LABORATOIRE DE GÉNIE CHIMIQUE
TOULOUSE • UMR 5503

Principle of Small Angles X-rays Scattering or SAXS, physical parameters (Rg, Volume, P(r))

P. Roblin

[a] *Université Fédérale Toulouse Midi-Pyrénées; INPT, UPS Laboratoire de Génie Chimique;*
118 Route de Narbonne, F-31062 Toulouse France
[b] *CNRS, UMR 5503, F-31062 Toulouse, France*

SAXS experiments in solution : global strategy to process the data

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 shape

Structure modeling

No structural information
- low resolution shape (ab initio calculation)
- fitting with numerical model (SASVIEW)

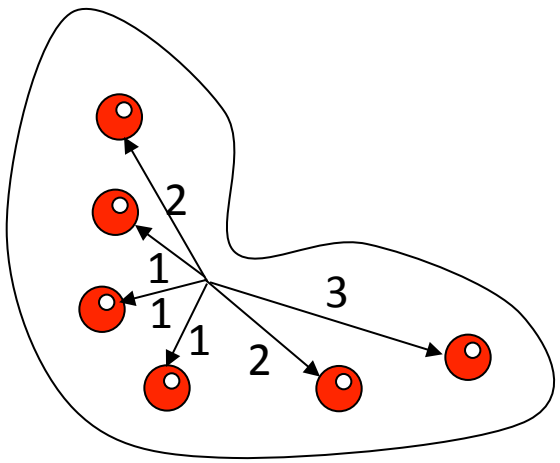
3D Structures available
- theorical curves calculation and data comparison

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

What do we mean by "size" ?

Radius of gyration Rg

The radius of gyration is the root-mean-square, mass-weighted average distance the scatterers from the center of mass of the object



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

$R_g = \sqrt{3.333} = 1.82$ u.a.

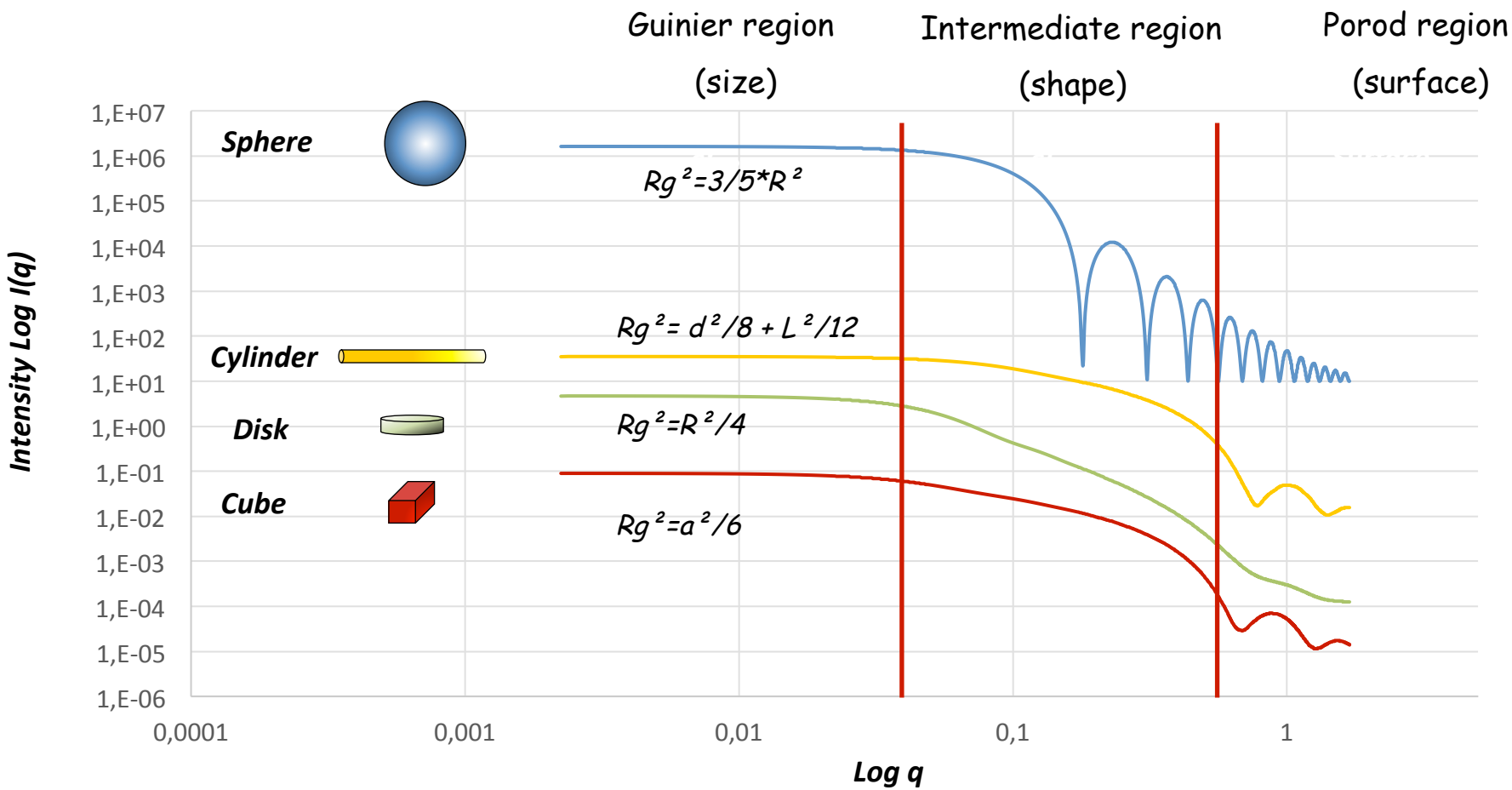
$$Rg^2 = \int V^{\blacksquare} r^2 (\rho_{object}(\vec{r}) - \rho_{solvent}) d\vec{r} \; / \int V^{\blacksquare} (\rho_{object}(\vec{r}) - \rho_{solvent}) d\vec{r}$$

r Position vector from center of mass

$\rho_{object}(\vec{r}) - \rho_{solvent}$ Variation contrast between object and solvent

The radius of gyration can give an idea of the compactness of the object and depends on the volume AND the shape of the particule

Scattering curves and Rg calculation from simple geometric forms



At the low q, the curves coming from the different shapes follow the same variation giving a plateau (Guinier region), whereas a medium q the decrease of the intensity depends of the shape of the particule

At the high q, the intensities of the four models decrease with the same slope because we define previously the objects with a define surface without density fluctuation.

Asymptotic behaviour at small angles : Guinier law

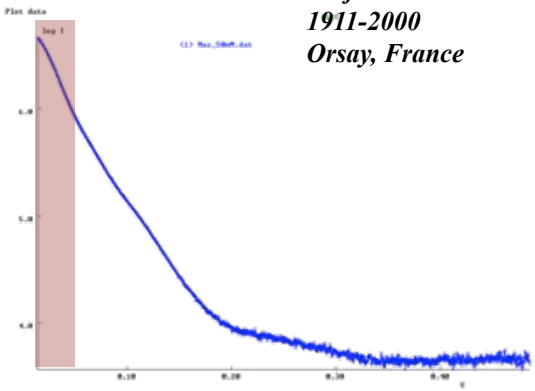
Close to q=0, the scattering intensity of a particle can be described by a Gaussian curve.
The Guinier law is equivalent of a linear variation of Ln(I(q)) vs q^2 (Guinier plot), providing Rg and I(0). The validity domain actually depends on the shape of the particle and is around q < 1.3/Rg for a globular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

Prof. André Guinier
1911-2000
Orsay, France

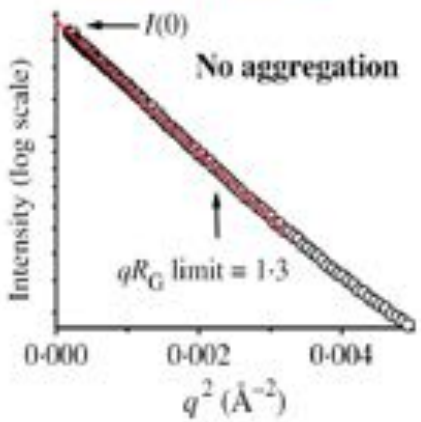
$$I(q) = I(0)\exp\left(\frac{-q^2 Rg^2}{3}\right) \quad \text{or} \quad Ln[I(q)] = Ln[I(0)] - \frac{q^2 Rg^2}{3}$$

