis and from P(r) is a useful cross-check.

A FEW EXPERIMENTAL CONSIDERATIONS

Set-up for BioSAXS at beamline SWING

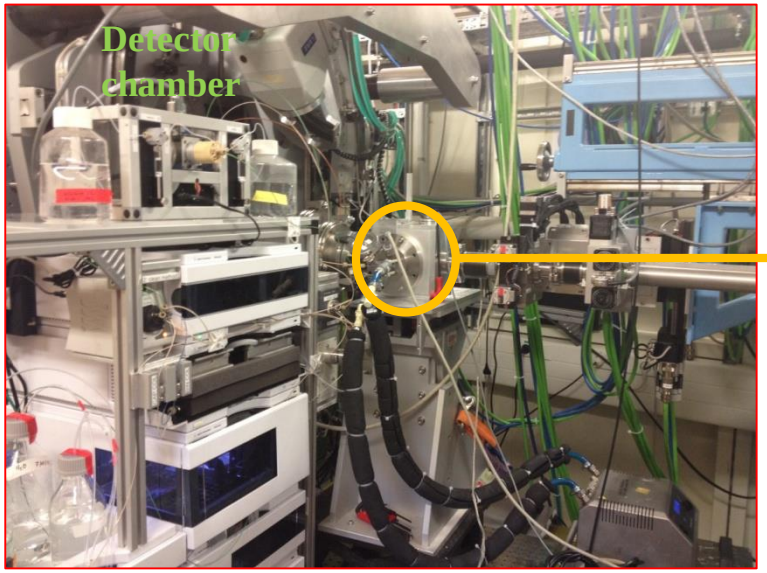


Sample environment dedicated to the biology :

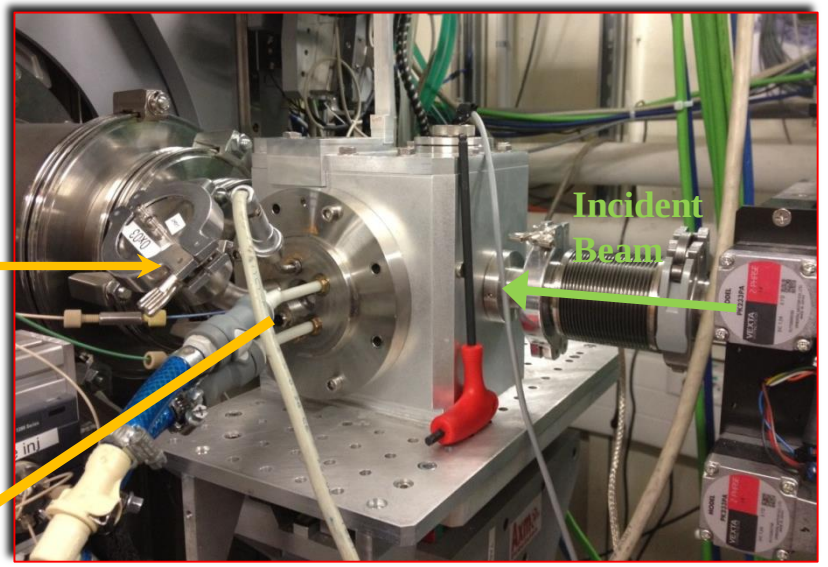


Set-up for BioSAXS at Beamline SWING

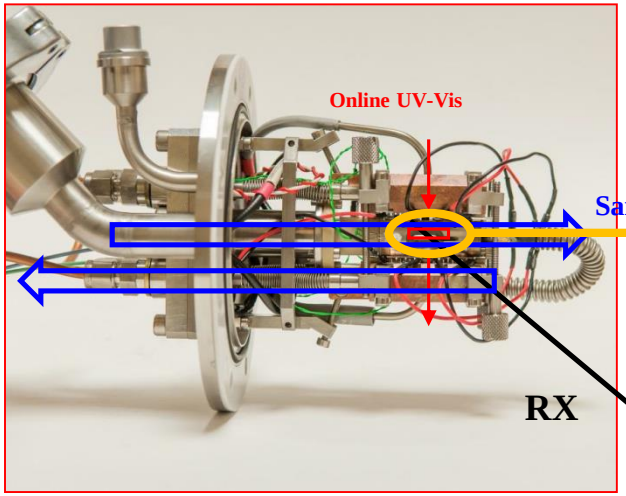
G. David and J. Pérez (2009), J. Appl. Cryst



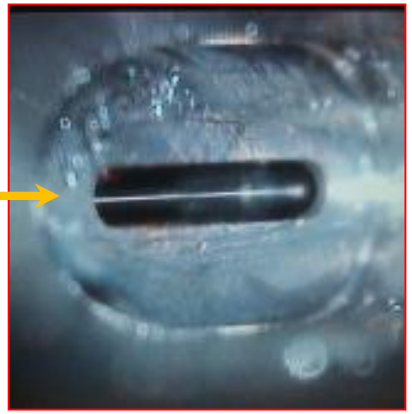
SEC-SAXS



BioSAXS Vacuum chamber



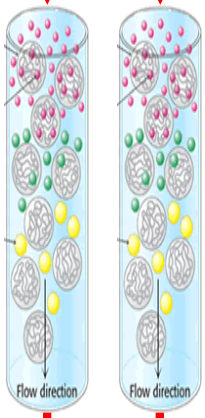
Details of the BioSAXS cell



Quartz capillary

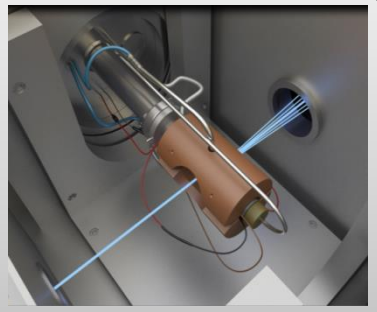
Bio-SAXS schematics

Samples for SEC-SAXS:
10µl (>20mg/ml) to 80 µl for (>1.0mg/ml)



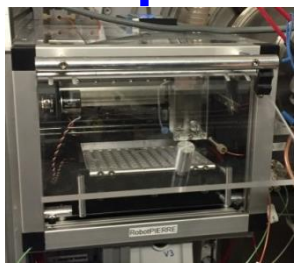
MALLS detector (18 angles+QELS)

UV detector (8 wavelengths)



SAXS

RI detector



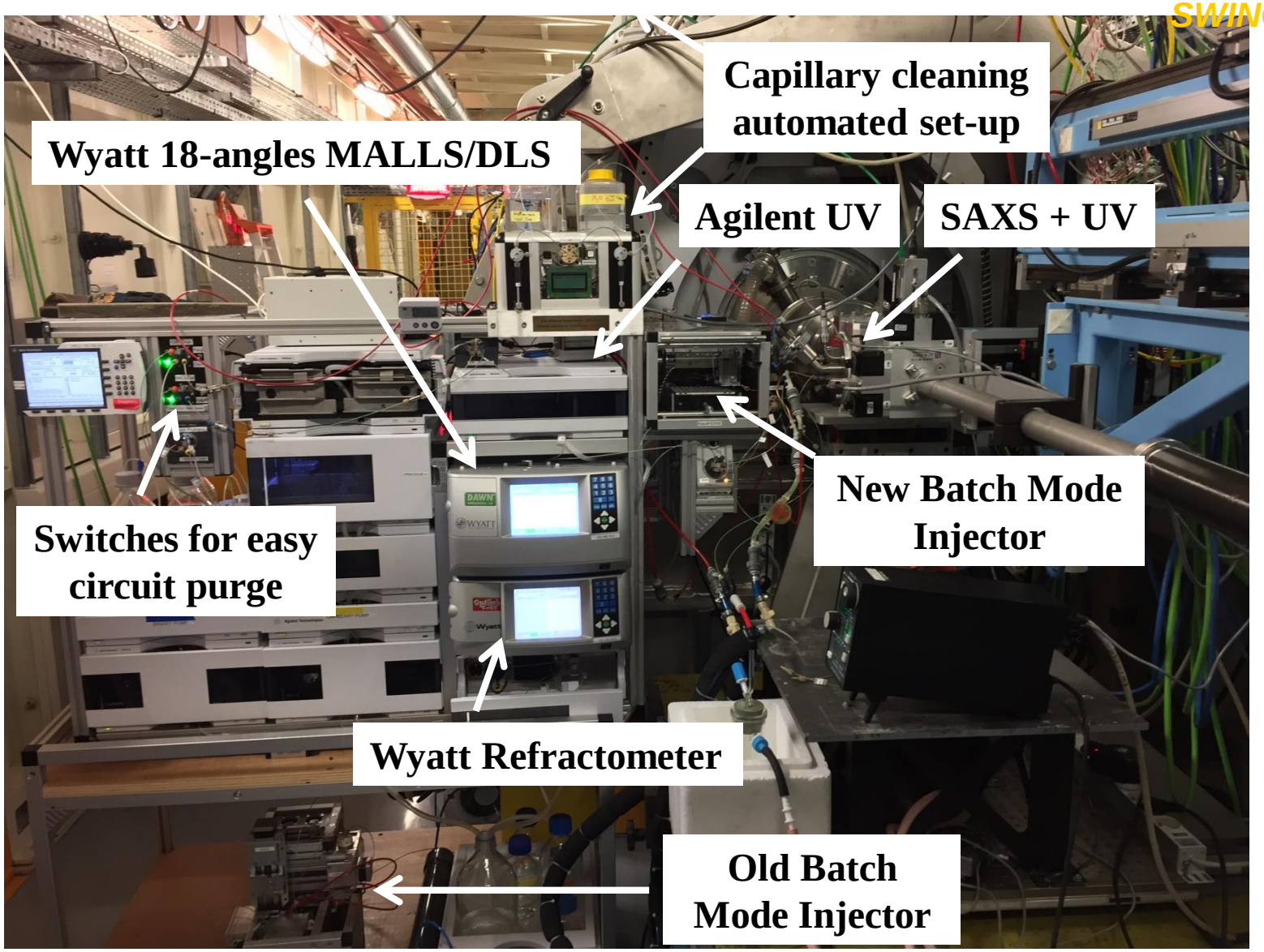
Samples for BATCH-SAXS:
10µl (>1 mg/ml) to 50 µl for (>0.1 mg/ml)
Less than 3 minutes per sample.

Automated process for injection and detection (passerelle GUI):

- Automated injection for both Batch and SEC-SAXS
- Users just need to define a list of: Sample names, Rack Positions, Volumes and Column circuits
- Multiple injections for both Batch and SEC-SAXS can be programmed

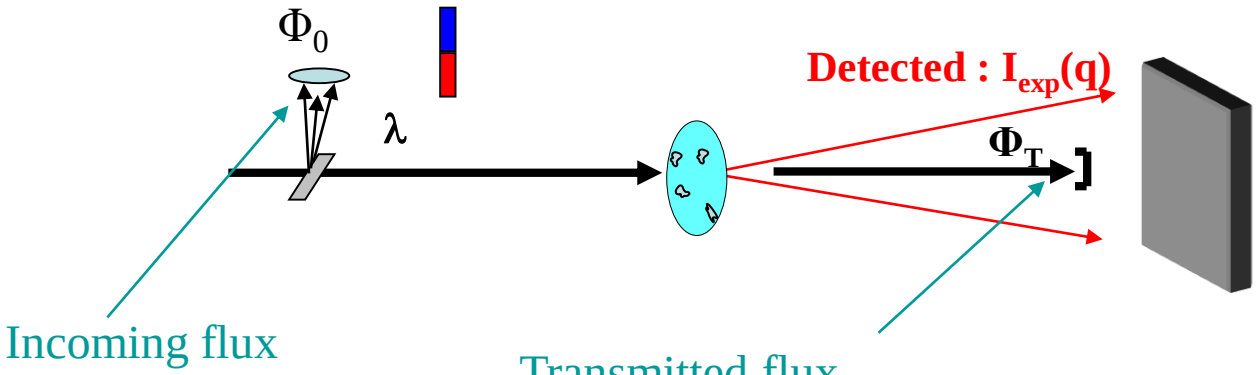
⇒ 46 proteins recorded using only SEC-SAXS in 21 hours.

