A few aspects of BioSAXS

(Solution X-ray Scattering from Biological Macromolecules)

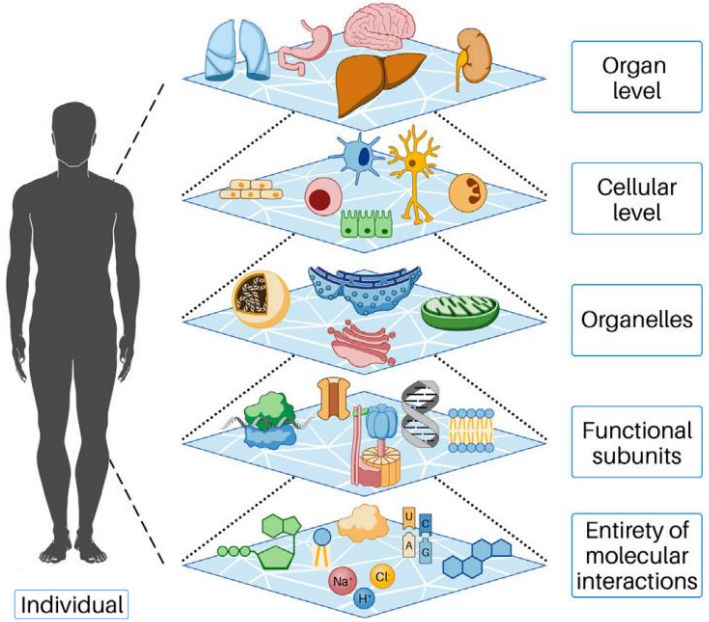
Javier Pérez

Beamline Manager

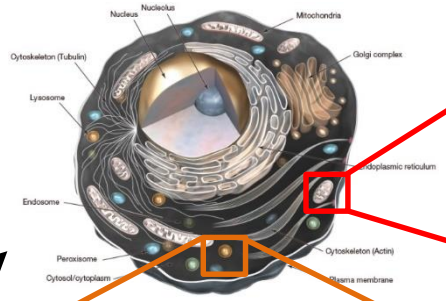
Beamline SWING, Synchrotron SOLEIL, Saint-Aubin, France

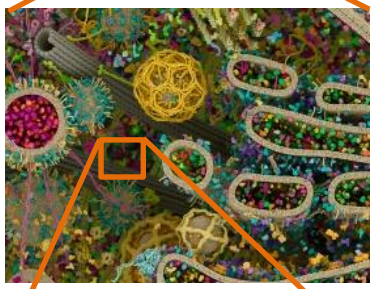
Proteins are the functional workhorses of life

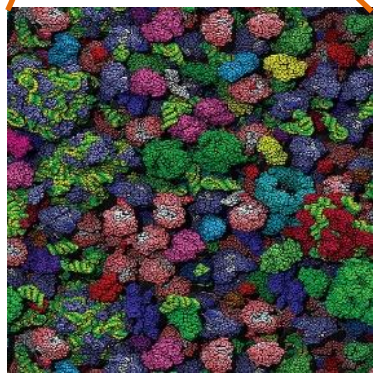
Health & Well-Being



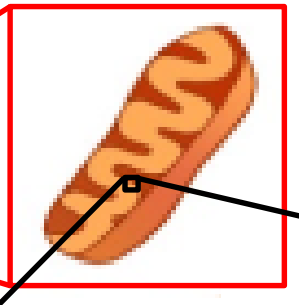
Hanselmann RG and Welter C (2022). *Front. Cell Dev. Biol.*





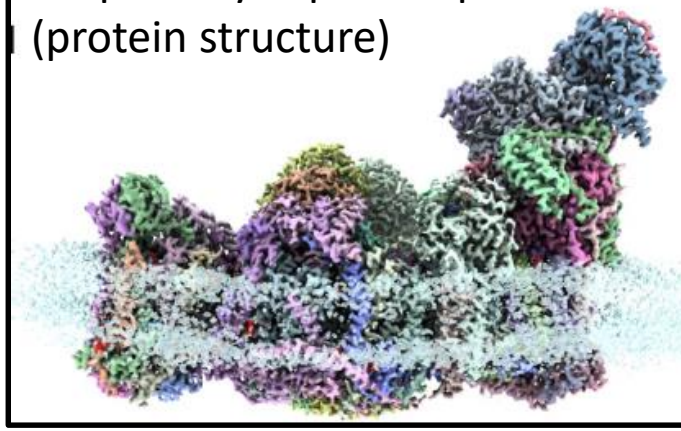


McGuffee SR, Elcock AH, 2010
PLoS Comput Biol 6(3)



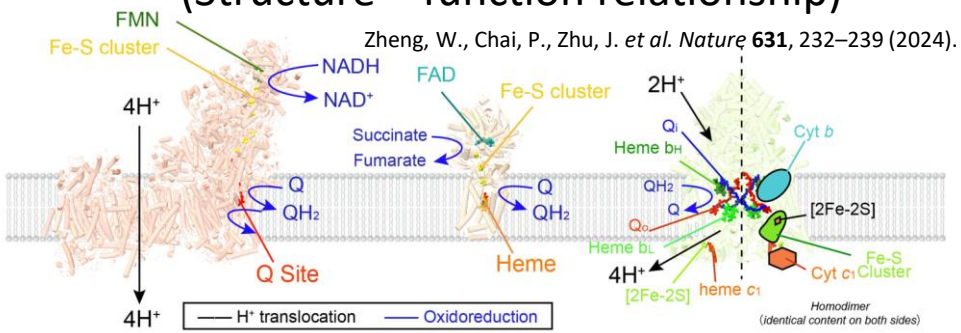
Mitochondria
(Main cellular role : energy production)

Respiratory supercomplex I
(protein structure)



Energy production cycle
(Structure – function relationship)

Zheng, W., Chai, P., Zhu, J. *et al. Nature* **631**, 232–239 (2024).



Life functions result from the collective behavior of many molecular parts interacting and reacting with each other.

Functions and interaction modes of proteins and other biomolecules arise from their three-dimensional structure → aim of Structural Biology.

High resolution techniques : Protein Crystallography, NMR, Cryo-EM

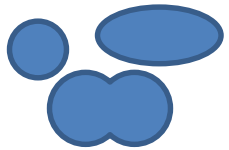
What may BioSAXS yield?

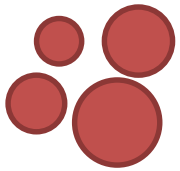
Shape (P(r))

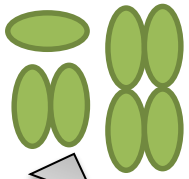
Size (Radius of gyration)

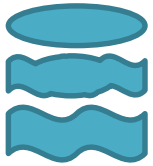
Mass / Oligomerisation
(~10% precision)

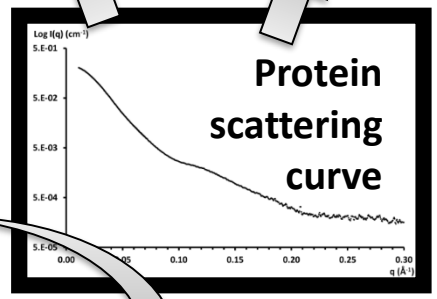
"Global flexibility"



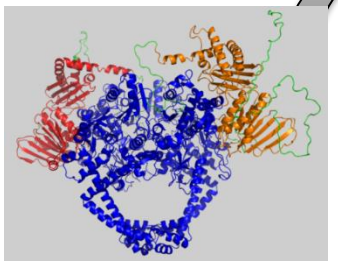




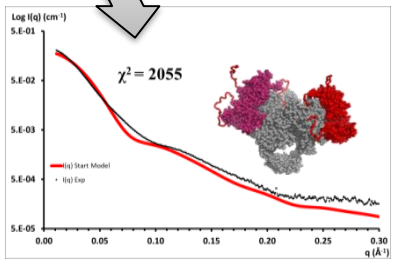




Atomic model from:
- MX
- NMR
- Cryo-EM
- Alphafold

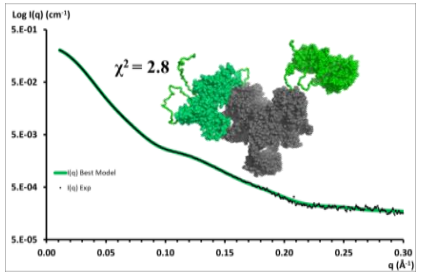


Petrella et al. 2019



Atomic Model comparison & validation

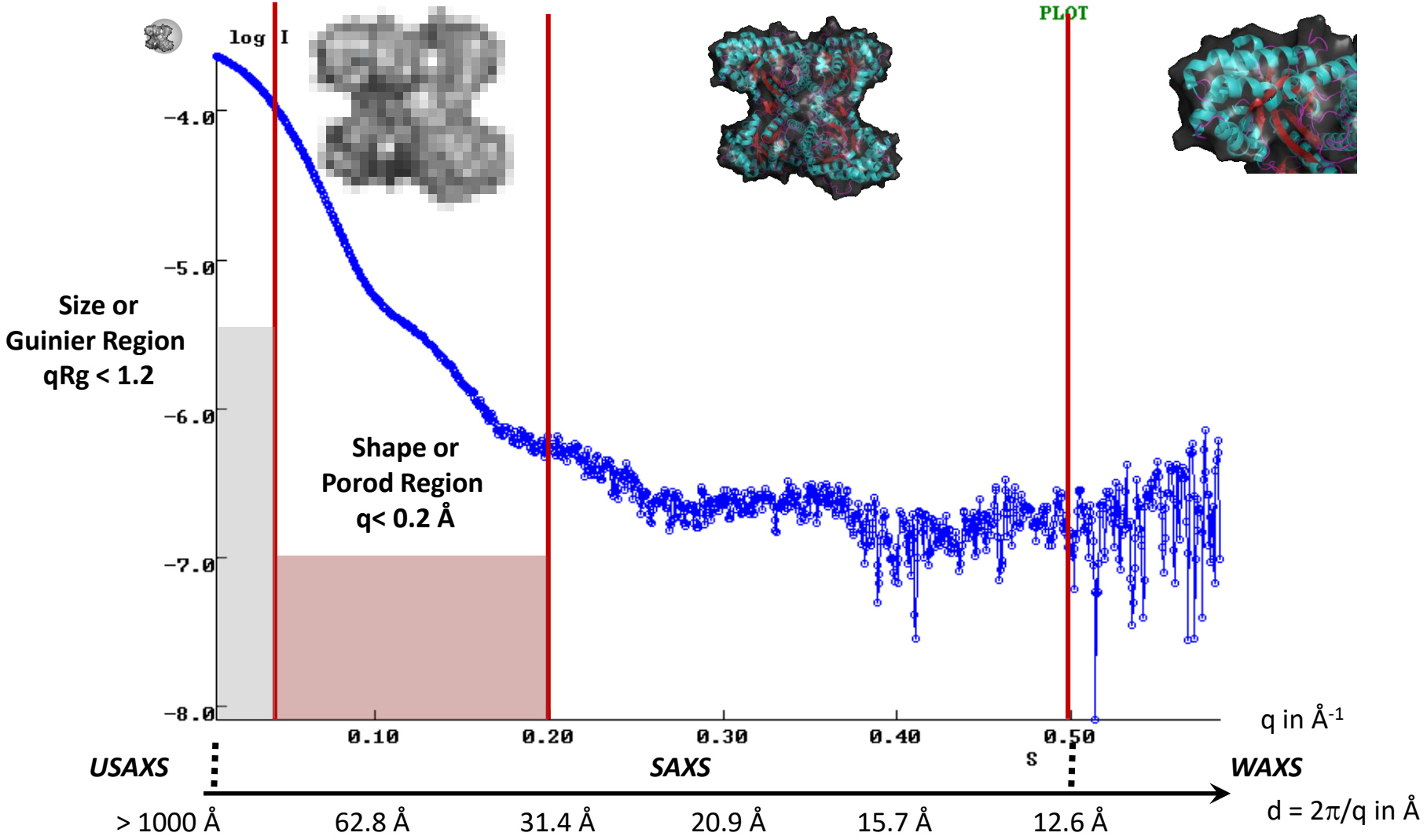
Molecular Modeling



Model "refinement"