Extrapolated intensity at origin

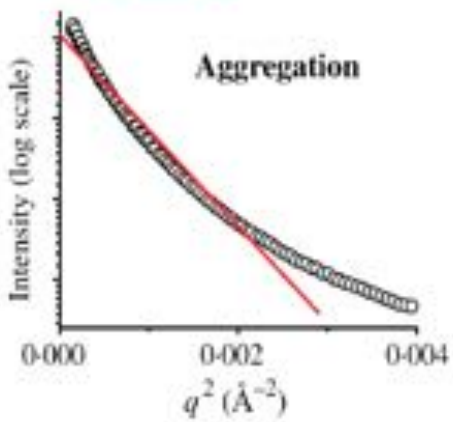
Radius of gyration



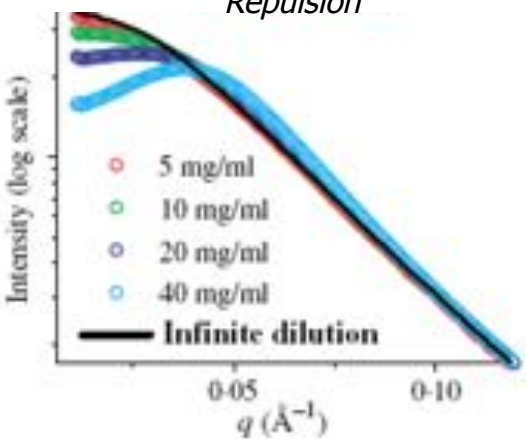
Mono-dispersed



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 :

Mass concentration (g/cm^3)

Absolute intensity (cm^{-1})

Scattering contrast per mass (cm/g)

$$I(0) = MW.[C].\Delta\rho_m^2/N_A$$

Molecular mass (g/mol ou kDa)

Avogadro Number (6,022.10^{23} mol^{-1})

$\Delta\rho_m = [\rho_{M,prot} - (\rho_{solv} \cdot \upsilon)]r_0$ *with*

- $\rho_{M,prot}$ *is the number of electrons per mass of dry protein* $= 3{,}22.10^{23}$ *e.g*$^{-1}$
- ρ_{solv} *is the number of of electrons per volume of aqueous solvent* $= 3{,}34.10^{23}$ *e.cm*$^{-3}$
- υ *is the partial specific volume of of the protein*
- r_0 *is the classical radius of electron* $= 2{,}8179.10^{-13}$ *cm*

Typically for protein : M (kDa) = 1500 * I(0) (cm^{-1}) / C (mg/ml) with $\Delta\rho_m \approx 2 \,.\, 10^{10}$ cm/g

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 (with water measurment where $I(q)_{H20}$ = 0,01632 cm^{-1})

Determination of the mass from a reference

The scattering at zero angle, I(0) is proportional to the molecular weight of the macromolecule, and the concentration and contrast of the macromolecule in solution. If a reference sample of known molecular weight and concentration is measured, it can be used to calibrate the molecular weight of any other scattering profile with known concentration :

Absolute intensity measured (cm^{-1})

Molecular mass of standart (g/mol ou kDa)

Molecular mass (g/mol ou kDa)

$$MW_m = (I(0)_m / c_m)(MW_{sdt} / I(0)_{sdt} / c_{std})$$

Absolute intensity of standart (cm^{-1})

Mass concentration of sample (g/L)

Mass concentration of sample (g/L)

Can be highly accurate for similar standards and samples under the same conditions.

The reference standard should have the same scattering contrast as the sample (i.e.. is in a similar buffer).

The standard and sample should be similar shapes (i.e. the same partial specific volume).

Asymptotic behaviour at larges angles : Porod law

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

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

$$\varphi = N\cdot V_{obj} / V$$

Number of objects

Object volume

Solution volume

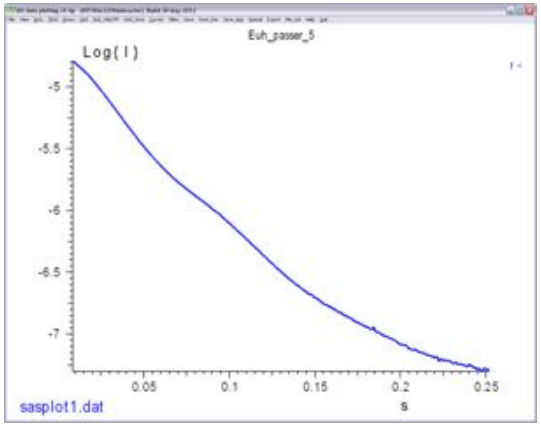
Volumic concentration

Porod invariant

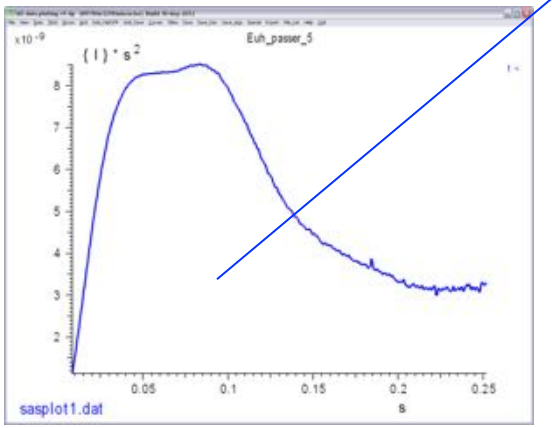
Does not depend on shape, only on contrast

The Porod Invariant is the integral of this curve

Representation : Log I(Q) vs Q



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



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

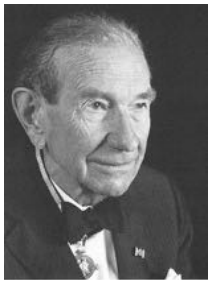
$$V_{obj} = \frac{2\pi^2 \cdot I(0)}{\mathbf{Q}}$$

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

Asymptotic behaviour at larges angles : Porod law

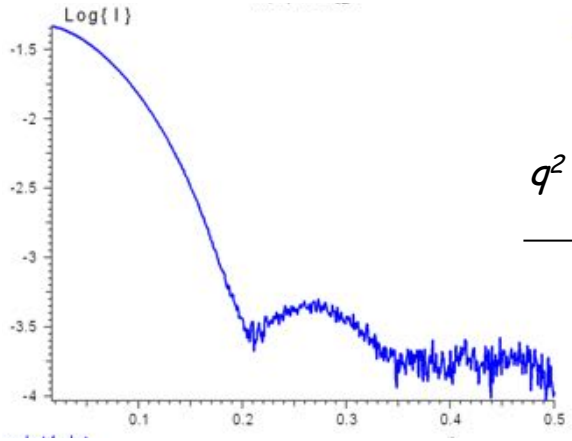
SAXS provides a sensitive means to ***evaluate the degree of compactness*** of a polymer:

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

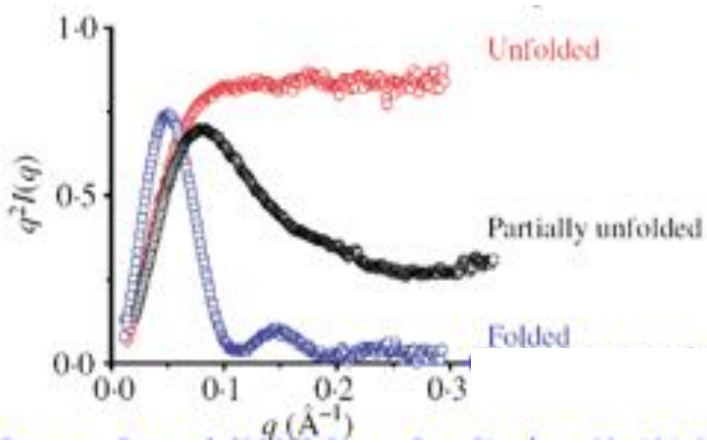


Prof. Otto Kratky
1902-1995
Graz, Austria

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



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



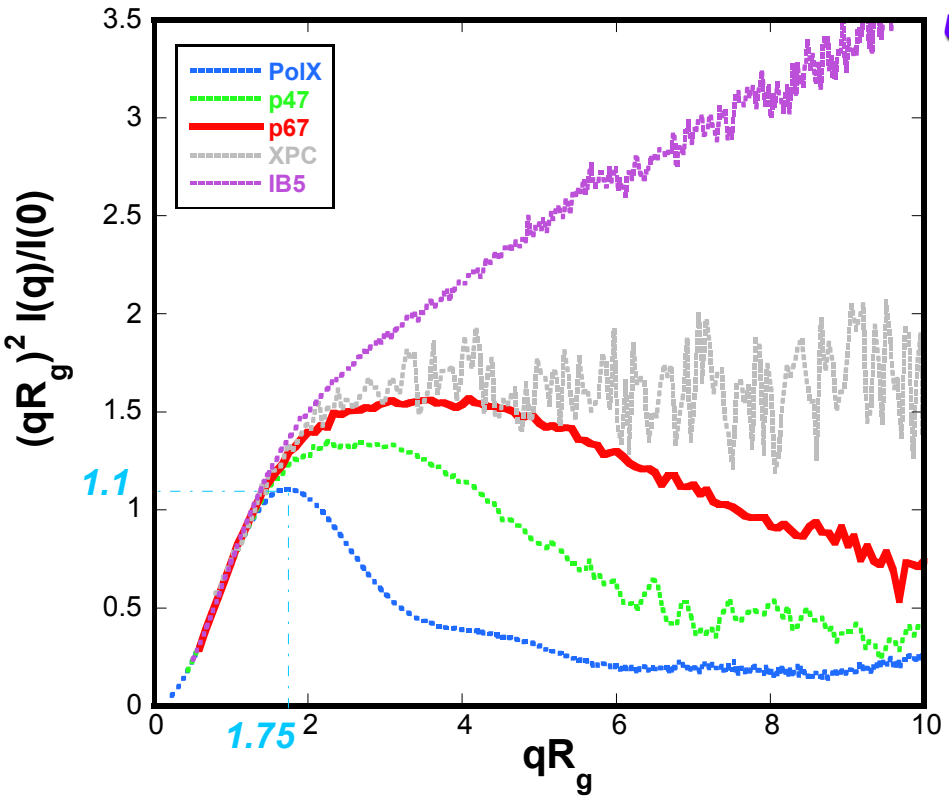
Putnam, D., et al. (2007) Quart. Rev. Biophys. 40, 191-285.

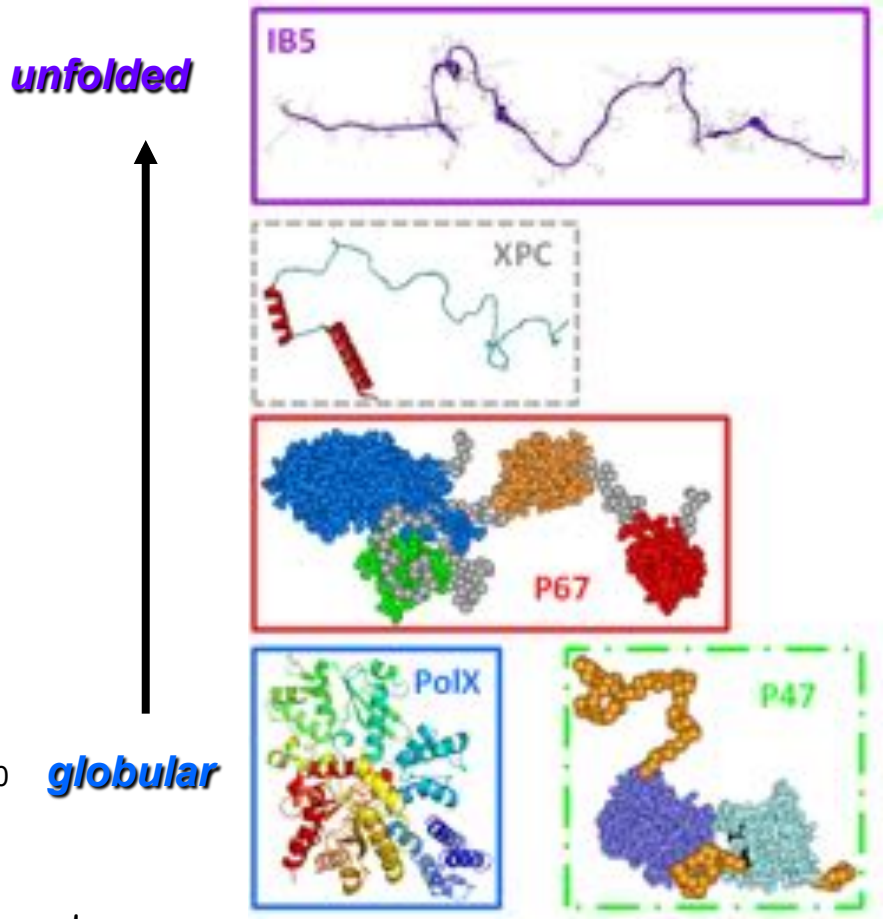
Folded particle : ***bell-shaped curve*** (asymptotic behaviour in $I(q) \sim q^{-4}$)

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





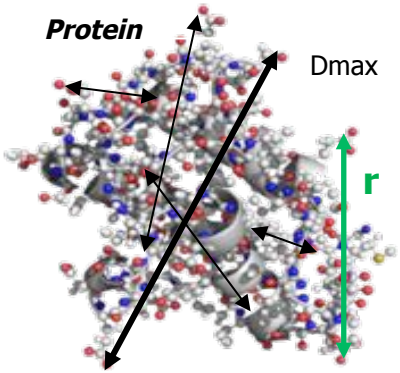
The bell shape vanishes as folded domains disappear and flexibility increases.

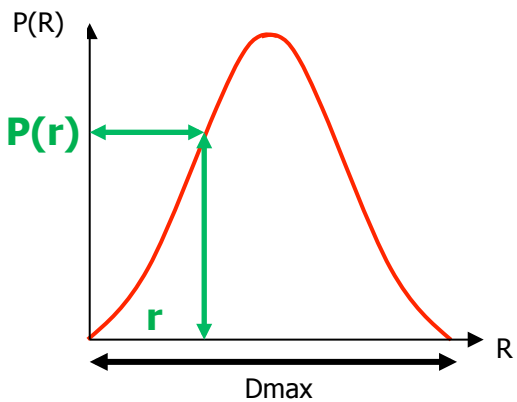
The curve increases at large q as the structure extends.

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

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

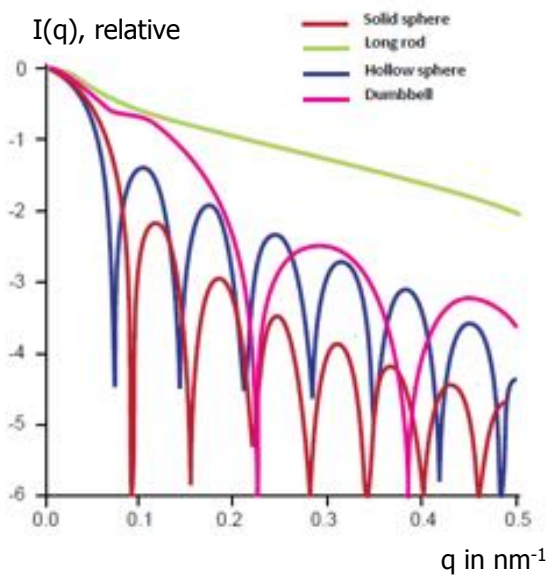
The scattered intensity I(q) can be written with the distribution function P(r). P(r) function is calculated with indirect Fourier transform applied to the scattered intensity I(q). The both curves contain the same information.

