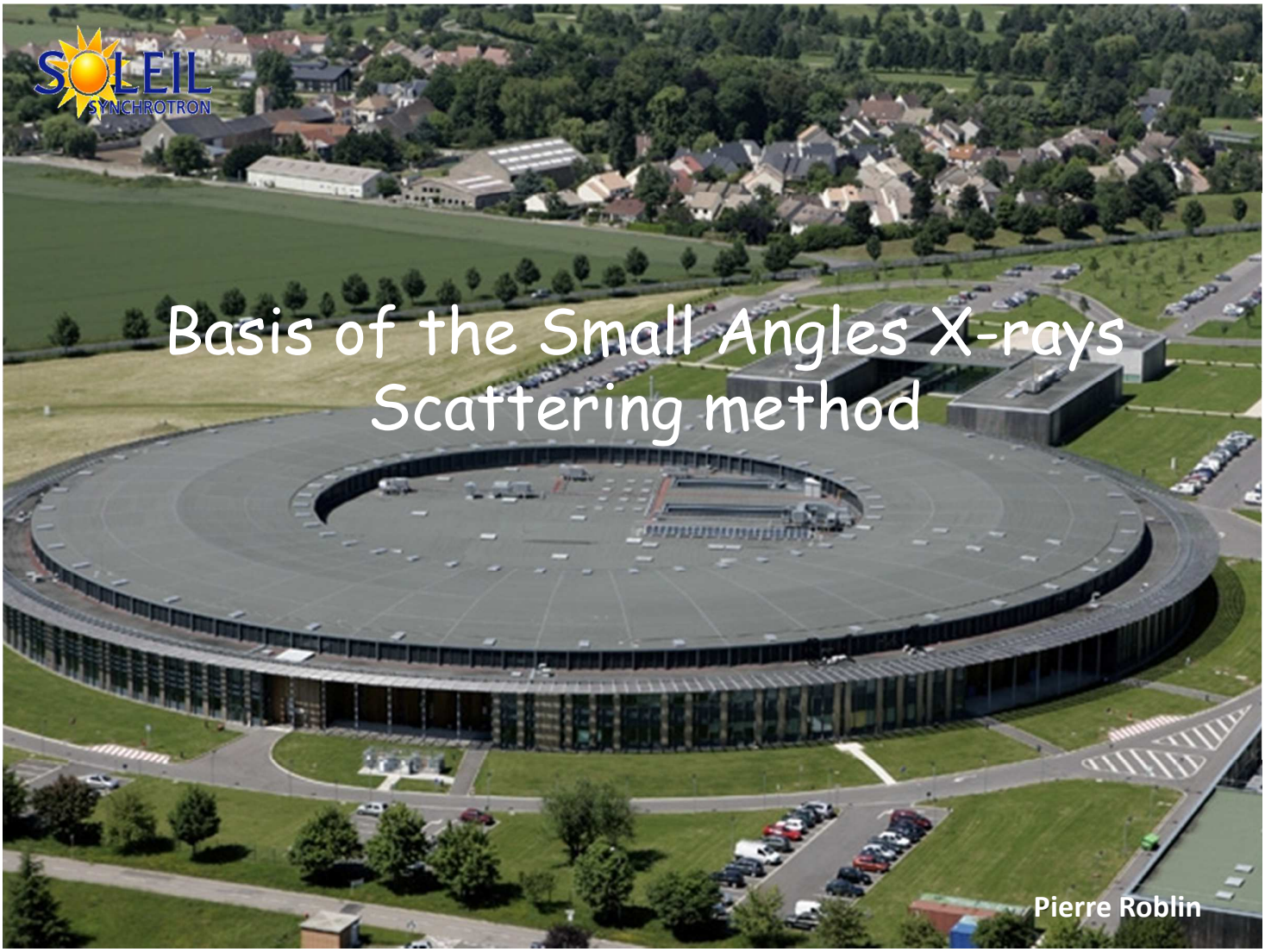


Pierre Roblin

<u>1997-2001</u> : Maitrise de biologie structurale Université Paul Sabatier

- *Résolution de structures par cristallographie des rayons X : β-Lactamases (Laurent Maveyraud, IPBS équipe biocristallographie J.P. Samama)*

<u>2001-2003</u> : Institut National des Sciences Appliquées de Toulouse. Bio-ingénierie des protéines option biologie structurale

- *RMN 2D sur des peptides de synthèse (James H. Davis, Université de Guelph, Ontario) et (Virginie Gervais, IPBS équipe RMN structurale A. Milon)*

<u>2002-2003</u> : DEA de biologie structurale Université Paul Sabatier

- Résolution de structures par cristallographie des rayons X : leucotoxine S. aureus (IPBS équipe biophysique structurale L. Mourey)

<u>2003-2007</u> : Thèse de biologie structurale Université Paul Sabatier

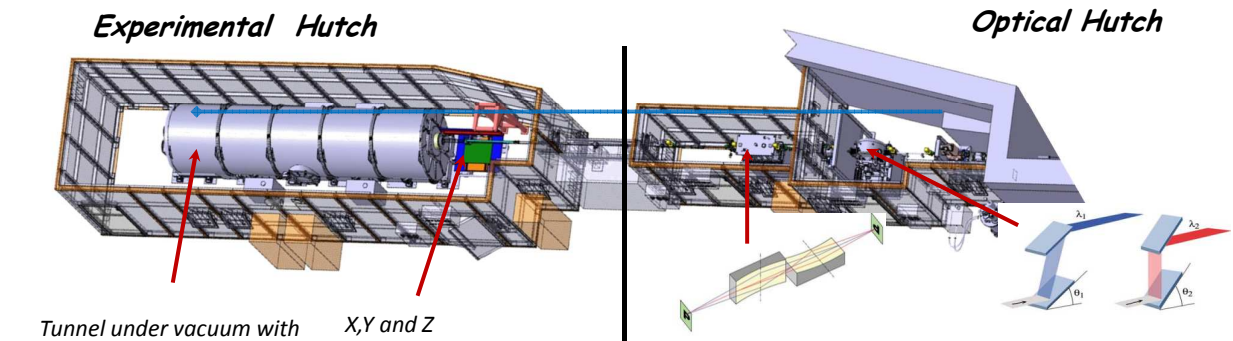
- RX + SAXS : polykétide synthase Myco. Tub (IPBS équipe biophysique structurale L. Mourey) et (D. Svergun EMBL Hambourg)

<u>2007-2009</u> : Post doc L.E.B.S. Gif sur Yvette

- RX + SAXS : complexe actine (Louis Renault équipe MF Carlier)

<u>Depuis 2009</u> : Ingénieur de recherche INRA, mise en disponibilité au synchrotron SOLEIL, ligne SWING

Set-up for BioSAXS at SWING Beamline

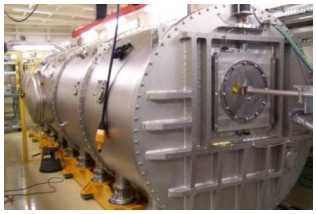


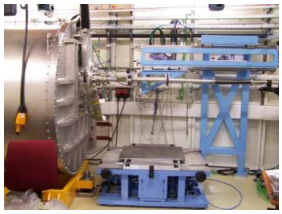
Tunnel under vacuum with X-rays detector

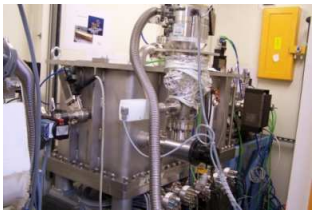
X,Y and Z motorized table

Double cristals mirror

Monochromator

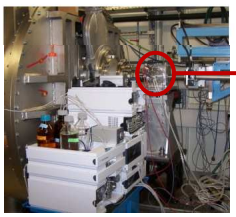




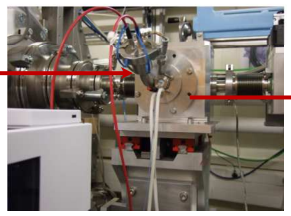


Sample environment dedicated to the biology :

