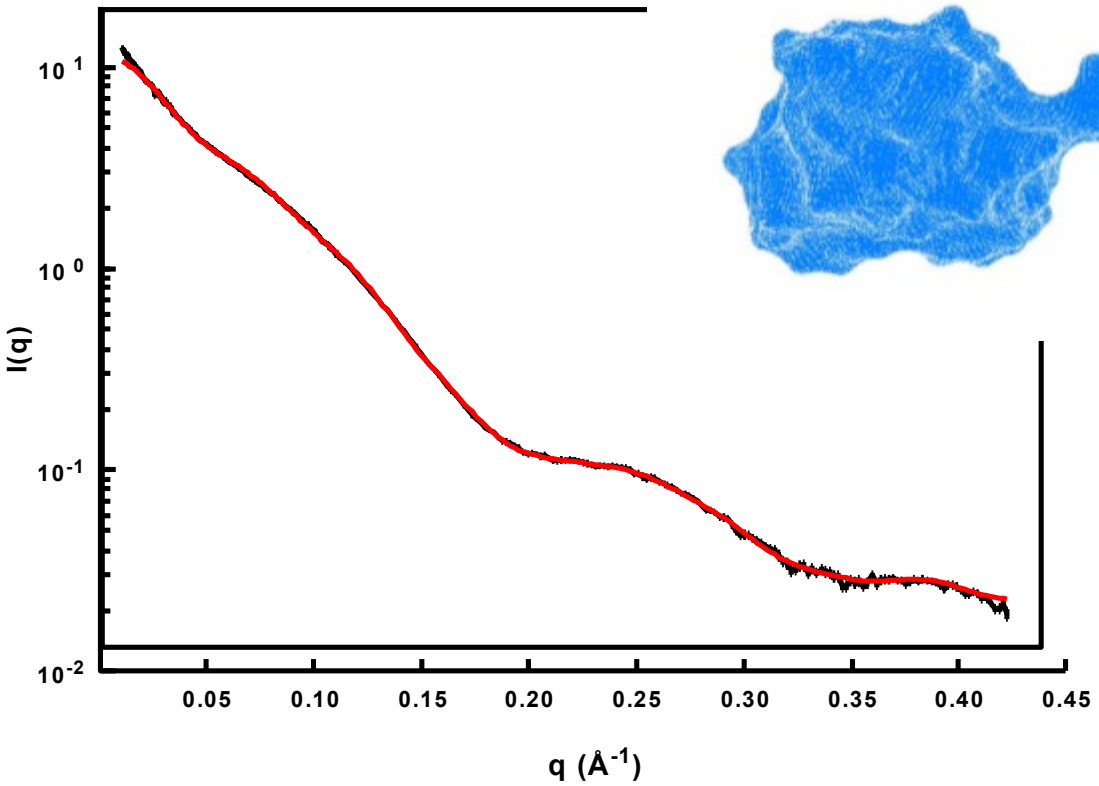
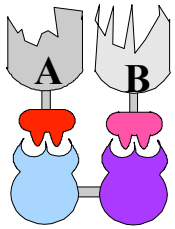
The basis of Small Angle X-ray Scattering and comparison to X-ray crystallography

advantages and limits

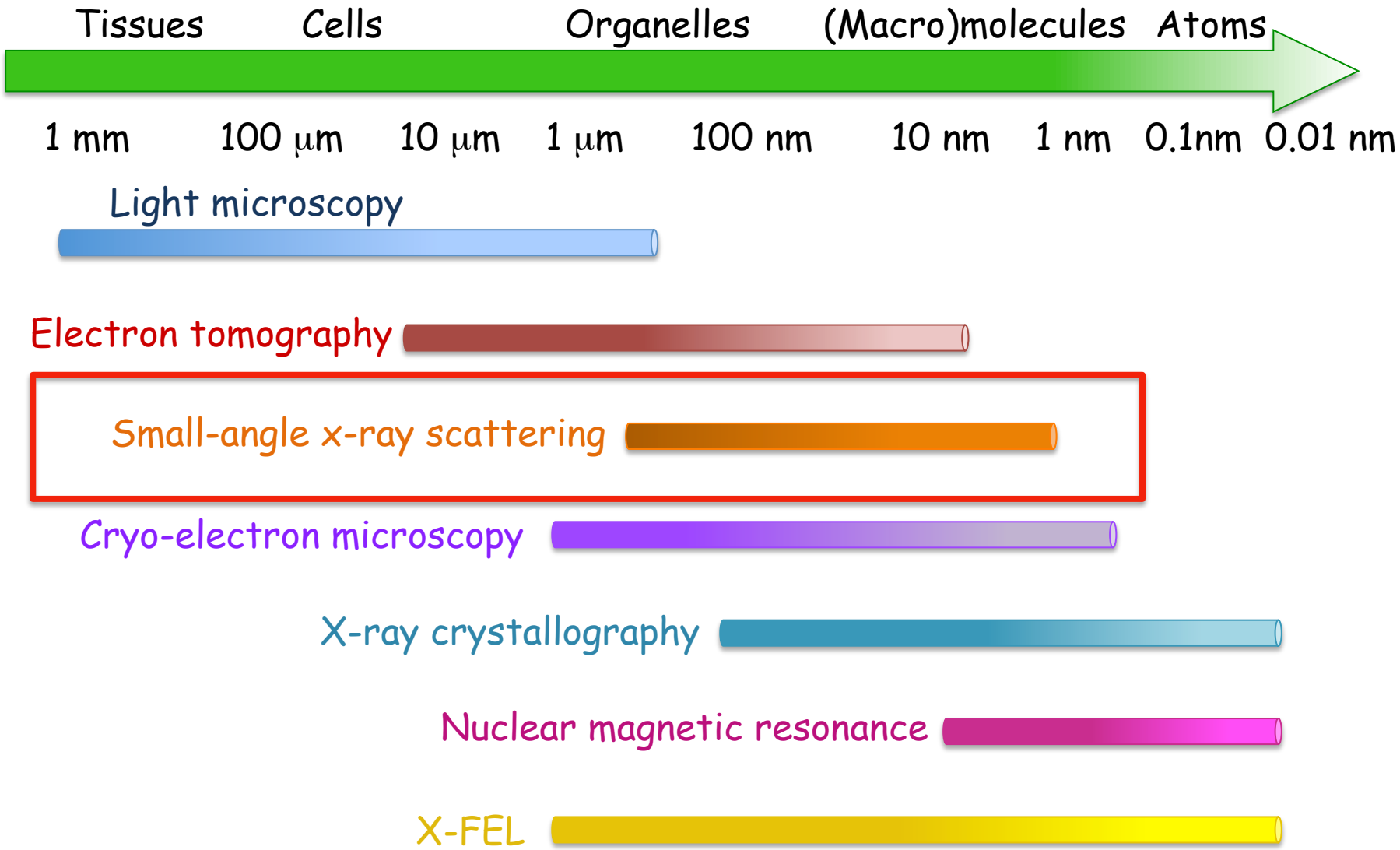


Plan of presentation

- Position of SAXS with respect to other structural methods
- Data acquisition and experimental setup
- Brief theory and principles of small angle X-ray scattering
- What do we measure?
- Data interpretation : modeling structures into envelopes
- comparison (and complementarity) to crystallography quality control, advantages and limits



Overview of structural methods



Principles of Small Angle X-ray Scattering in solution

Structural information obtained from a scattering curve

- biophysical parameters (**size** and **shape type**)
- **molecular mass**, **oligomerization state** and **volume**.

Biophysical informations derived directly from the SAXS curve

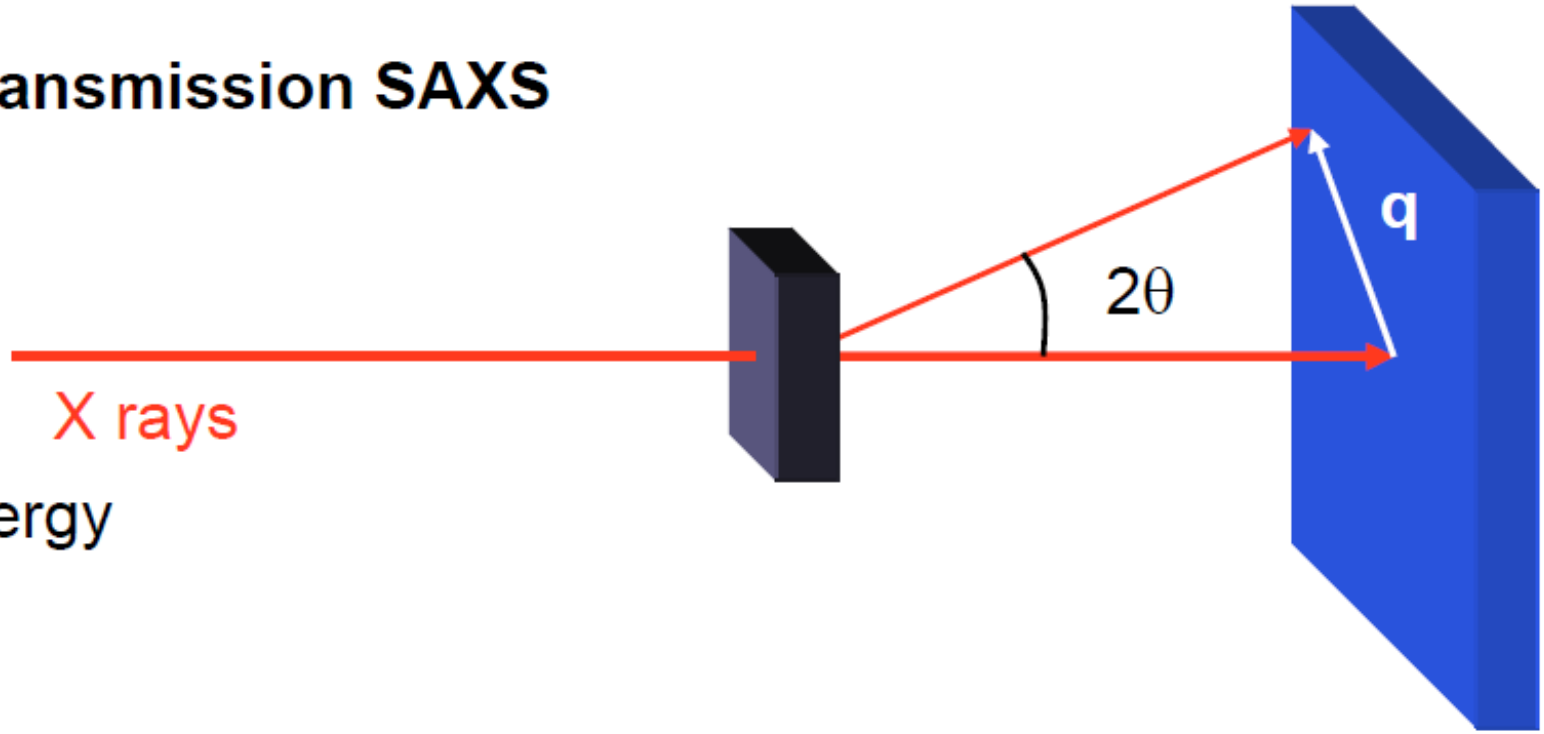
- possible **low resolution** molecular shape (ab initio methods)
- direct comparison with high resolution model
- possible model of (un)structured missing parts
- rigid body of complex

3D structural modeling
→ compatible models with SAXS data

NOT a unique model,
NO electron density map.

- Structural heterogeneity
- Flexibility of polypeptide chain
- low compactness
- non recombinant protein, or too large for RMN and too small for EM

SAS in transmission mode with 2D detector



Photon energy

$E = hc/\lambda$

Origin of diffusion

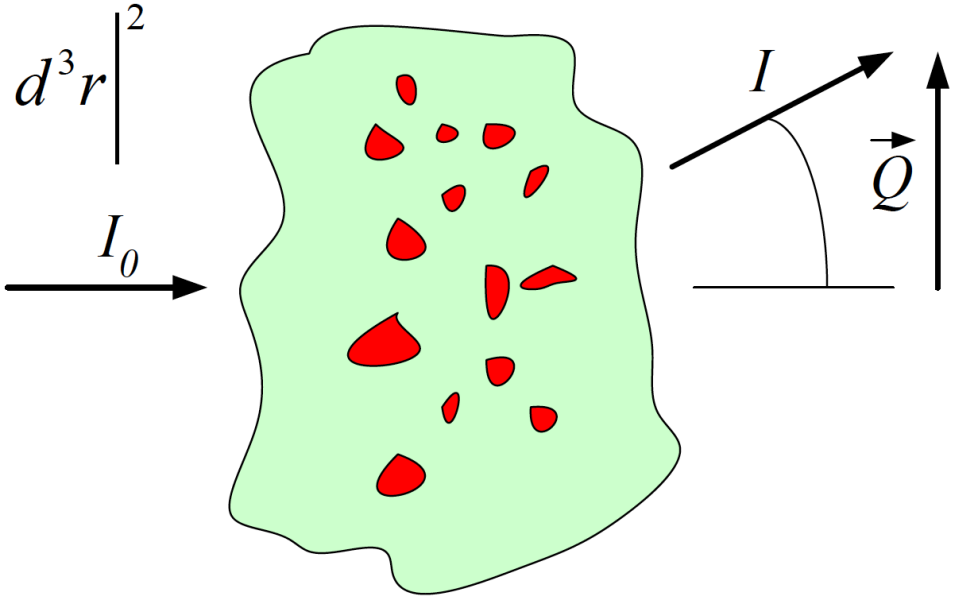
the diffusion arises from heterogeneity of density
of scattering lengths between macromolecules and the surrounding solvent

while the diffusing material and its solvent are homogenous

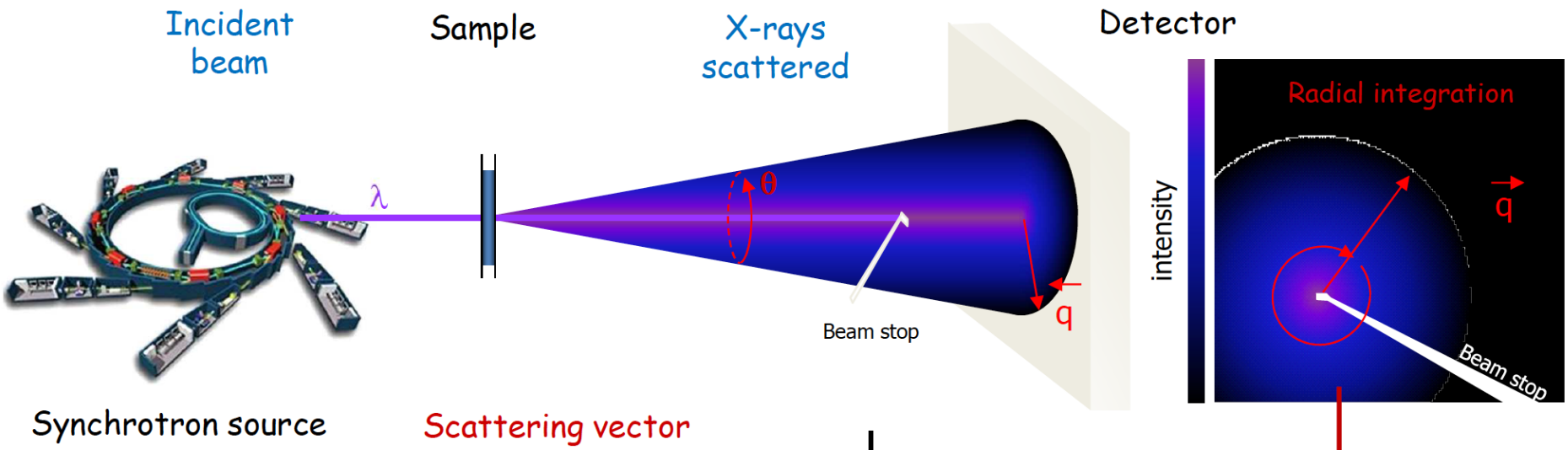
the waves are scattered once or not at all (no multiple diffusion)