A typical BioSAXS curve from a protein

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

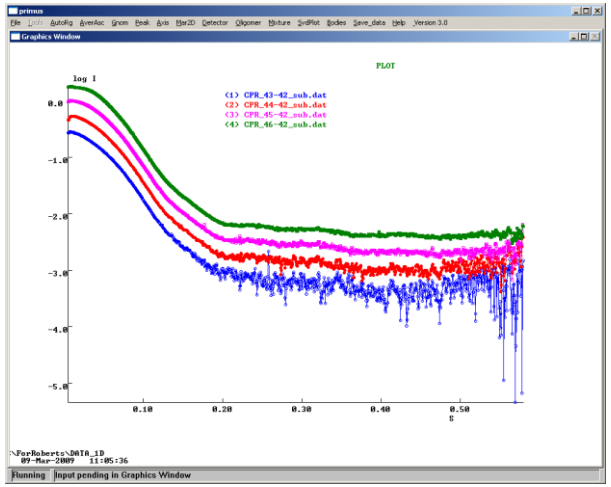


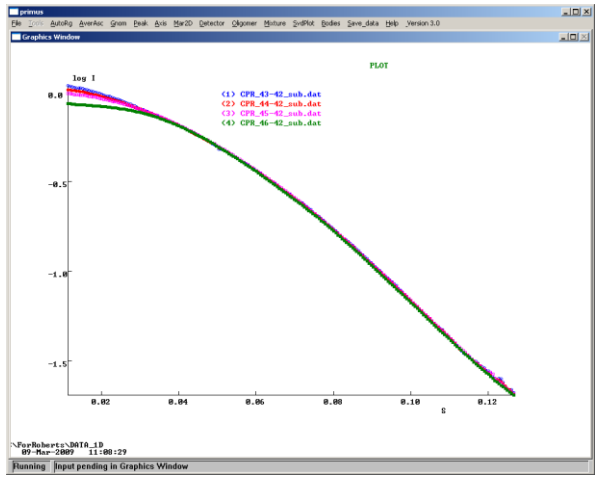
Evaluation of the solution properties

Checking monodispersity and absence of structure factor for the sample under study is crucial for non erroneous data interpretation

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

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





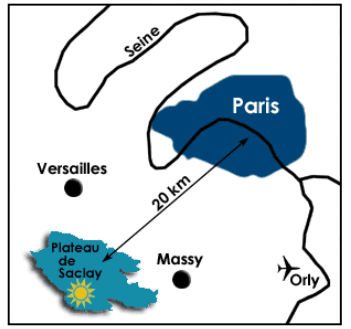
Current activities at beamline SWING

https://www.synchrotron-soleil.fr/en/beamlines/swing

Since 2008

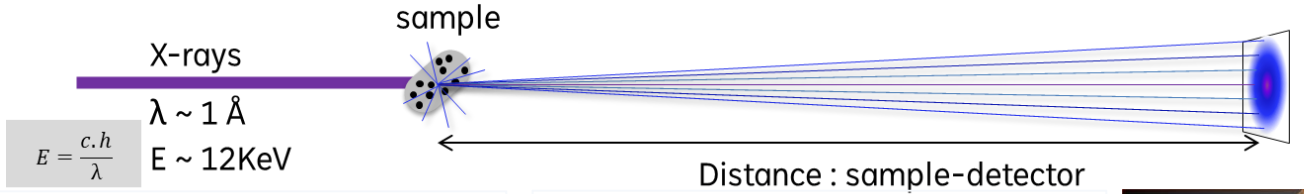
- **S**mall and **W**ide angle X-ray scatter**ING**

From macromolecule to material.



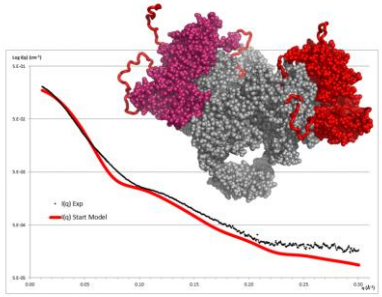
- Hard X-rays Energy (U20 undulator): 5 keV to 16 keV (wavelength: 2.5 Å to 0.75 Å)
- Usual beam size (FWHM): 25-400 (H) x 10-100 (V) µm²
- Structural information from non-crystalline samples (scale: nm to µm)

- BioSAXS
- « Classical » SWAXS
- µSAXS mapping
- PXCT (ptycho-tomography)



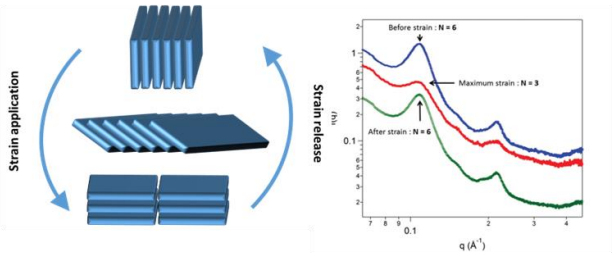
$$E = \frac{c.h}{\lambda}$$

Protein conformations



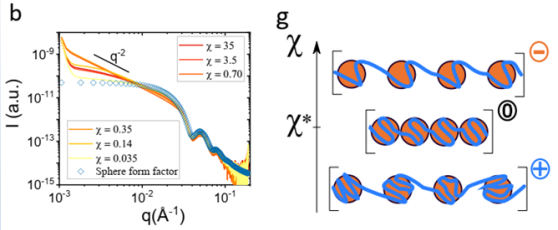
Roudenko O., Thureau A. & Pérez J., March 2018
Petrella, S., et al. (2019). Structure, 27, 579–589.

Deformation under strain

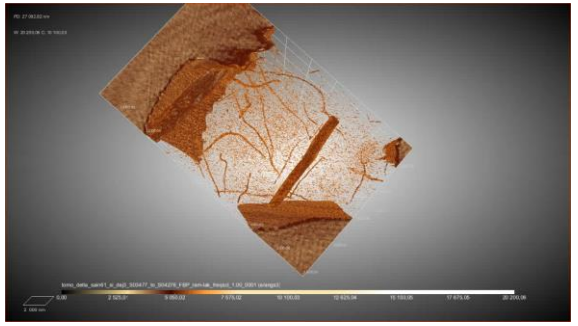


Beaudoin, E., *et al.* (2017), *Nanoscale*, 9(44): 17371-17377

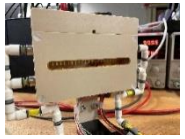
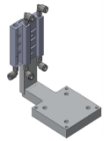
Colloidal growth mechanism

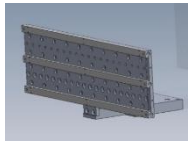
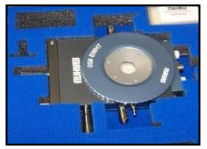


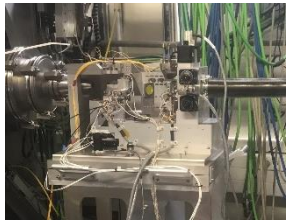
Voisin, F., *et al.* (2023), *Journal of Colloid and Interface Science*, 630: 355-364.