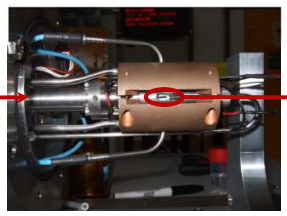
HPLC as autosampler



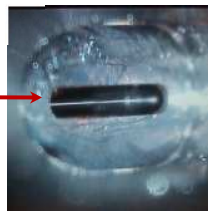
HPLC on the beamline



SAXS cell under vacuum

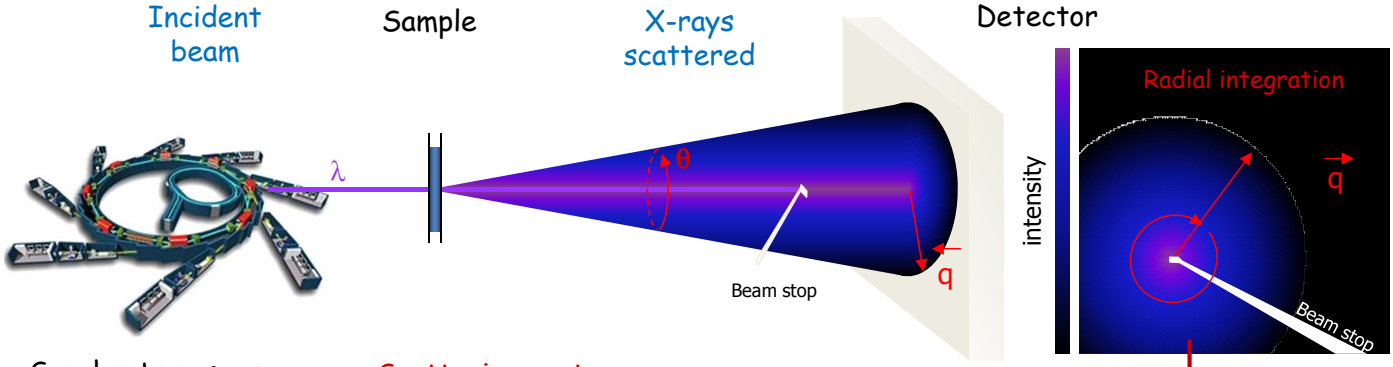


Details of the SAXS cell



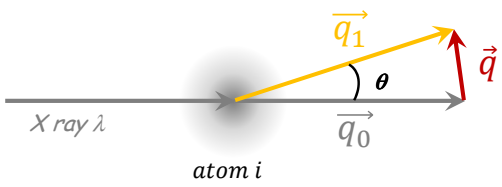
Quartz capillary

Principle of Small Angle X-rays Scattering experiences



Synchrotron source

Scattering vector



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

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

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

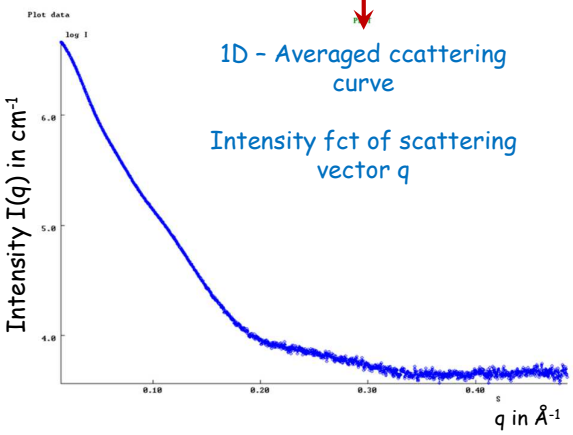
$\vec{q}$ scattering vector

$\vec{q_0}$ vector of incident wave

$\vec{q_1}$ vector of scattered wave

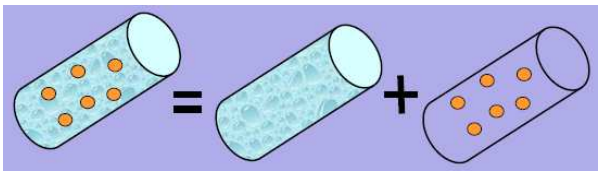
1D – Averaged ccattering curve

Intensity fct of scattering vector q

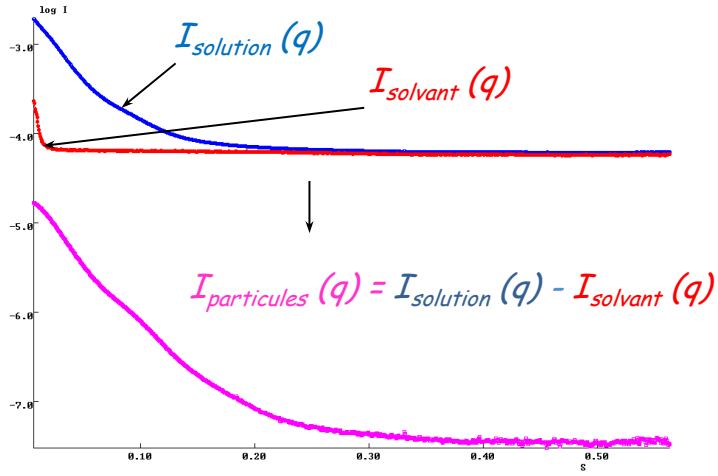


Basis of SAXS

Method sensible to the difference between the electronic density of the particule and the solvent (contrast)



$I_{solution}(q)$ $I_{solvant}(q)$ $I_{particules}(q)$



Structural information obtained from scattering curve

- biophysical parameters (size and shape of the object)
- molecular weight, oligomerization state and volume

Biophysical informations calculated directly from the SAXS curve

- *low resolution molecular shape calculation with ab initio method*
- *comparison with high resolution model*

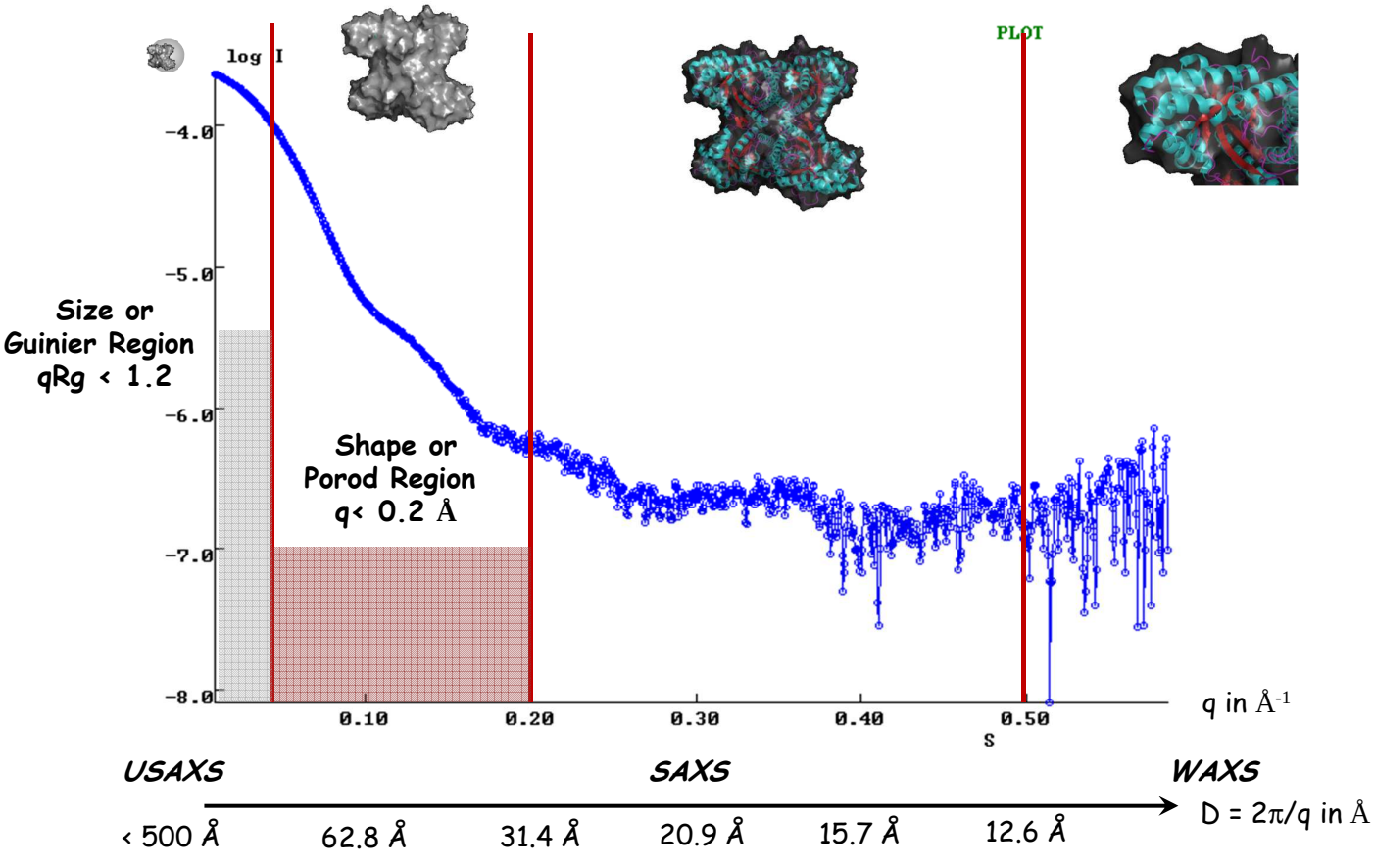
3D structural informations

- *molecular modeling of unstructured missing part*
- *molecular modeling rigid body of complex*