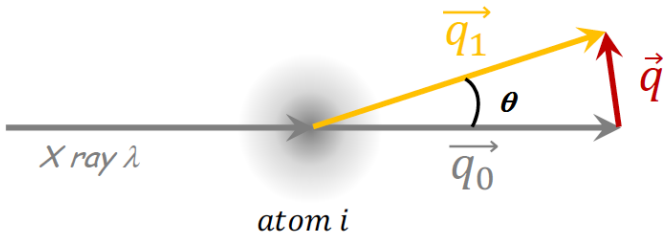
$$I(Q) \propto \frac{d\Sigma}{d\Omega}(Q) = \left| V_p^{-1} \int_{V_p^{-1}} \rho(r)\, e^{-Q\cdot r}\, d^3 r \right|^2$$

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



The principle of X-ray diffusion





$$\vec{q} = \vec{q}_1 - \vec{q}_0$$

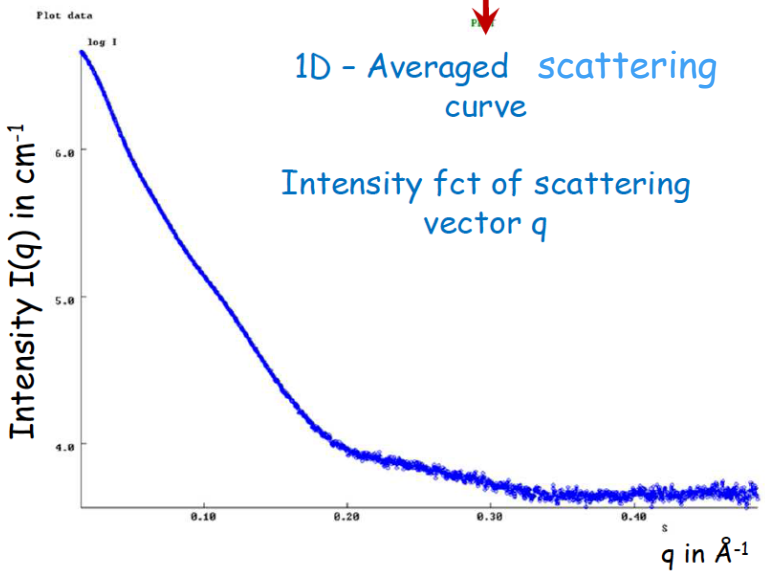
$q = 4\pi \sin\theta / \lambda$

$s = 2\sin\theta / \lambda$

$\vec{q}$ scattering vector

$\vec{q_0}$ vector of incident wave

$\vec{q_1}$ vector of scattered wave



1D - Averaged scattering curve

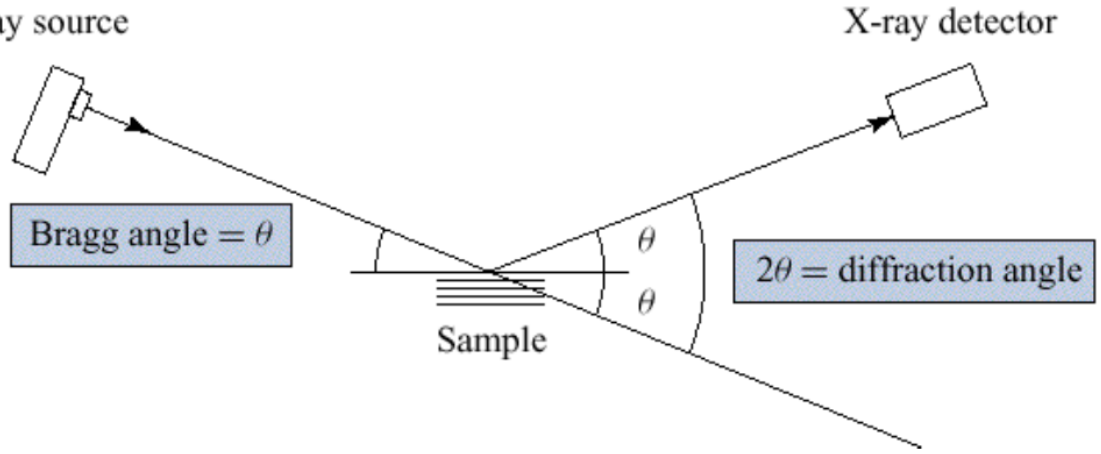
Intensity fct of scattering vector q

What are we measuring?

Sizes measured

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

$$d = \frac{2\pi}{Q}$$

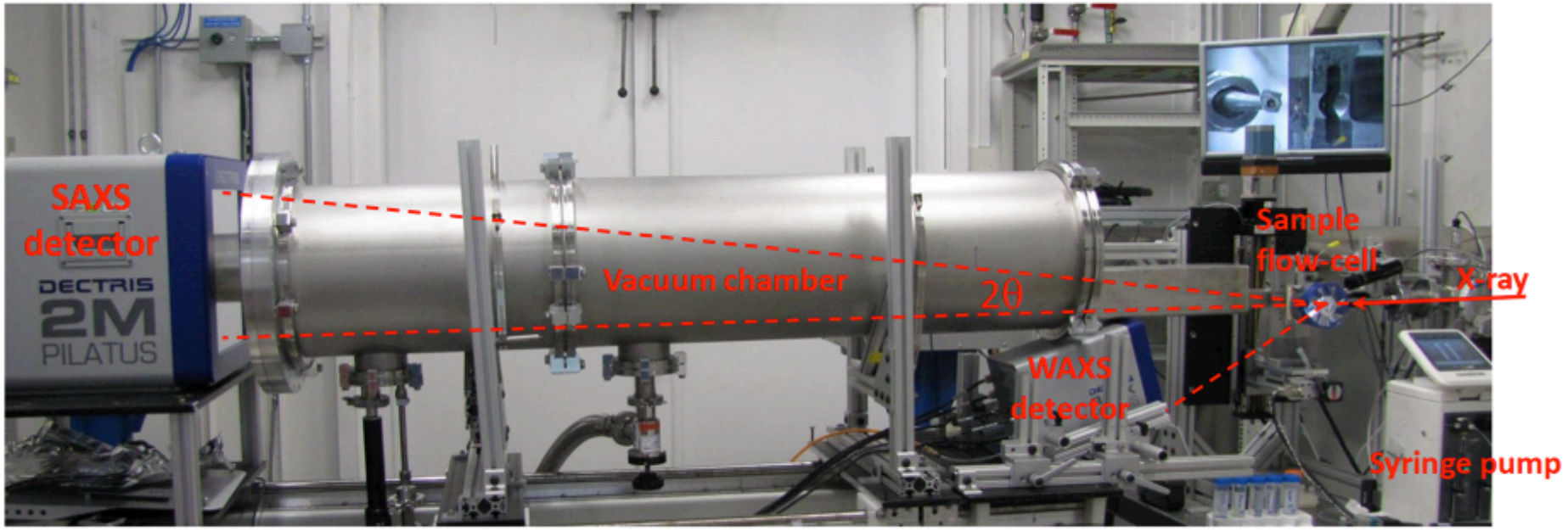


Q [A^{-1}]	D [nm]	2Θ [deg] 12keV
1	0.6	10
0.1	6	1
0.01	60	0.1
0.001	600	0.01
0.0002	3000 (3 um)	0.001

Detector dynamic range is important – Intensity ~ q^{-4}

Experimental setup Argon National Lab USA

SAXS/WAXS setup at 12ID-B at APS



Important parameters to set up an experience

- (minimum 50 microL) ⇒ 2-10 mg of protein

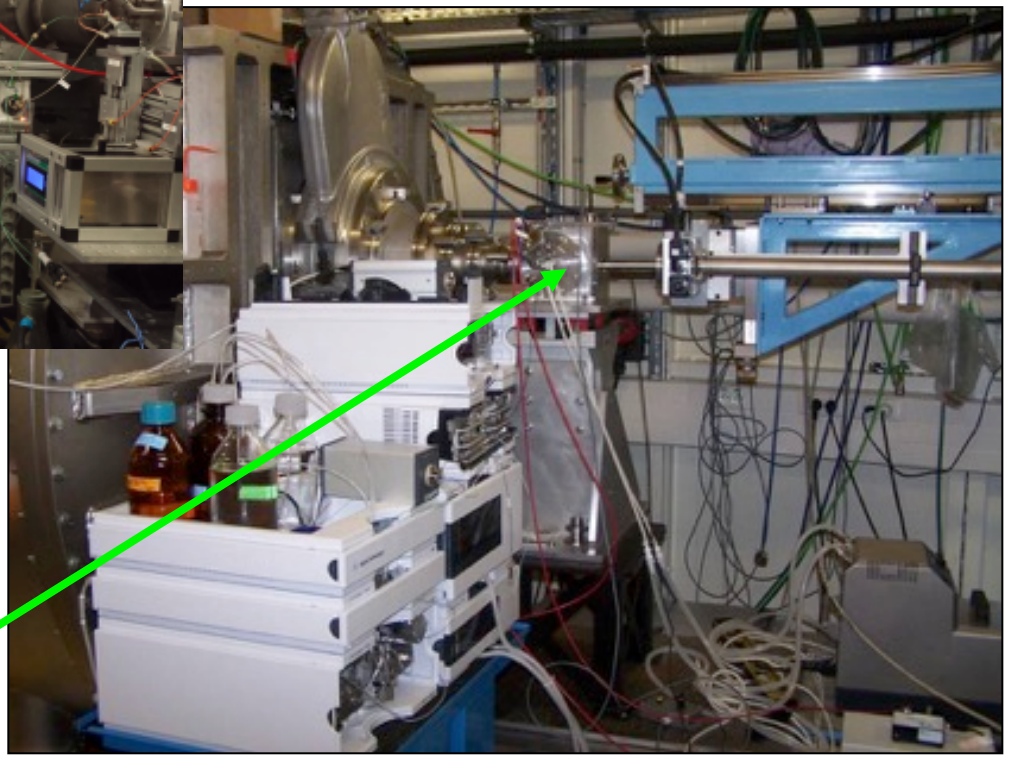
- Dialysis buffers + radio protectant (DTT (TCDE), glycerol, etc.)

- HPLC SEC-SAXS: exposure time 30 x 500ms

Detector ⟵ 4m ⟶

Sample

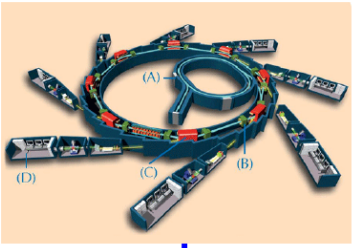
SWING Soleil



Experimental conditions

SAXS

SOLEIL SYNCHROTRON



Crystallography

PROXIMA-1

SOLEIL SYNCHROTRON

Gamme d'énergie	Between ~5 and 17 keV
Résolution en énergie	~2 eV
Source	In-vacuum U20 undulator. Source Size (sigma, µm): 388 (H) x 8.1 (V) Source Divergence (sigma, µrad): 14.5 (H) x 4.6 (V)
Optiques	Diaphragm at 11.7 m (1x0.5 mm^2) Fixed exit DCM Si_{111}at 20 m Fixed incidence focusing KB at 22.5 m Sample position : 30 - 32 m Detector / Sample Distance : 0.5 – 8 m
Environnements d'échantillon	X / Z precision table Stopped flow device for chemistry Online HPLC for proteins in solution (SAXIER project) High throughput sampler for proteins in solution (SAXIER project) Couette Cell (collaboration with LPS, Orsay) GISAXS chamber Automatic sample changer (50 samples, thermostated) Linkam heating stage THMS600 ...
Taille du faisceau	450x20 $µm^2$ FWHM dans la cabane expériences
Flux sur l'échantillon	8.10^{12}ph/s @7keV, 8.10^{11} ph/s @16keV (à 400 mA de courant anneau)
Détecteurs	SAXS : PCCD170170 (AVIEX), Gain > 3ADU/ph, Noise≈2ADU WAXS (2014) : Détecteur à pixels hybrides
Chambre de détection	Vide primaire, positions du détecteur SAXS : - 0.20 / + 0. 20 m (horiz), -0.20 / +0.20 (vert), 0.5 m / +6 m (le long du faisceau).