HPLC – MALLS/QELS – SAXS – RI online



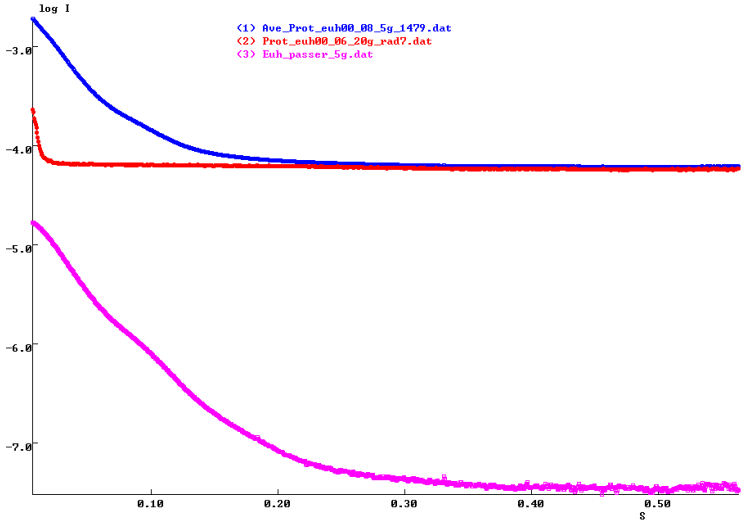
Transmission and buffer measurements are crucial

- Transmission
  - The experimental scattering intensity must be normalised by transmitted intensity.
  - Transmission intensity must be measured with high accuracy (~ 0.1 %).



- Buffer
  - Buffer and protein samples must be measured in the same cell for correct subtraction of parasitic background arising from slits and holder walls.
  - The buffer in the buffer sample must be identical to that of the protein sample (dialysis, SEC, …).

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

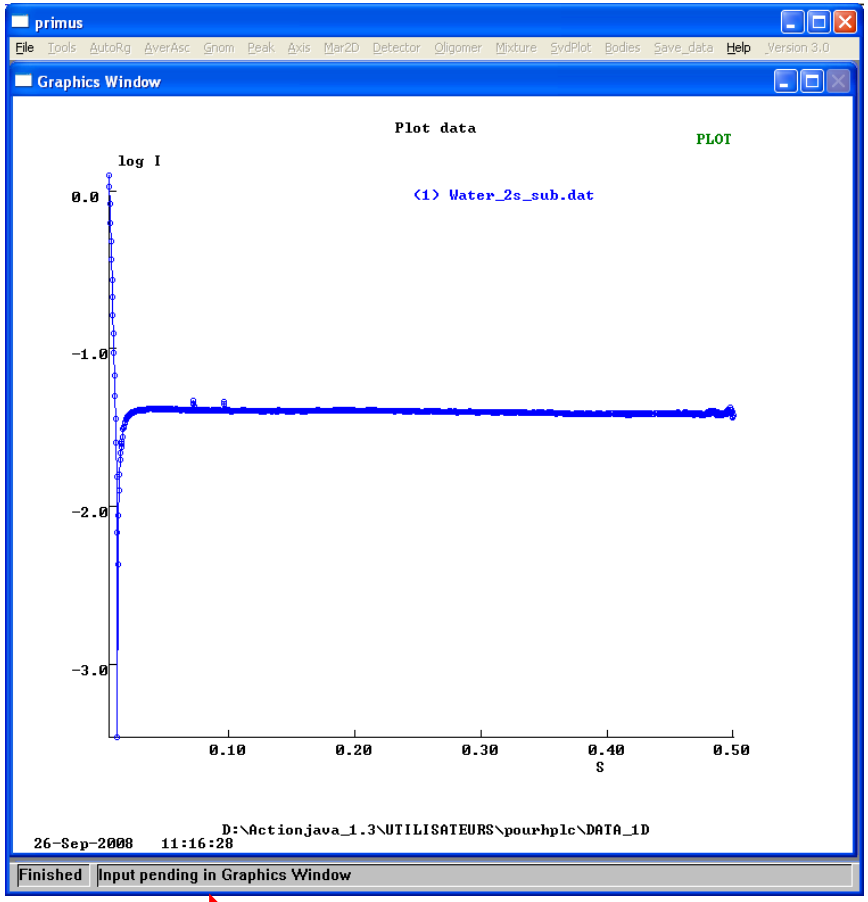


Calibration of the set-up using water scattering

Liquid scattering (theory): $I(q) = \text{constant at small } q = r_0^2 Z^2 \rho_A^2 \cdot kT\kappa_T$

Molecular density — Isothermic compressibility

$I_{H2O,theory} = 0.0163\ cm^{-1}$



Water is used as primary reference to get the absolute intensity scale

- Capillary diameter =1.6 mm
- Average of 2 frames of 2s
- Empty capillary subtracted
- Normalized by solid angle
- Normalized by transmitted intensity

Example:

$I_{H2O,exp} = 0.042$ Exp. Units

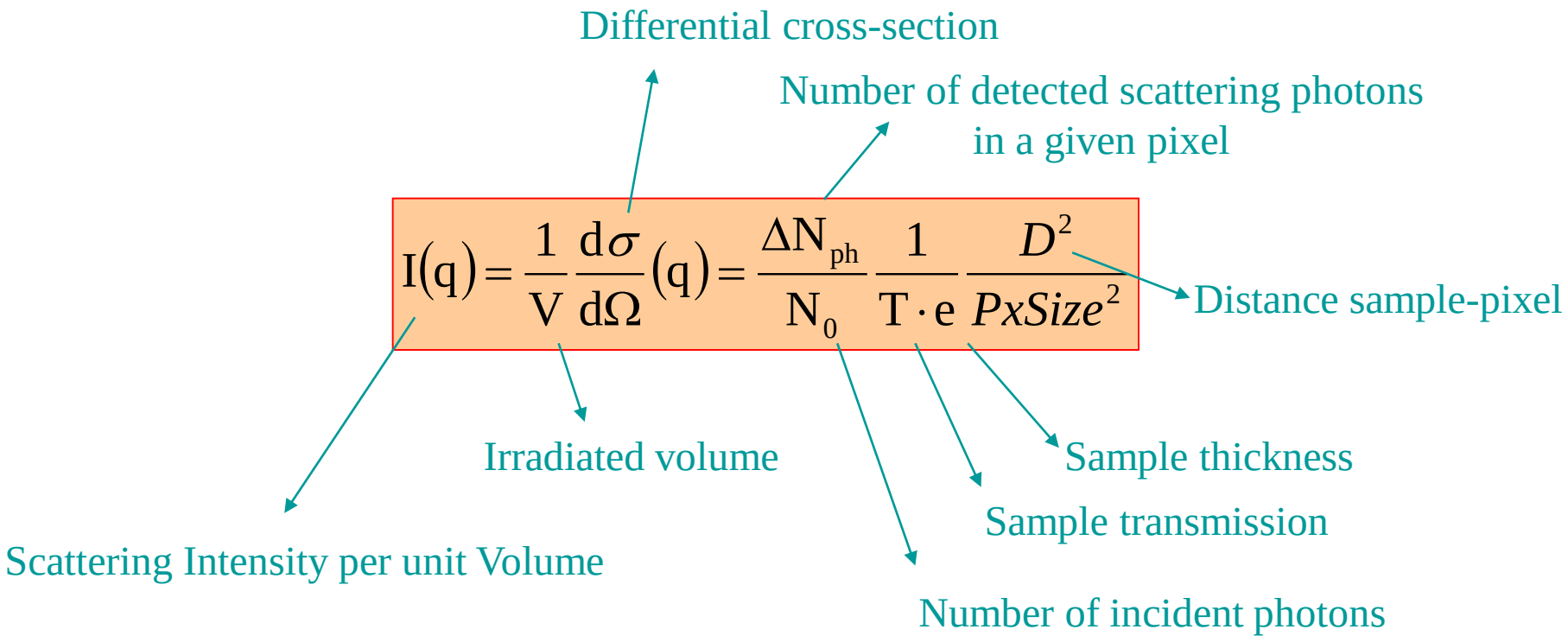
$I_{H2O,exp} = K_{exp} * I_{H2O,theory}$

→ Here : $K_{exp} = 2.56$ Exp.Units / cm^{-1}

For any sample in that capillary : $I_{theory}(cm^{-1}) = I_{exp} / K_{exp} = I_{exp} / 2.56$

Particles in solution

Relation between the the number of measured photons ΔN_{ph} on a given pixel of the detector, making a solid angle $\Delta\Omega$, and the Scattering Intensity per unit volume :