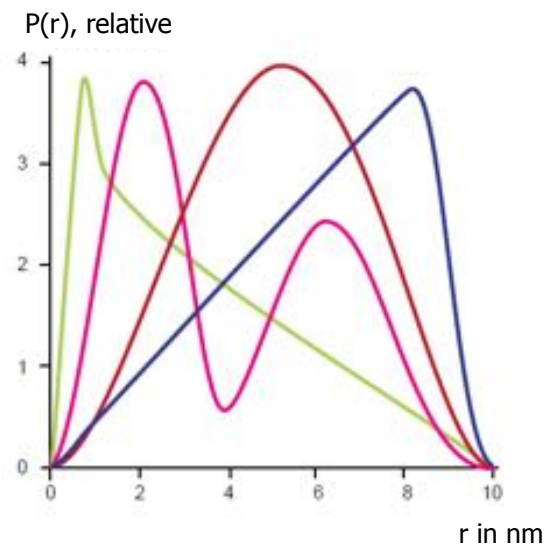
$$I(q) = 4\pi\, r_e^{\,2} \varphi \int_0^{D\max} P(r) \frac{\sin(qr)}{qr} dr$$

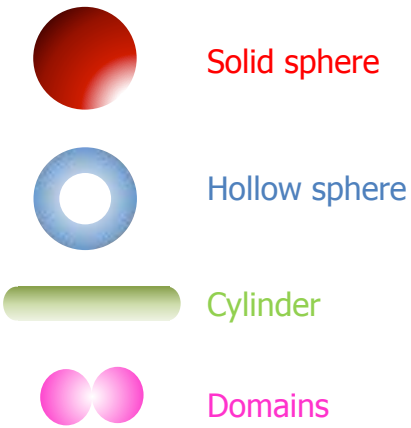
$$P(r) = \frac{r^2}{2\pi^2 \varphi\, r_e^{\,2}} \int_0^{\infty} q^2 I(q) \frac{\sin(qr)}{qr} dq$$

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

LABORATOIRE DE GÉNIE CHIMIQUE
TOULOUSE • UMR 5503

How to use SAXS data to retrieve structural information

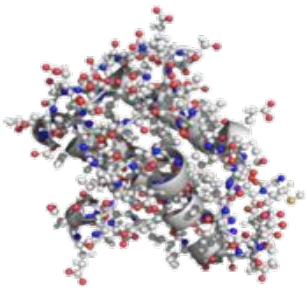
P. Roblin

[a] *Université Fédérale Toulouse Midi-Pyrénées; INPT, UPS Laboratoire de Génie Chimique;*
118 Route de Narbonne, F-31062 Toulouse France
[b] *CNRS, UMR 5503, F-31062 Toulouse, France*

SAXS for 3D structure reconstitution

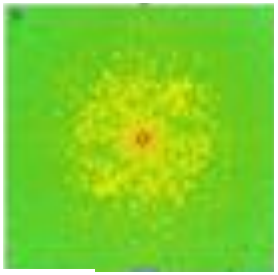
The 1D SAXS profile is the Fourier transform of the 3D structure. Contrary to the direct scattering calculation, the inverse problem cannot be solved analytically, i.e., no "inverse Debye" formula can be constructed to yield 3D position coordinates from scattering data.

Real space 3D Molecule



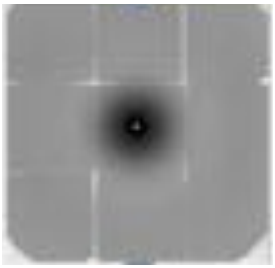
Fixed orientation → **Lose phase**

Reciprocal space 2D anisotropic image



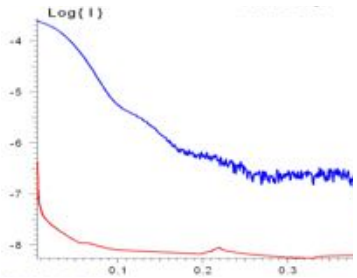
Averaged orientation → **Lose orientation**

Reciprocal space 2D isotropic image



Radial averaging →

1D profile reciprocal space



$$I_m(q) = \sum_{j=1}^{N_A}\sum_{j=1}^{N_A} f_i(q) f_j(q) \frac{\sin(q d_{ij})}{q d_{ij}}.$$

Debye Formula

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

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 ($r_0 \ll R$)

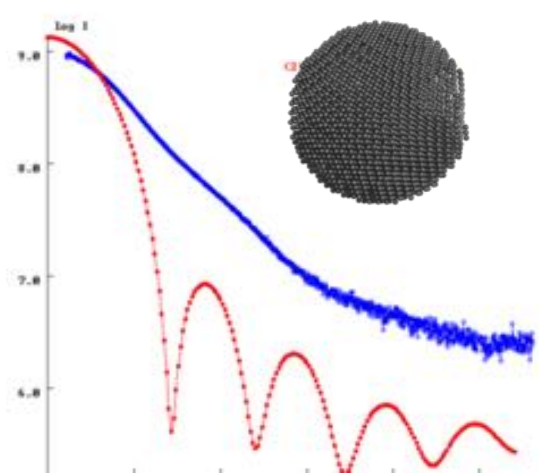
The number of the "dummy atom" $N \approx (R/r_0)^3$

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)

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

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

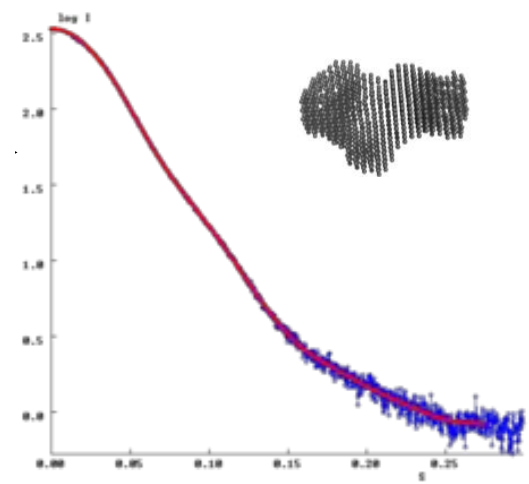
A)



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

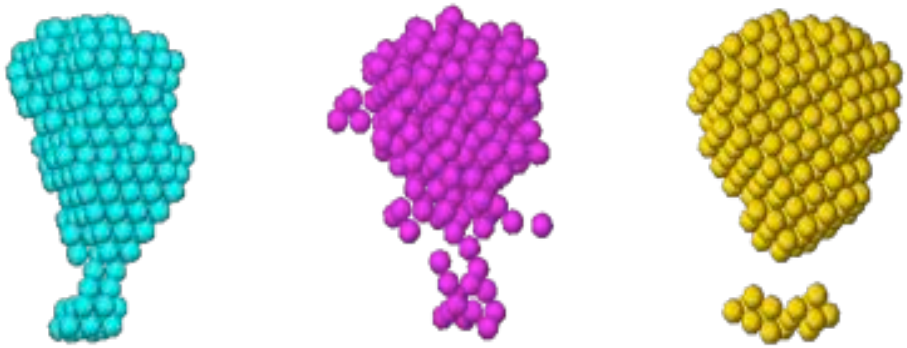
Obtaining 3D shapes from SAXS data is a defined problem that could be solved by introducing additional information to reduce ambiguity of interpretation

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

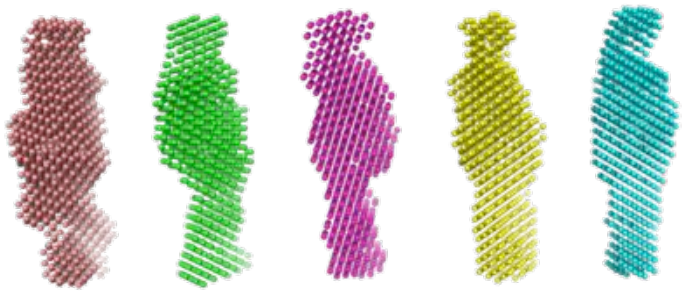
Disconnected model

3D shape reconstructions from SAXS data with DAMMIN

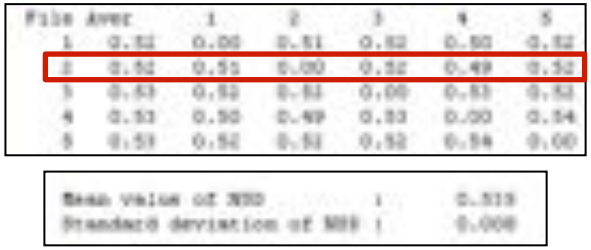
DAMMIN : necessit to perform a serie of run (10-50) to compare the different shape obtained with the same data.