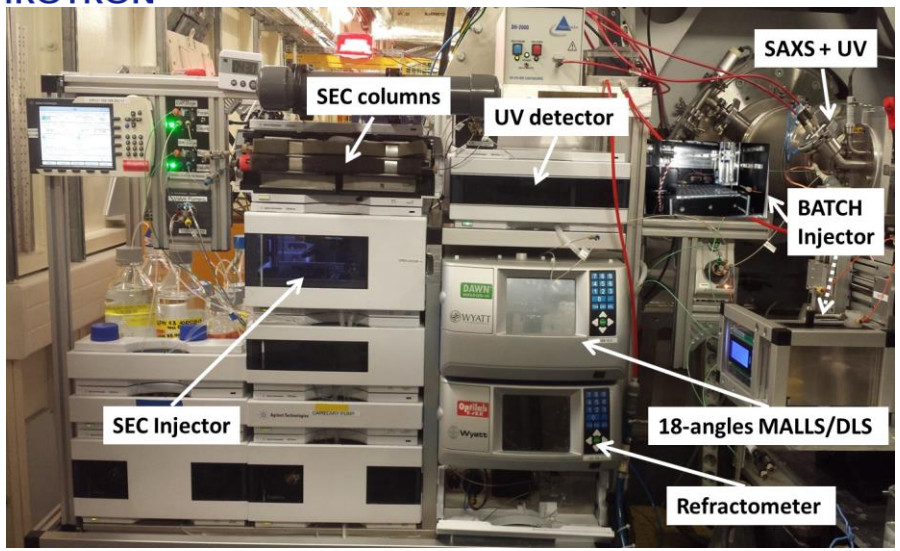
Dentin high resolution density map

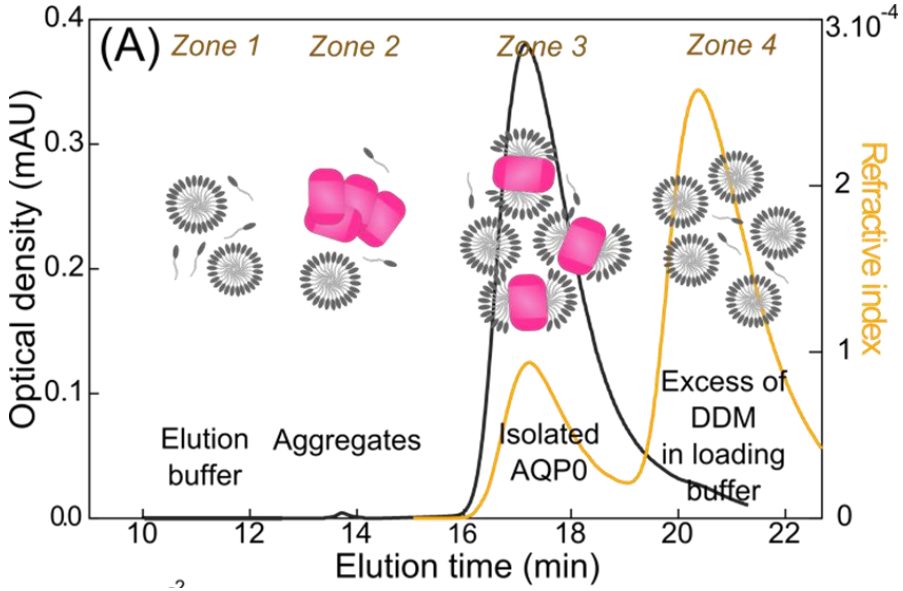



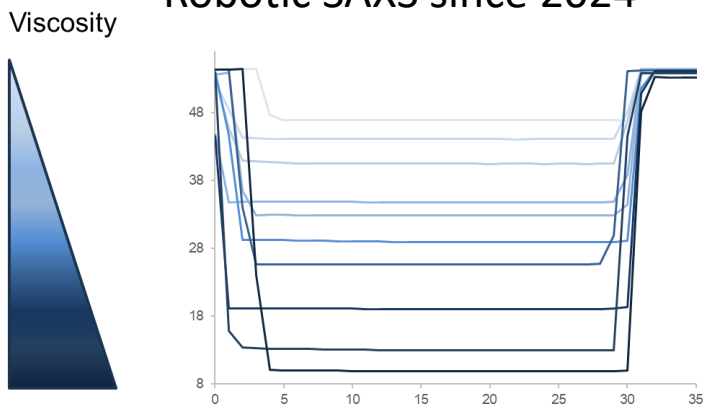
BioSAXS at SWING beamline



Ensure monodispersity with SEC-SAXS since 2008

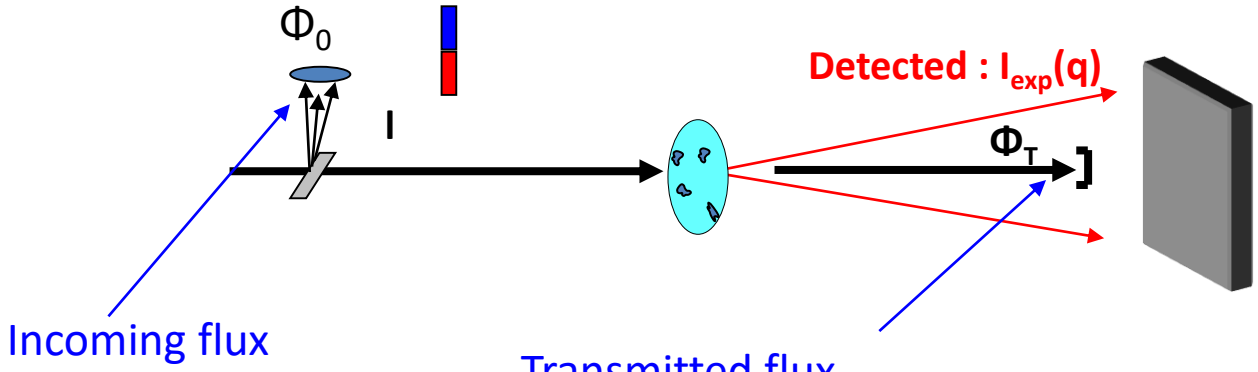


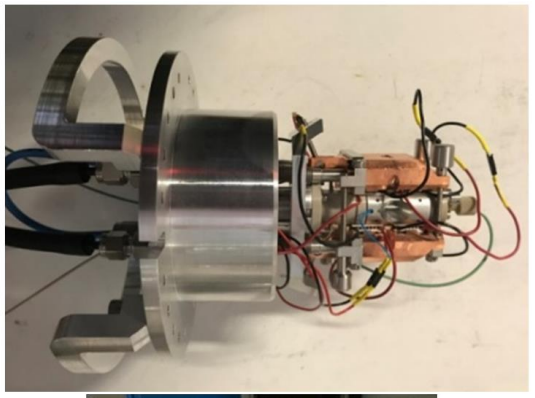
Robotic SAXS since 2024

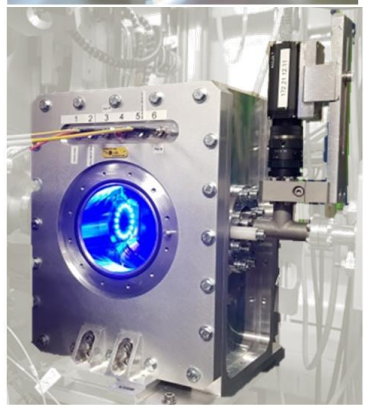


Transmission and buffer measurements are crucial

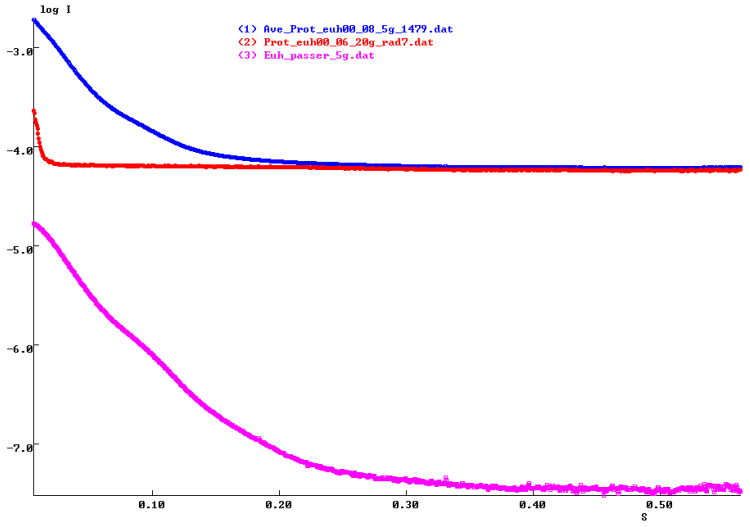
- Transmission
  - The experimental scattering intensity is normalised by transmitted intensity.
  - Transmitted intensity must be measured with high accuracy (~ 0.1 %).







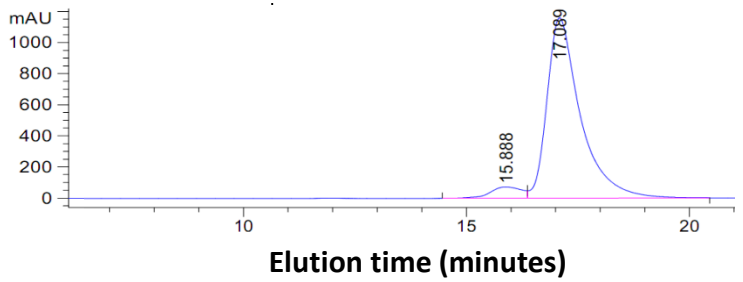
- Buffer
  - Buffer and protein samples must be measured in the same cell for correct subtraction of parasitic background arising from slits and holder walls.
  - The buffer in the buffer sample must be <u>the same</u> that of the protein sample (dialysis, SEC, …).



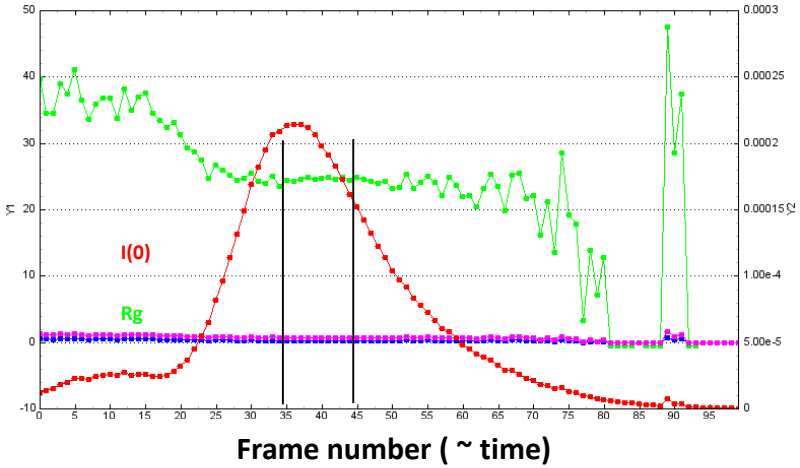
$$I_{particles}(q) = I_{sample}(q) - I_{buffer}(q)$$

Remove aggregates using SEC-SAXS

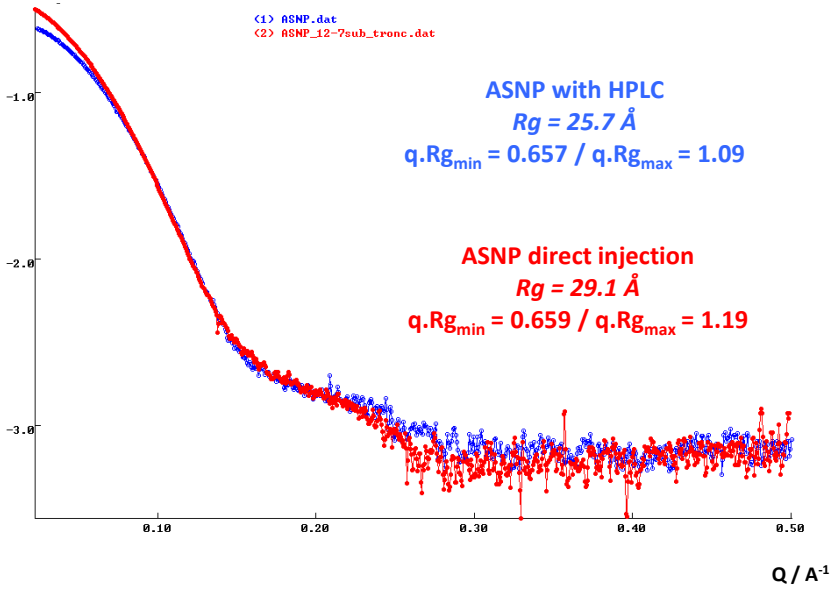
ASNP elution profile, monitored by UV absoprtion at 280 nm



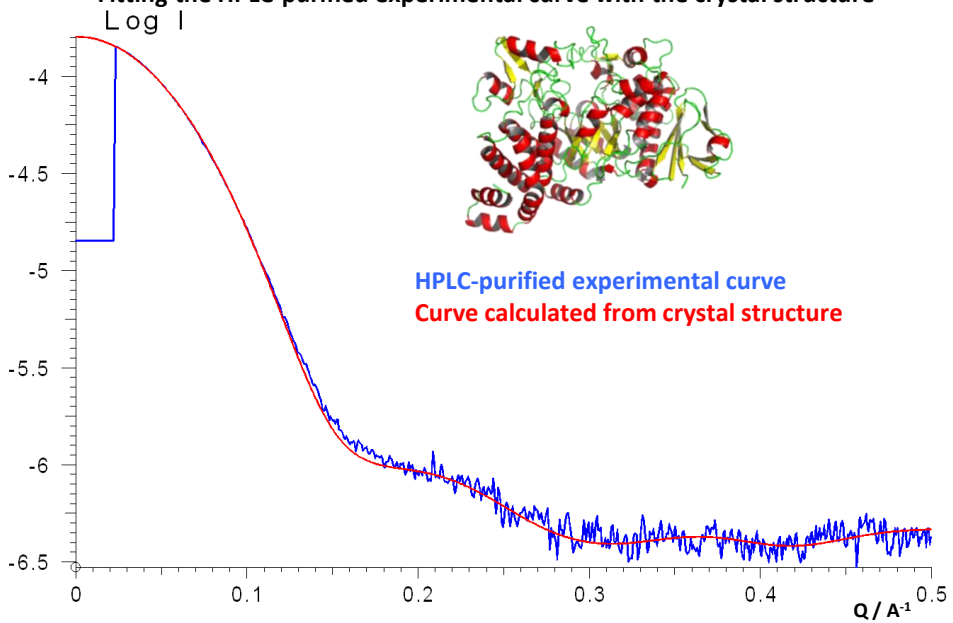
I(0) and Rg determined for each SAXS frame during elution