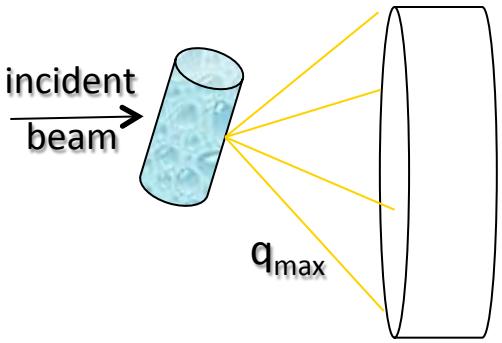
Données techniques

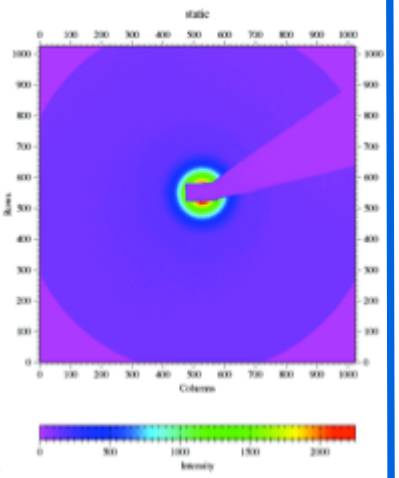
Energy range	Between 5 KeV and 15 KeV
Energy Resolution	~0.000075 (Si 311) - ~0.0002 (Si 111)
Source	U20 In-vacuum undulator
Flux @ first optical element	White beam – depends on undulator settings
Optics	Kirkpatrick-Baez pair of bi-morph mirrors plus channel cut cryogenically co monochromator crystal
Sample Environment	3 circle κ goniostat (10 µm sphere of confusion) 6 axis robot sample changer (MSC/Rigaku ACTOR) Oxford Cryosystems cryostream Si drift diode energy dispersive detector plus MCA for fluoresc measurements
Beam size at sample	Variable between 100x100 $µm^2$ to 250x250 $µm^2$
Flux on sample	> 2.0 e+12 Phot/s/0.02%bw for 500 mA stored current.
Detectors	ADSC Q315r
Polarization	Linear

Comparison of methods

SAXS

average of conformation in solution

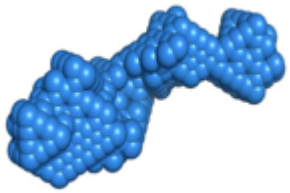




detector

data treatment

Analyses
ab initio
molecular modeling

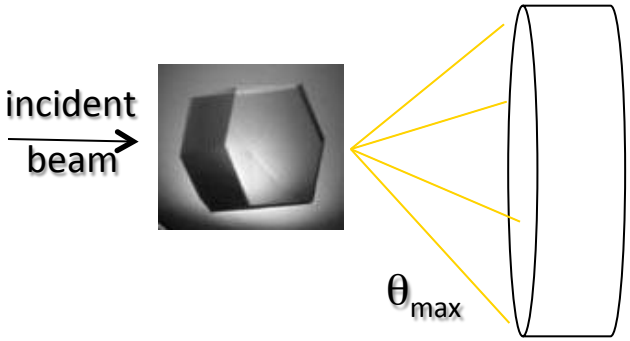


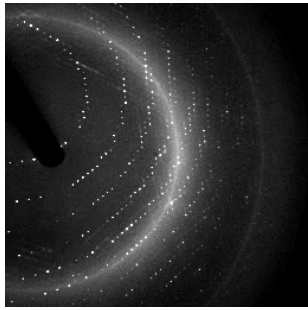
envelope, shape

Limiting factors : aggregation, heterogeneity

crystallography

average of atomic crystal structure

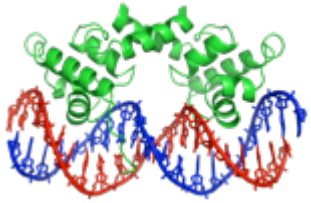




detector

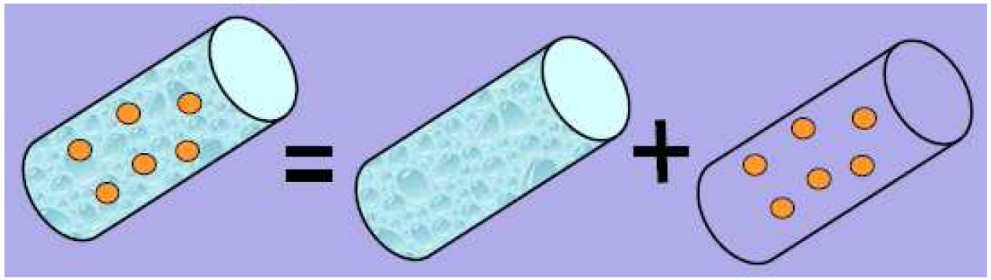
data treatment

Phasing
molecular replacement
MAD, SAD, MIR…



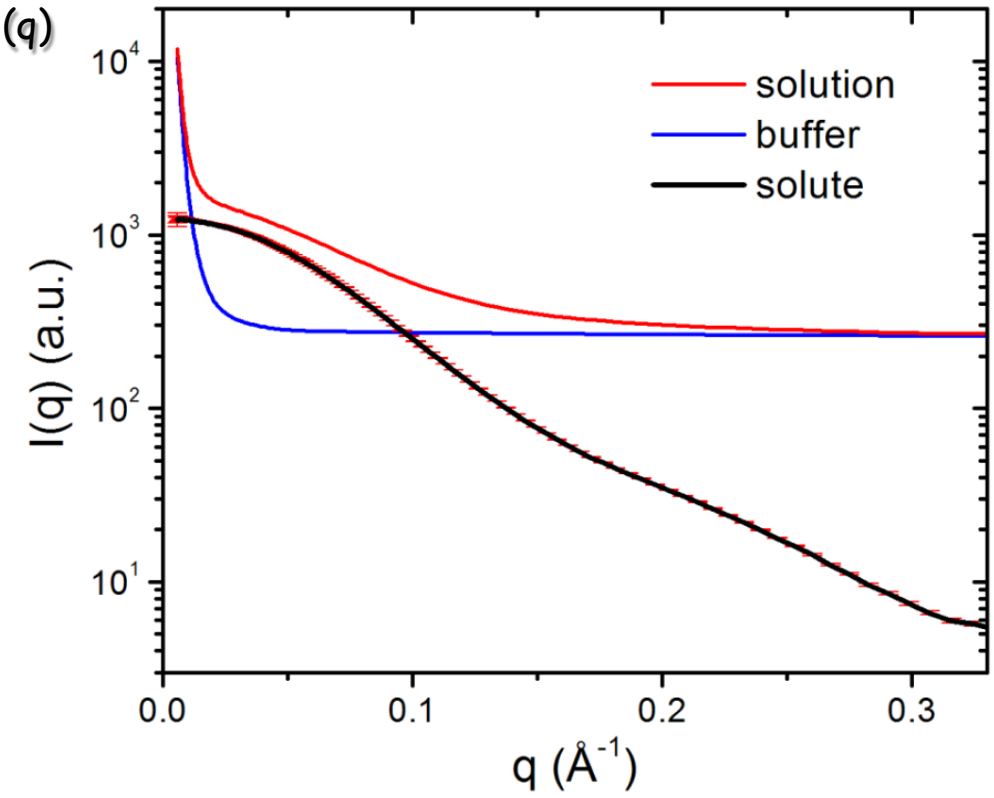
3D atomic structure

Limiting factors : diffracting crystals, phasing signal



1D SAXS profiles

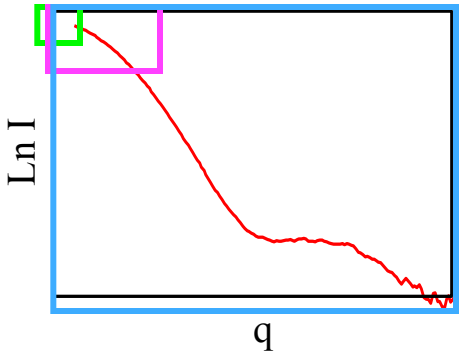
$I_{solution}(q) = I_{buffer}(q) + I_{solute}(q)$

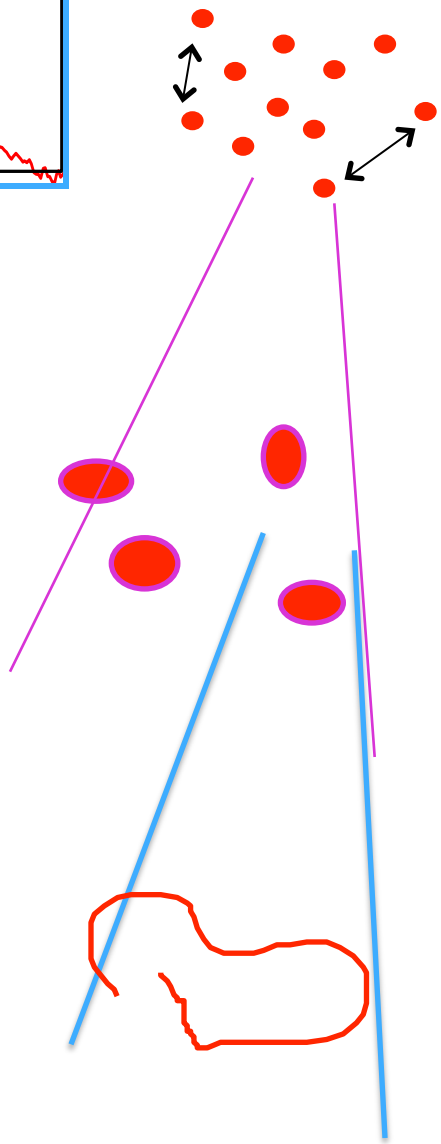


The measurement of diffusion often contains a lot of 'noise'

the diffusion vector q :

reciprocal space, inverse of a distance





q=0 (determination of I(0))

molecular weight (Mw), oligomerisation state and intermolecular interactions in solution

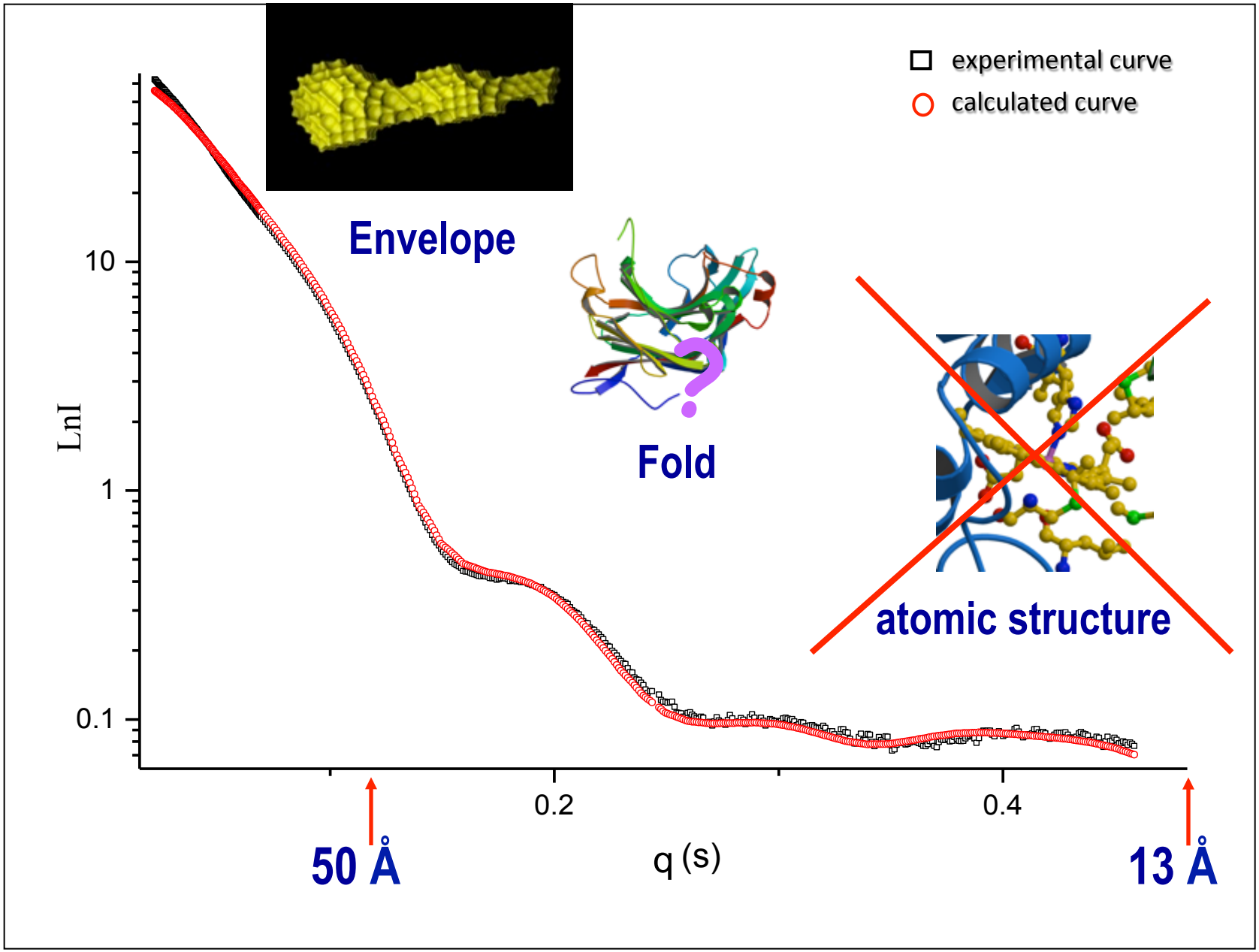
Region of small q

Particle dimensions (Radius of giration)

Region of larger q

Particle overall shapes (polypeptide chain conformations in solution)

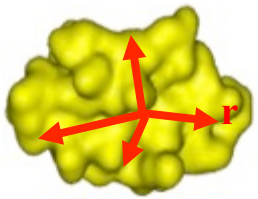
Information from small angle scattering



Data Analysis : Guinier law

The radius of gyration

$$R_g^2 = \frac{\int_V \rho(\vec{r})\, r^2 \quad d^3\vec{r}}{\int_V \rho(\vec{r}) \quad d^3\vec{r}}$$



⇒ **Mean square of atomic distances from center of gravity,**

(weighted by electron density ρ(r))

Close to q=0 the scattering of a particel can be described by a Gaussian curve

Guinier Law:

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

Prof. André Guinier
1911-2000
Orsay, France

The Guinier law is equivalent to a linear variation of Ln(Iq) vs q^2 → Guinier plot
A linear regression on the experimental Guinier plot directly provides Rg and Io

$$\text{Ln } I(q) = \text{Ln } I(0) - q^2 R_g^2 / 3$$

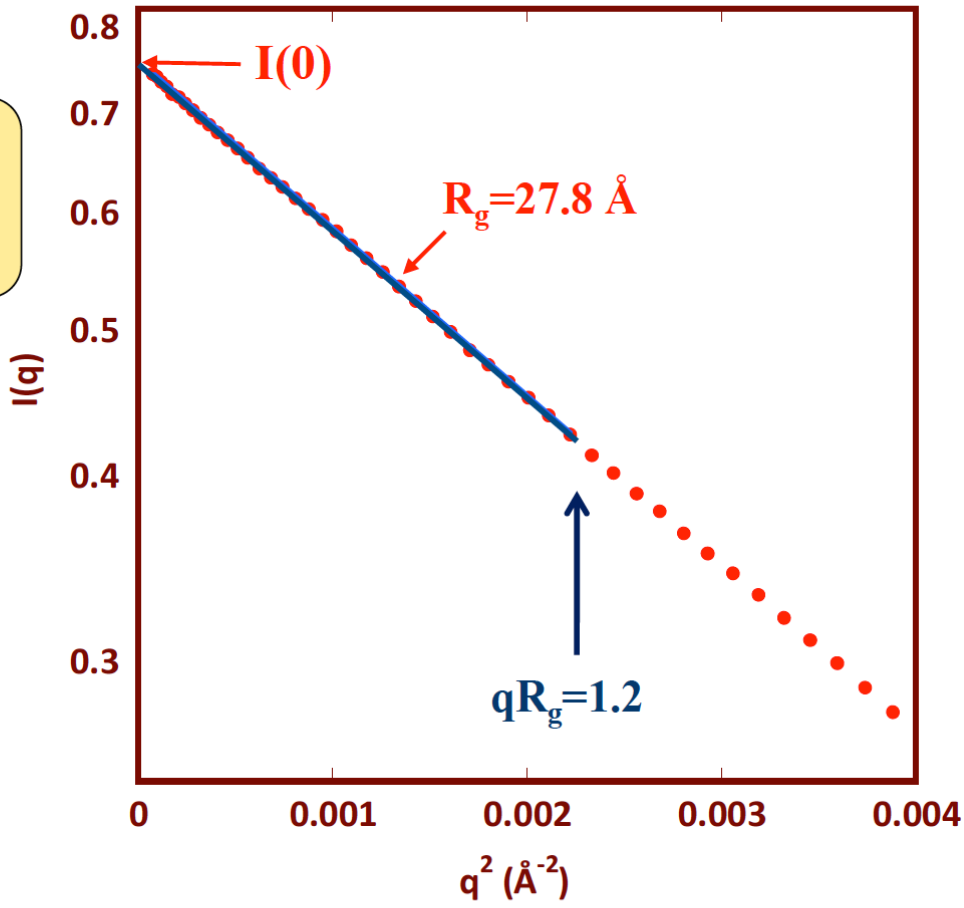
Data Analysis : Guinier law

Guinier analysis

R_g → size

I(0) → mol mass / oligomerisation state

$$ln[I(q)] \cong ln[I(0)] - \frac{R_g^2}{3} q^2$$



Validity range :

$0 < qR_g < 1$ for a solid sphere

$0 < qR_g < 1.3$ rule of thumb for a globular protein

ideal monodispersed

$I(0) \propto c.M_w/N$ (Cste)

⇒ Determination of *average dimension* of the particle

⇒ Determination of its *molecular mass*

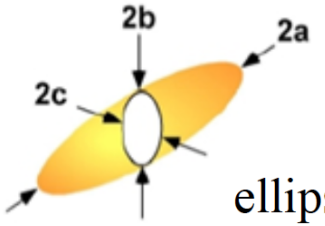
R_g : Mean square of atomic distances from center of gravity, (weighted by electron density ρ(r))

- In polymer physics, the radius of gyration is proportional to the root mean square distance between the parts of the object:

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

Gaussian chain

$$\langle R_g^2 \rangle = N \frac{l^2}{6}$$



ellipsoid

$$R_g^2 = \frac{1}{5}(a^2 + b^2 + c^2)$$

Sphere

$$R_g^2 = \frac{3}{5} R^2$$

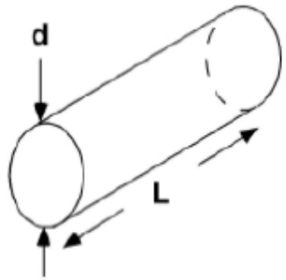
Thin rod

$$R_g^2 = \frac{L^2}{12}$$

Thin disc

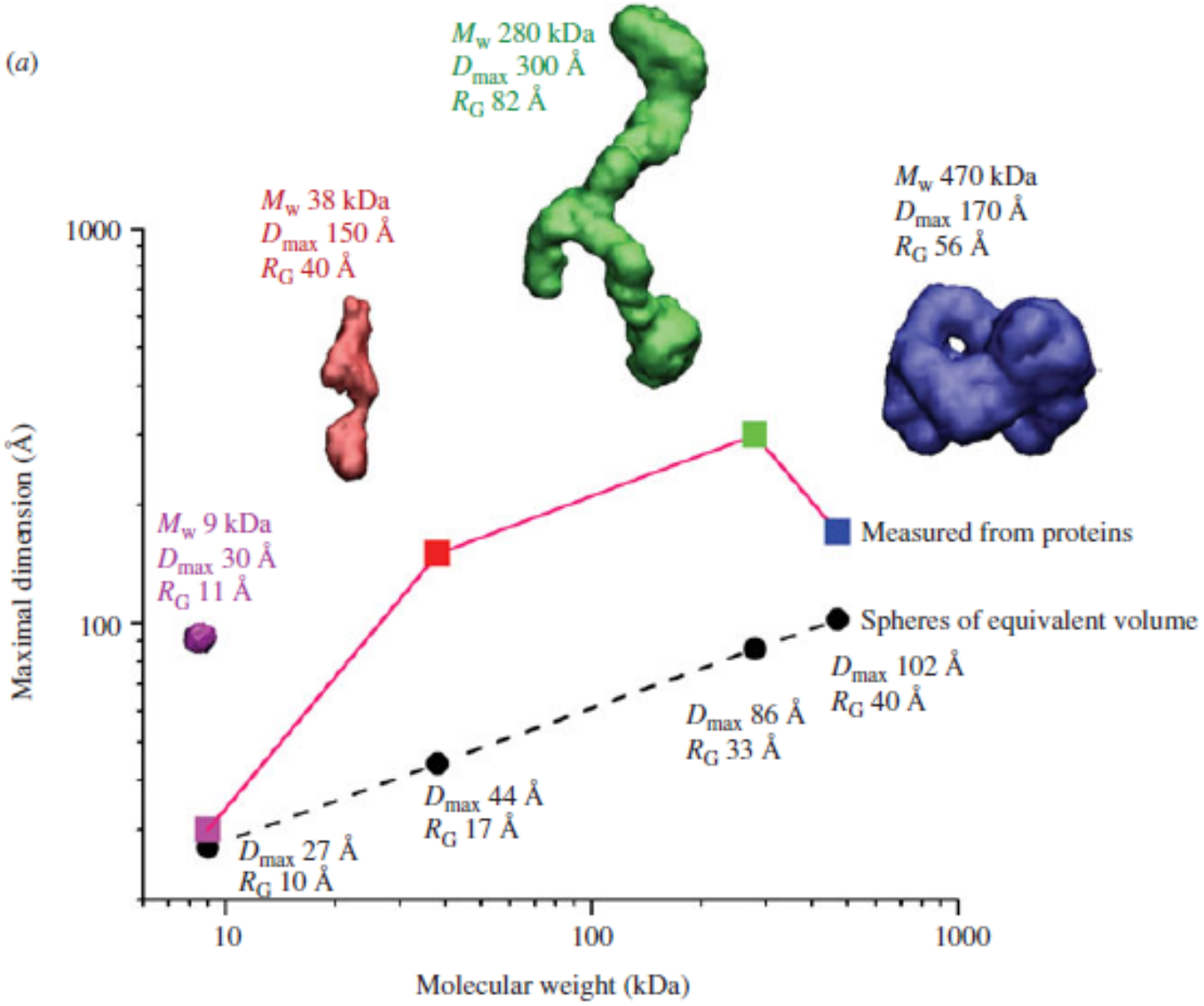
$$R_g^2 = \frac{R^2}{2}$$

Cylinder



$$R_g^2 = \frac{R^2}{2} + \frac{L^2}{12}$$

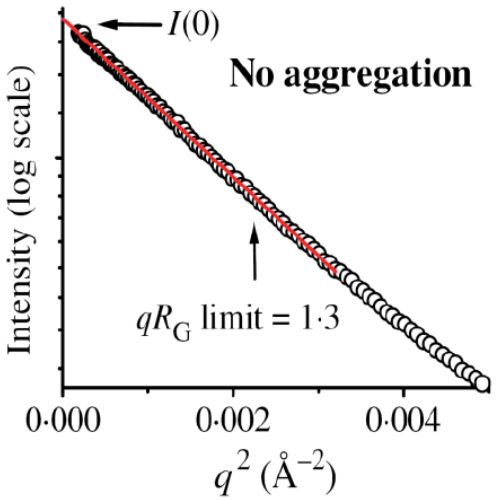
Relation of R_G to molecular weight (Mw) – roughly linear ONLY for spherical proteins



Guinier Plot: interactions & sample condition

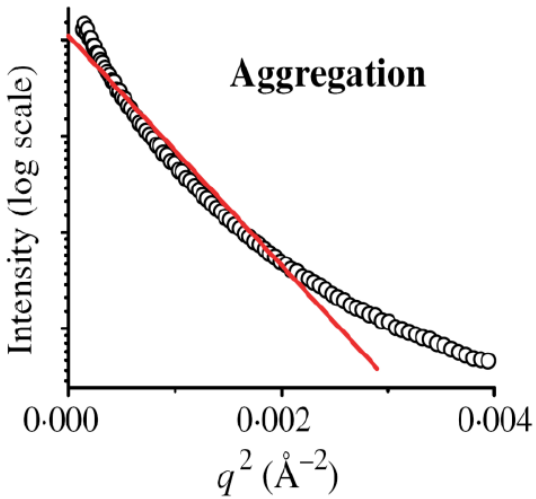
Mono-dispersed

Normal / linear



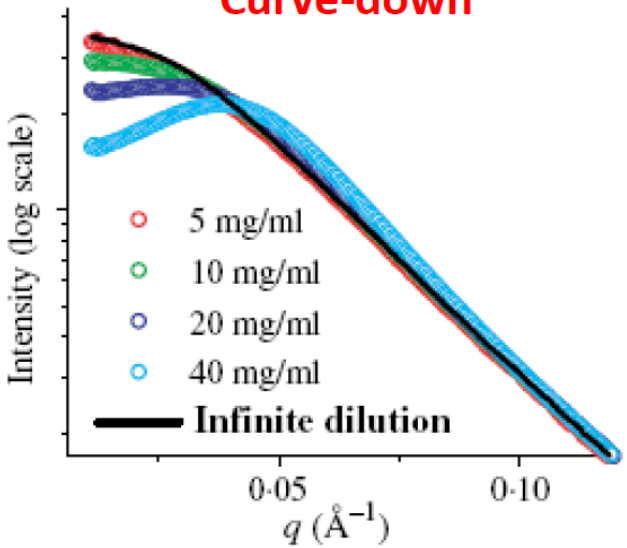
Poly-dispersed aggregates

Curve-up



Repulsion

Curve-down

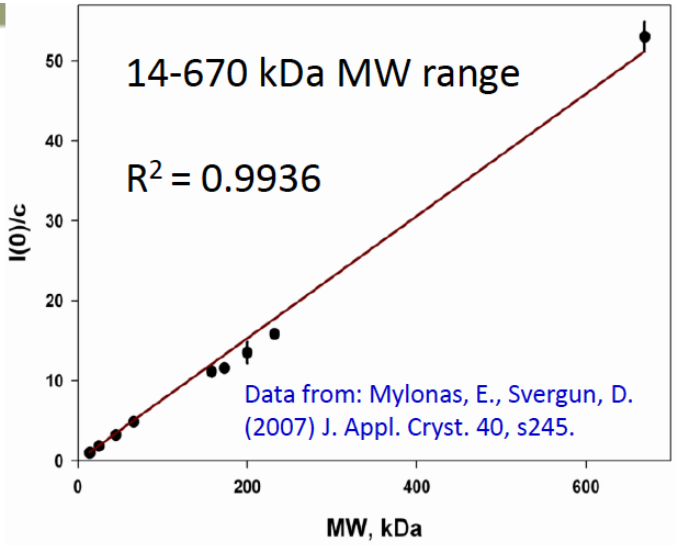


Figures from: Putnam, D., et al. (2007) Quart. Rev. Biophys. 40, 191-285.