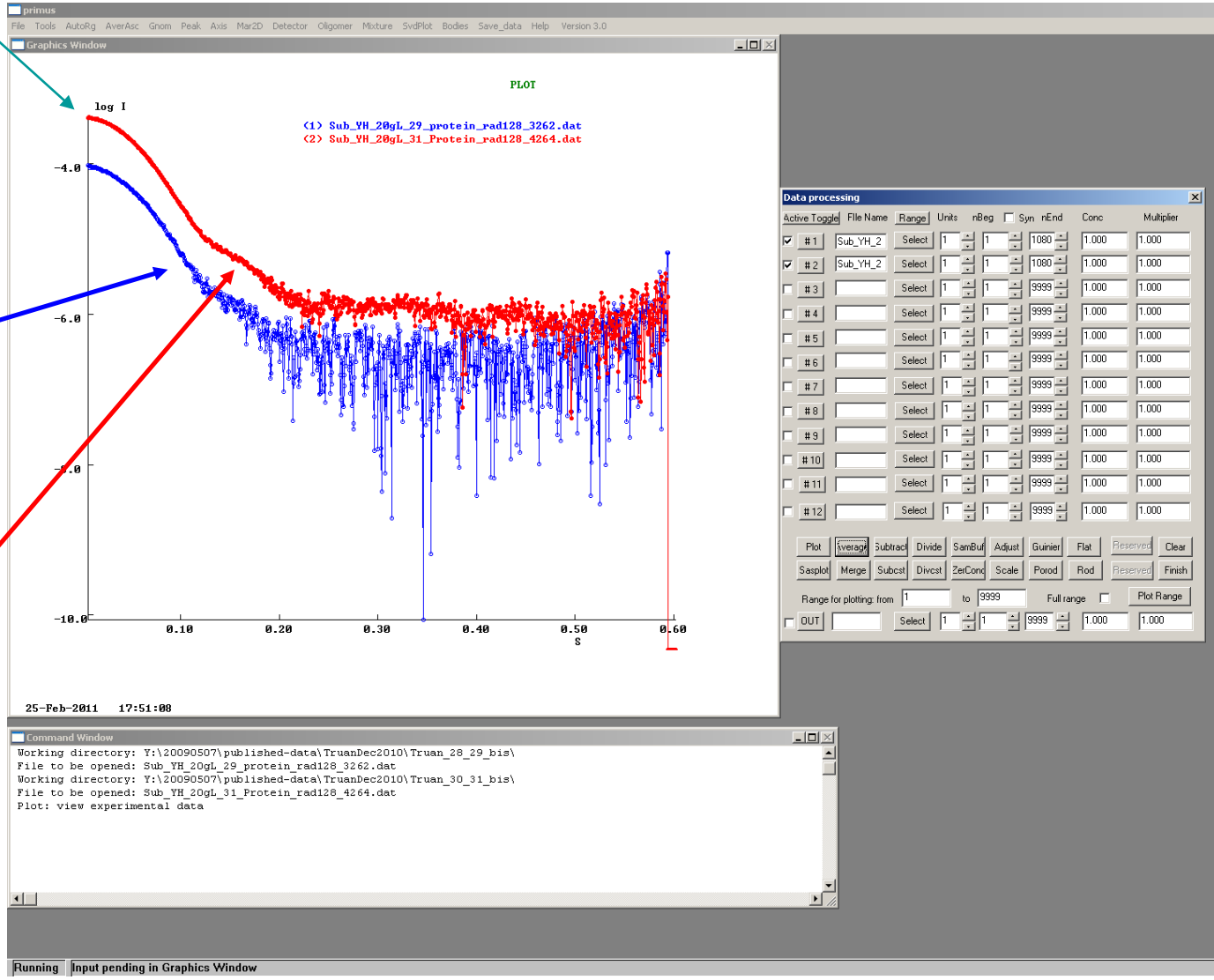


Merge low c (@ low q) and high c (@ high q) curves

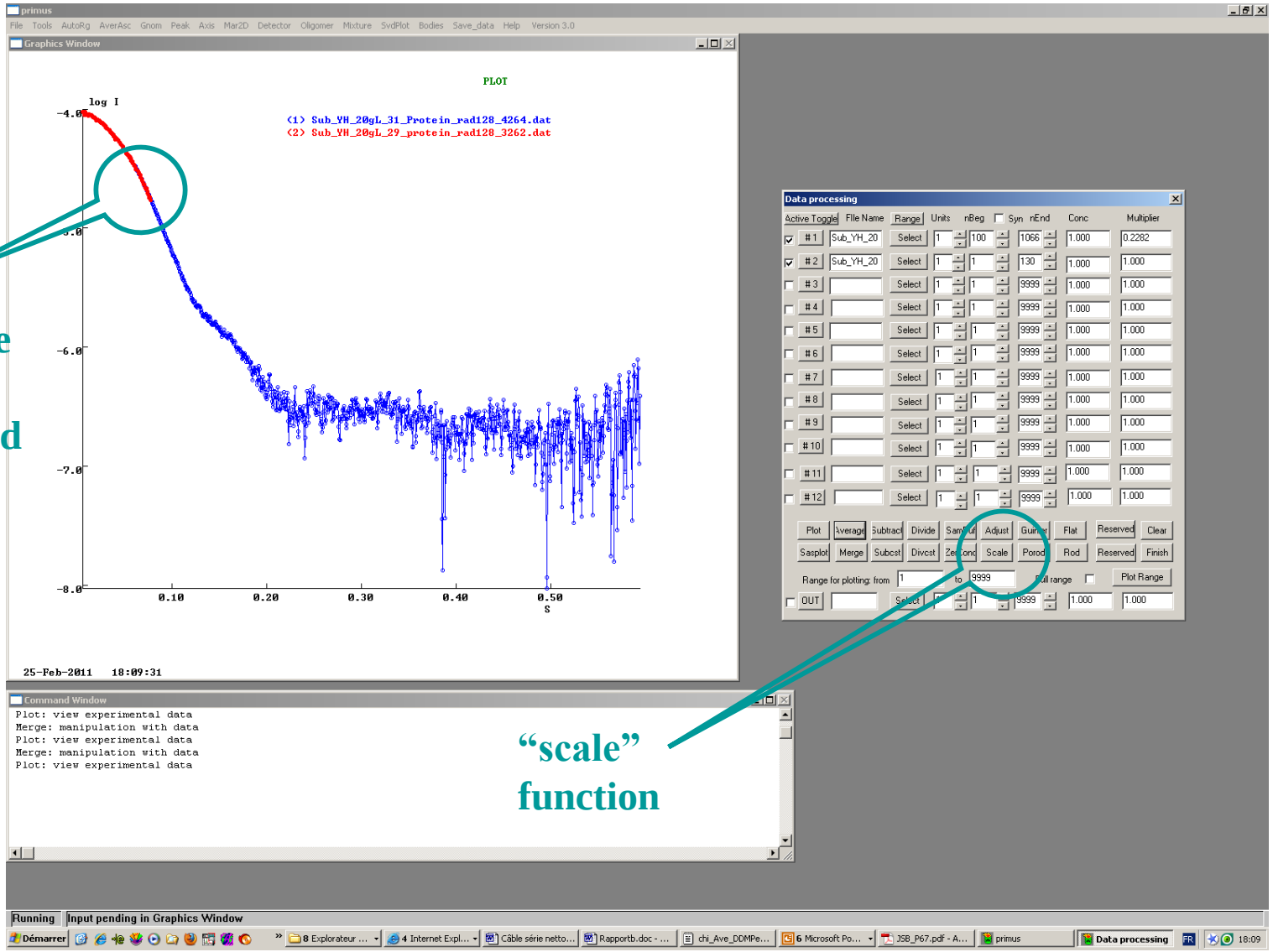
Here, slight repulsive interactions alter the concentrated curve at small angles

small angle data using the lowest concentration curve or an extrapolation to zero concentration from a series of dilute solutions (correction of interparticle effects)

larger angle data using the most concentrated solution



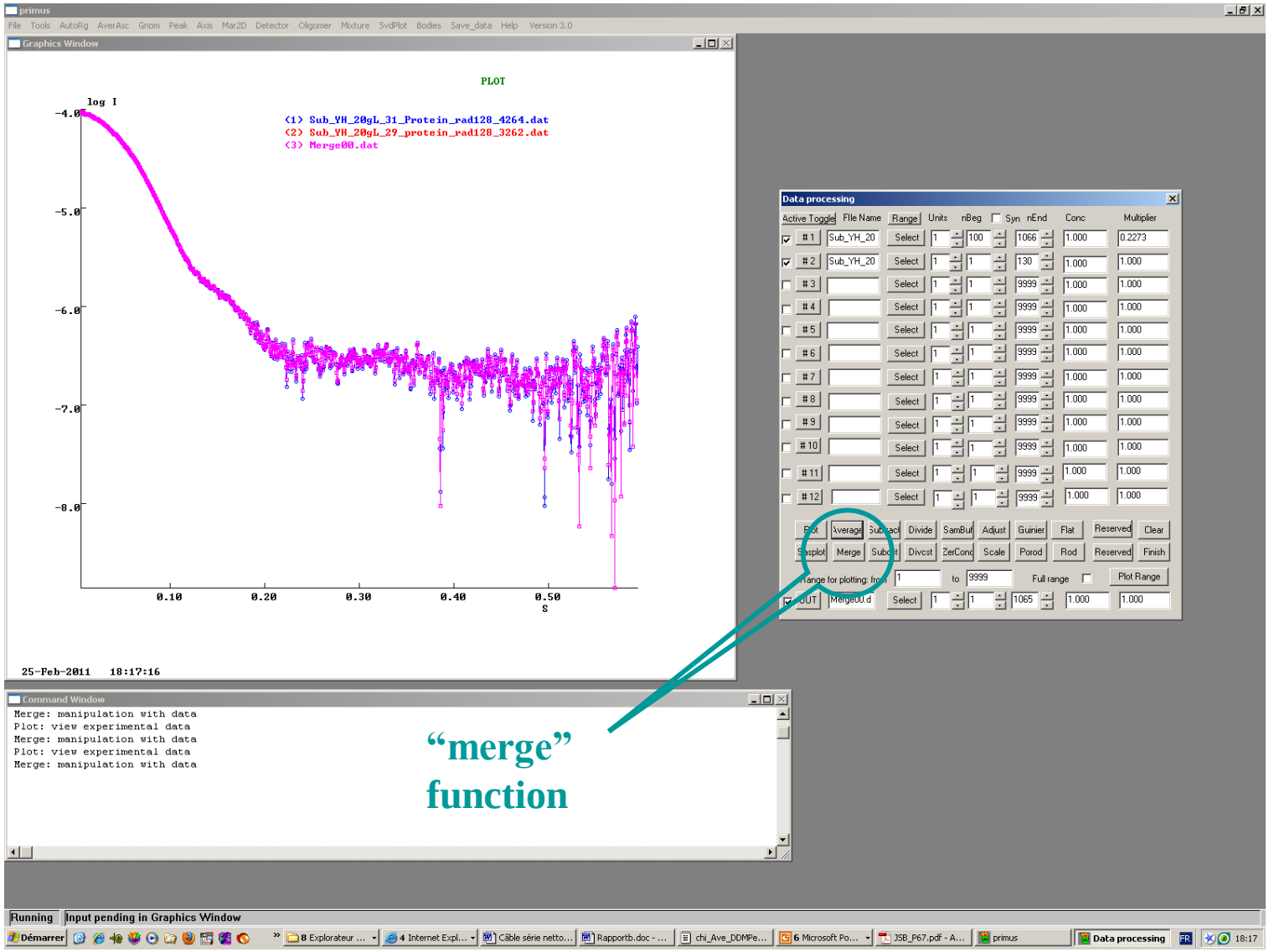
PRIMUS: merging data



The common range should be as restricted as possible to avoid adding noise

“scale” function

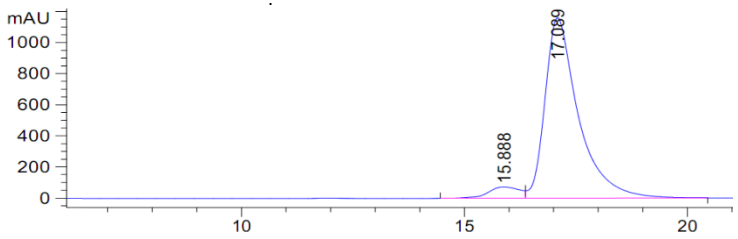
PRIMUS: final merged curve



2nd case : the solution is a slow equilibrium or an unwanted mixture

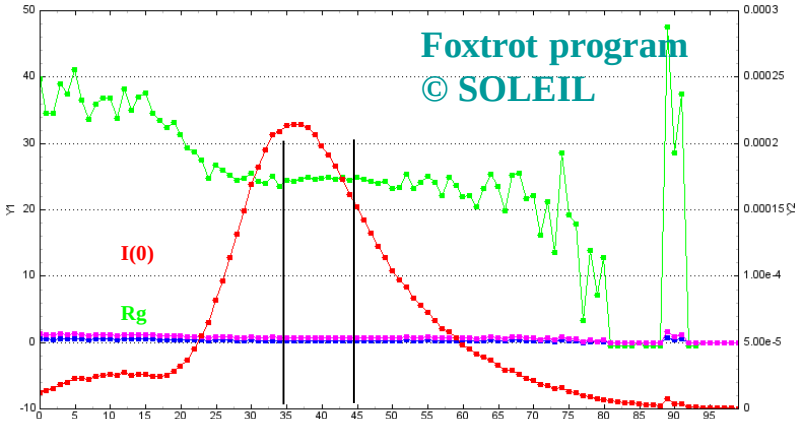
- Use on-line HPLC data collection (typ 50 µl)