SAXS data compatible model NOT a structure

A typical SAXS solution curve of protein

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



Asymptotic behaviour at small angles : Guinier law

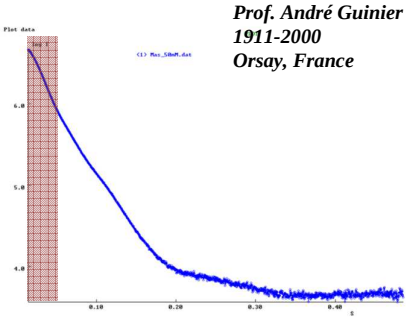
Close to q=0, the scattering intensity of a particle can be described by a Gaussian curve.
The Guinier law is equivalent of a linear variation of Ln(I(q)) vs q^2 (Guinier plot), providing Rg and I(0). The validity domain actually depends on the shape of the particle and is around $q < 1.3/Rg$ for a globular shape.

Prof. André Guinier
1911-2000
Orsay, France

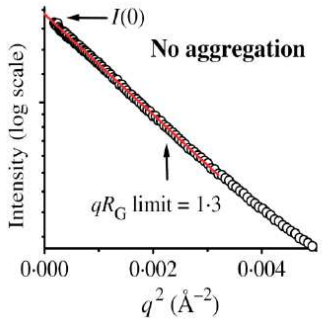
$$I(q) = I(0)\exp\left(\frac{-q^2 Rg^2}{3}\right) \quad \text{or} \quad Ln[I(q)] = Ln[I(0)] - \frac{q^2 Rg^2}{3}$$

Extrapolated intensity at origin

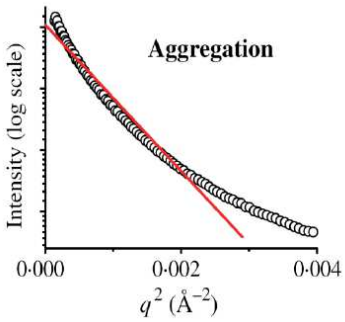
Radius of gyration



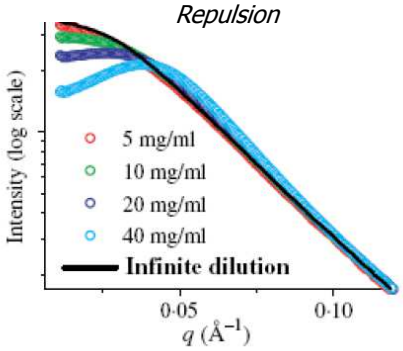
Mono-dispersed



Poly-dispersed aggregates



Repulsion



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

Determination of the mass from Guinier law

From extraplated intensity at the origin I(0), the molecular mass can be determined with the following equation :

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

Classical electron radius

Mass concentration

Protein specific volume

Electronic density contrast

M_{exp}(kDa) = I (0) *N_A /[c $\Delta\rho_m^2$]

with $\Delta\rho_m \approx 2 \cdot 10^{10}$ cm/g for proteins and I(0) in cm^{-1}

I(0) gives an independent estimation of the molar mass of the protein
(only if the mass concentration, c, is precisely known …)

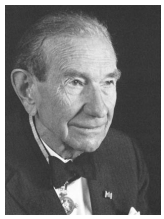
I(0) must be expressed in absolute units cm^{-1} necessiting a calibration before.

Typically : M (kDa) = 1500 * I(0) (cm^{-1}) / C (mg/ml)

Asymptotic behaviour at larges angles : Porod law

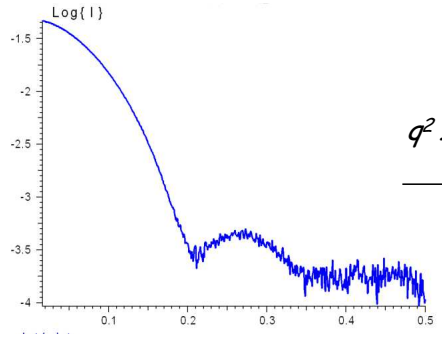
SAXS provides a sensitive means to ***evaluate the degree of compactness*** of a protein:

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

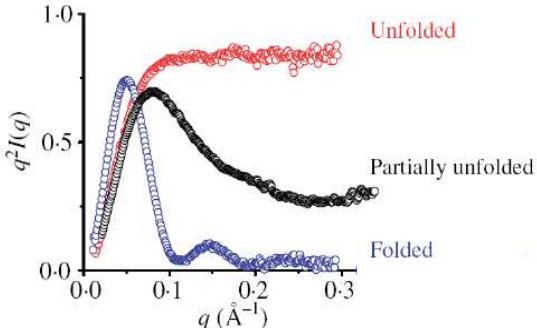


Prof. Otto Kratky
1902-1995
Graz, Austria

<u>This is most conveniently represented using the so-called Kratky plot:</u>



$q^2 I(q)$ *versus* q



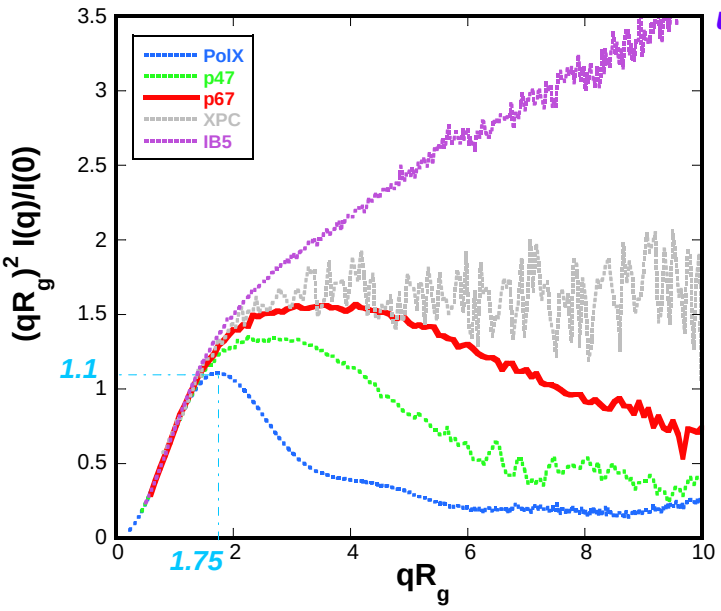
Putnam, D., et al. (2007) Quart. Rev. Biophys. 40, 191-285.

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

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

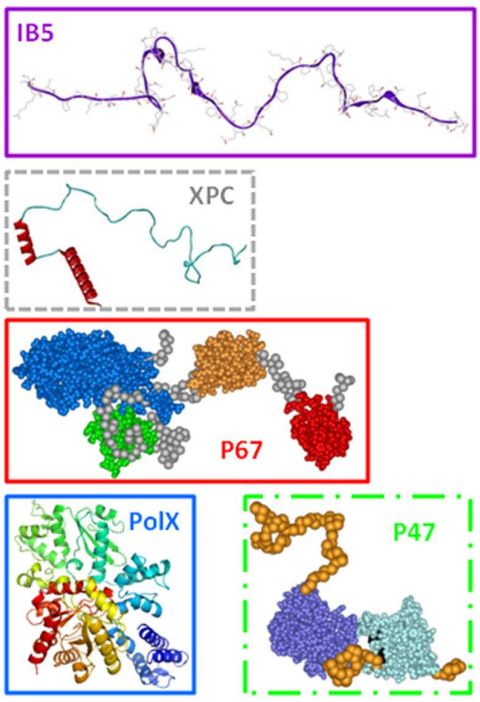
Extended polymer chain : ***increase*** at large q-values (asymptotic behaviour in $I(q) \sim q^{-1.x}$)

Dimensionless Kratky Plots of unfolded proteins



unfolded

globular



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

The curve increases at large q as the structure extends.