Forward scattering I(0) measures molecular weight

$$MW = \frac{N_A I(0)/c}{(\Delta\rho_M)^2}$$

I(0) at absolute scale; *c* mg/ml
Scattering length density diff: $\Delta\rho_M = \rho_M - \rho_s$
$N_A = 6.02*10^{23}$



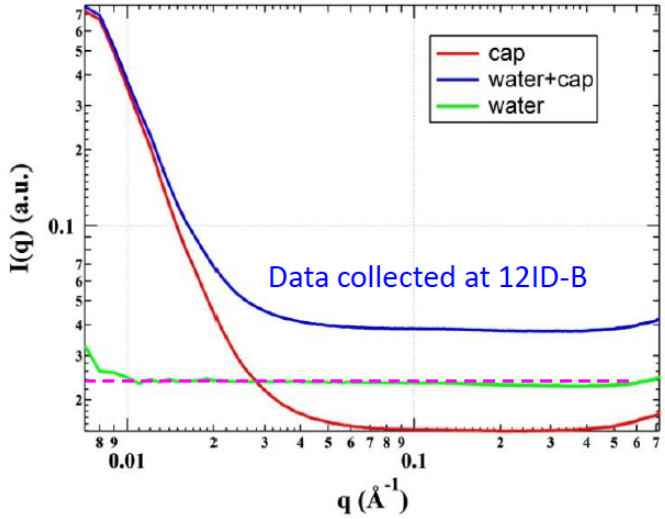
- Absolute scale: calibrate with water or other standard
  - Water: 1.632×10^{-2} cm^{-1} at 293K

For protein on average: $\Delta\rho_M = 2.086*10^{10}$ cm^{-2}

$$MW = 1385 \times I(0)/c$$

MW in kDa; *I(0)* in cm^{-1}, *c* in mg/ml

Orthaber, Bergmann and Glatter, J. Appl. Cryst. (2000). 33, 218



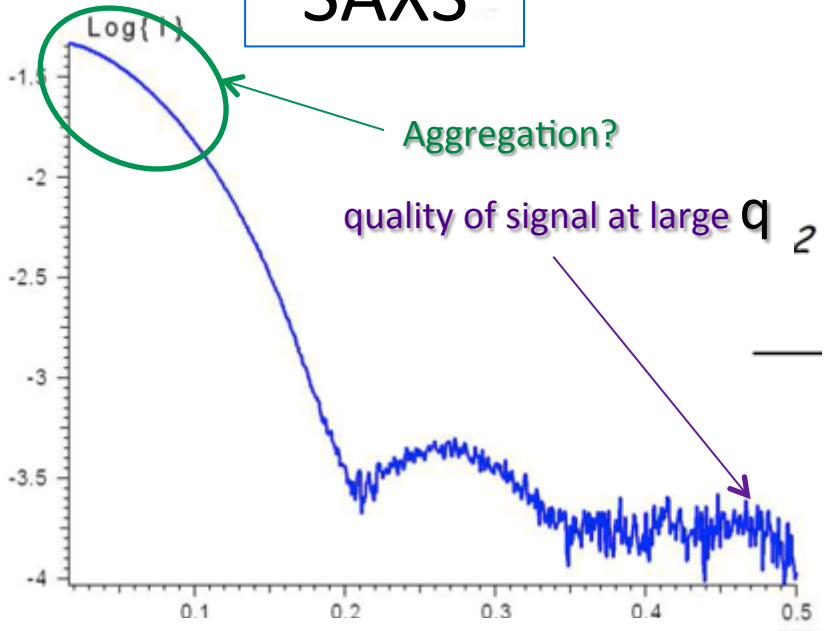
- Relative scale: secondary standard
  - Lysozyme, BSA, etc

$$MW_p = I(0)_p / C_p \frac{MW_{st}}{I(0)_{st}/C_{st}}$$

➢ Standard should has **same nature** of the molecules to determine, and with close MW. Using multiple standards is suggested.

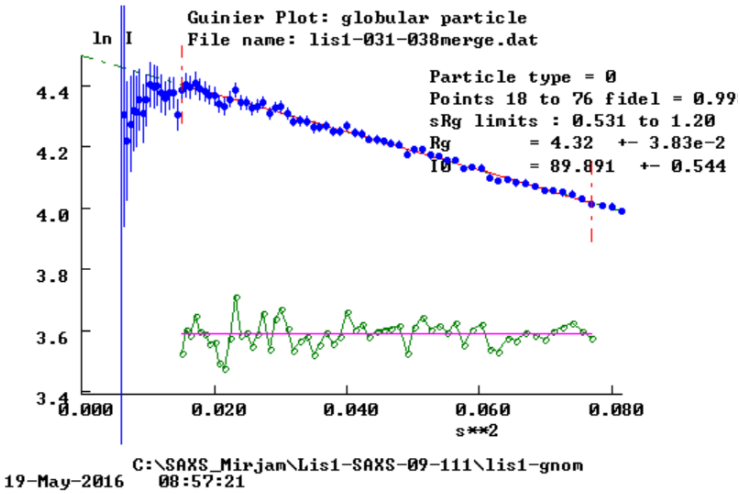
Direct information from scattering

SAXS



Aggregation?

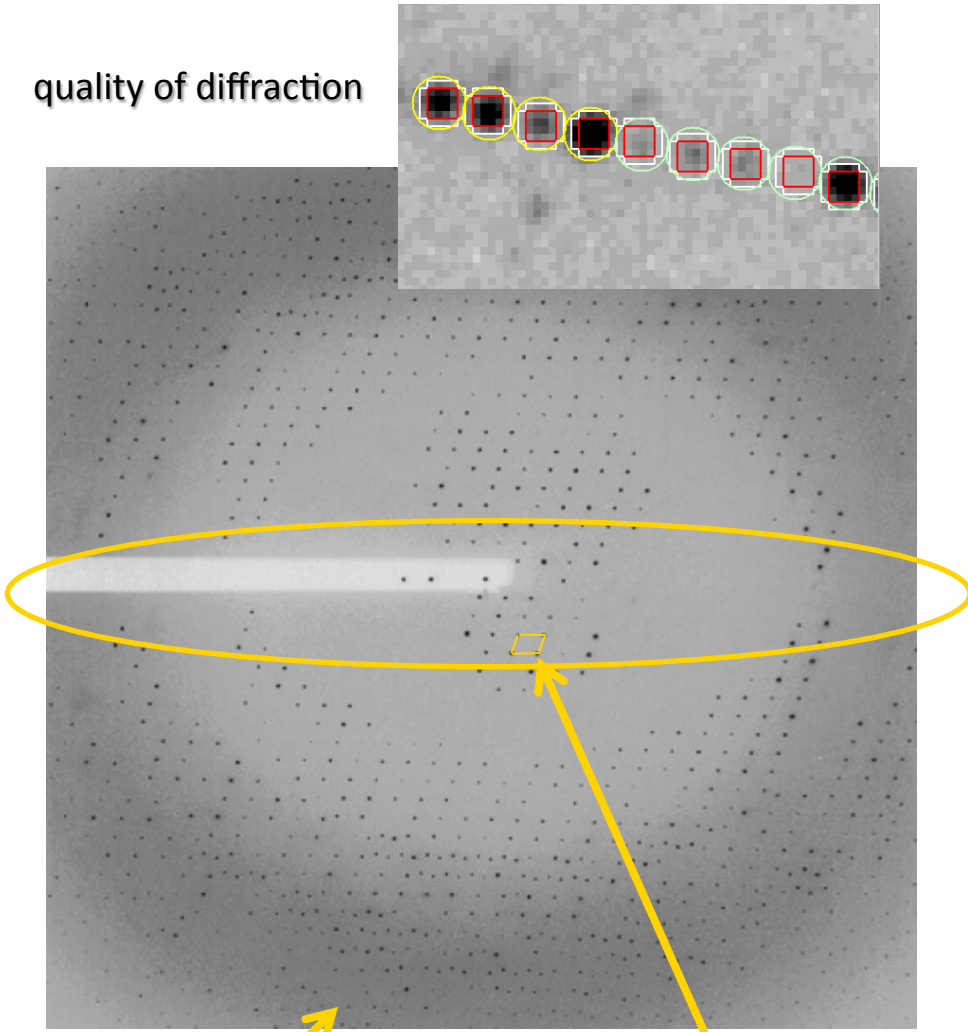
quality of signal at large q^2



Guinier plot : R_g and M_w

crystallography

quality of diffraction

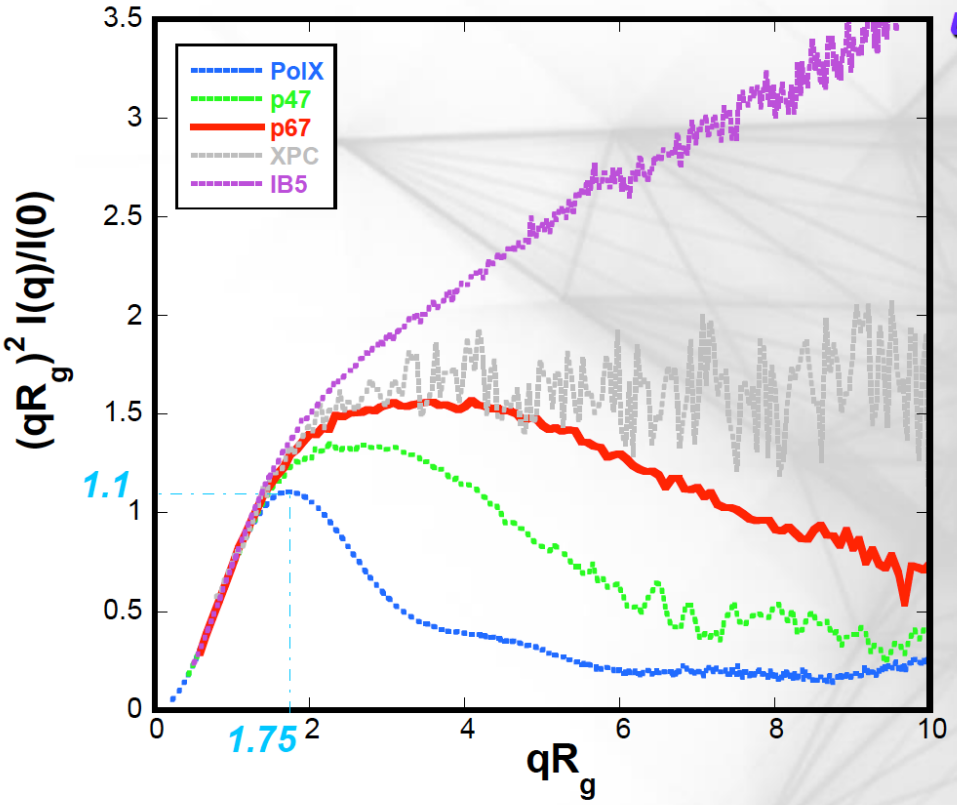


resolution limit 3.1 Å

unit cell & space group

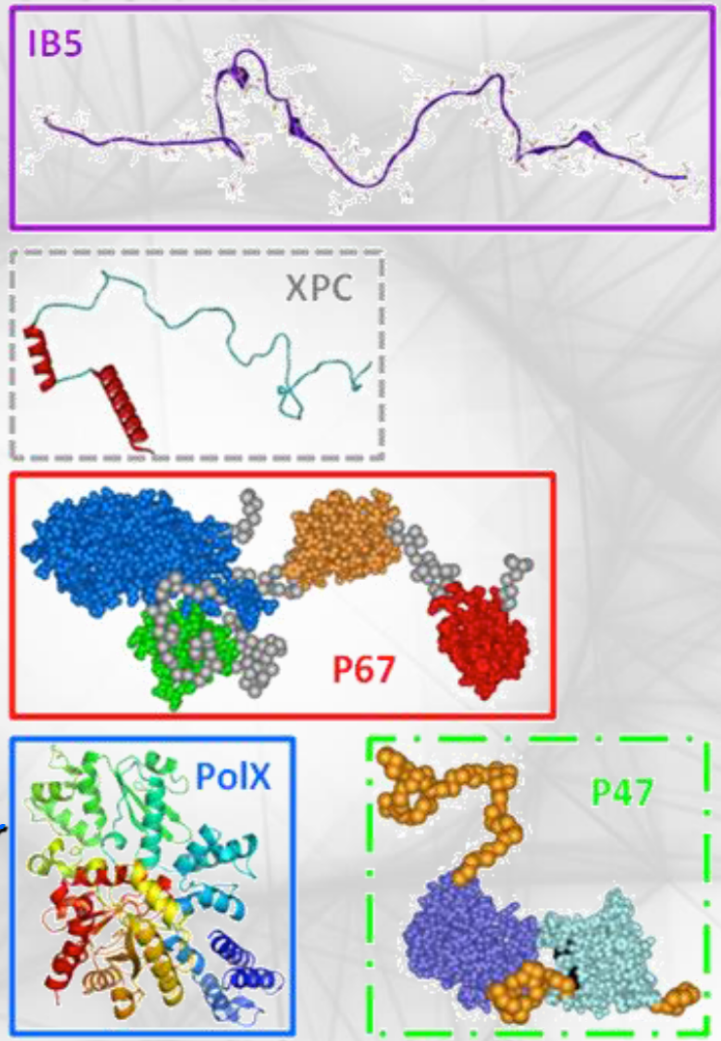
Dimensionless Kratky-plot of (partially) unfolded proteins

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.

Porod law

Hypothesis : the particle has a well defined interface with the surrounding buffer and a uniform electron density

$$\mathbf{Q} = \int_0^\infty I(q)\cdot q^2 dq = 2\pi^2\phi\cdot(1-\phi)\cdot r_e^2(\Delta\rho)^2$$

Porod invariant

Protein volumic concentration

$$\phi = \frac{\mathrm{N}\cdot \mathrm{V}_{obj}}{\mathrm{V}}$$

Number of proteins

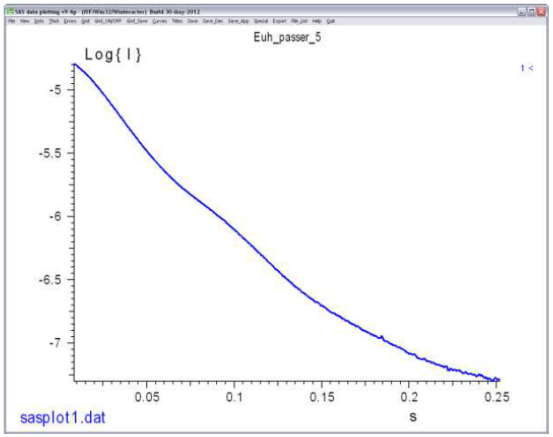
Protein volume

Solution volume

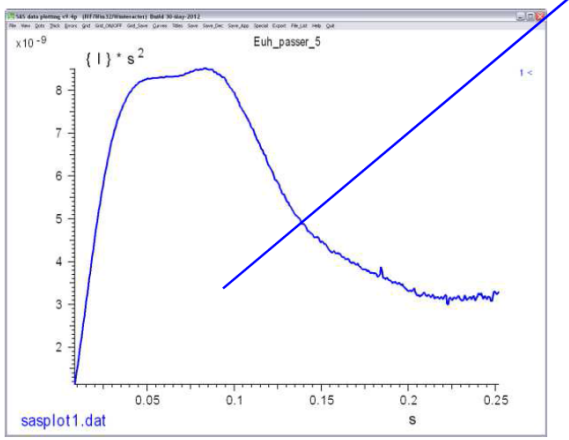
Does not depend on shape, only on contrast