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

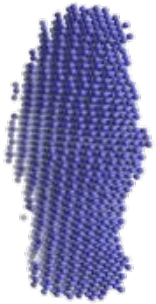


File Aver	1	2	3	4	5
1	0.52	0.00	0.51	0.52	0.50	0.52
2	0.52	0.51	0.00	0.52	0.49	0.52
3	0.51	0.52	0.51	0.00	0.51	0.52
4	0.53	0.50	0.49	0.53	0.00	0.54
5	0.51	0.52	0.51	0.52	0.54	0.00

Mean value of NSD	:	0.513
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

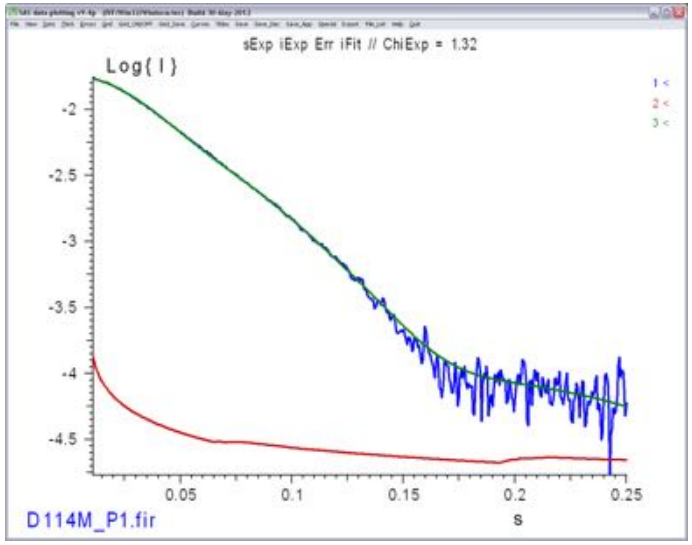
Be careful with <u>damfilt.pbd</u> because $I_{damfilt}(q) \neq I_{exp}(q)$

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 $C\alpha$ 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.

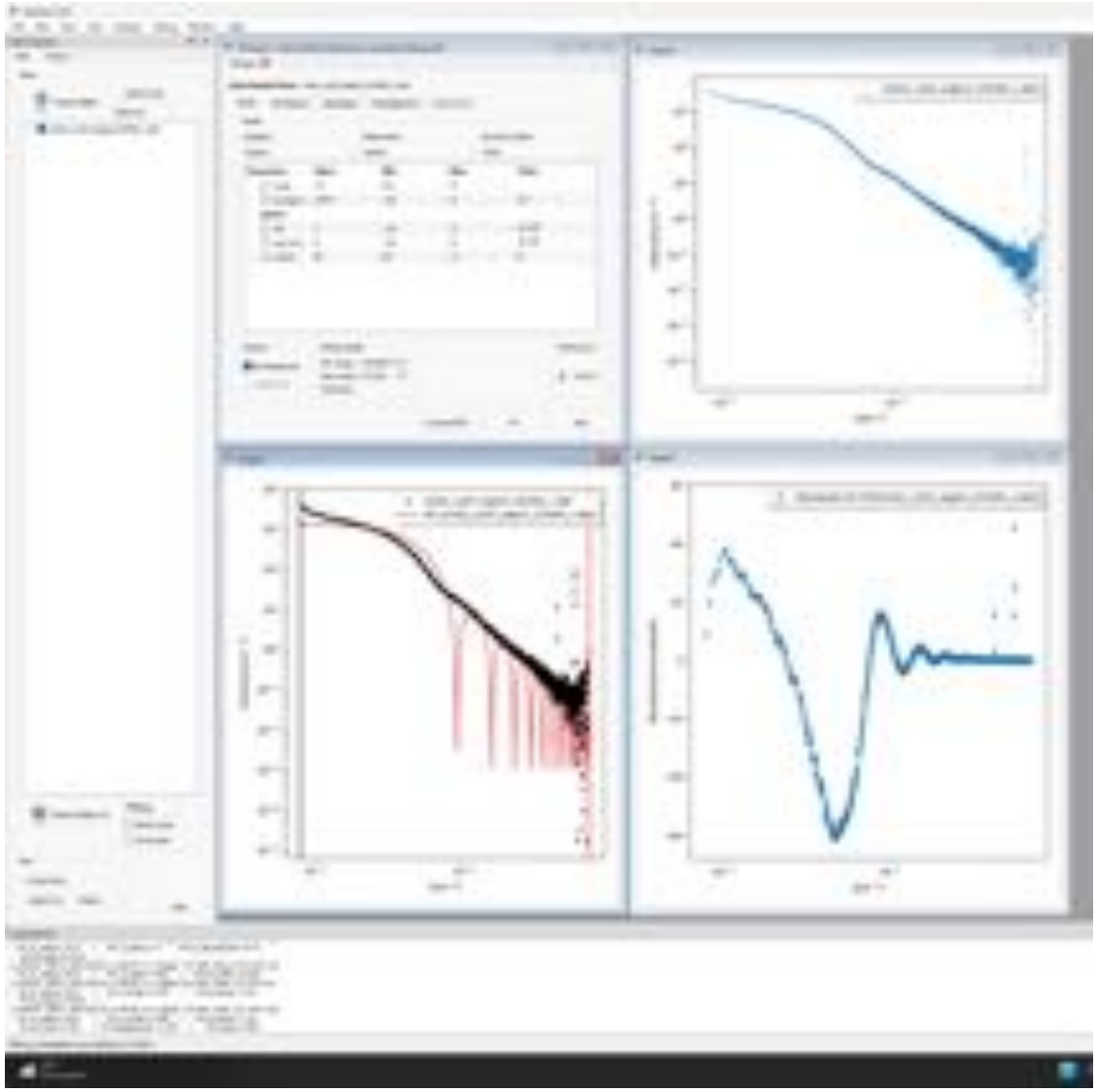
Example of program to perform ab initio molecular modeling

DAMMIN/DAMMIF/GASBOR : ATSAS online (https://www.embl-hamburg.de/biosaxs/atsas-online/)

DENFER: Ligne SWING –SOLEIL (https://www.synchrotron-soleil.fr/fr/lignes-de-lumiere/swing)

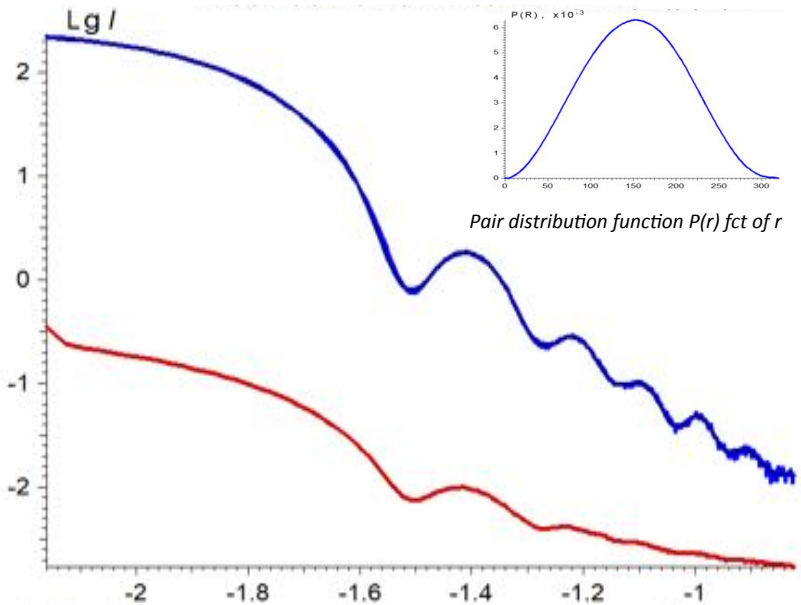
Fitting from numerical model – example with SASVIEW

SASVIEW https://www.sasview.org/



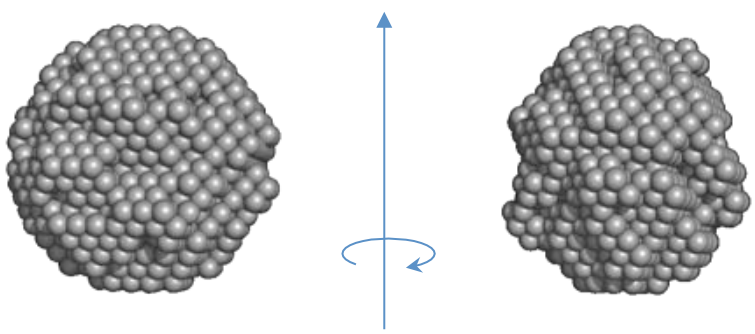
Ab initio molecular modeling and model calculation on same data