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

Asymptotic behaviour at larges angles : Porod law

Hypothesis : the particle has a well-defined interface with the surrounding buffer and a uniform electronic 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

$$\phi = \frac{N\cdot V_{obj}}{V}$$

Number of proteins

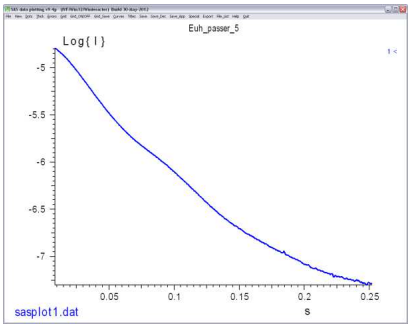
Protein volume

Solution volume

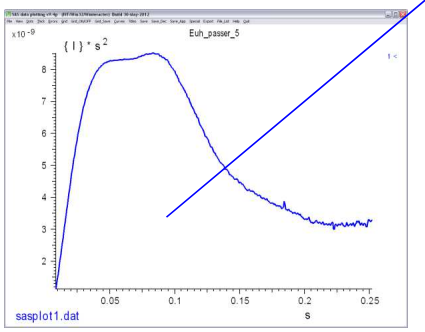
Does not depend on shape, only on contrast

The Porod Invariant is the integral of this curve

Representation : Log I(Q) vs Q



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



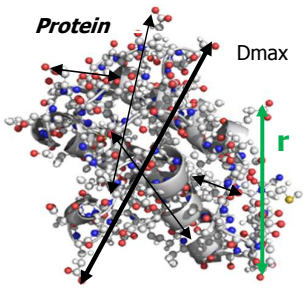
$$\mathbf{Q} = \int_0^{\infty} I(q)\cdot q^2 dq$$

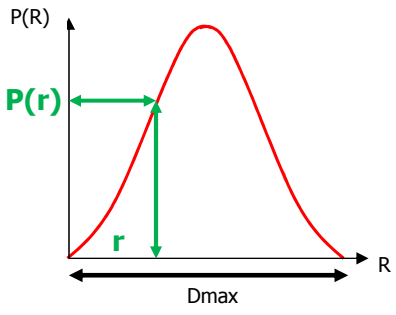
$$V_{obj} = \frac{2\pi^2\cdot I(0)}{\mathbf{Q}}$$

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

Back to real space : Distance Distribution Function

The distance distribution function represent the distribution of the distance between each atoms pair. The size of the particule is limited and satisfied the conditions where P(r=Dmax) = 0 and P(r=0) =0





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

$$R_g^2 = \frac{\int_0^{D\max} r^2 P(r)dr}{2\int_0^{D\max} P(r)dr}$$

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

This alternative estimate of R_g makes use of the whole scattering curve, and is much less sensitive to interactions or to the presence of a small fraction of oligomers. Comparison of both estimates : useful cross-check

Back to real space : Distance Distribution Function

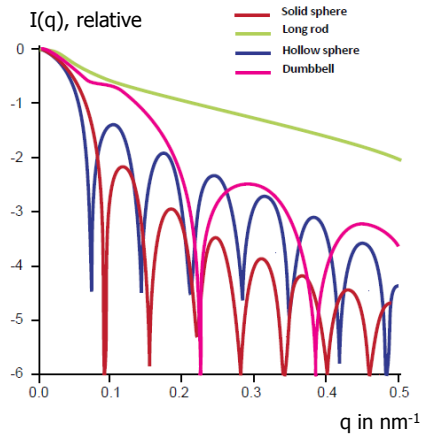
The scattered intensity I(q) can be written with the distribution function P(r). P(r) function is calculated with indirect Fourier transform applied to the scattered intensity I(q). The both curves contain the same information.

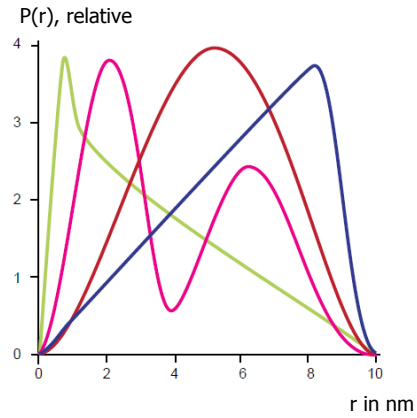
$$I(q) = 4\pi r_e^2 \varphi \int_0^{D\max} P(r) \frac{\sin(qr)}{qr} dr$$

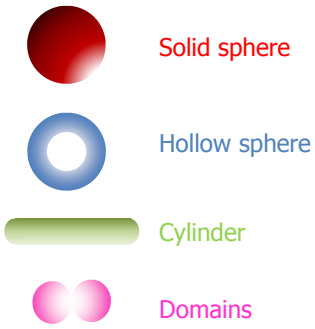
$$P(r) = \frac{r^2}{2\pi^2 \varphi r_e^2} \int_0^{\infty} q^2 I(q) \frac{\sin(qr)}{qr} dq$$

with Dmax as maximal distance in the particule

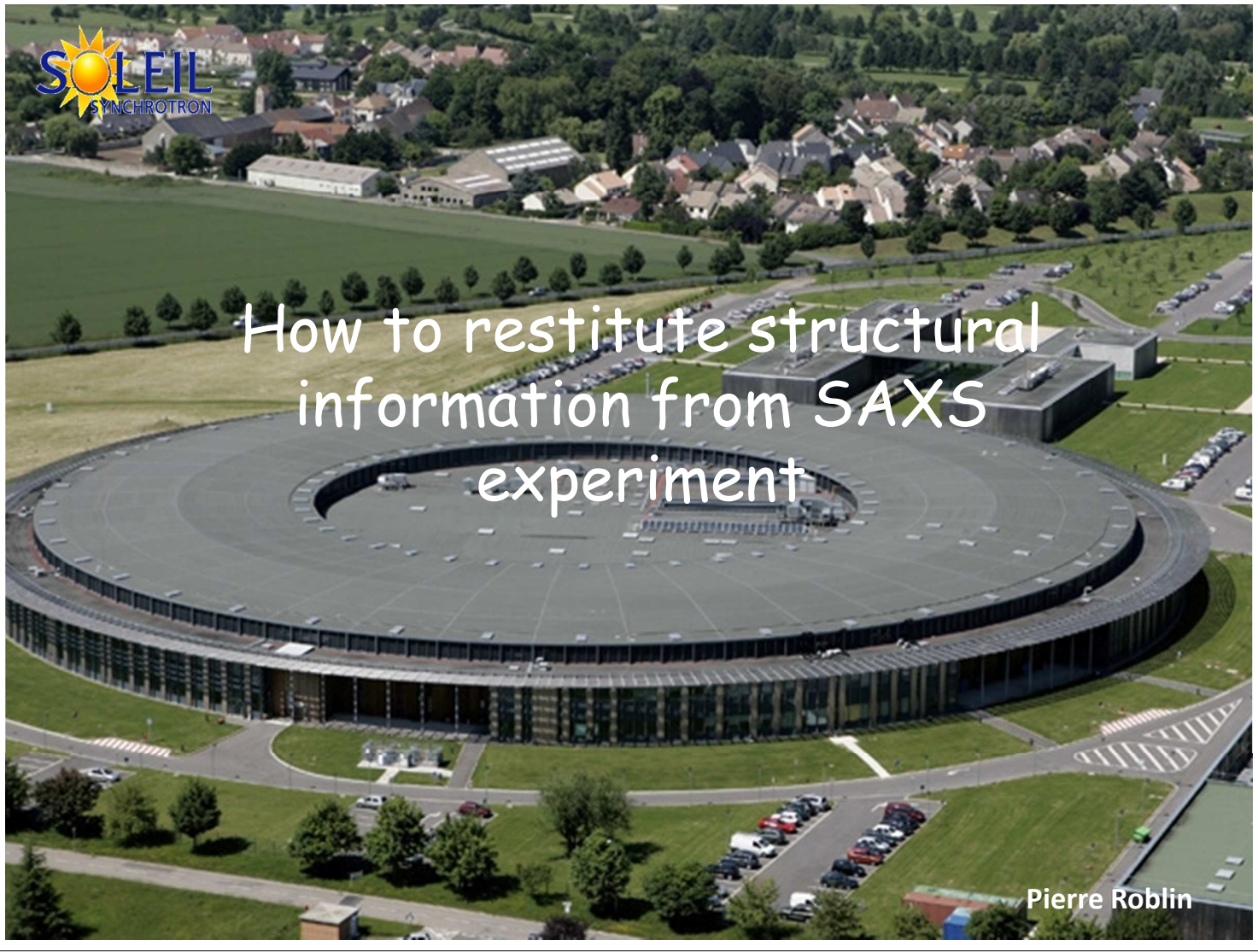
However, direct calculation of P(r) from I(q) is made difficult and risky by $[q_{min}, q_{max}]$ truncation and data noise effects.