The Porod Invariant is the integral of this curve

Representation : Log I(Q) vs Q



Kratky representation : $I(Q)\cdot Q^2$ vs Q



$$\mathbf{Q} = \int_0^\infty I(q)\cdot q^2 dq$$

$$\mathrm{V}_{obj} = \frac{2\pi^2\cdot I(0)}{\mathbf{Q}}$$

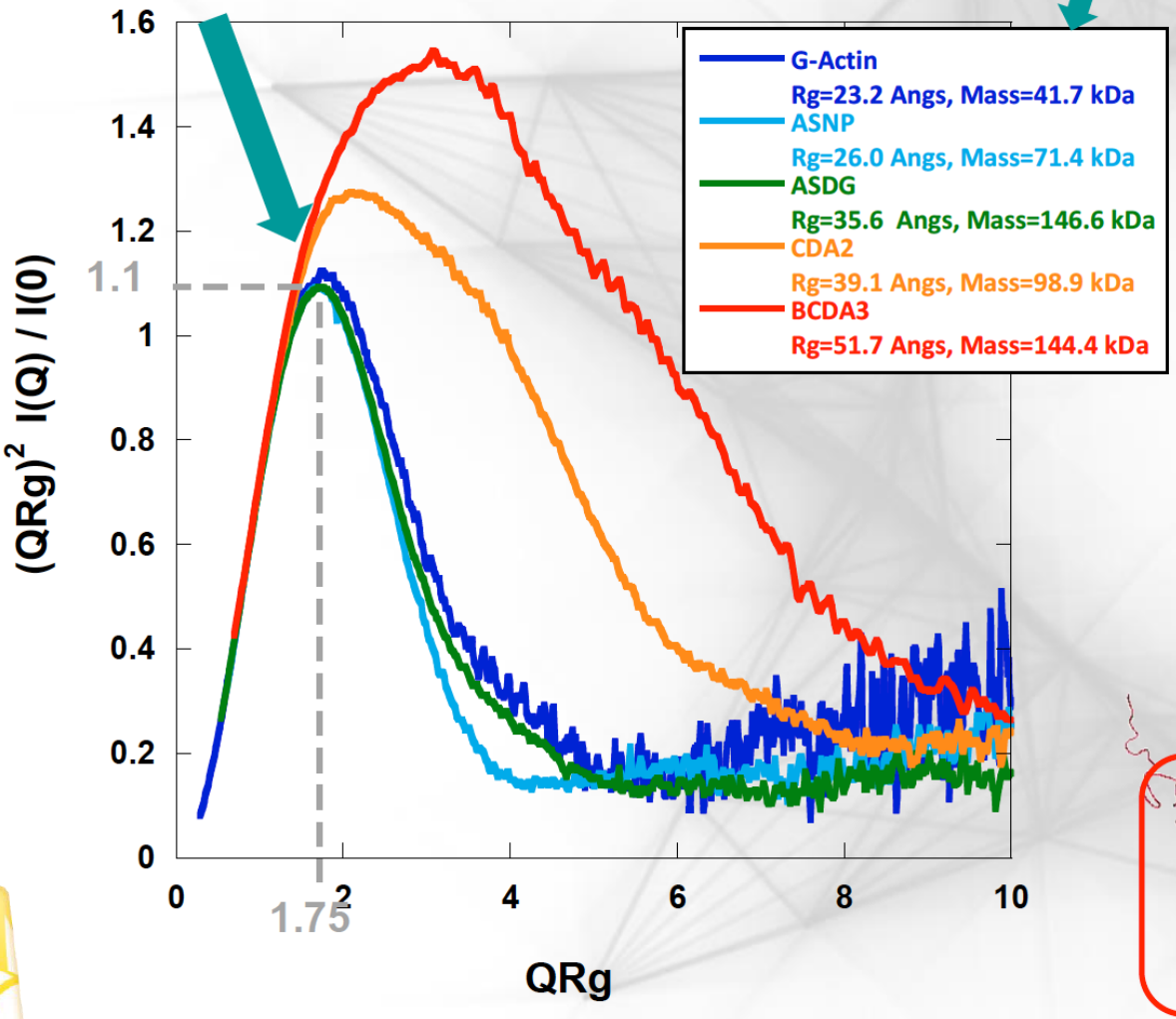
- Valid for diluted systems
- Does not require absolute units
- Not valid for unfolded objects
- Not precise

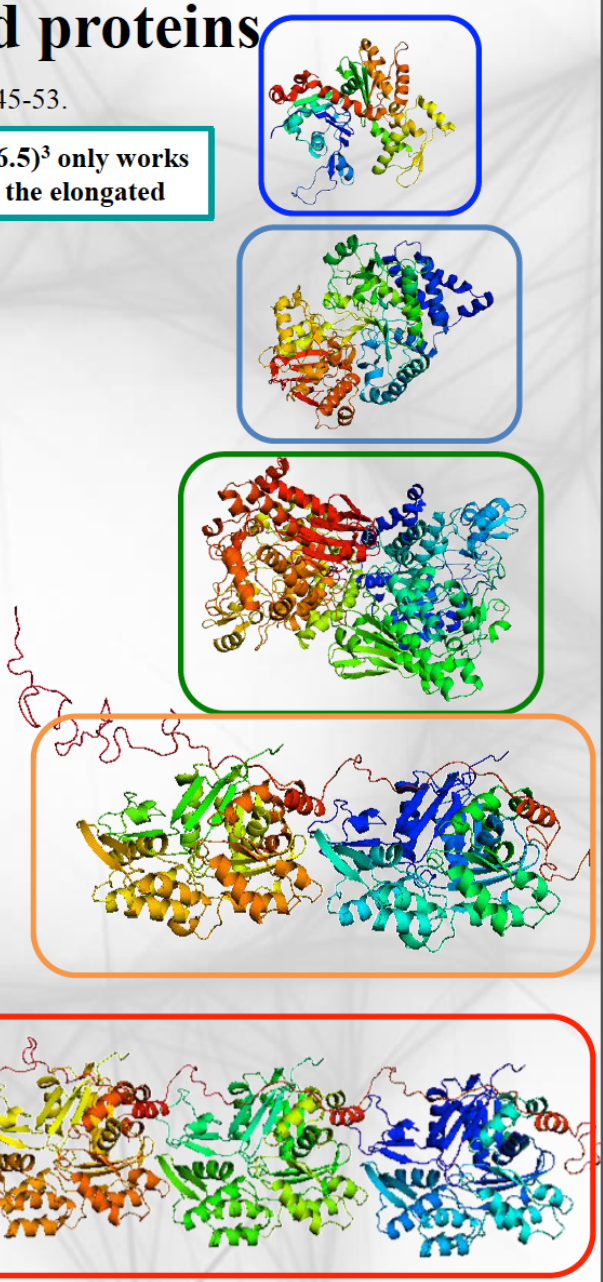
Dimensionless Kratky Plots of folded proteins

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



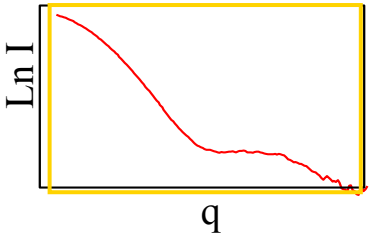


The maximum value on the dimensionless bell shape tells if the protein is globular.

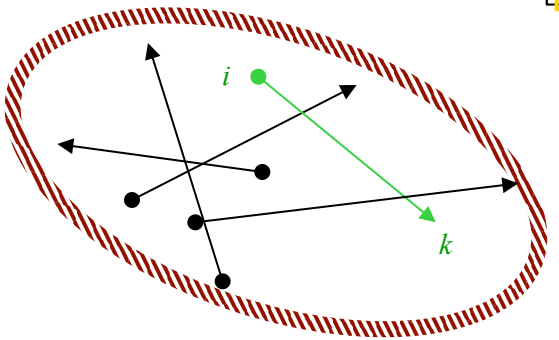
courtesy by Aurelien Thureau

Back to real space: distance distribution function

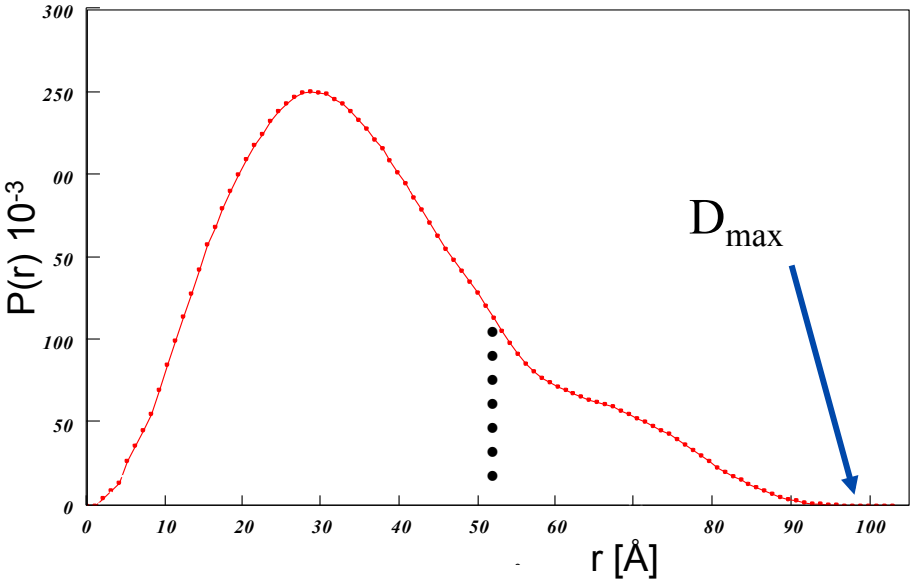
=> Transformation into real space with a Fourrier transform F (I(q)) :



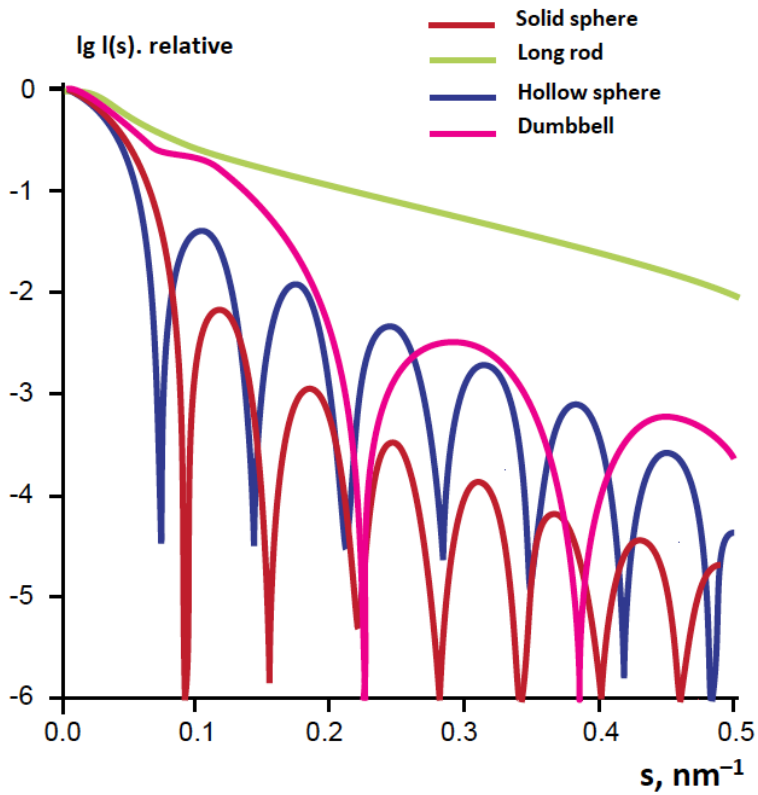
$$P(r) = \mathcal{F}(I(q)) = \int I(q)\, e^{-i\vec{q}.\vec{r}}\, dq$$

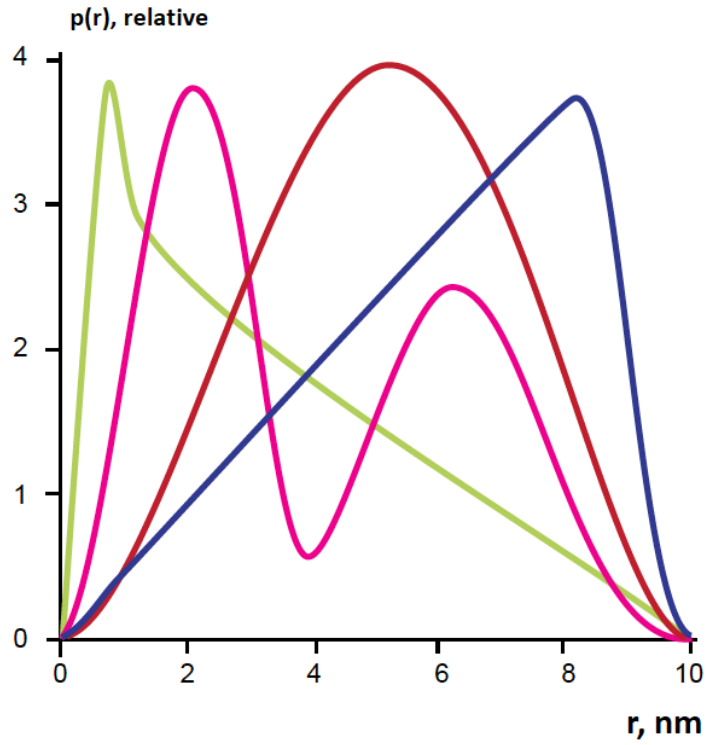


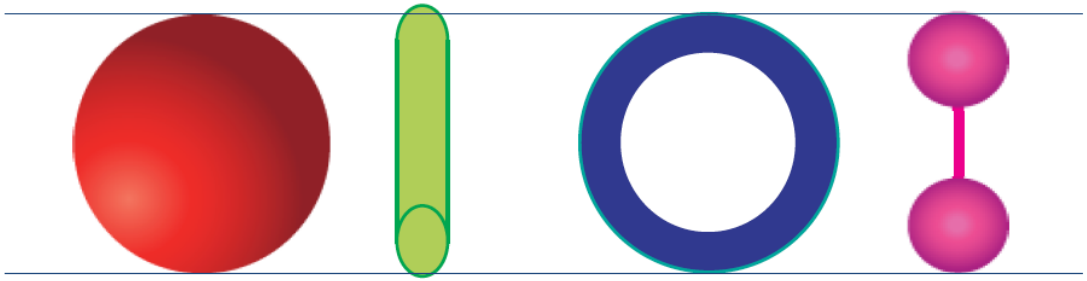
- Function of distribution of distances between all atomes.
- Histogram of all existing intramolecular distances