ASNP elution profile, monitored by UV absoprtion at 280 nm



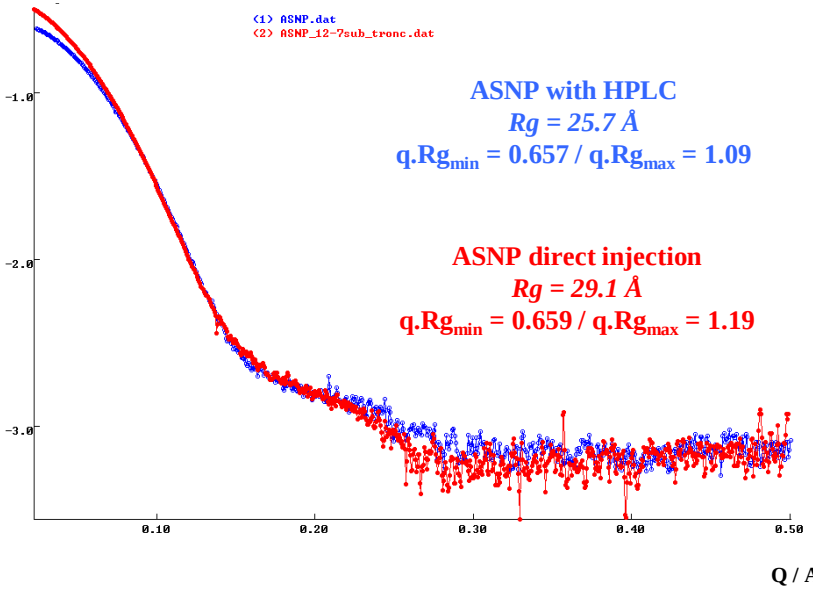
Elution time (minutes)

I(0) and Rg determined for each SAXS frame during elution

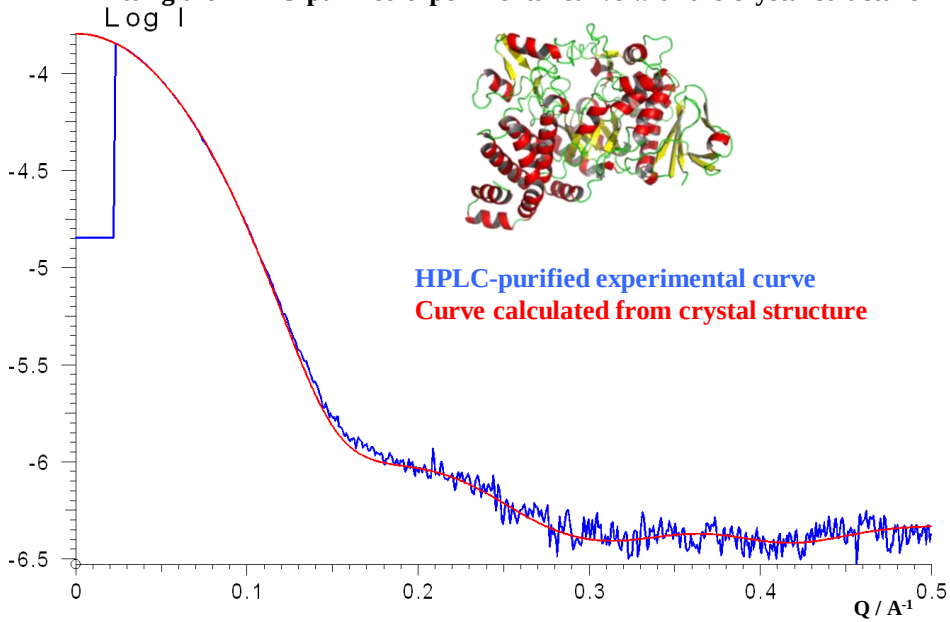


Frame number (~ time)

Comparison between HPLC-purified and Direct injection curves

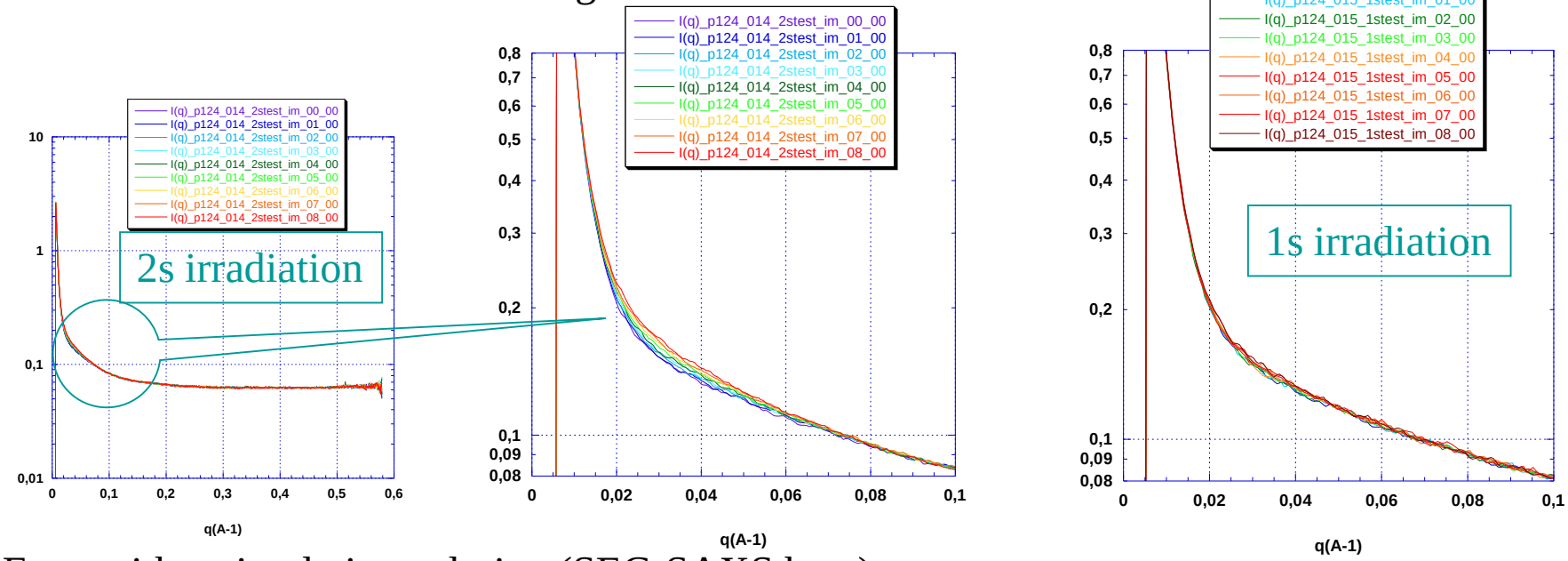


Fitting the HPLC-purified experimental curve with the crystal structure

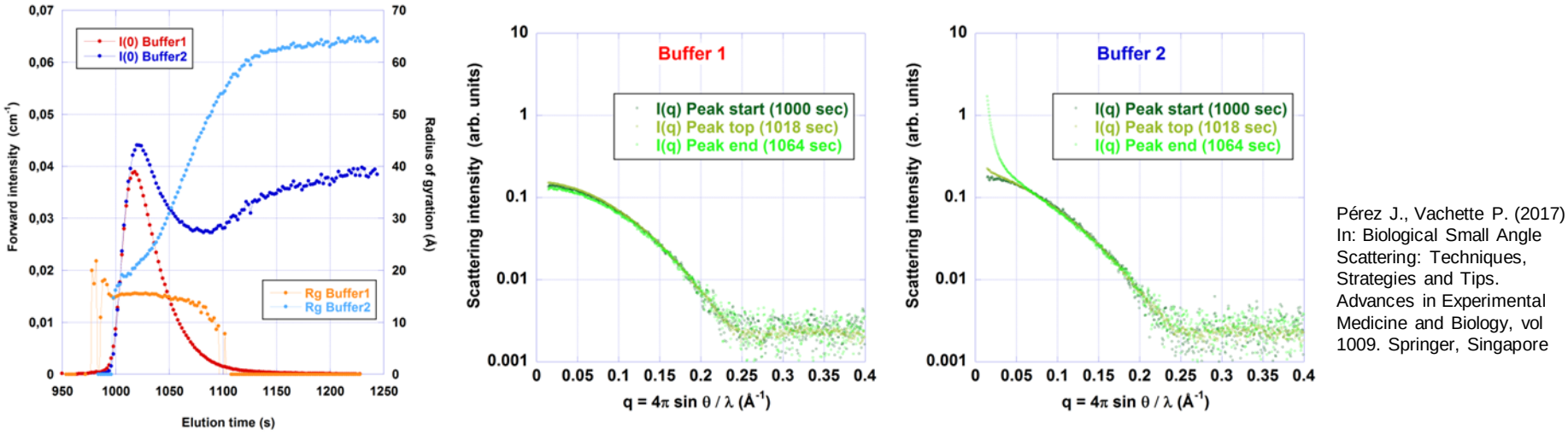


Unwanted Radiation effects

Data collection without circulating the solution



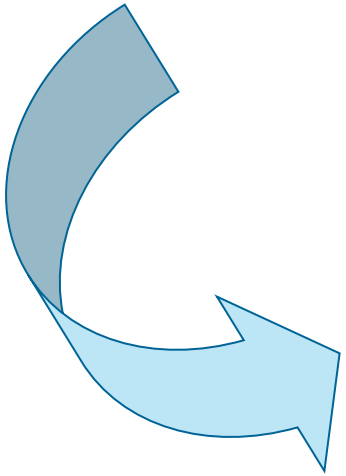
Even with a circulating solution (SEC-SAXS here)



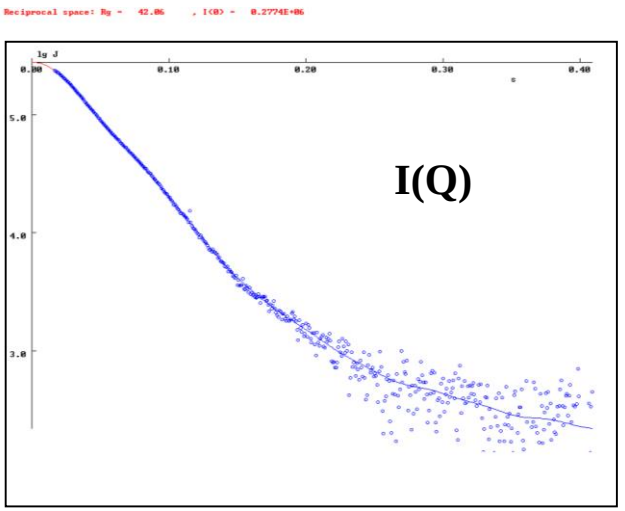
Pérez J., Vachette P. (2017) In: Biological Small Angle Scattering: Techniques, Strategies and Tips. Advances in Experimental Medicine and Biology, vol 1009. Springer, Singapore