Ab initio calculations



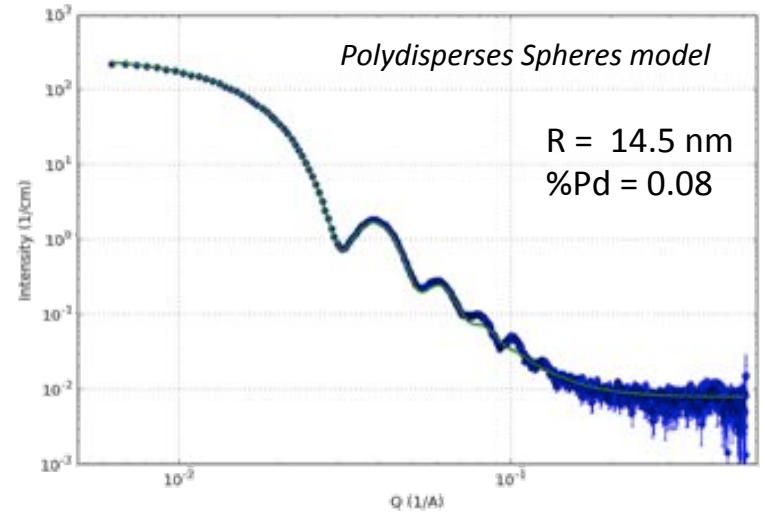
Pair distribution function P(r) fct of r

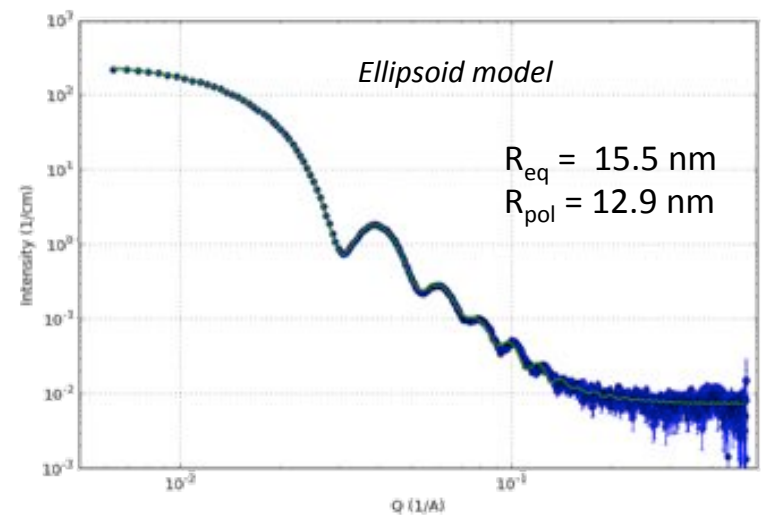
Beads model provided by Dammin



90°/z

Sasview calculations





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 :

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

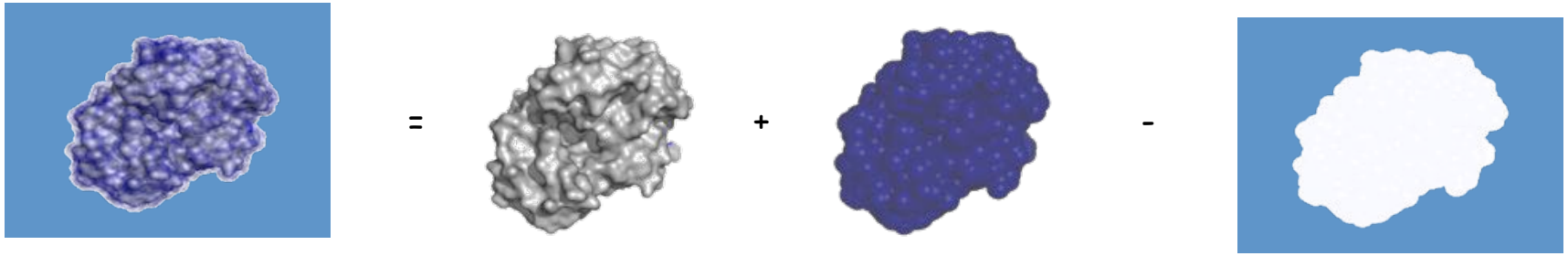
$F_i(q)$, $F_f(q)$, Form factor of atom i and atom j

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

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

$A_s(q)$ = scattering from excluded volume

<u>In CRYSOL program</u>, 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 $\delta\rho$

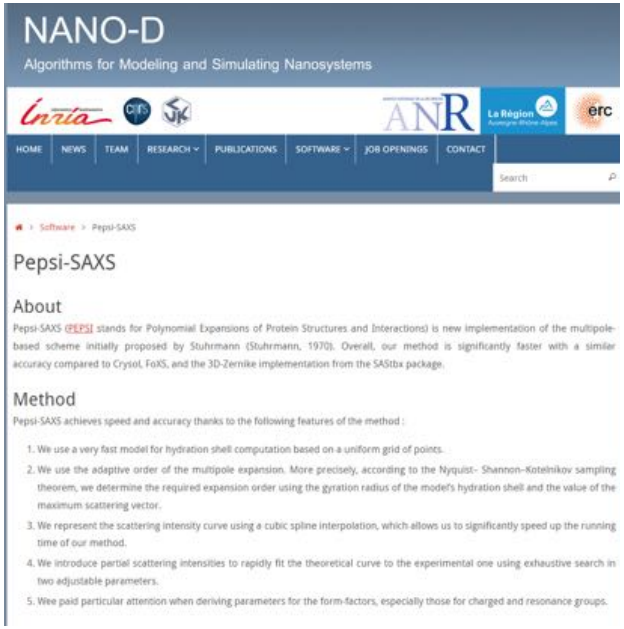
$$\chi^2 = \frac{1}{N-1} \sum_{i=1}^{N} \left[\frac{I_{\exp}(q_i) - scale * I_{calc}(q_i)}{\sigma_{\exp}(q_i)} \right]$$

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

Example of program to calculate scattering curve from 3D structure

CRYSOL: ATSAS package (https://www.embl-hamburg.de/biosaxs/software.html)

Pepsi-SAXS: NanoD -team (https://team.inria.fr/nano-d/software/pepsi-saxs/)



FoXS: https://modbase.compbio.ucsf.edu/foxs/

Rigid body modeling against SAXS data

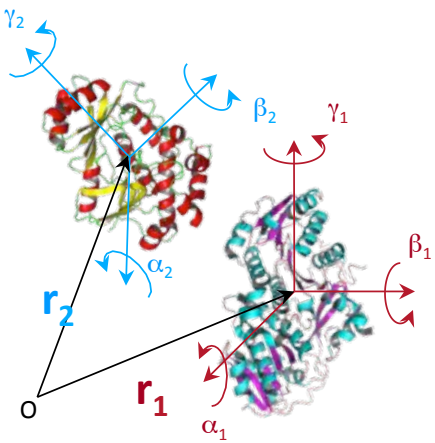
SASREF : when atomic structures of domains are known, but no their mutual organization

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.

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

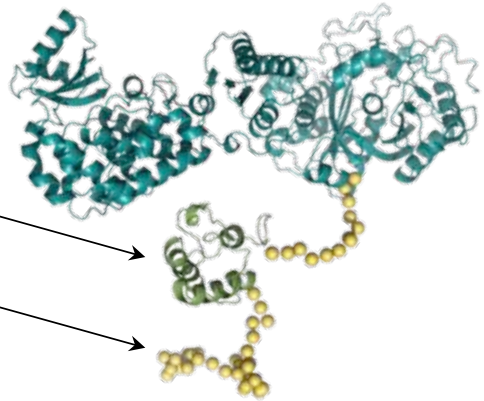
Petoukhov & Svergun (2005). Biophys. J., 89, 1237-1250.

Rigid body modeling 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

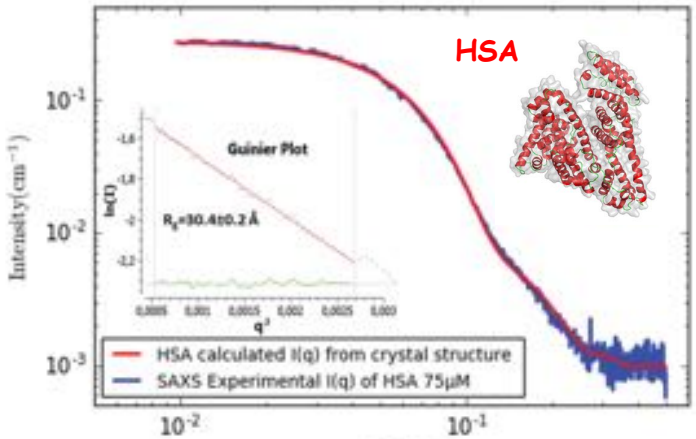
The algorithm of minimization is the same used with SASREF with a penalty function including the steric clashes Pcross, the dihedral angle Pang and Pdih, and the compactness of the loop Pext. The possibility to give information about contacting residues from other experiences is also added.