Comparison between HPLC-purified and Direct injection curves

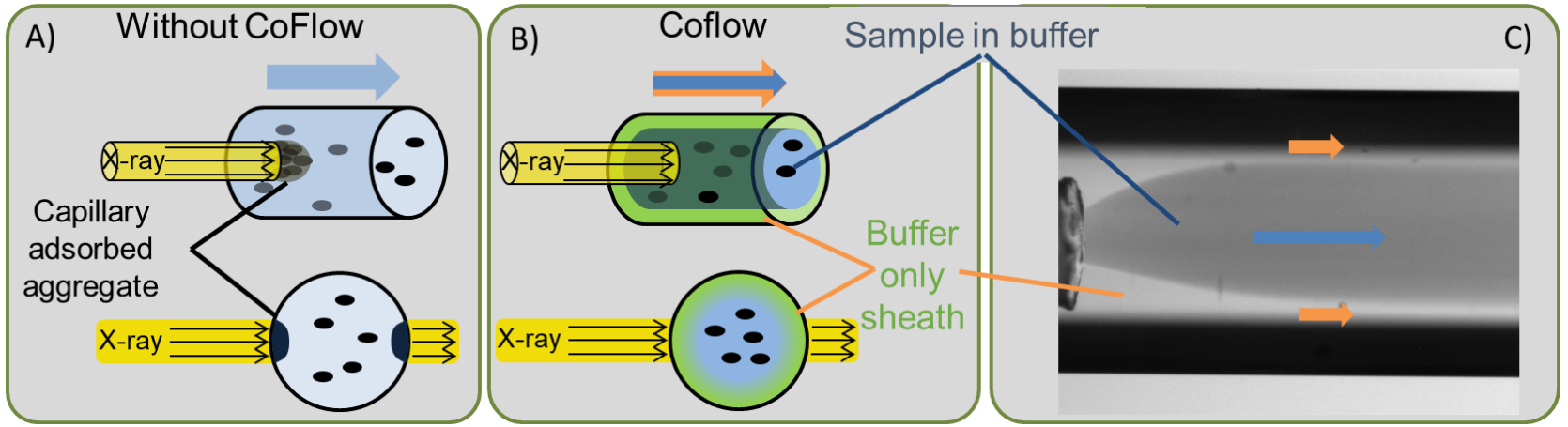


Fitting the HPLC-purified experimental curve with the crystal structure

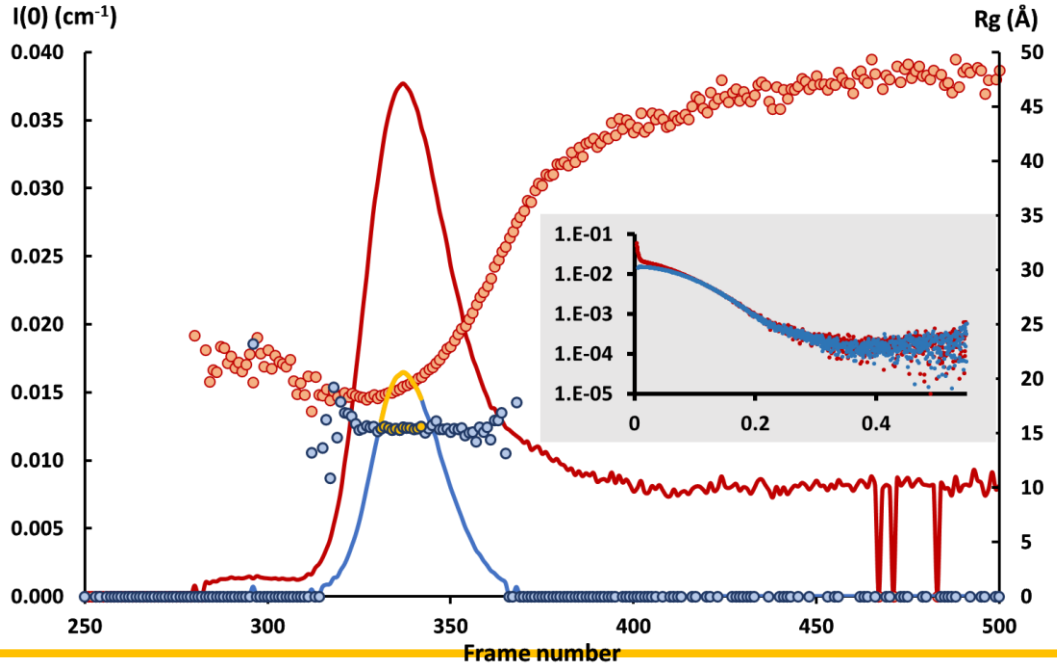


Unwanted Radiation effects

Co-flow set-up



Original idea:
Kirby, N *et al.* (2016). Acta Cryst. D72, 1254–1266.



Mass retrieval from Guinier analysis

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

Prof. André Guinier
1911-2000
Orsay, France

Absolute Unit : cm^{-1}

Classical electron radius

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

$$Rg^2 = \frac{\int_V r^2\left(\rho_{prot}(\vec{r}) - \rho_{buf}\right)d^3\vec{r}}{\int_V \left(\rho_{prot}(\vec{r}) - \rho_{buf}\right)d^3\vec{r}}$$

Mass concentration

Electronic density contrast

Protein specific volume

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

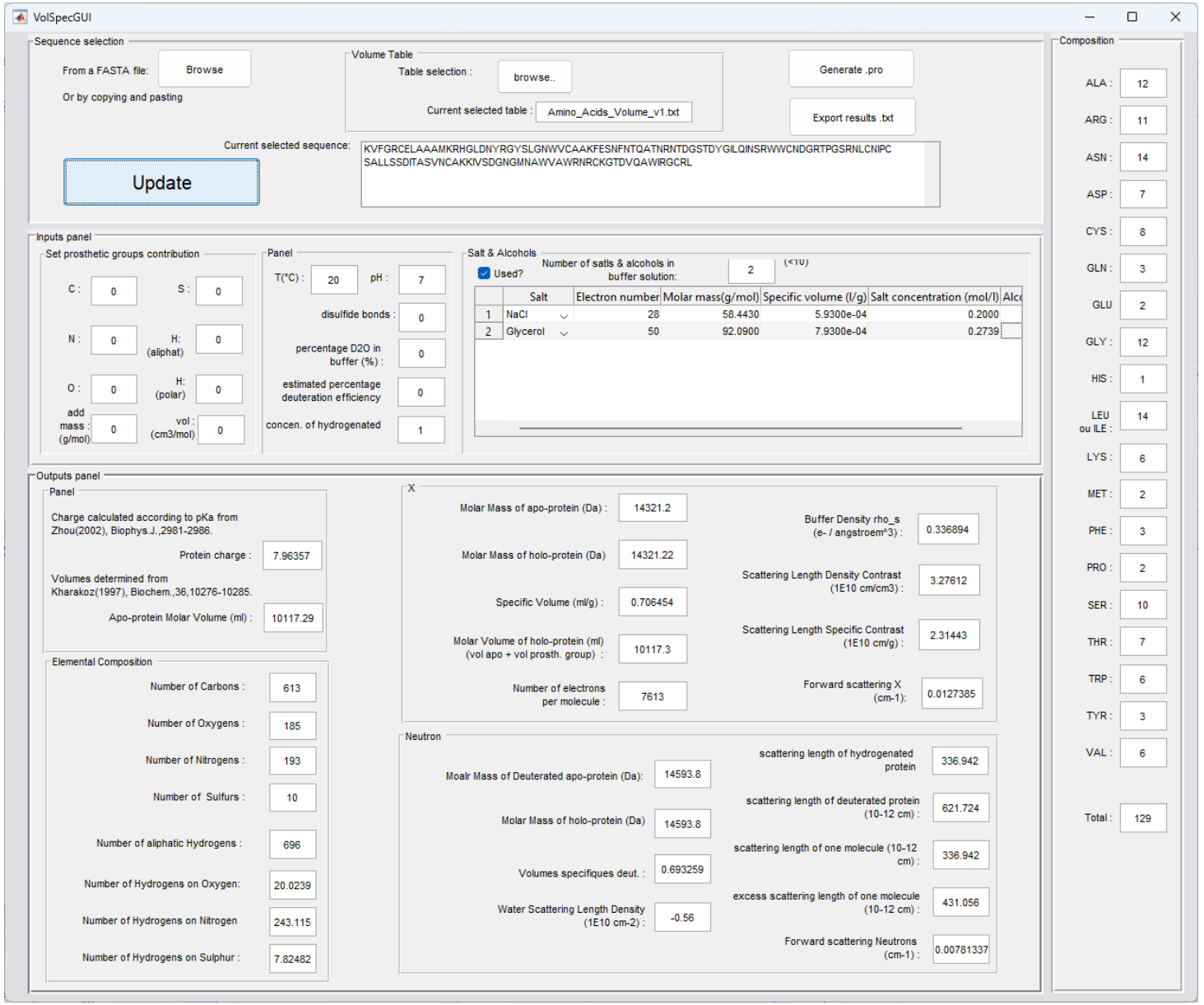
Typically :
M (kDa) = (1200 ~ 1600) * I0 (cm^{-1}) / C (mg/ml)

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

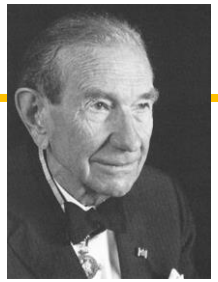
For globular proteins : $R_g\,(Å) \approx 6.5 * M^{\frac{1}{3}}, M\ in\ kDa$
For unfolded proteins : $R_g\,(Å) \approx 8.05 * M^{0.522}$

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

VolSpec : a tool to compute I_0 (cm^{-1}) from protein sequence and buffer comp.



Kratky Plot

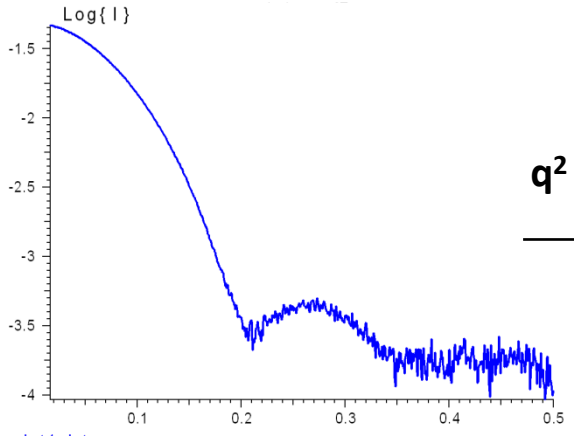


Prof. Otto Kratky
1902-1995
Graz, Austria

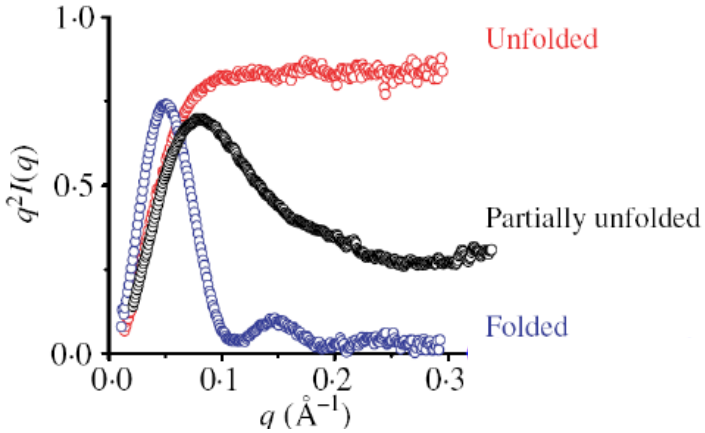
SAXS provides a sensitive means to ***evaluate the degree of compactness*** of a protein:

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

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



q^2 I(q) *versus* q



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

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

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

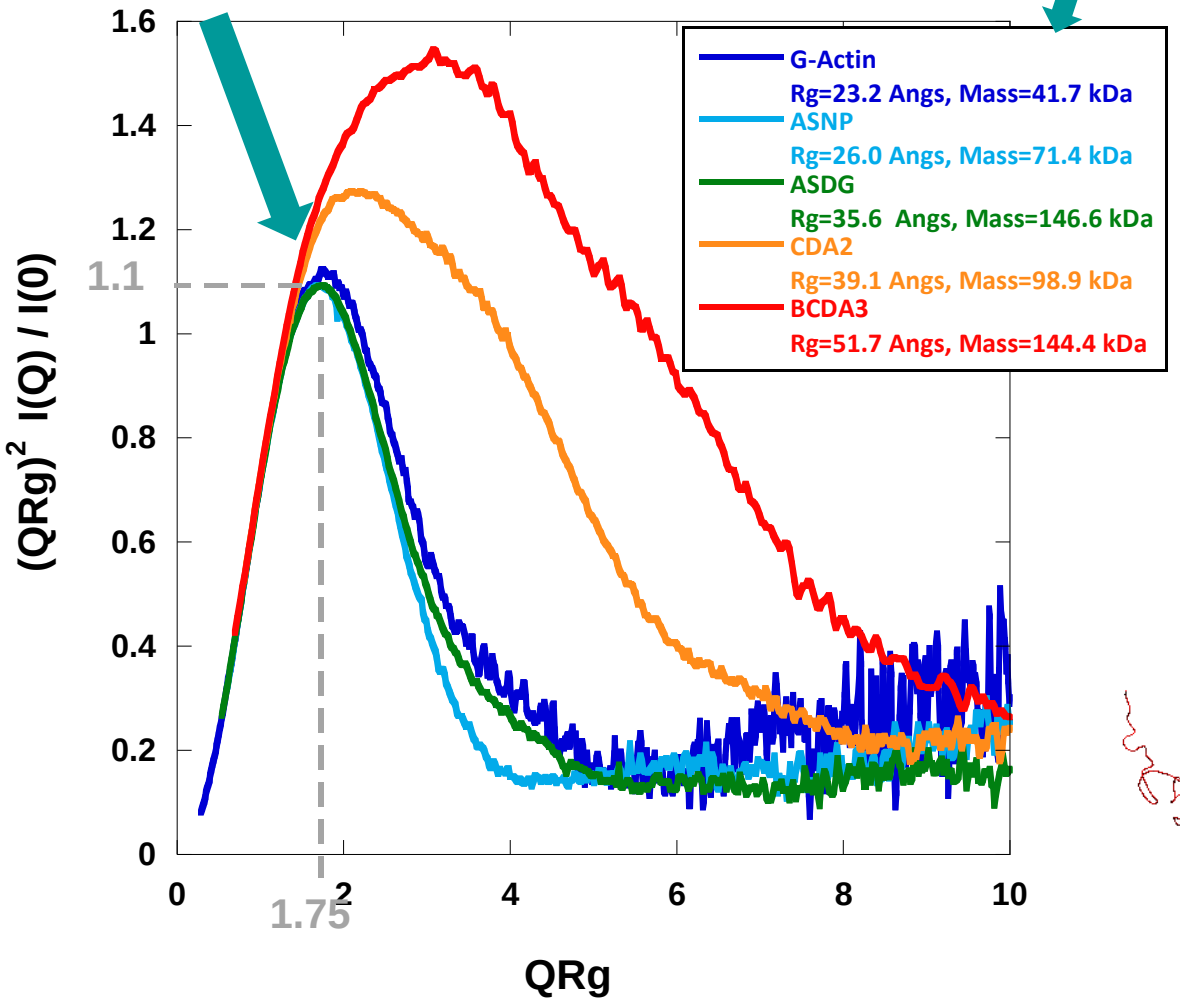
Extended polymer chain : *increase* at large q-values (asymptotic behaviour in $I(q) \sim q^{-1.x}$)

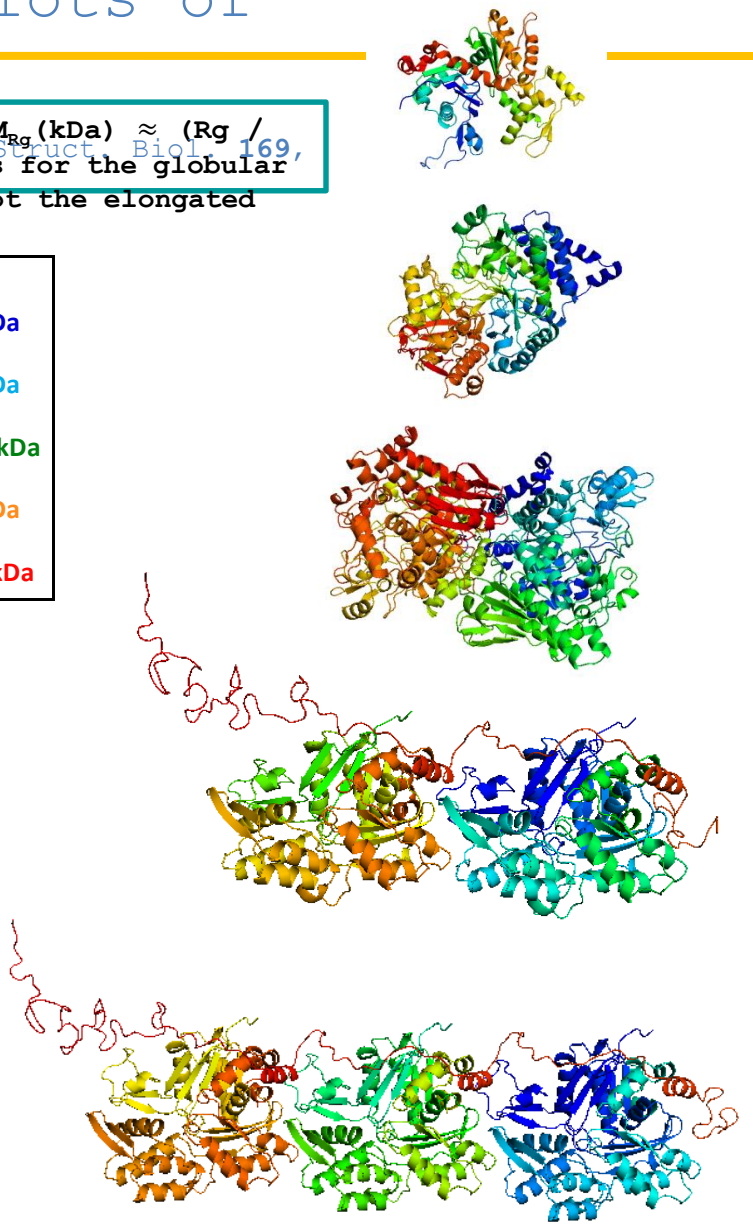
Dimensionless Kratky Plots of folded proteins

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

For globular structures, the plots fold into the same maximum

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





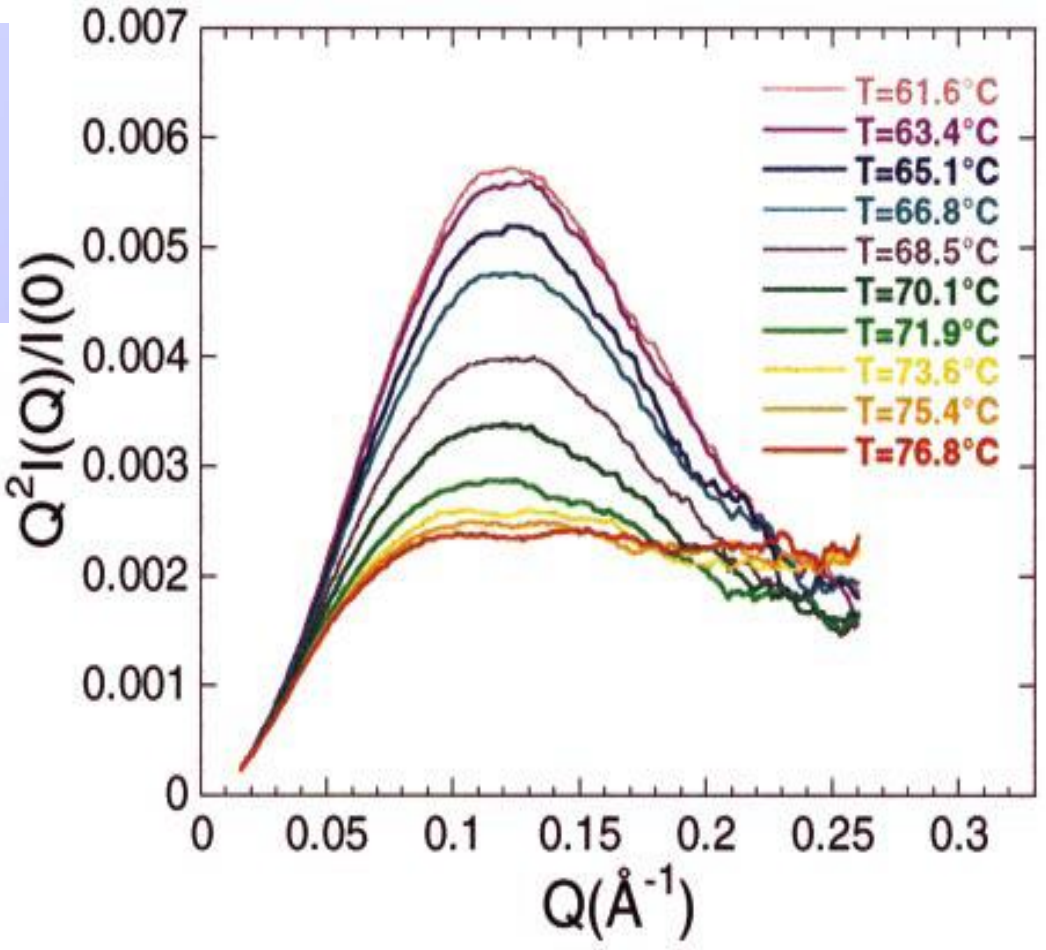
e position of the maximum on the dimensionless bell shape tells to what extent the protein is globula

Kratky Plot : NCS heat unfolding

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

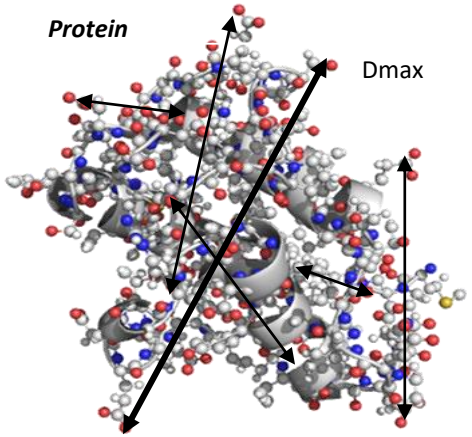
In practice, thin Gaussian chains do not exist.