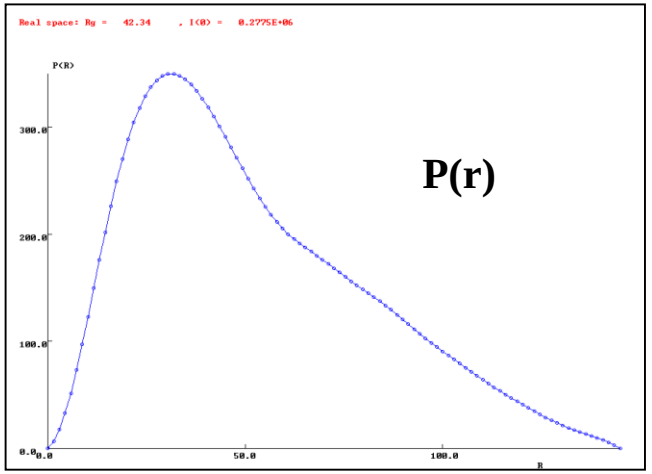
At this stage

We have gone from

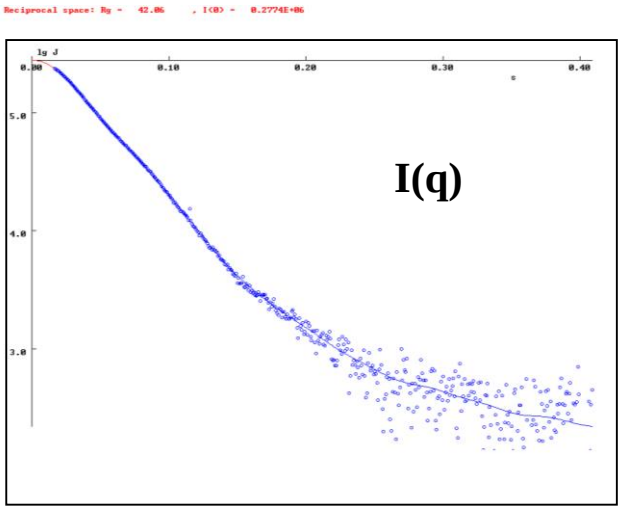


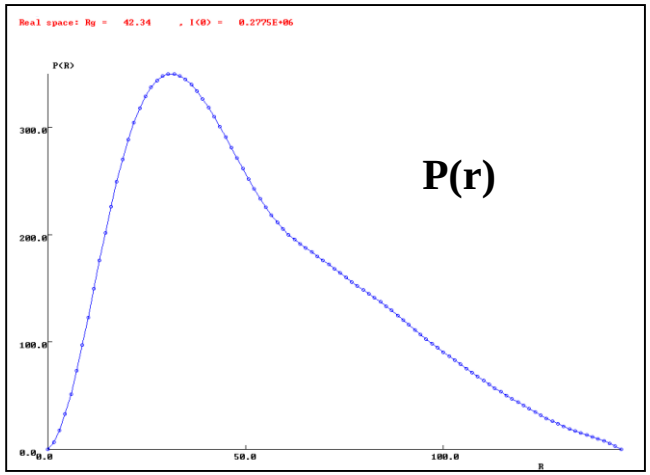
to



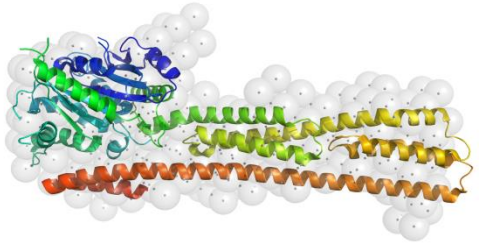


Now, we have to go from





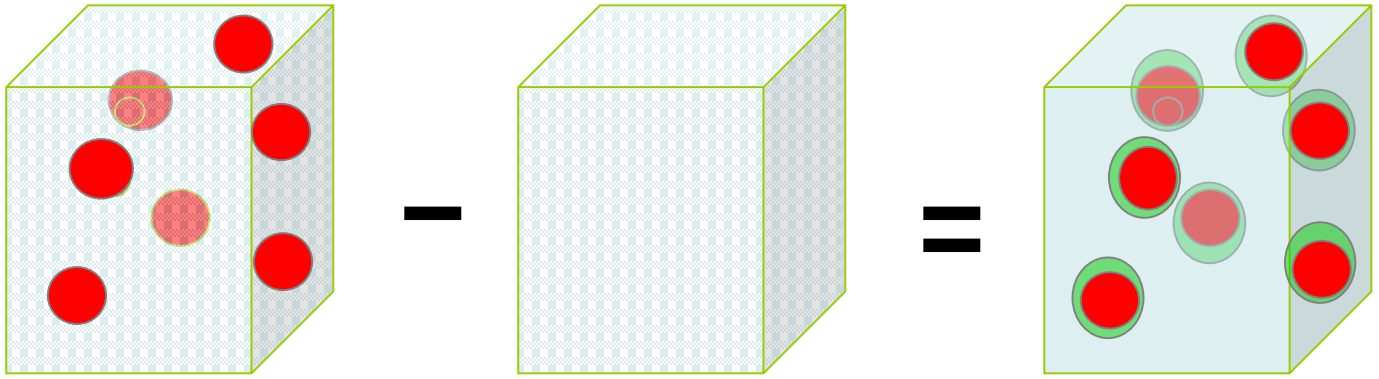
to



MODELLING

Slide slightly modified from Dmitri Svergun, EMBL Hamburg

Solvent scattering and contrast



$I_{solution}(s)$ $I_{solvent}(s)$ $I_{particle}(s)$

To obtain scattering from the particles, solvent scattering must be subtracted to yield the effective density distribution $\Delta\rho(r) = \rho(r) - \rho_0$, where ρ_0 is the scattering density of the solvent.

Further, the bound solvent density may differ from that of the bulk.

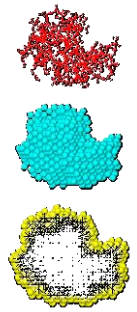
CRYSOL: from atomic coordinates to a SAXS curve

$$I_{th}(q) = \left\langle \left| A_a(\vec{q}) - \rho_s A_s(\vec{q}) + \delta\rho_b A_b(\vec{q}) \right|^2 \right\rangle_{\Omega}$$

$A_a(\boldsymbol{q})$ = *molecular scattering amplitude in vacuum*

$A_s(\boldsymbol{q})$ = *scattering amplitude from excluded volume*

$A_b(\boldsymbol{q})$ = *scattering amplitude from the hydratation shell, layer of arbitrary thickness 3Å*



In CRYSOL program, in order to gain computing time, I(q) is developed in a series of Bessel functions and spherical harmonics :

$$I_{calc}(q) = \sum_{l=0}^{L} \sum_{m=-1}^{l} \left| A_{lm}(q) - \rho_0 C_{lm}(q) + \delta\rho B_{lm}(q) \right|^2$$

The experimental scattering curves are then fitted using only 3 parameters in order to minimize the discrepancy χ :

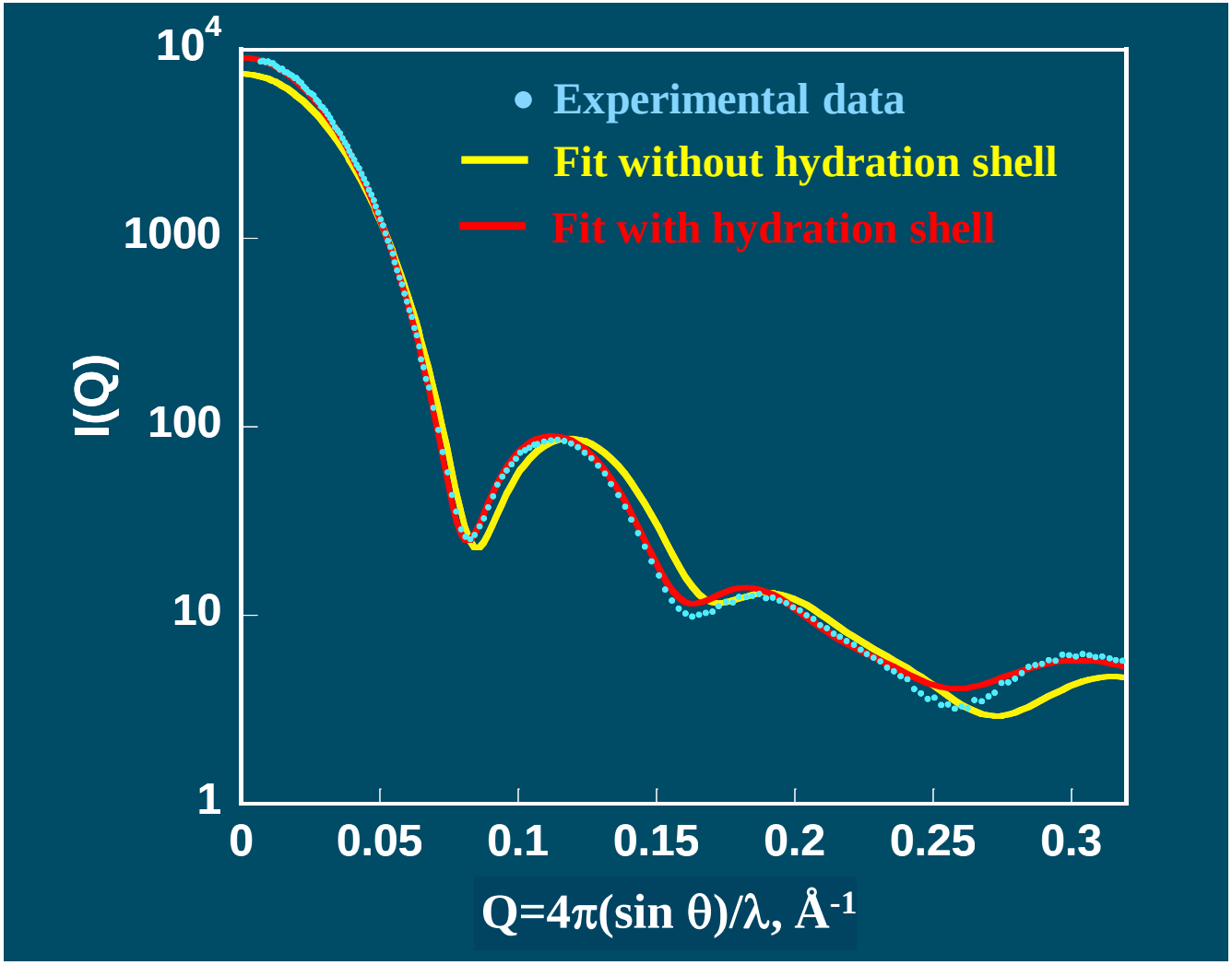
- the general scale of $I_{calc}(q)$
- the total excluded volume V, which is equivalent to modifying the average contrast ρ_0
- the contrast of the border layer $\delta\rho$