SAXS and P(r) of different forms

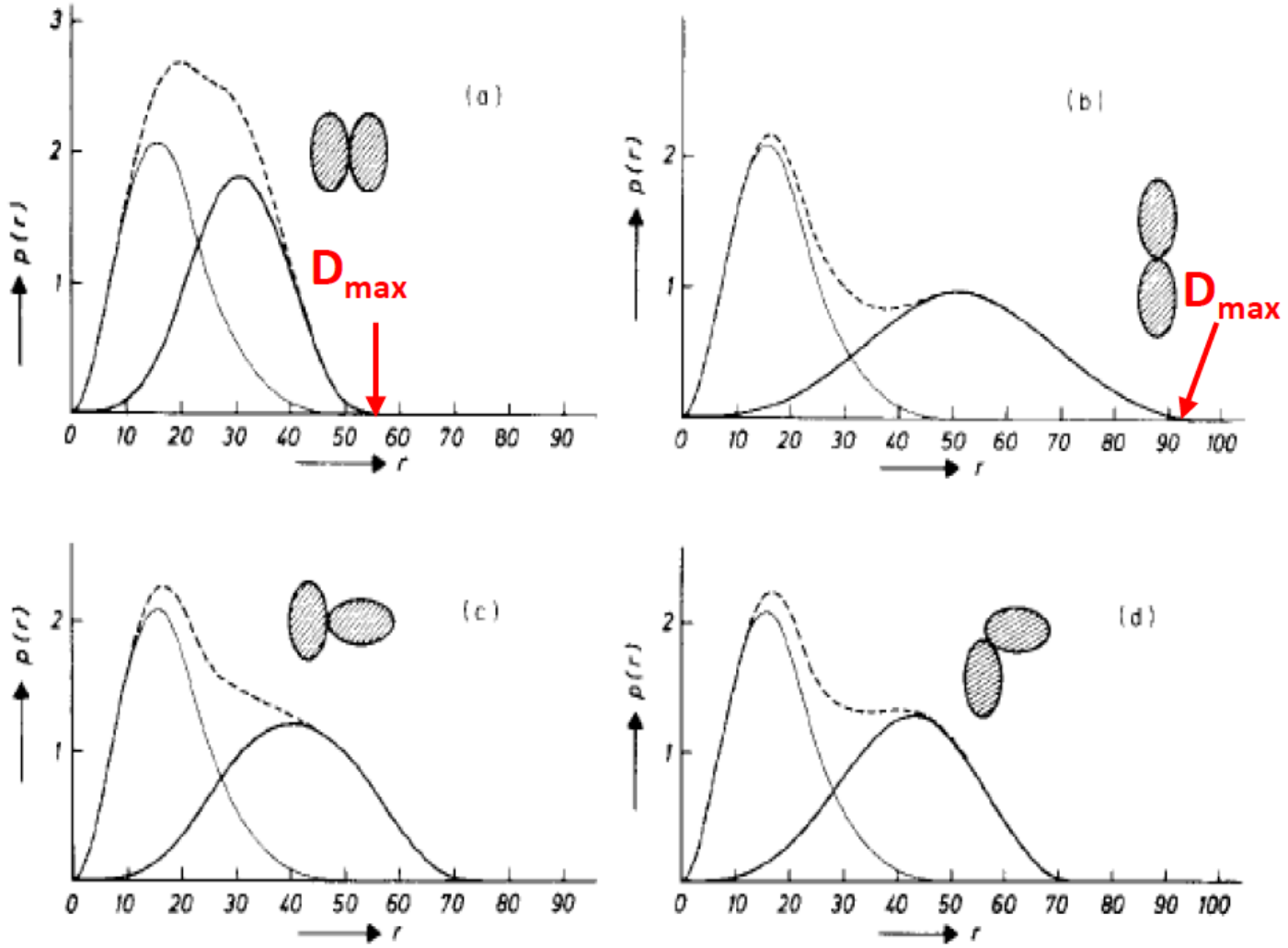






Adapted from: Svergun, D., Koch, M. (2003) Rep. Prog. Phys. 66, 1735-1782.

P(r) functions for different 'options' of a dimer



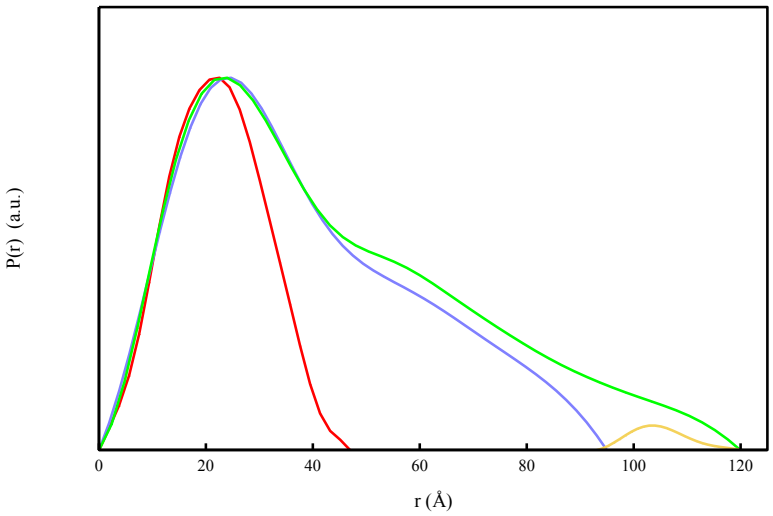
As well the form and the value of D_{max} vary for different options

Humicola insolens
EGV native and truncated

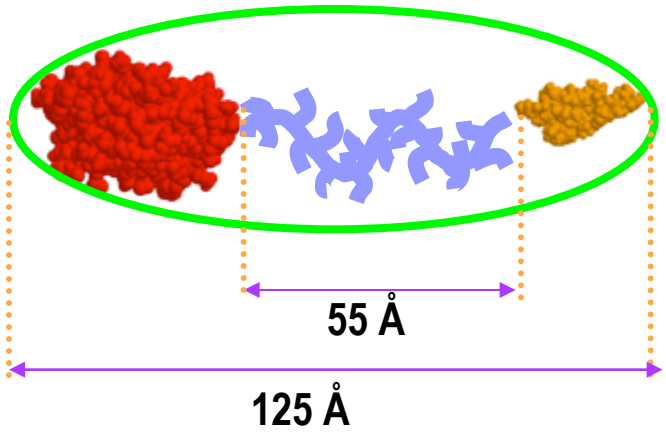
Mw = 38 kDa N = 210+36+38 aa

	Rg (Å)	D_{max} (Å)
Catalytic domain	17.3 ± 0.3	45 ± 5
EGV without CBD	30.0 ± 0.4	100 ± 10
EGV full lenfth	33.5 ± 0.5	125 ± 5
CBD	9.2	31

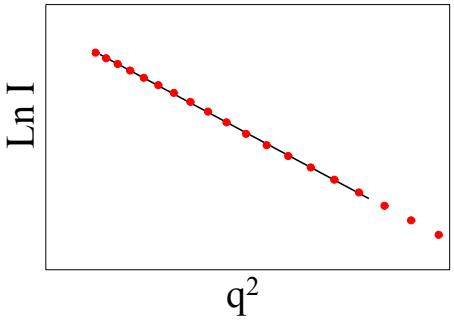
Distance distribution function



Linker : ≈ 0.7 residues/Å

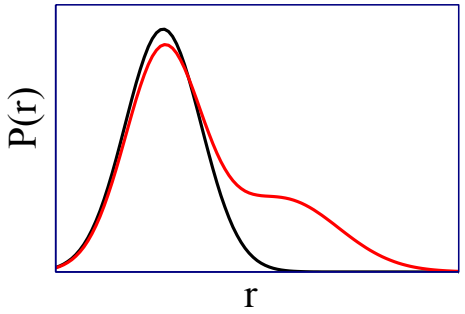


Small angle scattering



I vs q (ou s)

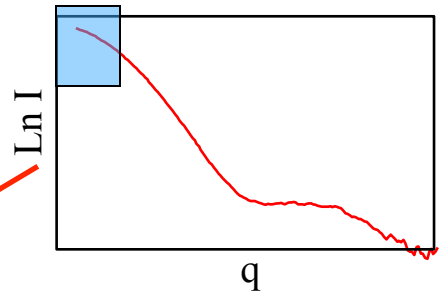
(q= 4π/λ sinθ)



Guinier approximation

$I(q) = I(0) \exp(-q^2 R_g^2/3)$

- Radius of giration
- I(0) Molecular weight



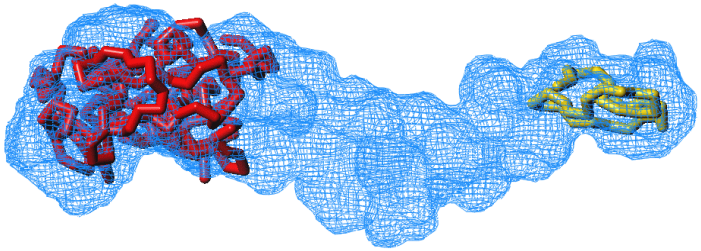
Distance distribution function P(*r*)

$= \int I(q) \exp(-2\pi i q r) dq$

- maximum distance D_{max}
- overall shape of the object

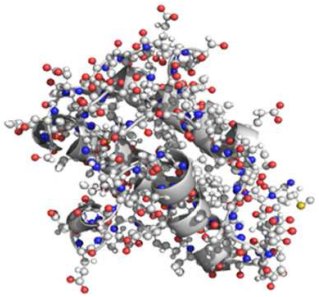
Form factor:

- Comparaison with crystal structure
- conformational modeling
- Calculation of envelopes or volumes



Reconstruction of the 3D structure from 1D SAXS is an ill-conditioned problem

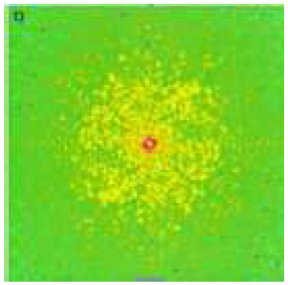
Real space 3D Molecule



Fixed orientation

Lose phase

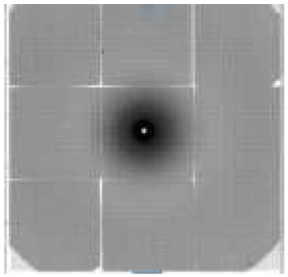
Reciprocal space 2D anisotropic image



Averaged orientation

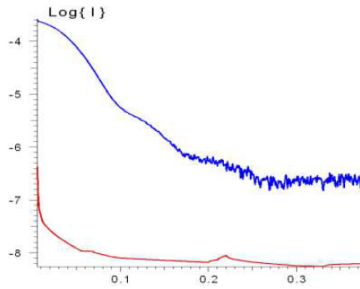
Lose orientation

Reciprocal space 2D isotropic image



Radial averaging

1D profile reciprocal space



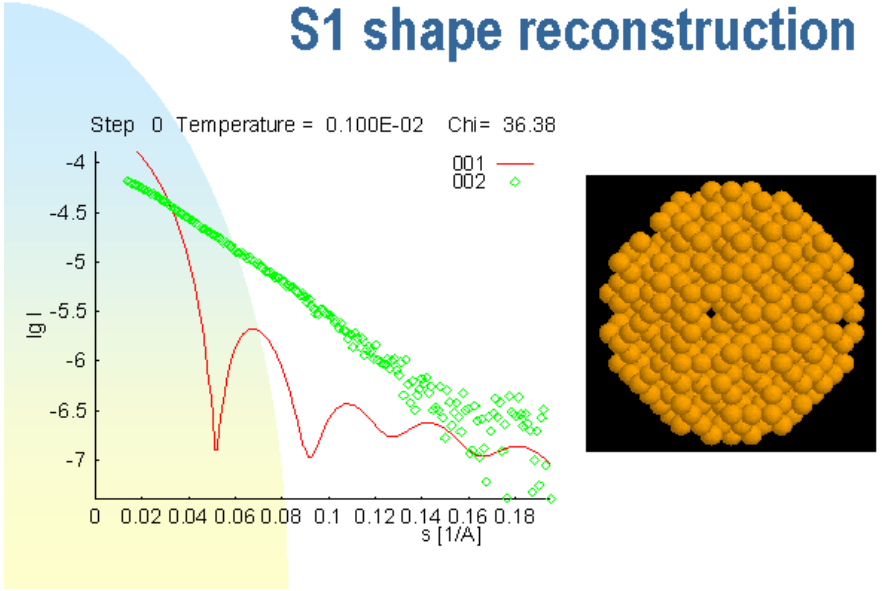
How to reconstruct the 3D structure from the 1D SAXS profile ?