The pair distribution function entirely depends on the shape of the particle



SAXS experiments : strategy

Data analysis

- *Guinier approximation*
  - Rg (size) and I(0) (mass and oligomeric state)
- *Distance distribution function p(r) :*
  - Dmax evaluation
  - Rg (size) and I(0) compatibility with Guinier approximation
  - Global form 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*
  - 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

From an atomic structure to a solution scattering pattern

The scattering pattern of a particule with an atomic structure resolved by crystallography or NMR can be solved analytically

Debye method to compute scattering of electrons from nuclear position :

$$I(q) = \sum_{i=1}^{M} \sum_{j=1}^{M} F_i(q) F_j(q) \frac{\sin(q.r_{i,j})}{q.r_{i,j}}$$

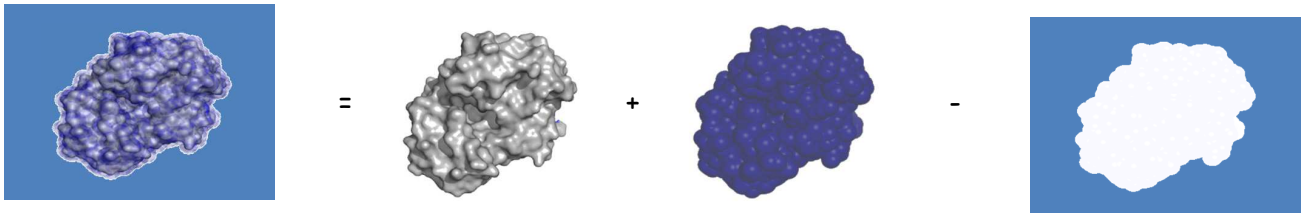
$F_i(q)$, $F_f(q)$, Form factor of atom i and atom j

M number of atom in the protein

Distance r between atom i and atom j

Approach computationally expensive and time-cost increases quadratically with the number of atom in the protein

The experimental scattering curves are obtained by substracting the contribution of the solvent. But the solvated molecules have a border of solvent bound with a diffusion density different from the disordered solvent



Molecule in solution | Molecule in vacuum | Hydrated molecule | Excluded volume

From an atomic structure to a solution scattering pattern

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

$A_a(q)$ = atomic scattering in vacuum

$A_a(q)$ = scattering from the hydratation shell, layer of thickness 3Å

$A_s(q)$ = scattering from excluded volume

In CRYSOL program, in order to gain computing time, I(q) is developped 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
- the contrast of the border layer δρ

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

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

From an atomic structure to a solution scattering pattern

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/

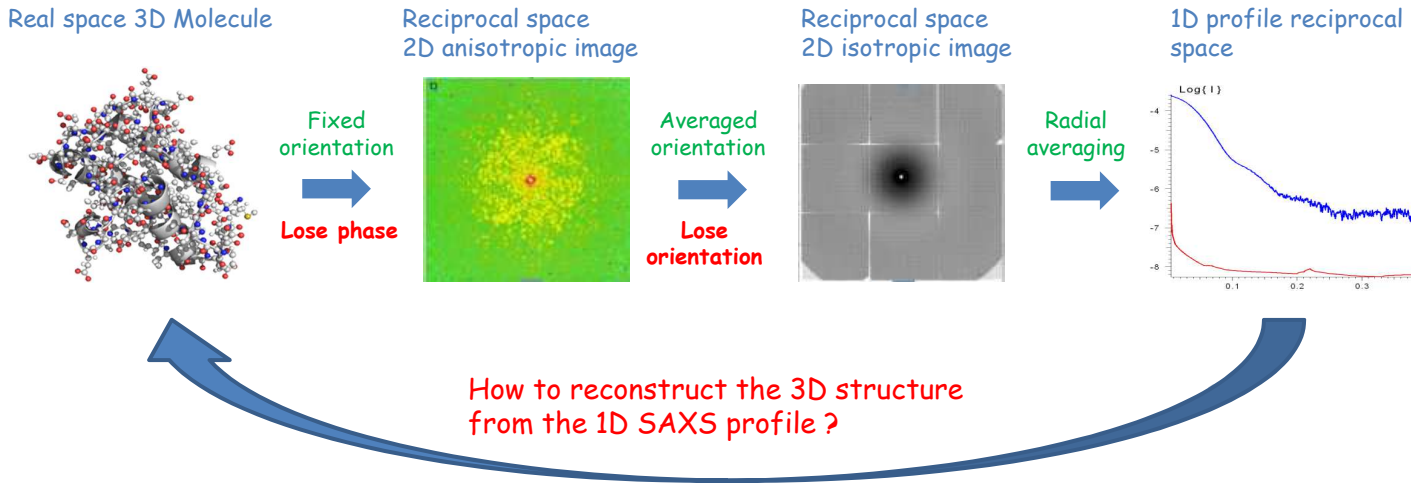
J. Bardhan, S. Park and L. Makowski (2009) **SoftWAXS**: a computational tool for modeling wide-angle X-ray solution scattering from biomolecules J. Appl. Cryst. 42, 932-943
A program to compute WAXS,
Upon request

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

SAXS for 3D structure reconstitution

The 1D SAXS profile is the Fourier transform of the 3D structure. Contrary to the direct scattering calculation, the inverse problem cannot be solved analytically, i.e., no "inverse Debye" formula can be constructed to yield 3D position coordinates from scattering data.



Ambiguity of the solution :

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

3D shape reconstructions from SAXS data with DAMMIN

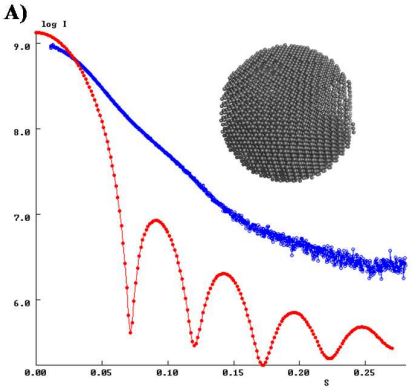
Ab initio shape modelling : nothing is known excepted the curve !

Principle of the method : any structure 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 ($r_0 \ll R$)

The number of the "dummy atom" $N \approx (R/r_0)^3$

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



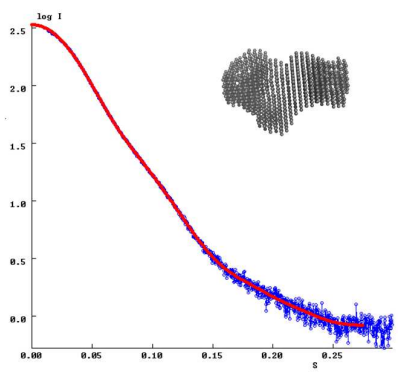
$$f(X) = \chi^2\left[I(q)_{\exp}, I(q,X)\right] + \alpha P(X)$$

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

After k iterations

X is a conformation of the system

P(X) is a penalty function



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

3D shape reconstructions from SAXS data with DAMMIN

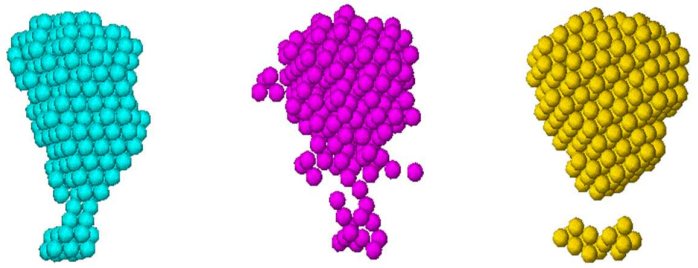
Obtaining 3D shapes from SAXS data is a defined problem that could be solved by introducing additional information to reduce ambiguity of interpretation

Introduction of the penalty function to limit the formation of discontinuous models or disjoint spheres

$$P(X) = 1 - \langle C(N_e) \rangle \qquad C(N_e) = 1 - \exp(-0.5 N_e)$$

N_e is the number of contact of a sphere with the neighboring spheres, N_e is equal to 12 in hexagonal lattice.