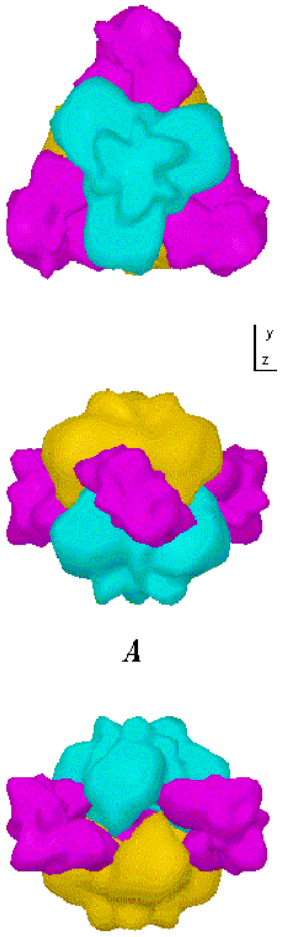
$$\chi^2 = \frac{1}{N-1} \sum_{i=1}^{N} \left[\frac{I_{\exp}(q_i) - scale * I_{calc}(q_i)}{\sigma_{\exp}(q_i)} \right]$$

Svergun , Barberato & koch (1995), J. Appl. Cryst., ***28****, 768*

Effect of the hydration shell

T state of *E. coli* allosteric ATCase





3D shape reconstructions from SAXS data with DAMMIN

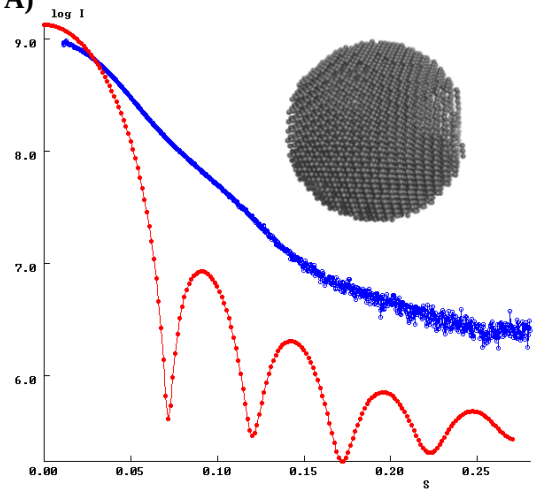
Ab initio shape modelling : nothing is known excepted the curve !

Principle of the method : any structure volume of homogeneous electronic density can be approximated at any resolution by a set of spheres of small enough diameter

Starting model = sphere with a radius R = Dmax/2 with N scattered beads ($r_0 << R$)
The number of the "dummy atom" $N \approx (R/r_0)^3$

Each sphere is associated to a position j and an index X_j corresponding to the type of the phase ($X_j = 0$ for the solvent and $X_j = 1$ for the molecule)

A)

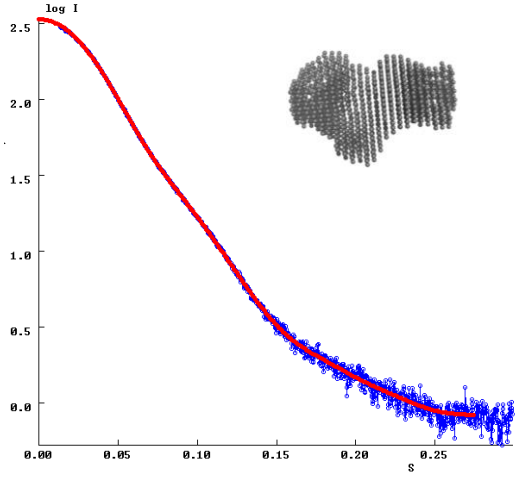


$$f(X) = \chi^2\left[I(q)_{\exp}, I(q,X)\right] + \alpha P(X)$$

X is a conformation of the system
P(X) is a penalty function

$$\chi^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left[\frac{I_{\exp}(q_i) - scale * I_{calc}(q_i)}{\sigma_{\exp}(q_i)}\right]$$

After k iterations



D. I. Svergun, M. Kozin, M. Petoukhov, V. Volkov (1999). Biophys J. 2879-2886.

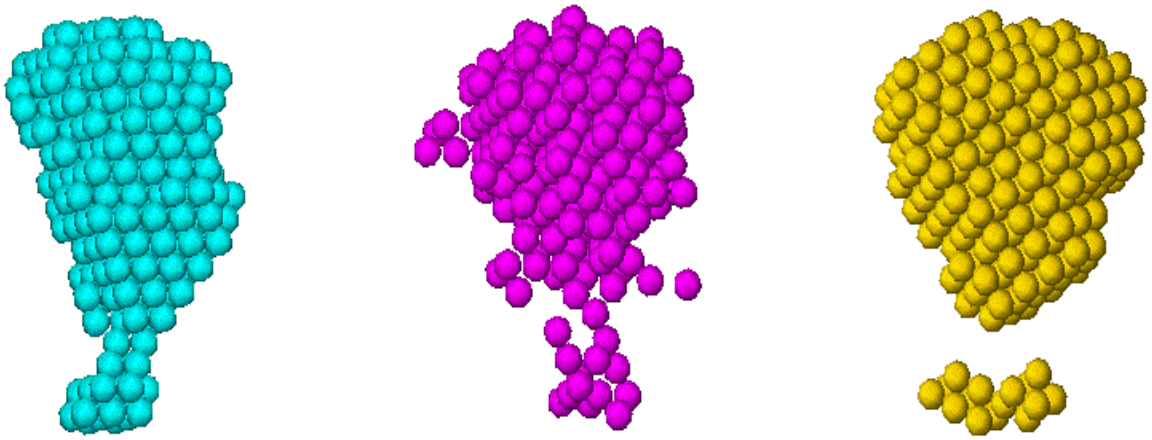
3D shape reconstructions from SAXS data with DAMMIN

- Obtaining 3D shapes from SAXS data is a ill-defined problem that can be **partially** solved by introducing additional information to **reduce** ambiguity of interpretation

- Using simulated annealing, finds a compact dummy atoms configuration X that fits the scattering data by minimizing

$$f(X) = \chi^2[I_{\exp}(s), I(s, X)] + \alpha P(X)$$

where χ is the discrepancy between the experimental and calculated curves, $P(X)$ is the penalty to ensure compactness and connectivity, $\alpha > 0$ its weight.



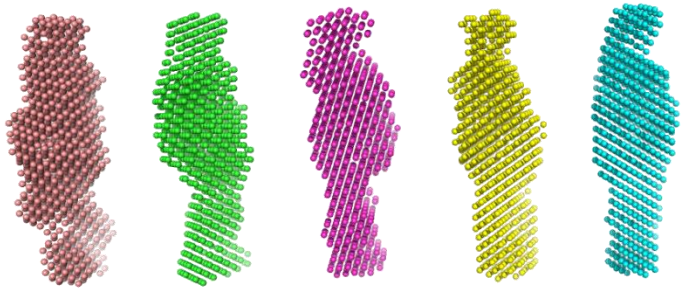
compact

loose

disconnected

3D shape reconstructions from SAXS data with DAMMIN

- A series of runs (10-50) are performed to compare the different shapes obtained from the same data.
- After the run, an optimal superposition of models is realized with the program suite DAMSEL and DAMSUP.
- The algorithm defines a criteria of similarity, called « Normalized Spatial Discrepancy » or NSD, which measures the agreement between any pair of models.
- Similar shapes results in NSD < 1, very similar shapes NSD ≈ 0.5



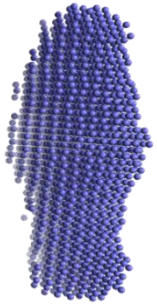
Shp1 Shp2 Shp3 Shp4 Shp5

File	Aver	1	2	3	4	5
1	0.52	0.00	0.51	0.52	0.50	0.52
2	0.52	0.51	0.00	0.52	0.49	0.52
3	0.53	0.52	0.52	0.00	0.53	0.52
4	0.53	0.50	0.49	0.53	0.00	0.54
5	0.53	0.52	0.52	0.52	0.54	0.00

Mean value of NSD : 0.535
Standard deviation of NSD : 0.008

Damsel.log

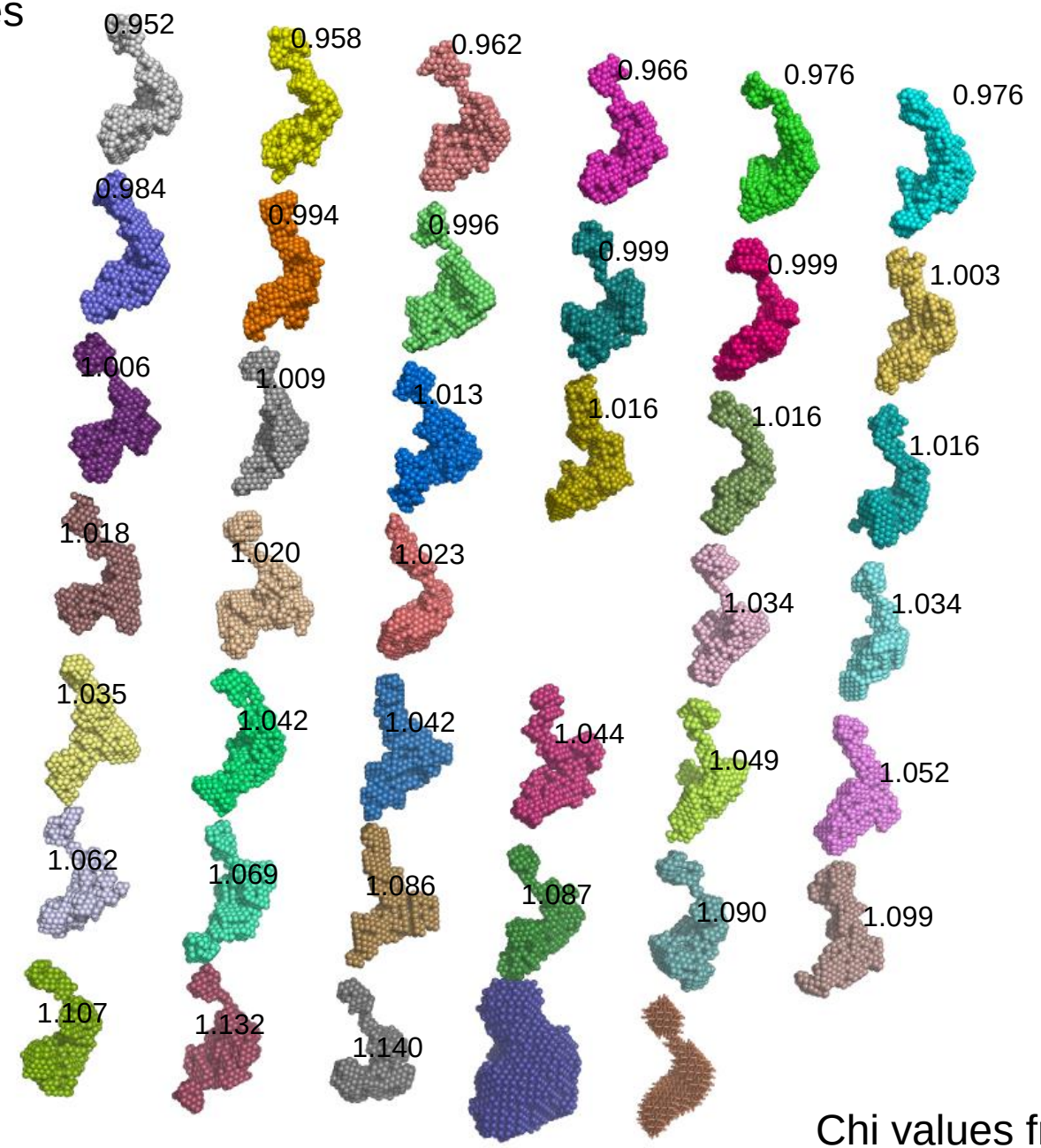
Damfilt
(intersection)