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

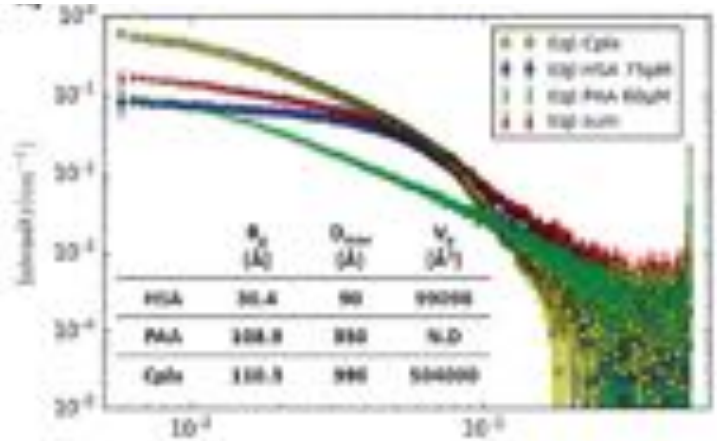
Petoukhov & Svergun (2005). Biophys. J., 89, 1237-1250.

Example of Rigid body modeling with non protein substrate

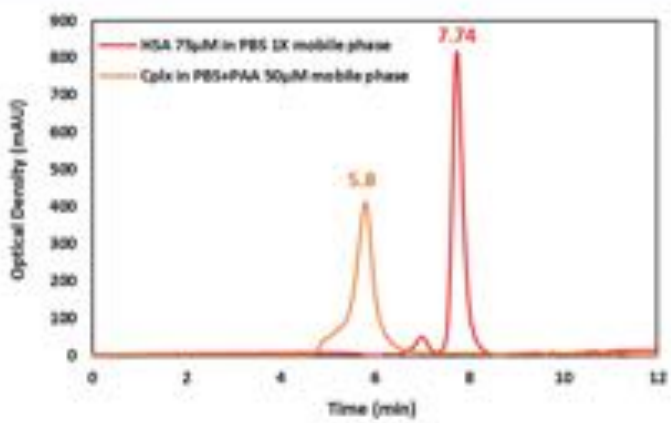
International ANR MOSAIC3D, R&D Center for Membrane Technology of Taïwan/LGC , Pierre Aimar, Patrice Bacchin, Christel Causserand, Pierre Roblin & PhD Charaf Merzougui



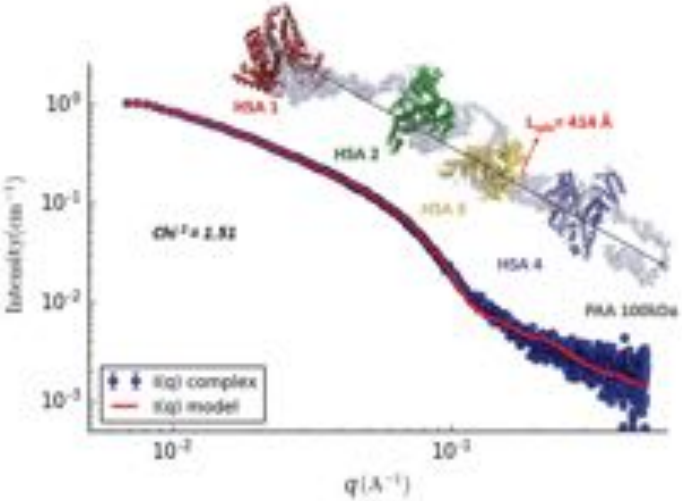
Experimental curve comparison of the HSA obtained on the SAXS laboratory and curve calculated from the structure of any PDB atom of the HSA



SAXS curve comparison obtained on the HSA and PAA mixture with the sum of SAXS curves of HSA and PAA



Exclusion chromatography profile performed with HSA in PBS buffer and HSA with PAA in PBS buffer + 5g / L PAA



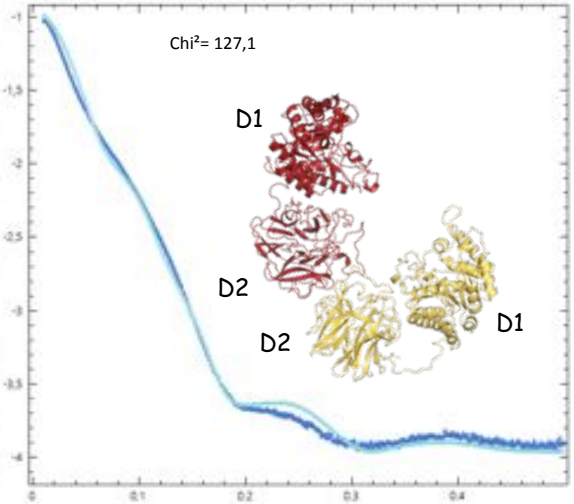
Molecular modeling in all atom of PAA-5HSA complex against SAXS data

Merzoughy & all : Pearl-necklace assembly of human serum albumin with the poly (acrylic acid) polyelectrolyte investigated using small angle X-ray scattering (SAXS) **Soft matter** *2020*

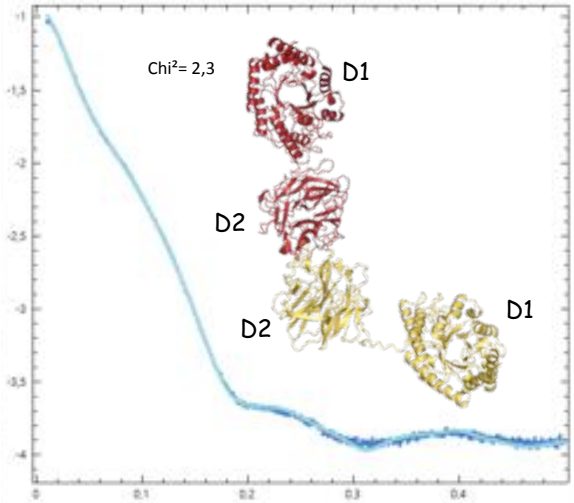
Rigid body modeling with missing loop against SAXS data

DADIMODO : Refining Atomic Models of Multi-Domain Protein Complexes from SAXS data (https://dadimodo.synchrotron-soleil.fr/submission)

Body1 : A 31-250 Body2 : A 275-605 , B 275-605 Body3 : B 31-250



Structural model created by dynamic molecular



Structural model calculated by Dadimodo

LABORATOIRE DE GÉNIE CHIMIQUE
TOULOUSE • UMR 5503

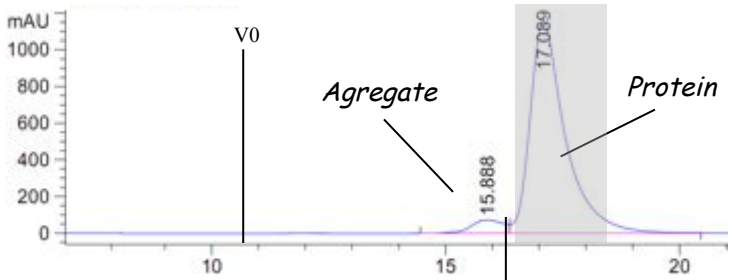
Principle of Small Angles X-rays Scattering or SAXS

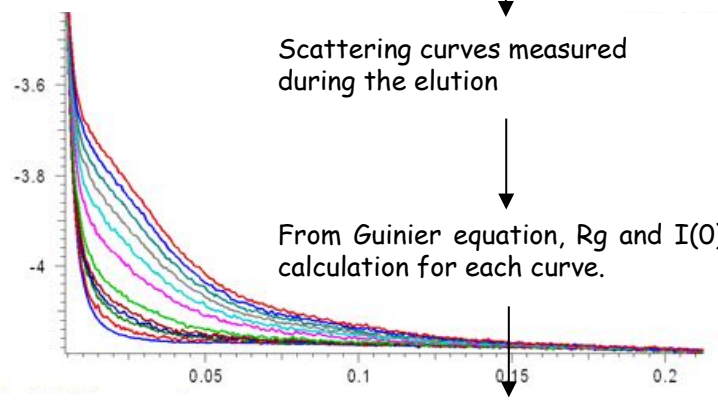
P. Roblin

[a] *Université Fédérale Toulouse Midi-Pyrénées; INPT, UPS Laboratoire de Génie Chimique;*
118 Route de Narbonne, F-31062 Toulouse France
[b] *CNRS, UMR 5503, F-31062 Toulouse, France*