- In this case where the sphere has a maximum contact $C(12) = 1$ so $P(X) = 1-1 = 0$ (no penalty)
- With disconnected or loose sphere where $C(0) = 0.002$ so $P(X) = 0.998$ (strong penalty)
- With sphere on the surface $N_e \approx 6$, $C(6) = 0{,}943$ so $P(X) = 0.057$ (low penalty)



Compact model

Loose model

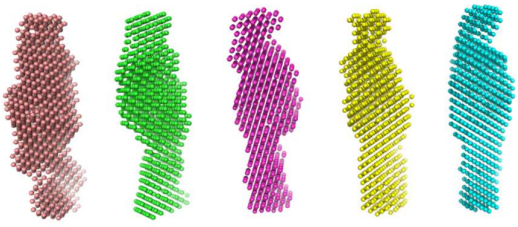
Disconnected model

3D shape reconstructions from SAXS data with DAMMIN

DAMMIN : necessit to perform a serie of run (20-50) to compare the different shape obtained with the same data.

After the run, an optimal superposition of models is realized with the program suite DAMSEL and DAMSUP.

The algorithm define a criteria of similarity, called « Normalized Spatial Discrepancy » or NSD, which measure the agreement between two models.
For similar shape NSD < 1, typically very similar shape NSD ≈ 0.5



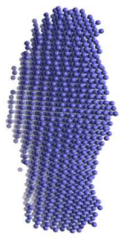
Shp1 Shp2 Shp3 Shp4 Shp5

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 (average)



Damaver (all superimposed)

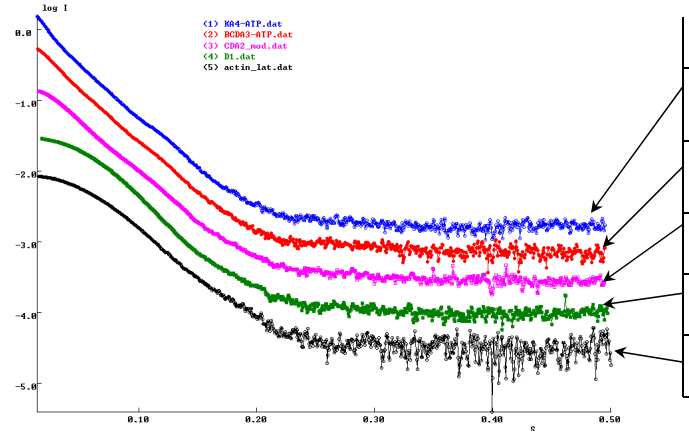
Model are conserved if the NSD < Mean of NSD + 2*standart deviation

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.pbd because $I_{damfilt}(q) \neq I_{exp}(q)$

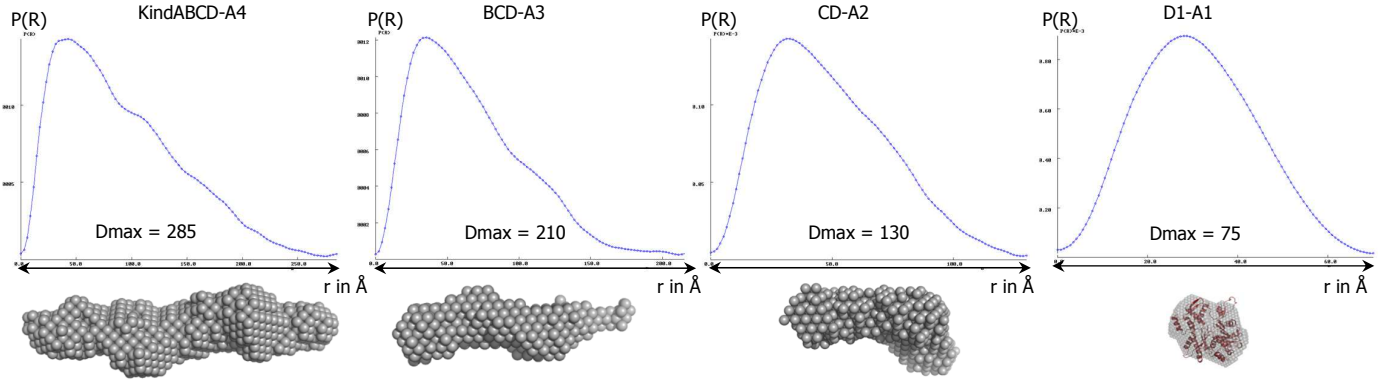
3D shape reconstructions from SAXS data with DAMMIN

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

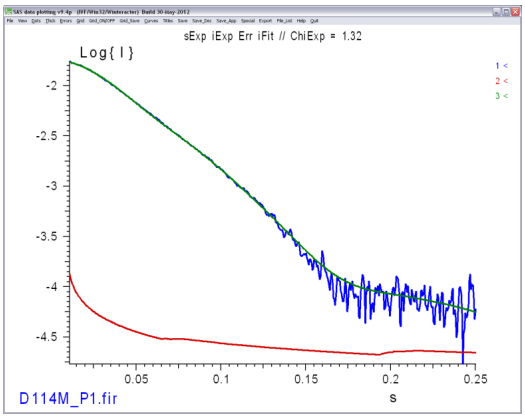


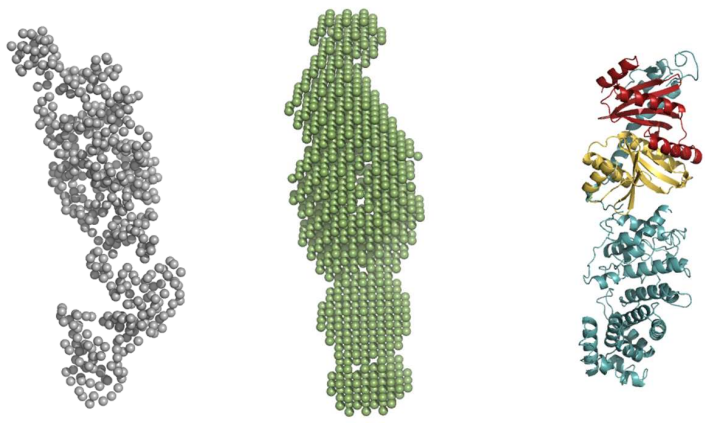
Ab initio model accounting for high resolution data

DAMMIN/DAMMIF : very low resolution because restricted portion of the data used ($q < 0.2$ Å^{-1}), and amguity of the models

GASBOR : a protein comprising N residues is represented by an ensemble of N spheres centered at the Cα positions.

An intial gas-like distribution of dummy residues is refined using Simulated Anneling to fit the data under constraints ensuring a final chain like distribution





GASBOR beads model | DAMMIF shape | High resolution structure

D. Svergun et al.(2001), Biophys. J., 80, 2946-2953.

Rigid body modeling against SAXS data

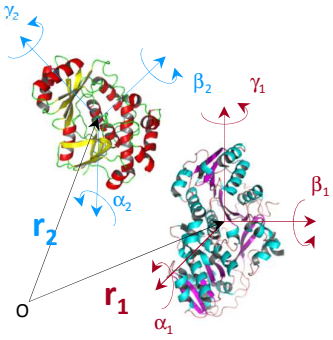
SASREF : when atomic structures of domains are known, but no their mutual organization

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

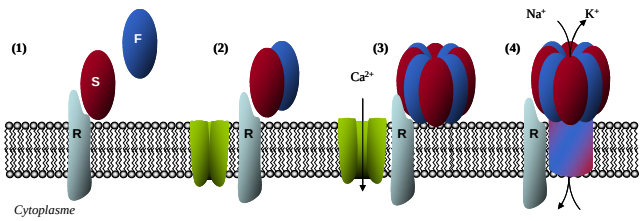
The algorithm of minimization is the same used with DAMMIN with a penalty function (interconnectivity of the subunits, the steric clashes) and possibility to give information about contacting residues from other experiences.