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

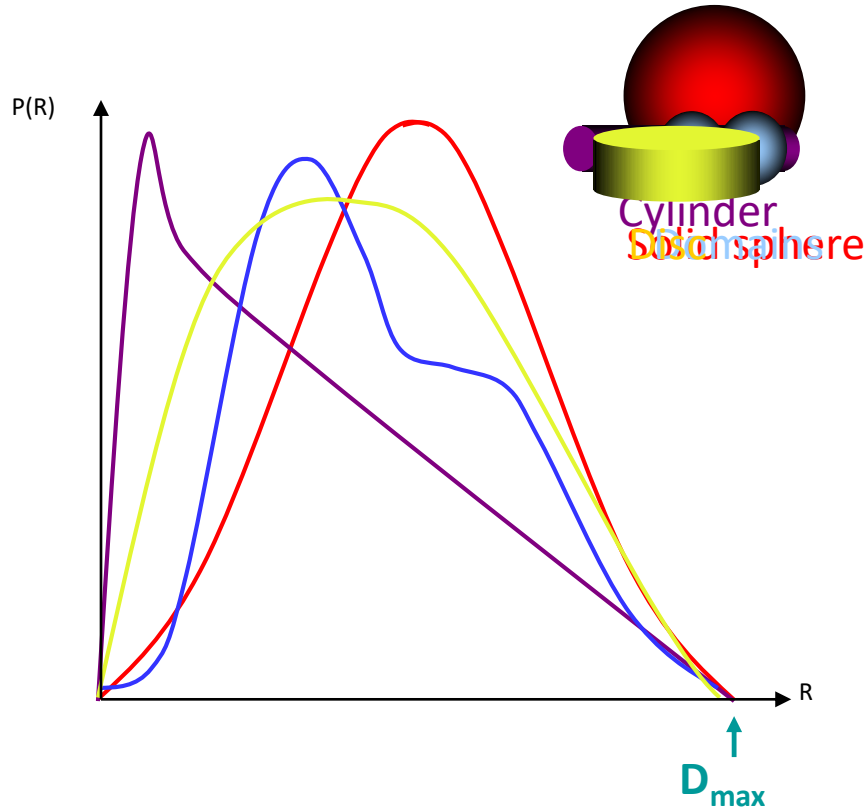


Pair Distribution Function p(r)

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



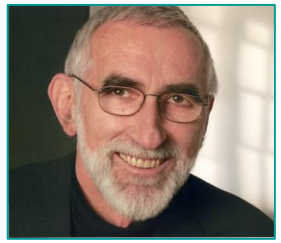
p(r) vanishes at r = D_{max}



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

Back-calculation of the Pair Distribution Function

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



Prof. Otto Glatter
Guinier Prize 2012
Graz, Austria

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

- D_{max} is proposed by the « user »
- A guess for p(r) is decomposed over [0, D_{Max}] by a linear combination of orthogonal functions

$$p_{calc}(r) = \sum_{1}^{M} c_n \, \phi_n(r)$$

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

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

- $\{c_n\}$ are optimized recursively

Dr. Dmitri Svergun
Guinier Prize 2018
Hamburg, Germany

Svergun (1988) : program "GNOM"

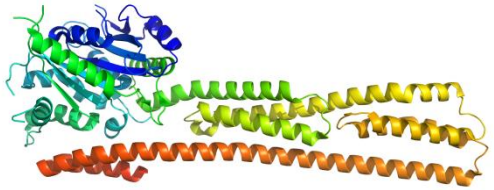
M ~ 30 - 100 ⇒ ill-posed LSQ ⇒ regularisation method

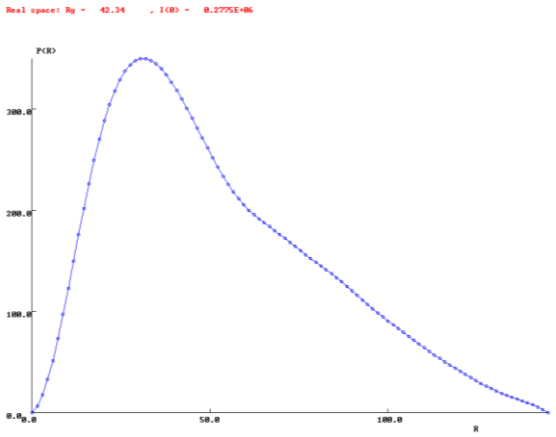
+ "Perceptual criteria" : smoothness, stability, absence of systematic deviations

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

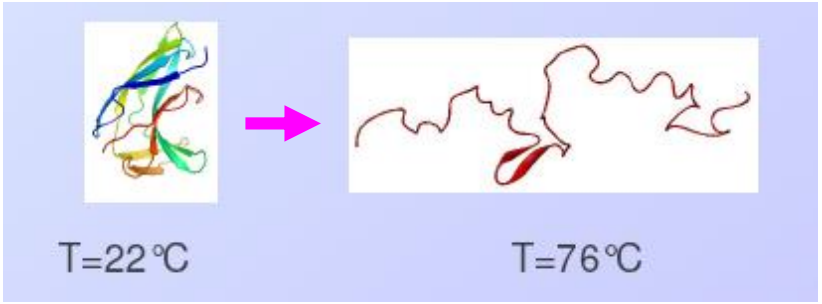
Experimental examples

GBP1

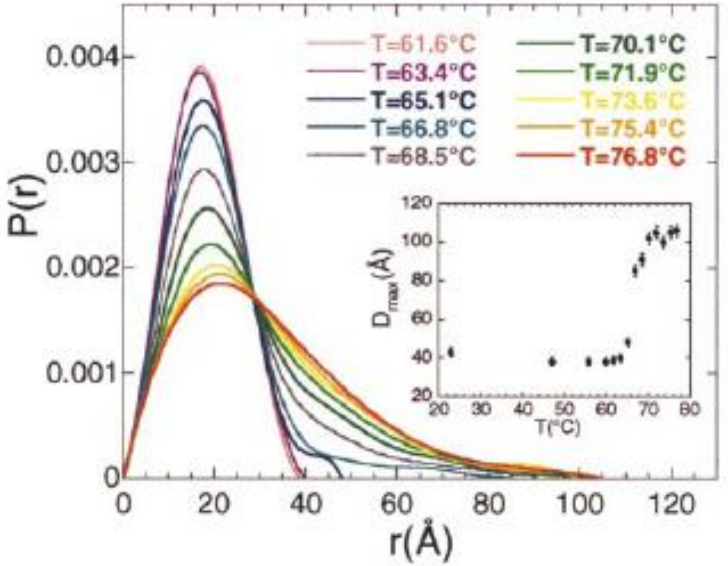




Heat denaturation of Neocarzinostatin

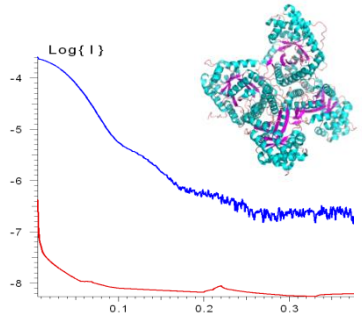


Pérez et al., J. Mol. Biol. (2001) 308, 721-743

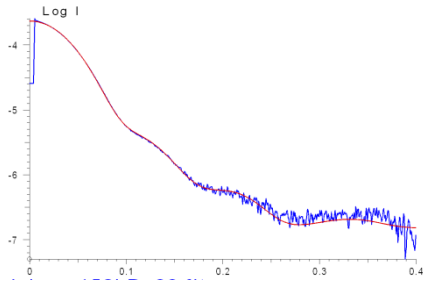


BioSAXS data modeling

- Theorical model or complete atomic structure available

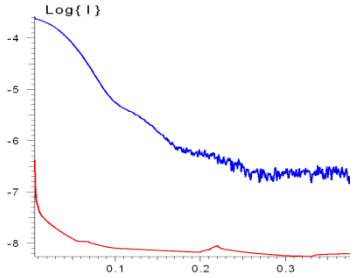


Validation/identification in solution

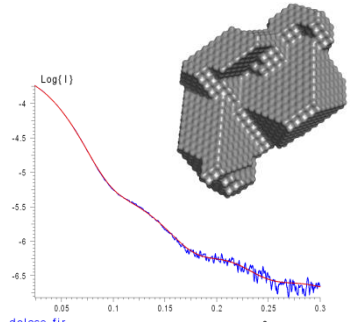


CRYSOL (Atsas)
FOXS
PepsiSAXS
WAXSiS

- Nothing known (except the curve)

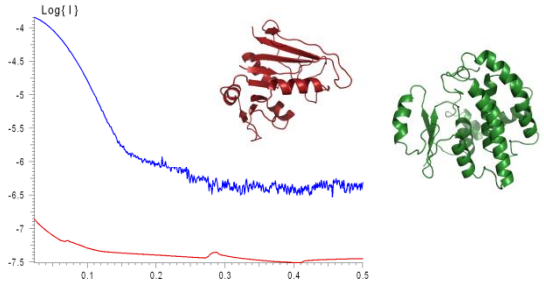


Ab initio models



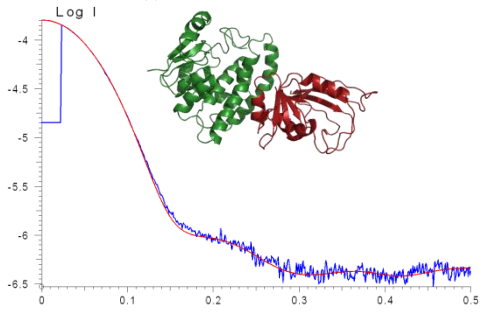
DAMMIN (Atsas)
DAMMIF (Atsas)
GASBOR (Atsas)
MONSA (Atsas)
DENFERT

- Structures of subunits available



Rigid body modeling of the complex and

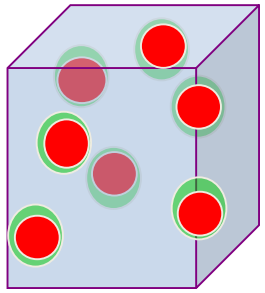
molecular modeling of the missing part



SASREF (Atsas)
BUNCH (Atsas)
CORAL (Atsas)
DADIMODO
BILBO-MD

From atomic coordinates to a SAXS curve: Crysol program

The bound solvent density differs from that of the bulk



$$I_{calc}(q) = \langle |A_a(\vec{q}) - \rho_S A_s(\vec{q}) + \delta\rho_b A_b(\vec{q})|^2 \rangle_\Omega$$

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

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

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

To gain computing time, $I_{calc}(q)$ is expanded in a series of Bessel functions and spherical harmonics.

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

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

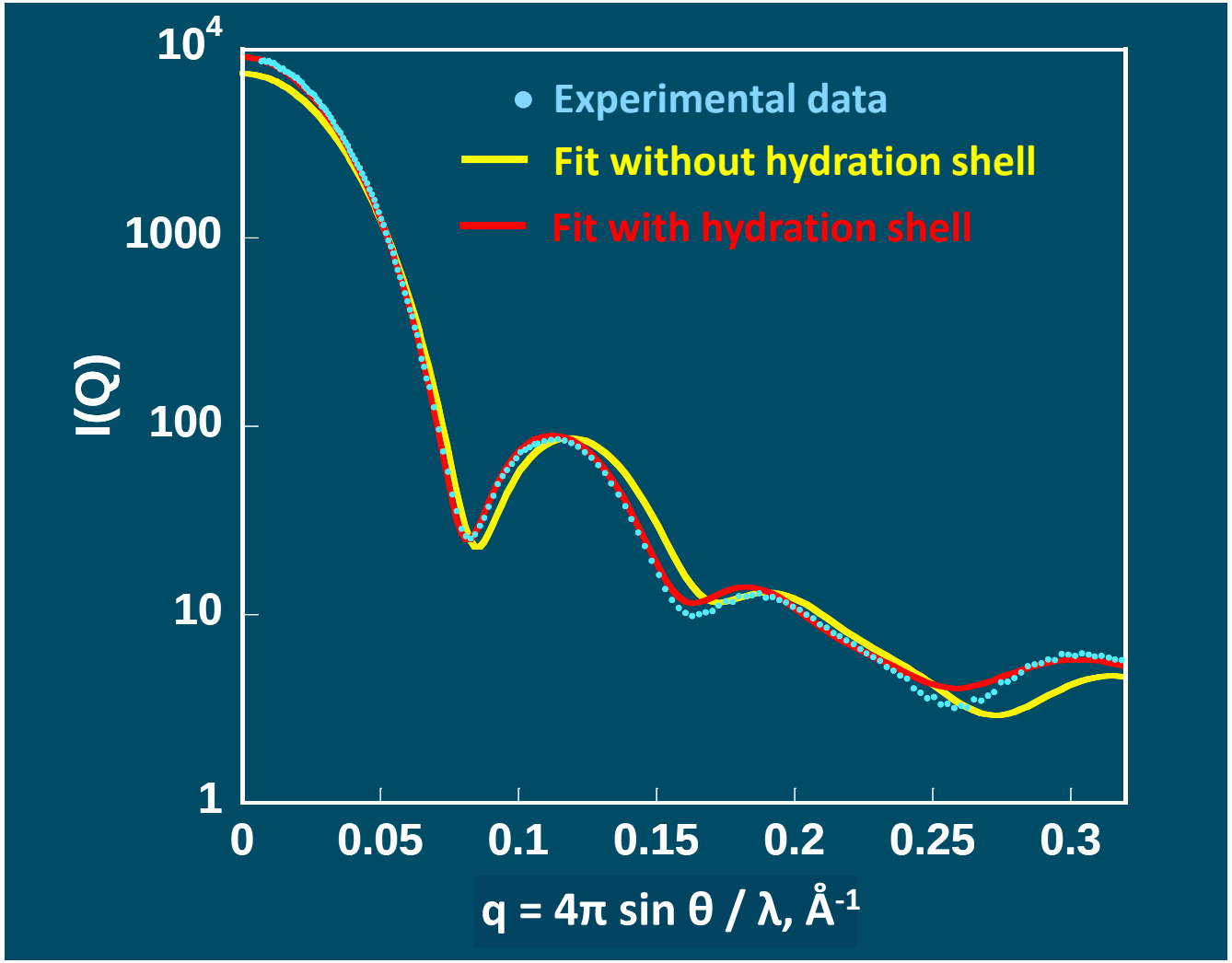
- the general scale of $I_{calc}(q)$
- the total excluded volume V, which is equivalent to adjusting the average electronic contrast
- the contrast of the border layer $\delta\rho$