Ambiguity of the solution :

One 3D structure → One SAXS curve **BUT** One SAXS curve → Many 3D structures

ab initio calculation of overall shape

➡ Fill envelope with dummy-atoms with a given diameter (3.8 to 10Å)

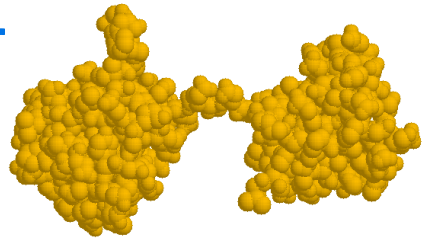


- **Global shape of macromolecule in solution**
- **Respective position of sub domains (known 3D structure)**

Programs DAMMIF, GASBOR (D. Svergun), DALAI_GA (F. Diaz)

Shape calculations

➡ Calculation of an envelope filled with pseudo-at of variable diameter (3.8 to 10Å)



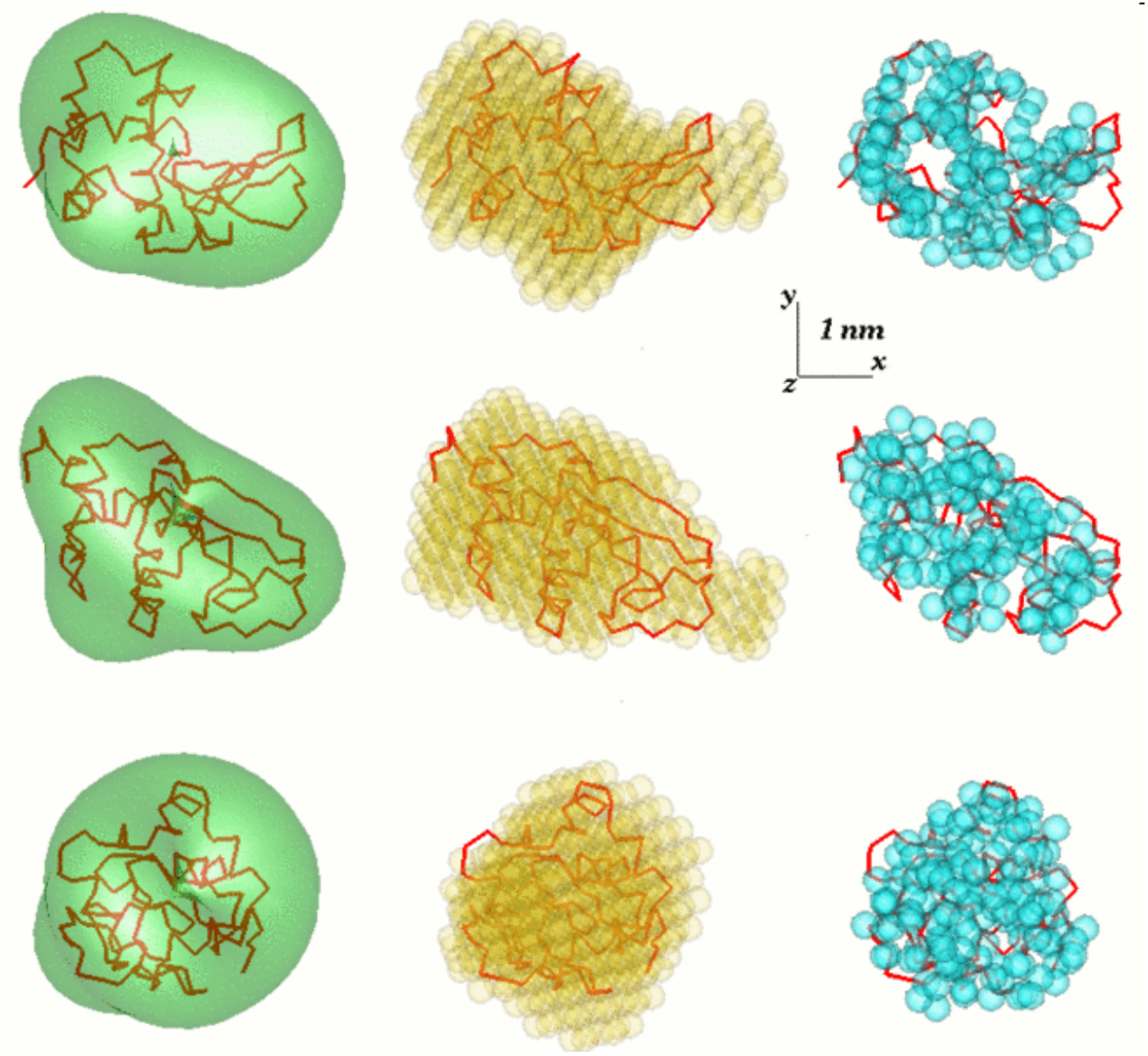
- DAMMIN/DAMMIF : Simulated annealing with constraints to minimize the surface
- GASBOR : Simulated annealing with distance constraints of close neighbors Ø=3.8Å

(D. Svergun)

- DALAI_GA : Genetic Algorithm without external constraints Ø depends on resolution

(F. Diaz)

http://www.embl-hambourg.de/ExternalInfo/Research/Sax/software.html


DALAI_GA
DAMMIN
GASBOR
y
1 nm
x
z

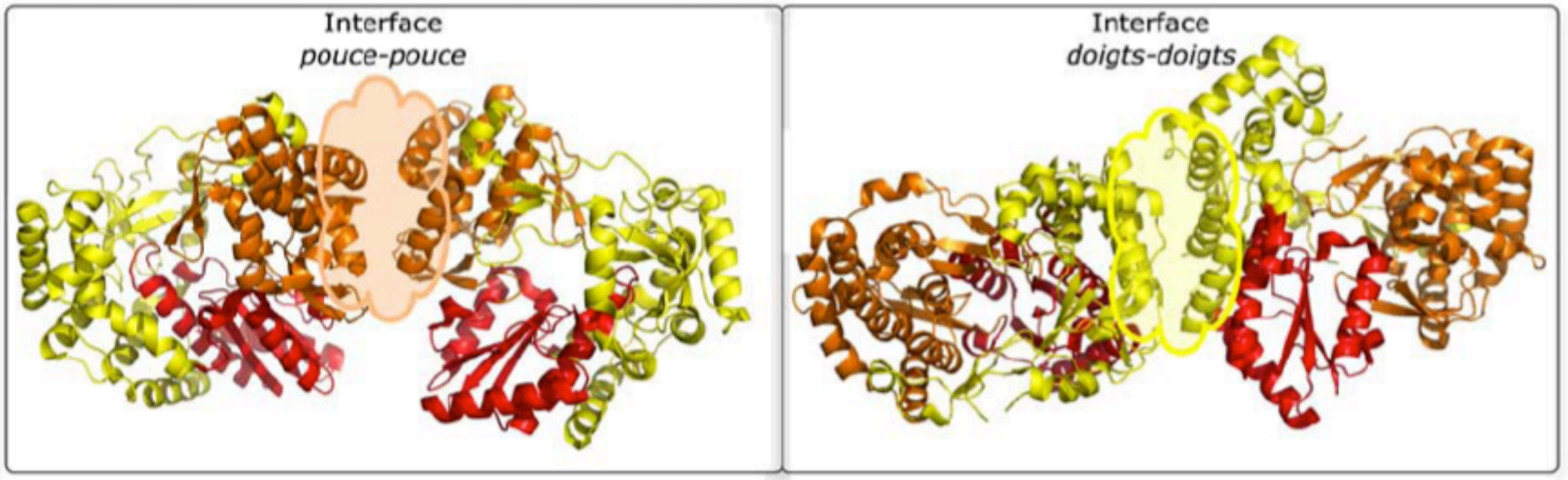
quaternary structure in solution and complexes

If 3D crystal structure of each sub-unit is known

⇨ Find the respective position of each sub-unit

- SASREF : **Rigid Body Modelling**

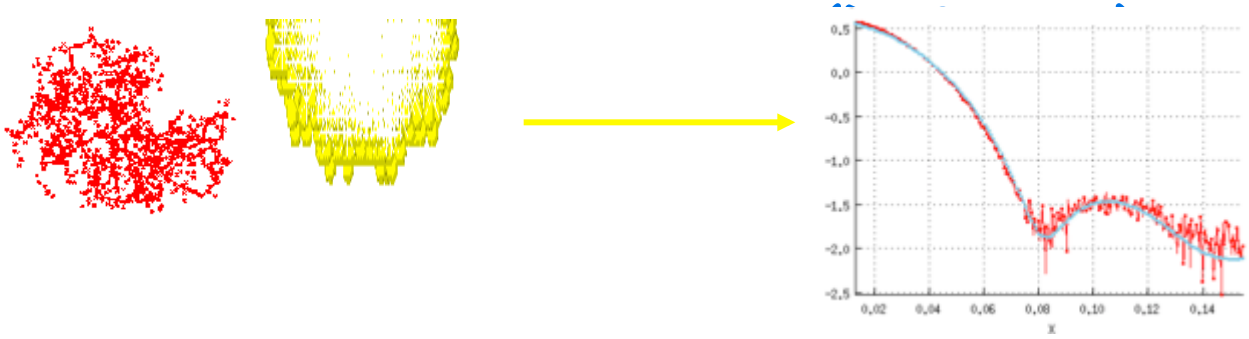
(D. Svergun)



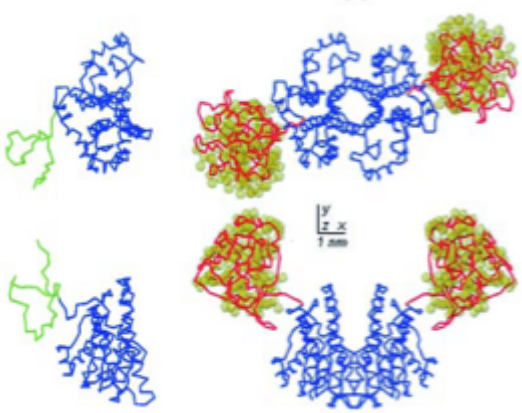
conformational flexibility

➡ Comparaison of solution structure to crystal structure

- CRYSOL: Calculate theoretical diffusion curve of a macromolecule taking in account the hydratation shell



➡ Calculate the conformation of a missing loop



- CREDO & BUNCH : position of C_α of a missing domain or portion of overall structure

(D. Svergun)

Comparaison of solution structure to crystal structure

.....more programs

Svergun D, Barberato C, and Koch M.H.J. (1995) **CRYSOL** - a program to evaluate x-ray solution scattering of biological macromolecules from atomic coordinates.
J. Appl. Cryst. 28, 768
Most popular for BioSAXS, stand-alone program, fit model to data, fast computational algorithm
http://www.embl-hamburg.de/biosaxs/atsas-online/crysol.php

Grishaev A, Guo L, Irving T, Bax A. (2010) **AXES** Improved Fitting of Solution X-ray Scattering Data to Macromolecular Structures and Structural Ensembles by Explicit Water Modeling. J. Am. Chem. Soc. 132, 15484-6.
Use explicit water modeling solvation layer, robust fitting approach
http://spin.niddk.nih.gov/bax/nmrserver/saxs1/

J. Bardhan, S. Park and L. Makowski (2009) **SoftWAXS**: a computational tool for modeling wide-angle X-ray solution scattering from biomolecules J. Appl. Cryst. 42, 932-943
A program to compute WAXS,
Upon request

Schneidman-Duhovny D, Hammel M, Sali A. (2010) **FoXS**: a web server for rapid computation and fitting of SAXS profiles. Nucleic Acids Res. 38 Suppl:W540-4.
Debye-like computation, web server based
http://modbase.compbio.ucsf.edu/foxs/

Zuo X, Zhang R, Tiede DM. **SolX**: A computer program for solution molecular x-ray scattering simulations . Photosynth Res. 2009 Nov-Dec; 102(2-3): 267-279.
Debye-like computation, Windows-based, can handle non-standard atoms/residues, for biomolecules and supramolecules
Upon request zuox@anl.gov, tiede@anl.gov

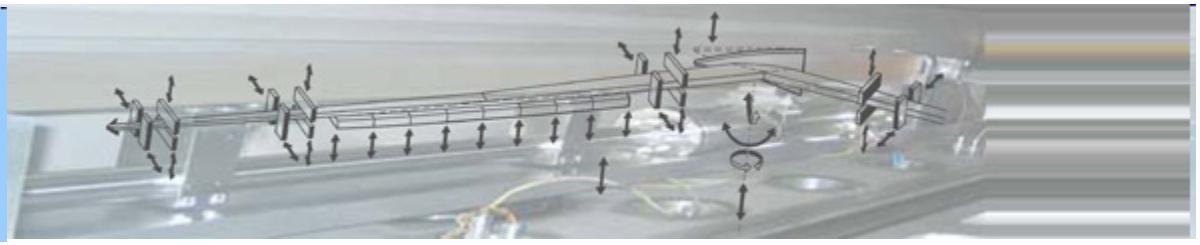
Group of Dmitri Svergun:

EMBL HAMBURG

Biological
Small Angle Scattering

http://www.embl-hamburg.de/biosaxs/

SAXS Forum:



http://www.saxier.org/

Comparison of methodological limits

SAXS

- ✓ provides envelope or overall shape

  average 3D conformations in solution

- ✓ needs pure samples
- ✓ large quantities
- ✓ supports irradiation with time
- ✓ does not form aggregates
- ✓ "low resolution" – no atomic information
- ✓ average of particles in solution

crystallography

- ✓ provides an atomic (average)

  3D structure of the crystallized molecule

- ✓ needs pure samples
- ✓ large quantities
- ✓ crystallizes
- ✓ produces Bragg diffraction
- ✓ no information of conformation in solution
- ✓ (almost) no information on dynamics

Complementarity of the methods

SAXS

crystallography

mean global conformation in solution

possibility to constrain modelisation of flexibility

solid state atomic 3D structure

allows to verify which oligomerisation is compatible with exp. SAXS curve

Ambiguity about oligomeric state

allows to analyze (estimate) the flexibility domain conformations in solution

Ambiguity about domain (or loop) organisation/conformations

Sign in

SASBDB

Small Angle Scattering Biological Data Bank

Search

Advanced search E.g. SASDBF4, Lyz, Nucleic Acids Res

Home | Browse | Submit data | About SASBDB | Help

SASBDB

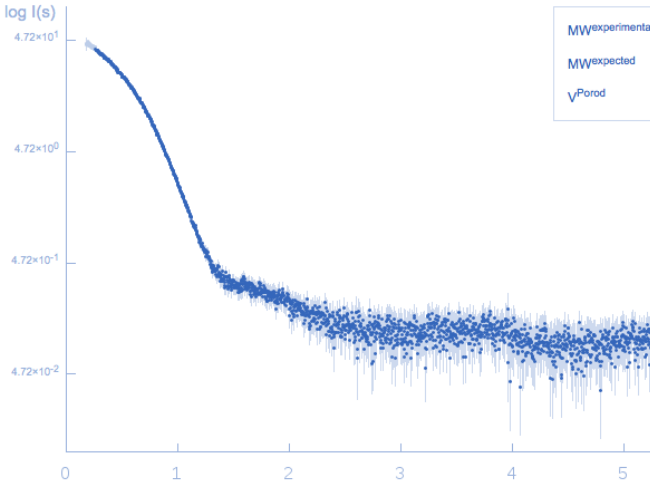
Structure of ThiM from Vitamin B1 biosynthetic pathway of Staphylococcus aureus - Insights into a novel pro-drug approach addressing MRSA infections.

Download files ▾

Drebes J, Künz M, Windshügel B, **Kikhney AG**, Müller IB, Eberle RJ, Oberthür D, Cang H, Svergun DI, Perbandt M, Betzel C, Wrenger C, Sci Rep 2016 Mar 10;6:22871 PubMed

SASDAX8 – Ribokinase ThiM

Ribokinase ThiM



$MW^{experimental}$	77	kDa
$MW^{expected}$	89	kDa
V^{Porod}	145.1	nm^3