SAXS characterization of macromolecule with coupled HPLC

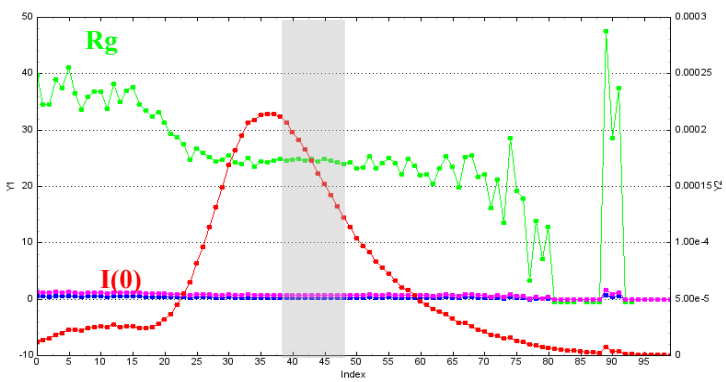
Typical experience performed with HPLC





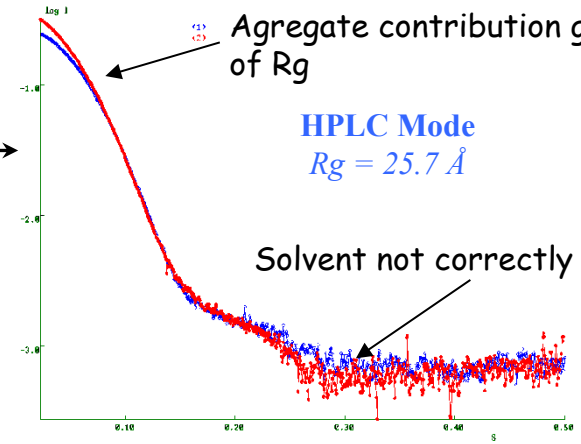
Scattering curves measured during the elution

From Guinier equation, Rg and I(0) calculation for each curve.



Selection of identical curves before data averaging

Comparison between curve obtained with direct injection and after HPLC method



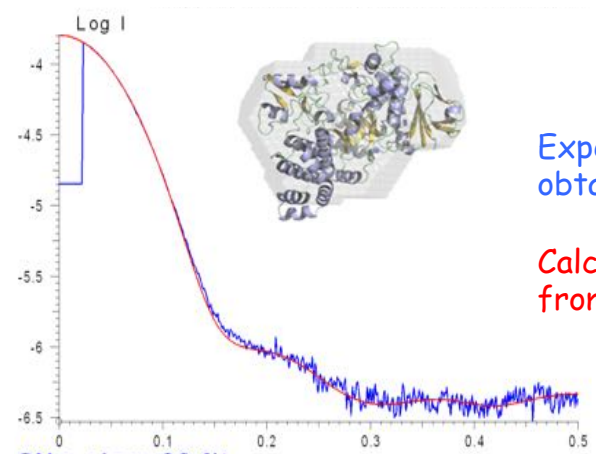
Agregate contribution giving a too high value of Rg

HPLC Mode
Rg = 25.7 Å

Direct Mode
Rg = 29.1 Å

Solvent not correctly substracted

Q en A^{-1}



Experimental curve obtained with HPLC

Calculated curve obtained from atomic structure

Q en A^{-1}

Column separation ensure monodispersity of the measured solution

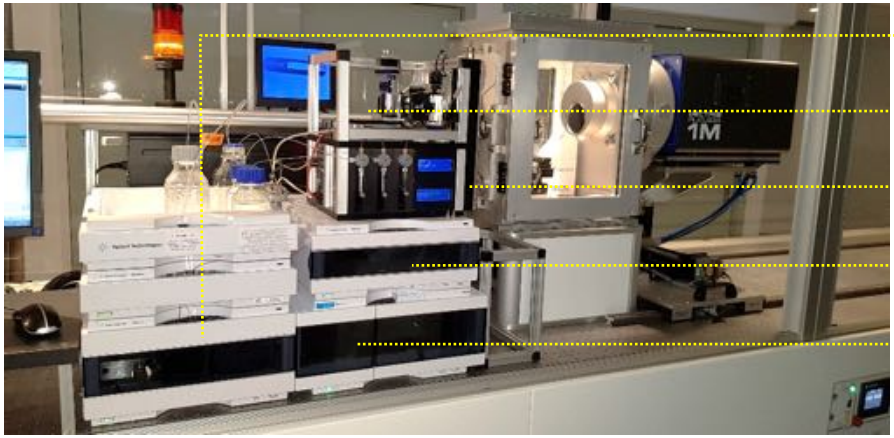
Perfect substraction of the solvent contribution

Homemade samples environnement developement

<u>Major problems with protein :</u>

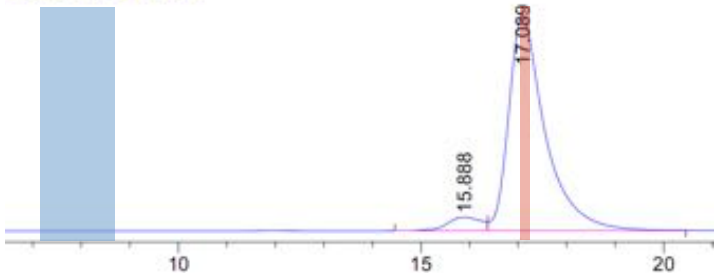
- Low contrast with the solvent and dilute system and strong propension to denaturation and agregation

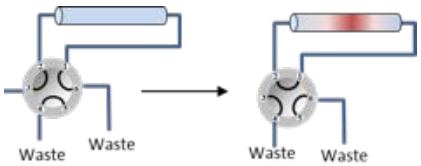
Solution consist to add a supplemntary step of purification with a size exclusion chromatography online



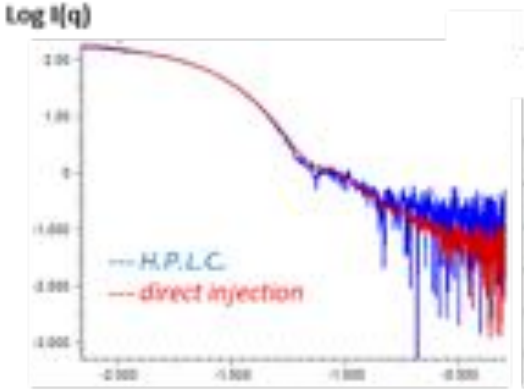
PUMP
↓
Autosampler
↓
S.E.C. column
↓
U.V. detector
↓
Refractometer
↓
Flow cell

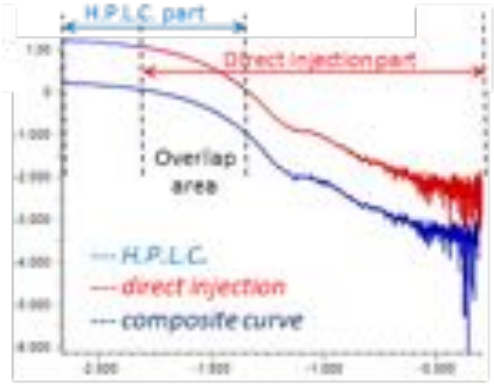
Elution profile of a protein

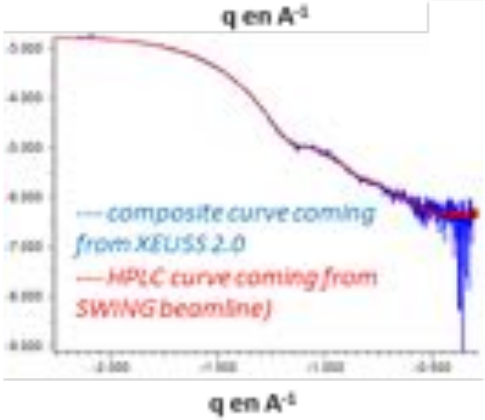




Valve for sample locking in the flow cell







How to deal with the complex with low affinity