Damaver
(all superimposed)

- Models are conserved if its NSD < Mean of NSD + 2*standart deviation
- The model with the lowest NSD is the shape which has the most similarities with other, and **can** be regarded as the most representative of envelopes in accordance with the SAXS data
- Be careful with damfilt.pdb because $I_{damfilt}(q) \neq I_{exp}(q)$

NSD values

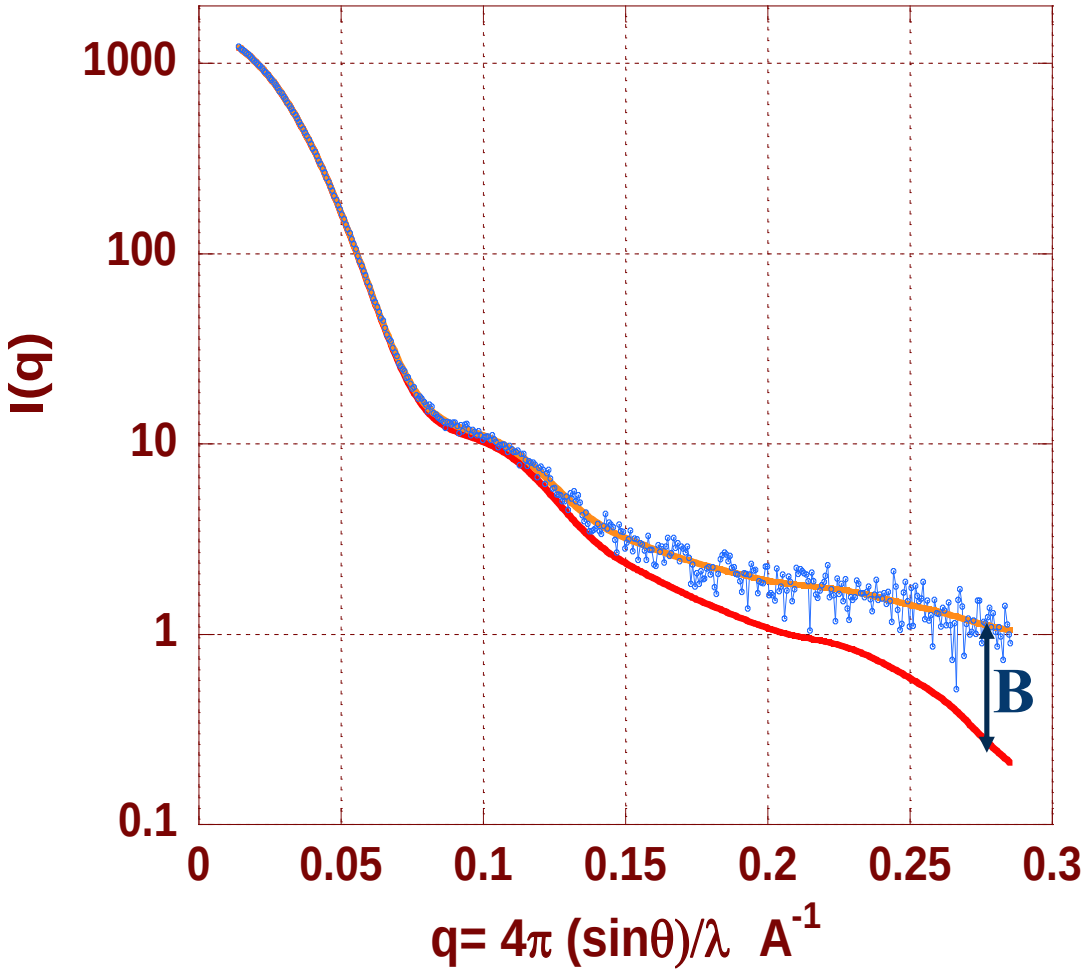


Chi values from 1.704 to 1.942

Be aware : “Porod law” is forced for ab initio shape determination

DAMMIN : shape determination
Model with uniform density

Fitting data with approximate q^{-4} high angle trend by subtracting a constant.



SASREF: Rigid body modeling against SAXS data

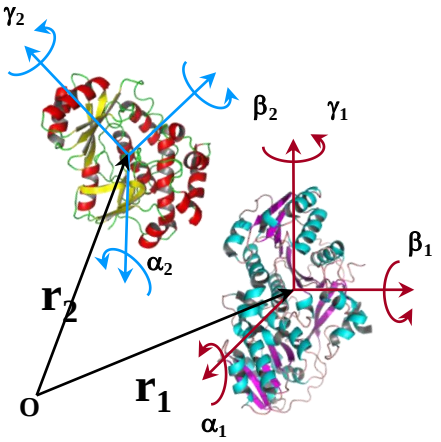
SASREF : when atomic structures of domains are known, but no their mutual organization

The objective is to find the relative orientation of each subunit with a correct agreement with the SAXS data of the complex

The scattering intensity I(q) of the complex is equal to the sum squared of the amplitudes of each subunit

$$I(S) = \left\langle \left| \sum_{k=1}^{K} A^{(k)}(\vec{S}) \right|^2 \right\rangle_{\Omega}$$

$$A^{(k)}(\vec{S}) = \exp(i.\vec{S}.\vec{r}_k) \prod (\alpha_k.\beta_k.\gamma_k)[C^{(k)}(\vec{S})]$$



The amplitude are calculated with CRYSOL from the high resolution structure of each monomer

The algorithm of minimization is the same used with DAMMIN with a penalty function (interconnectivity of the subunits, the steric clashes) and possibility to give information about contacting residues from other experiences.

$$f(X) = \sum_i \chi_i^2 + \alpha_{dist} P_{dist}(X) + \beta_{cross} P_{cross}(X) + \gamma_{cont} P_{cont}(X)$$

Petoukhov & Svergun (2005). Biophys. J., 89, 1237-1250.

DADIMODO : All-atom rigid body + linkers refinement *vs.* SAXS / NMR data

Collab : Christina Sizun & François Bontems (ICSN, Gif sur Yvette))

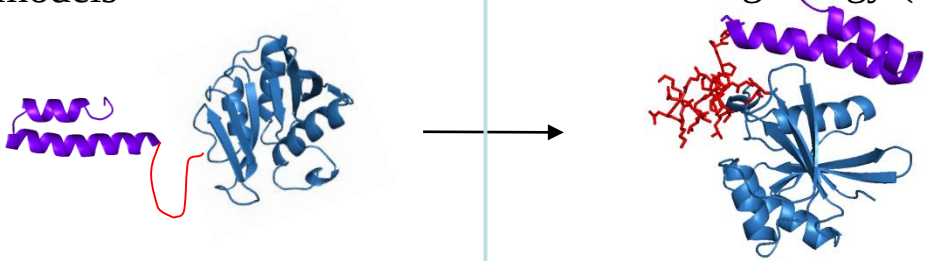
F. Mareuil, et al. (2007) *Eur Biophys J.*

Evrard et al. (2011), *J. Appl. Cryst.*

Modelling approach : complete atomic model

Full structure initiated with :

- Crystal or NMR domain structures
- Homology models



External information:

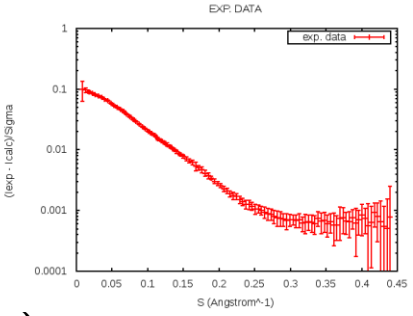
- Sequence
- Sub-parts moved as rigid-bodies (user-defined)
- A correct stereochemistry is maintained at all steps by minimizing energy (Amber 99 Force Field)

Experimental data:

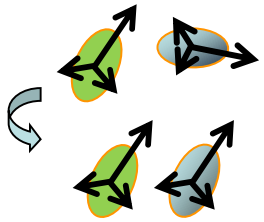
- SAXS
- NMR

  RDC

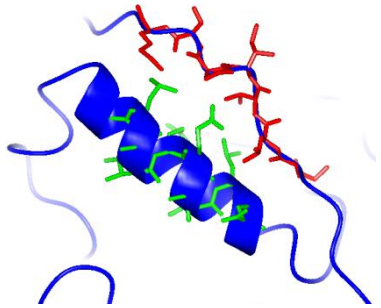
  ADR (chem. shift map.)



SAXS score



RDC score

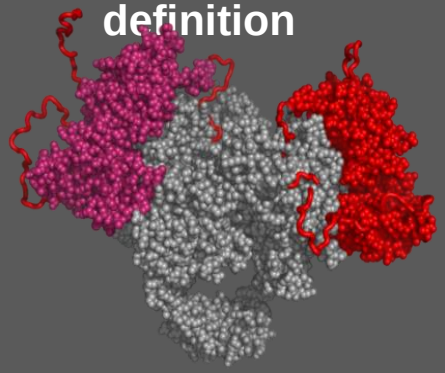


ADR score

Optimisation of the structure via a genetic algorithm

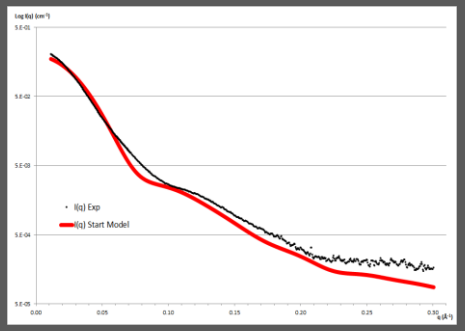
Dadimodo: refining domain orientations

Full atom Pdb file + rigid domains definition



+

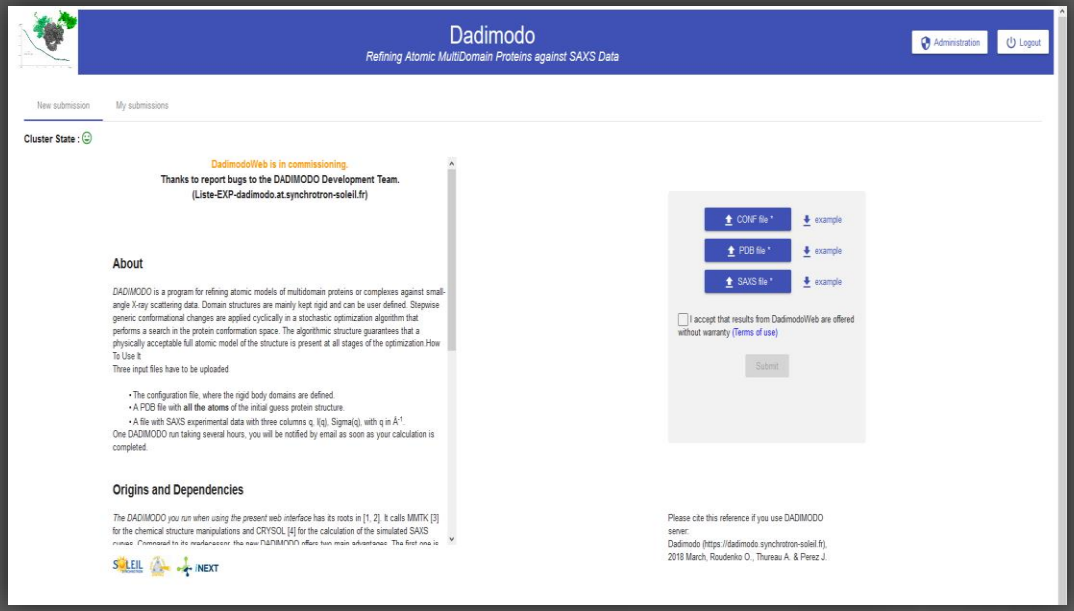
Saxs curve



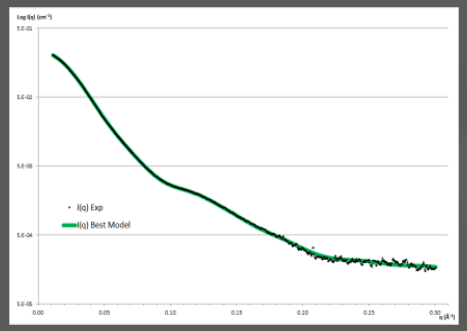
Petrella et al., 2019, Structure, *in press*