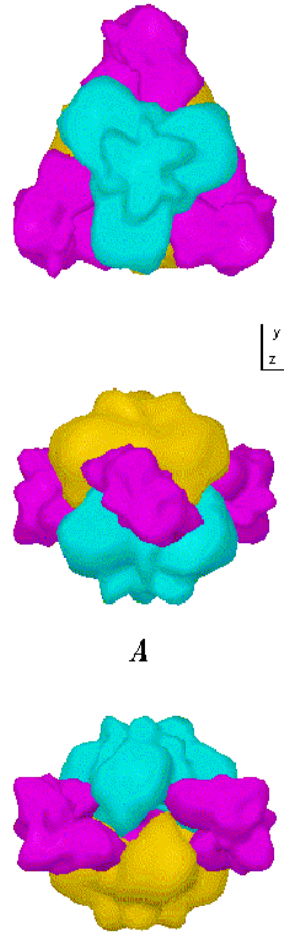
$$\chi^2 = \frac{1}{N-1} \sum_{i=1}^{N} \left[\frac{I_{\exp}(q_i) - scale * I_{calc}(q_i)}{\sigma_{\exp}(q_i)} \right]$$

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

Effect of the hydration shell

T state of *E. coli* allosteric ATCase





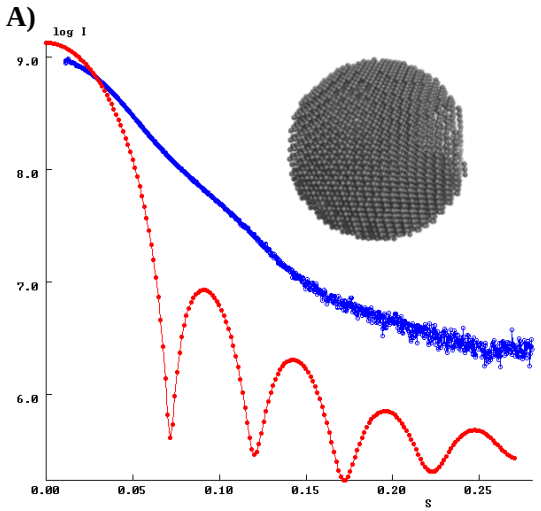
3D shape reconstructions from SAXS data with DAMMIN

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

Principle of the method : any structure volume of **homogeneous electronic density** can be approximated at any resolution by a set of spheres of small enough radius (r_b)

Starting model = sphere with a radius R = $D_{max}/2$ with N scattered beads (r_b << R)
The number of the "dummy atom" N ≈ $(R/r_b)^3$