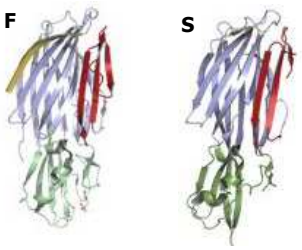
$$f(X) = \sum_i \chi_i^2 + \alpha_{dist} P_{dist}(X) + \beta_{cross} P_{cross}(X) + \gamma_{cont} P_{cont}(X)$$

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

Rigid body modeling against SAXS data

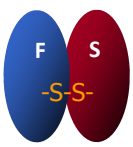
Example : structural characterization of covalent heterodimer

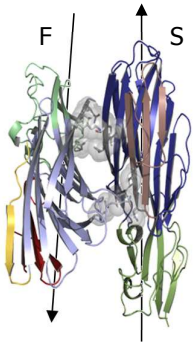




Structure of each subunits solved but no crystal of the hetero-octamer

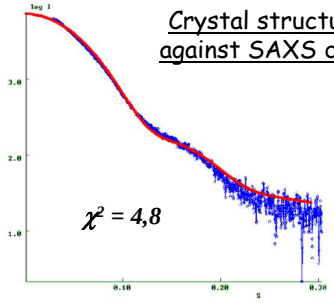
Conception of a covalent F-S heterodimer with the same biological properties as individual F and S monomers



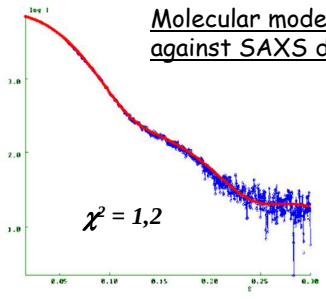


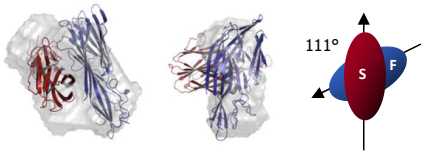
The structure solved by cristallogrphy is imcompatible with the formation of octamer

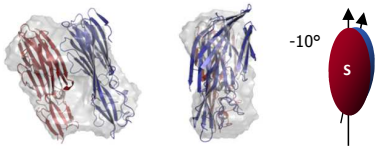
Crystal structure against SAXS data



Molecular modeling against SAXS data





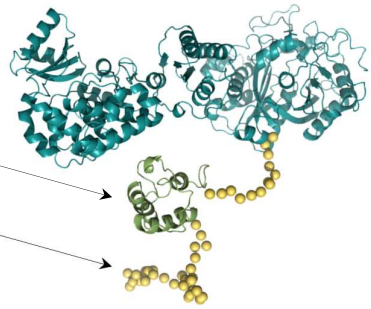


Rigid body modeling with missing loop against SAXS data

BUNCH and CORAL : quaternary structure analysis of multidomain protein

Combination of rigid body and ab initio modeling :

- position and orientation of rigid domains
- probable conformation of flexible linkers



$$f(X) = \sum_i \chi_i^2 + \alpha_{ang} P_{ang}(X) + \beta_{cross} P_{cross}(X) + \gamma_{dih} P_{dih}(X) + \delta_{ext} P_{ext}$$

As SASREF, the amplitude are calculated with CRYSOL from the high resolution structure of each monomer

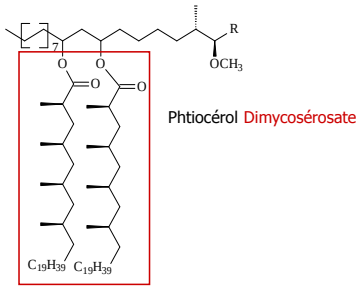
The algorithm of minimization is the same used with SASREF with a penalty function including the steric clashes Pcross, the dihedral angle Pang and Pdih, and the compactness of the loop Pext. The possibility to give information about contacting residues from other experiences is also added.

Flexibility → no unique structure !
NOT a structure but a SAXS data compatible model

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

Rigid body modeling with missing loop against SAXS data

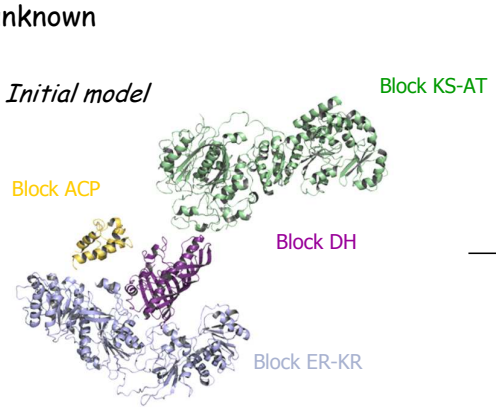
Example : structural characterization of multidomain protein MAS (mycoserosic acid synthase)

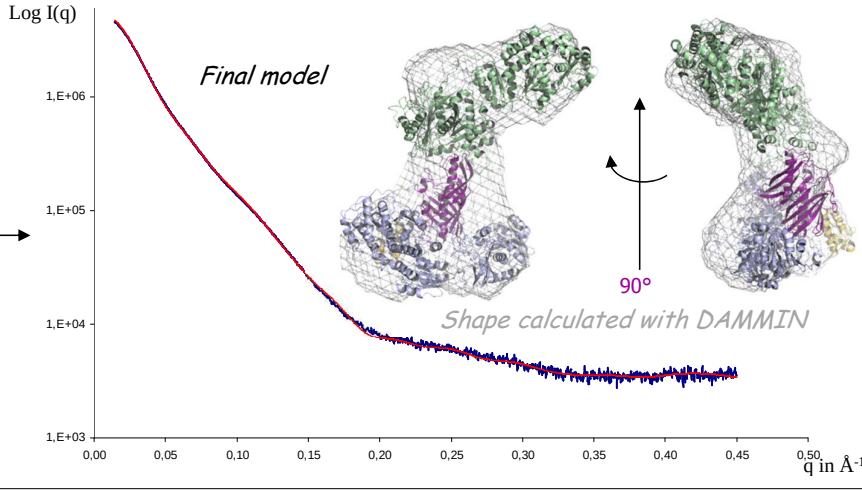


Phtiocérol Dimycosérosate

KS : ketosynthase domain
AT : acyl transferase domain
DH : dehydratase domain
ER : enoyl reductase domain
KR : keto reductase domain
ACP : acyl transferase domain

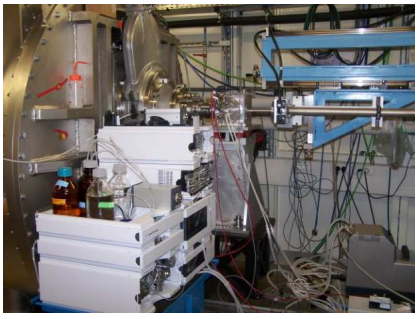
Six catalytic domains, each structure of the domain are known but he structure of entire protein is unknown

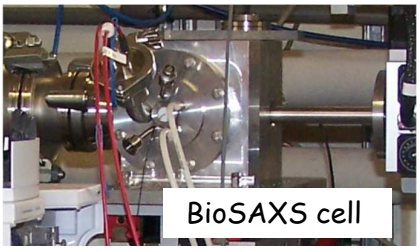




SAXS experiments : why use HPLC-SEC ?