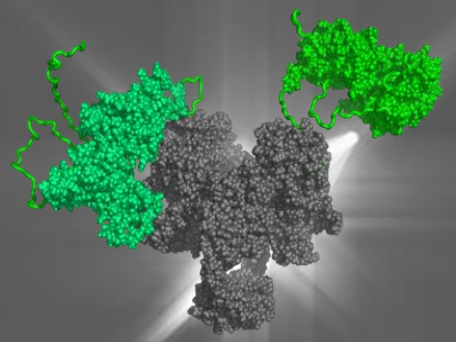
https://dadimodo.synchrotron-soleil.fr

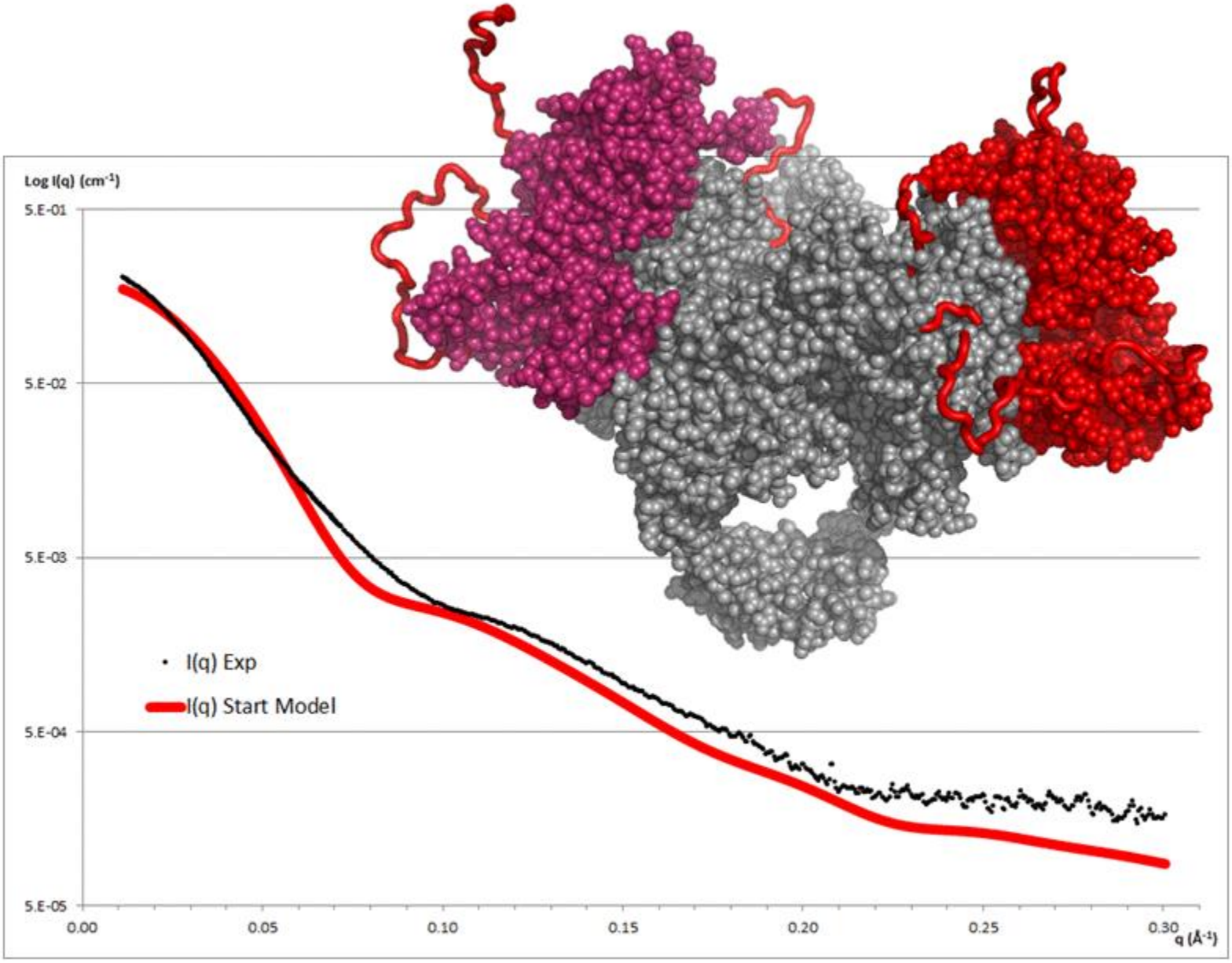
Roudenko O., Thureau A. & Pérez J., March 2018



5 SAXS compatible models

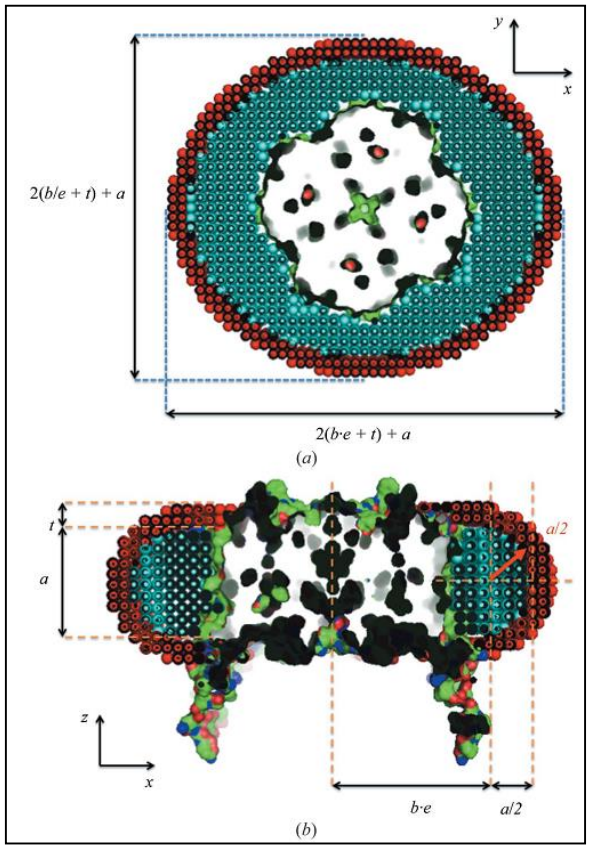




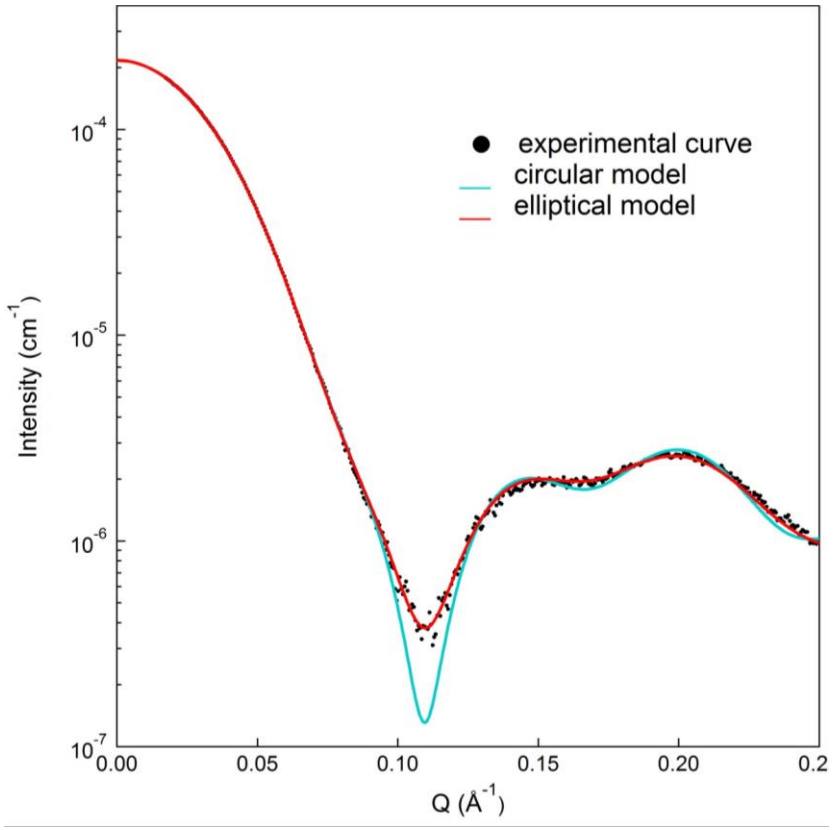

Log I(q) (cm-1)
5.E-01
5.E-02
5.E-03
5.E-04
5.E-05
0.00
0.05
0.10
0.15
0.20
0.25
0.30
q (Å-1)
I(q) Exp
I(q) Start Model

Memprot : a program to generate & optimize a detergent corona model

Pérez J. & Koutsioubas, A. (2014). Acta Cryst.D70



Algorithm of the *Memprot* program. The program essentially creates PDB files with the models made of the full-atom protein structure and the parameterized coarse-grained detergent corona, and *CRYSOL* is called to calculate the SAXS curves. An overall sorting on the χ value is performed to keep the best model.



Berthaud et al. (2012), *JACS*, 134 (24), 10080-10088

A FEW CONCLUSIONS

- **A scattering pattern can be accurately calculated from atomic coordinates, thereby providing a link between high resolution and BioSAXS work.**

- **Using SAXS patterns, a*b initio* methods can propose possible shapes of a macromolecule**

- **Based on all-atom partial structural information, stochastic algorithms may provide possible models for missing regions**

- **Analysis and modeling require a monodisperse and ideal solution, which has to be checked <u>independently</u>.**

SAXS is at his best when it is used to distinguish between several preconceived hypotheses.