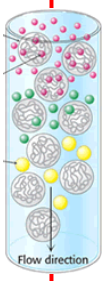
HPLC-SEC coupled to the beamline

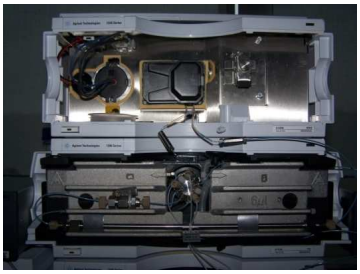


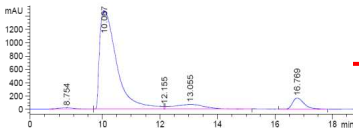


Autosampler

Separation column and UV detection

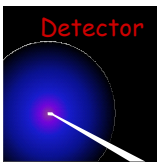


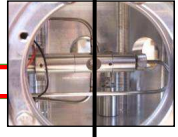




Sample for HPLC : 10µl (20mg/ml) to 80 µl for (2,5mg/ml)

Sample for direct injection : 20µl to 50µl with [protein] > 1mg/ml





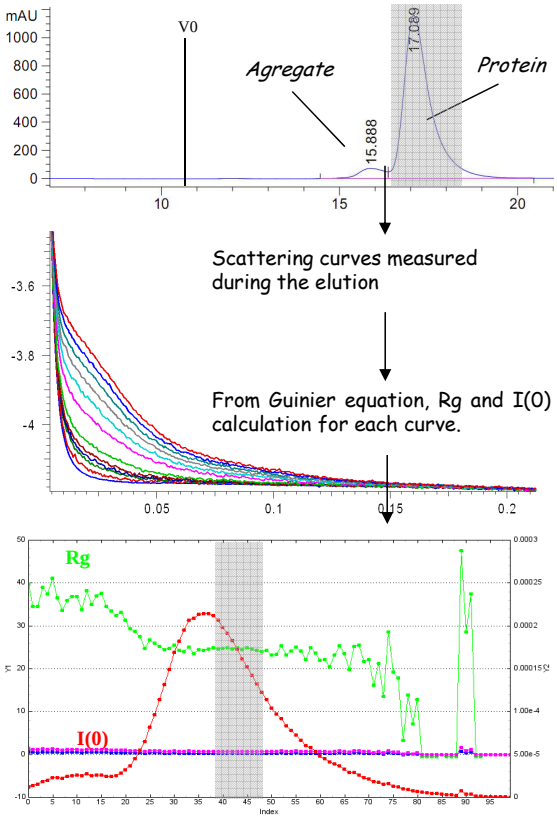
BioSAXS cell

X-rays

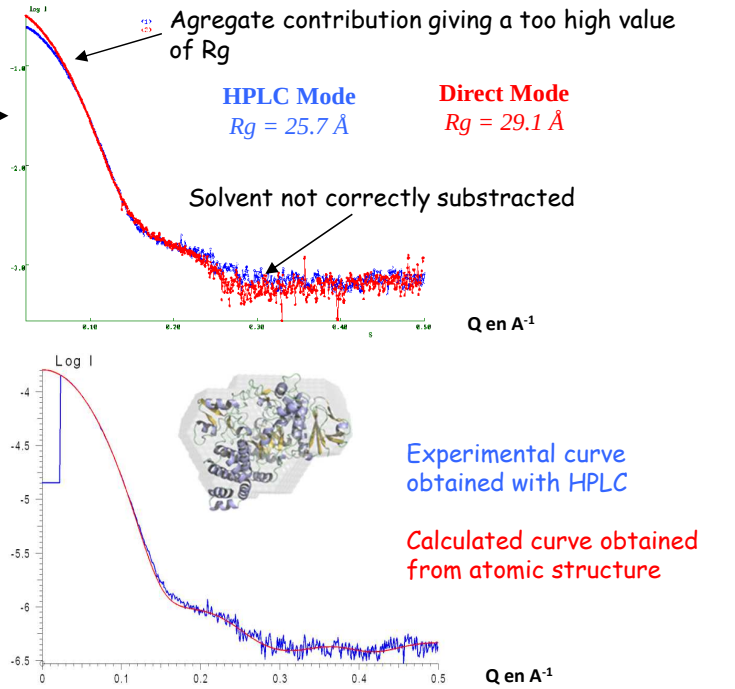
- Monodisperse solution by removing aggregation
- Oligomers separation
- Perfect substraction of the solvent
- Possibility to study protein complex with low Kd

SAXS characterization of macromolecule with coupled HPLC

Typical experience performed with HPLC



Comparison between curve obtained with direct injection and after HPLC method

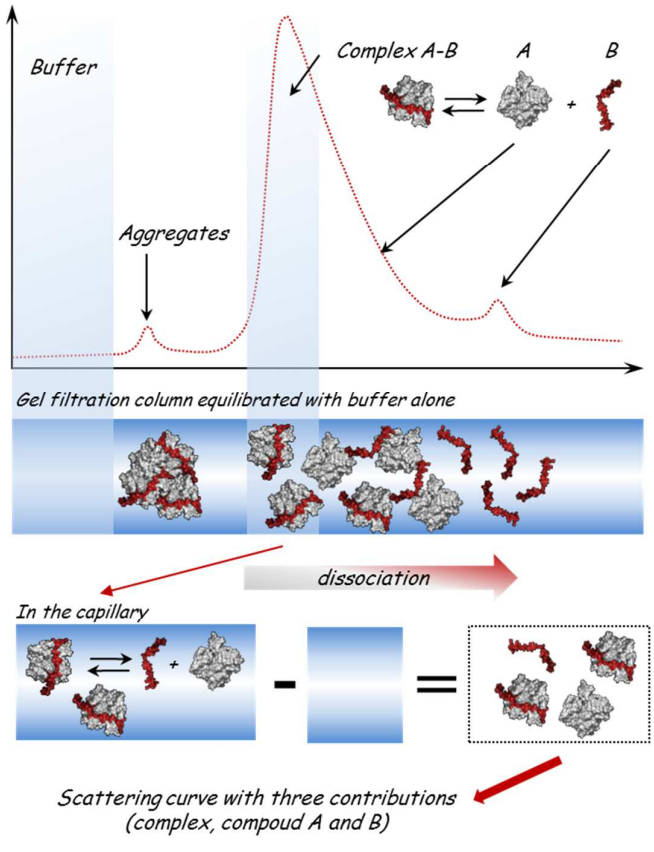


Column separation ensure monodispersity of the measured solution

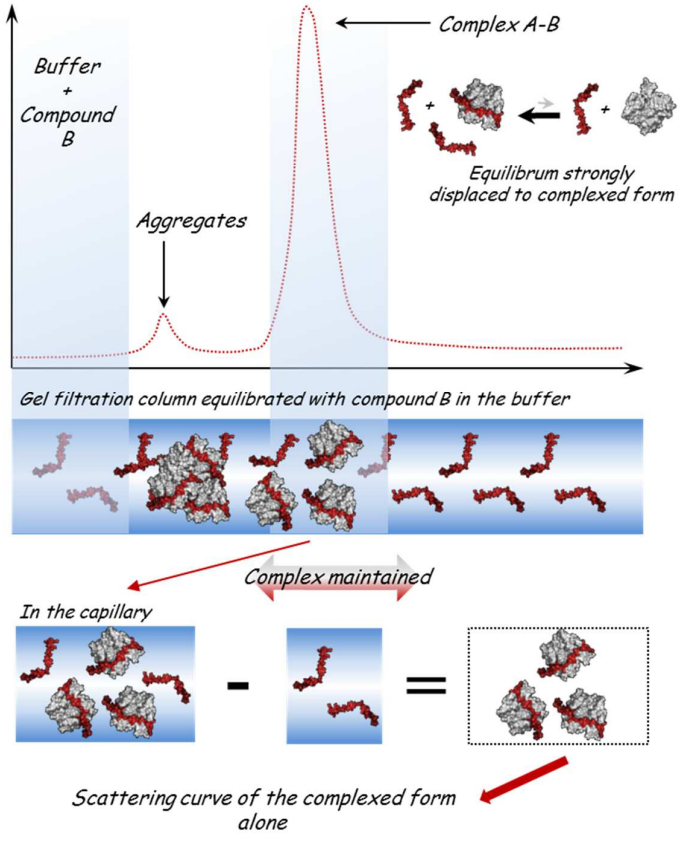
Perfect substraction of the solvent contribution

How to deal with the complex with low affinity

HPLC with no compound in the buffer



HPLC with compound in the buffer



Conclusion

To perform SAXS experiment don't forget :

- analysis and modeling require a monodisperse and ideal solution (~50 to 100µl up to 2mg/ml)
- structure of the protein or homolog , the subunits or different parts of the complex must be known
- missing parts (internal loop or N or C-ter part) represent less than 10% of the mass

- because of the ambiguity of the solution, don't forget to valid your model with other technics that can constrain the field of possibilities (EM,NMR, mutagenesis…)

IN

SAXS

OUT