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

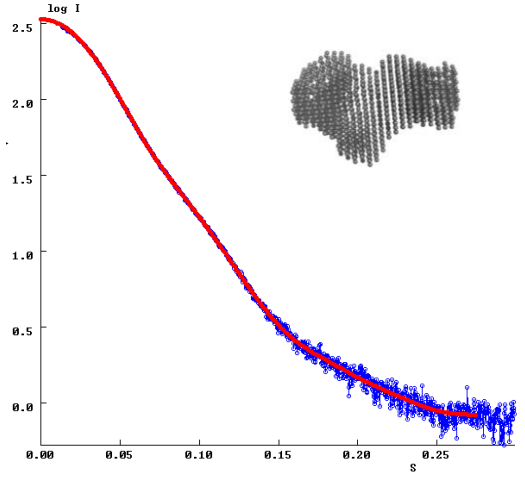


$$f(X) = \chi^2\left[I(q)_{\exp}, I(q,X)\right] + \alpha P(X)$$

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

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

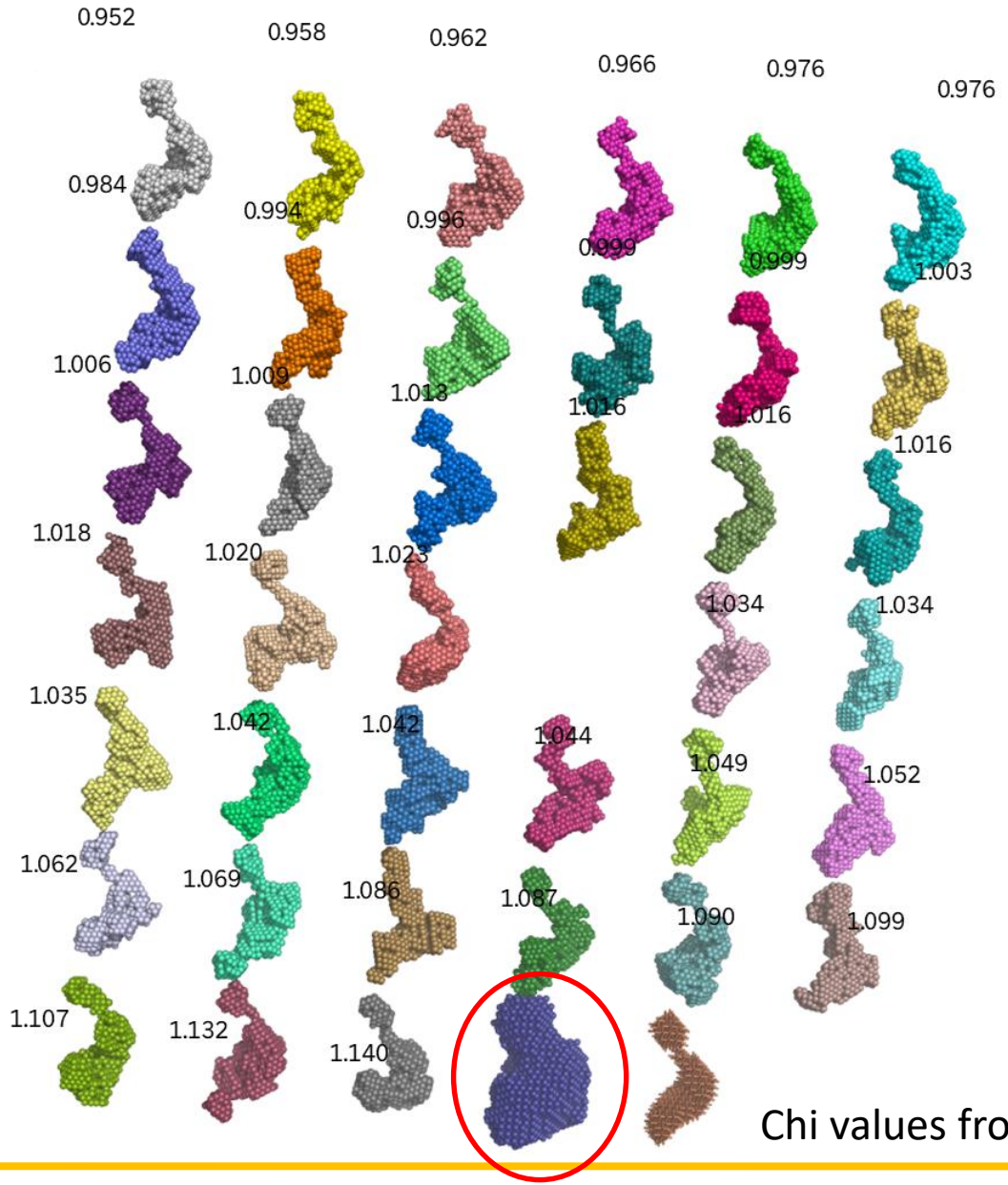
After k iterations



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

3D shape reconstructions from SAXS data with DAMMIN

NSD values



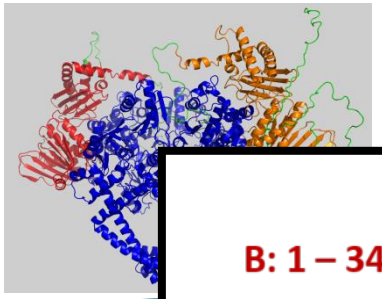
Chi values from 1.704 to 1.942

High-resolution information from domains only: Dadimodo Web Server

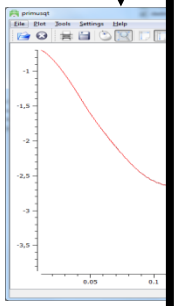
https://dadimodo.synchrotron-soleil.fr

3 input files needed to launch Dadimodo on the Web Server

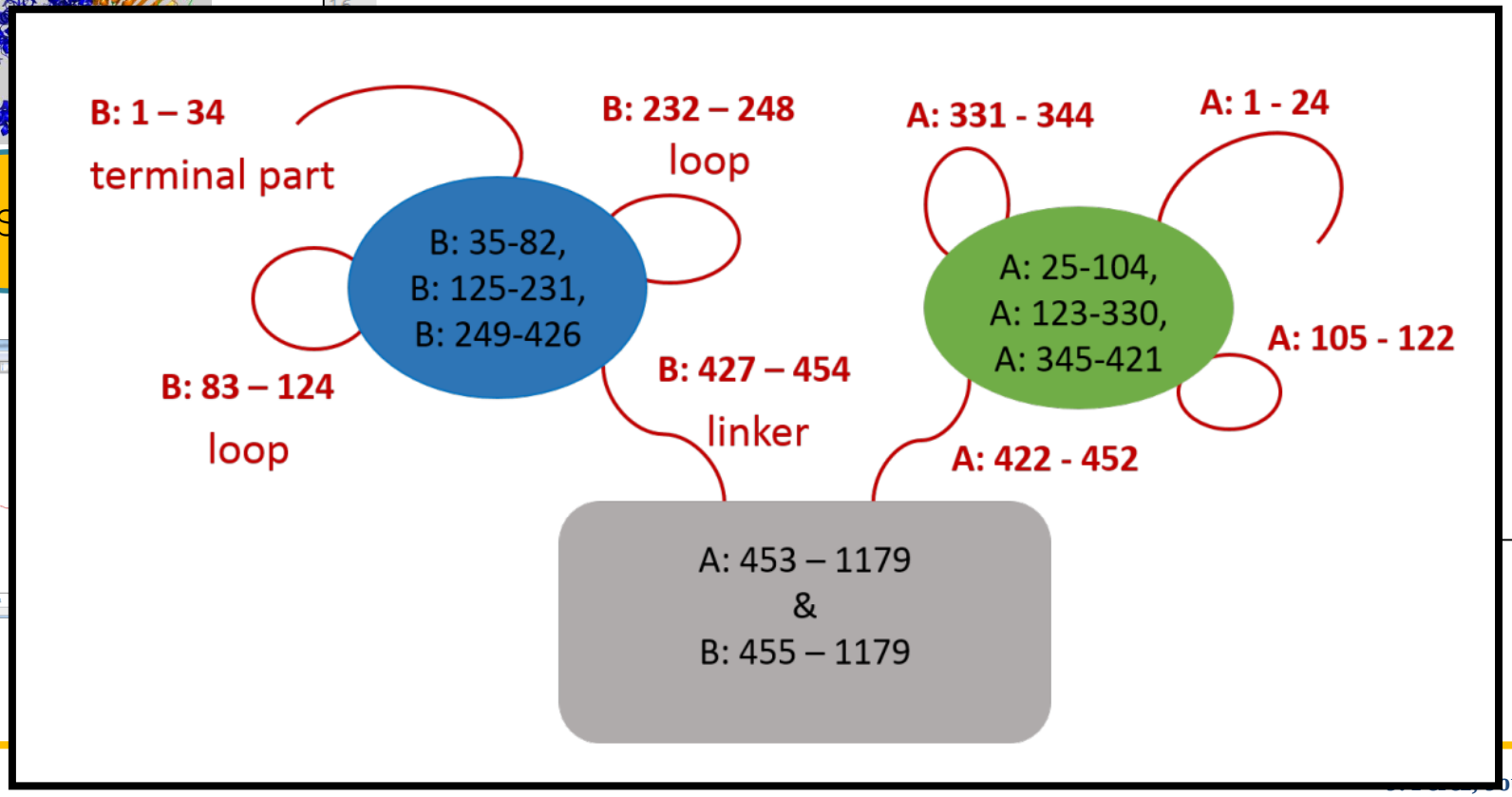
Complete PDB file



SAXS



```
6  ########################## USER DEFINED PARAMS ##########################
7
8  [structure] ; ------------------ STRUCTURE ------------------
9
10 # Define here the rigid bodies in the same nomenclature as your pdb
11 # Syntax: rigid_body =  list of expressions 'chain: first_residue - last_residue' separated by comma
12 #
13 body1 = A: 25-104, A: 123-330, A: 345-421
14 body2 = A: 453-1179, B: 455-1179
15 body3 = B: 35-82, B: 125-231, B: 249-426
```



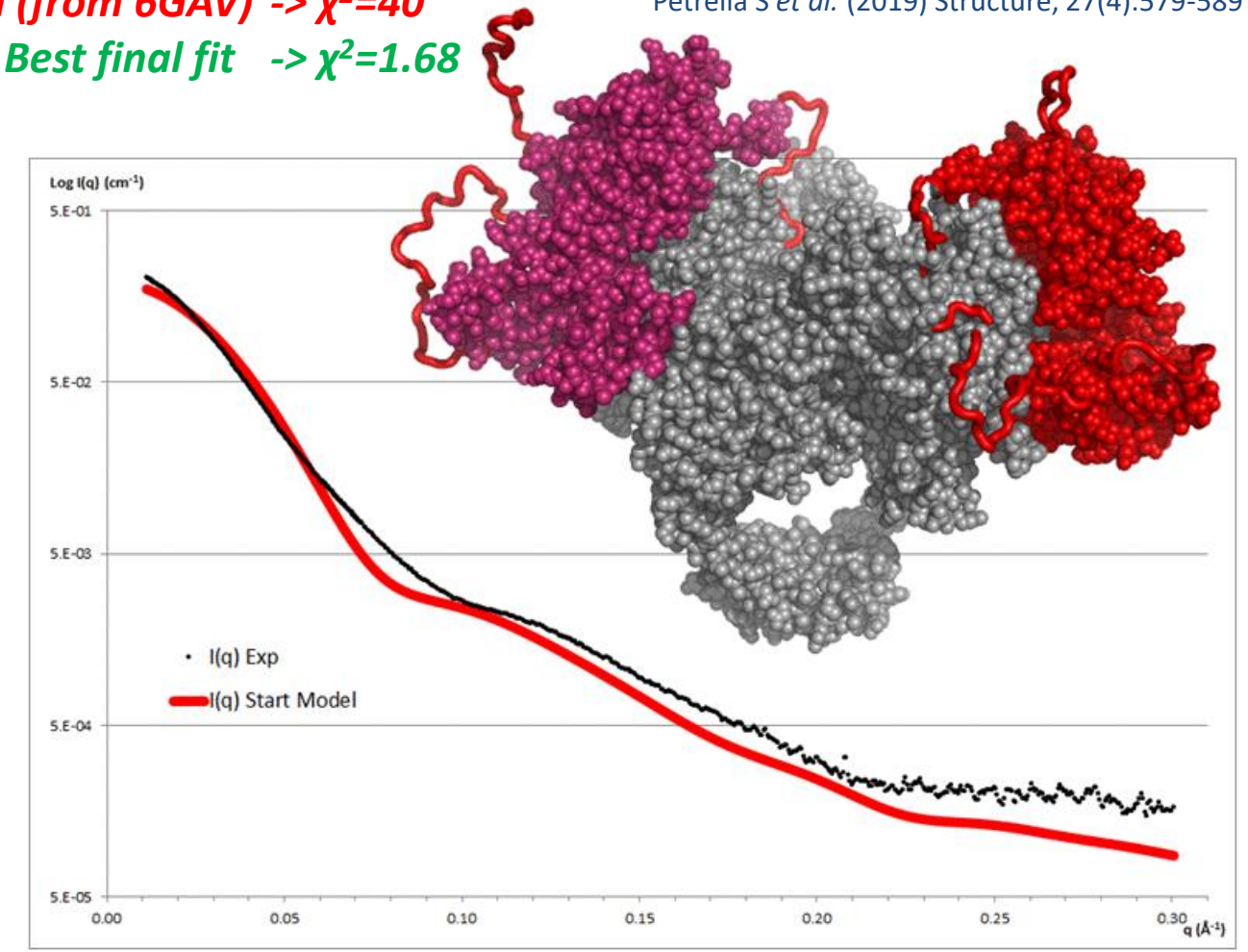
Rigid-Body refinement using Dadimodo

Mycobacterium tuberculosis DNA Gyrase

Petrella S *et al.* (2019) Structure, 27(4):579-589

Start model (from 6GAV) -> χ^2=40

Best final fit -> χ^2=1.68

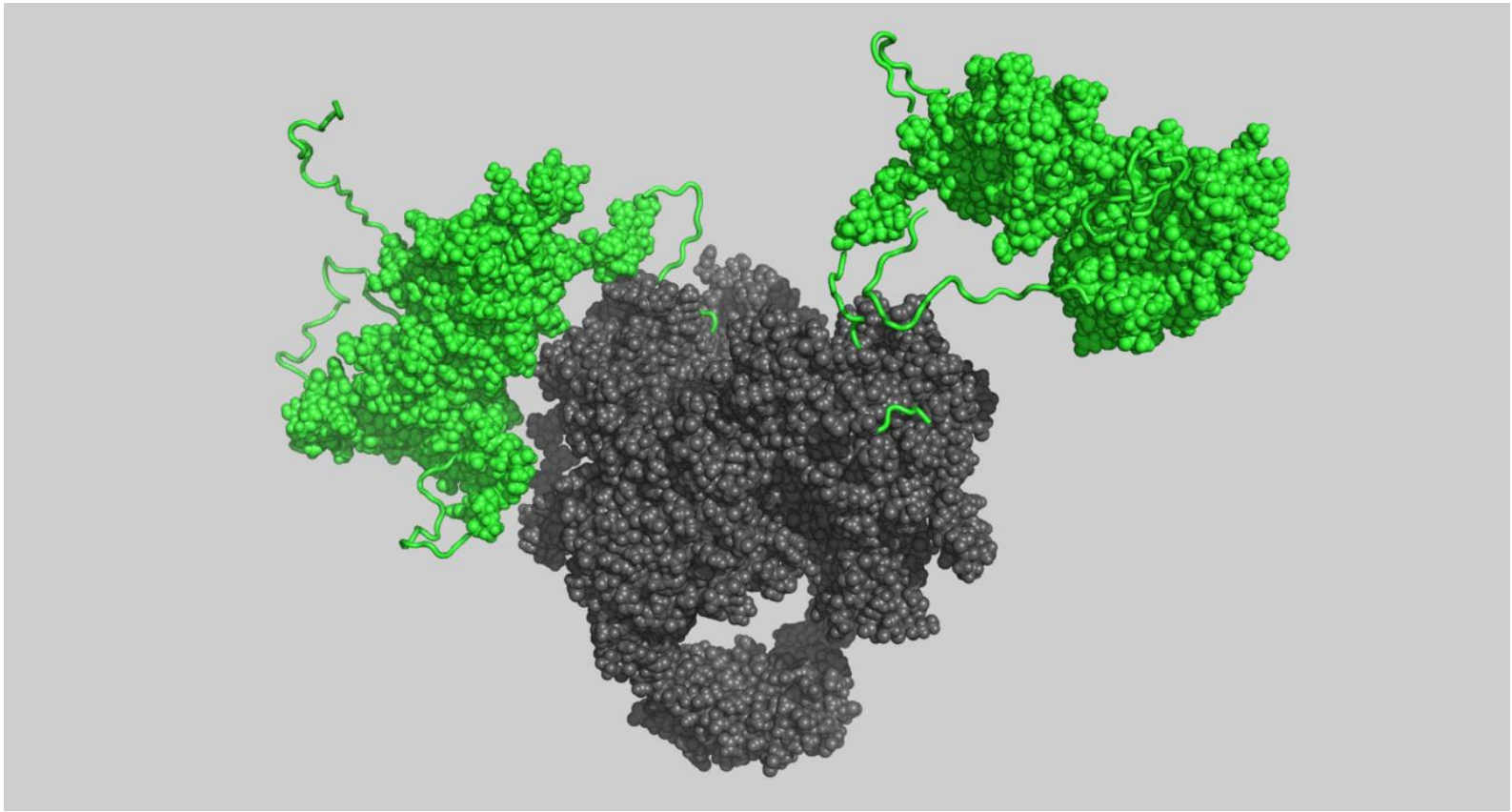


Rigid-Body refinement using Dadimodo

Mycobacterium tuberculosis DNA Gyrase

5 best final fits : 1.68 < χ^2 < 1.76

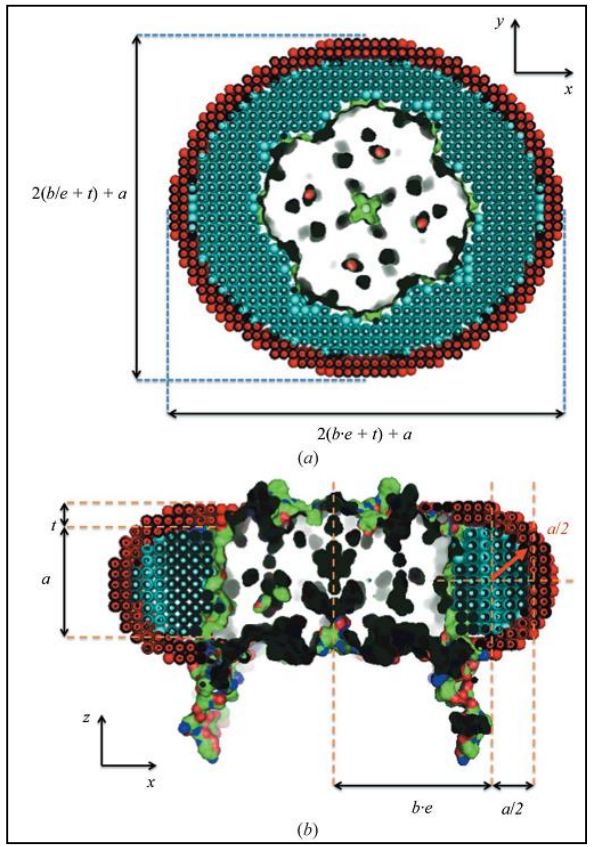
Petrella S *et al.* (2019) Structure, 27(4):579-589



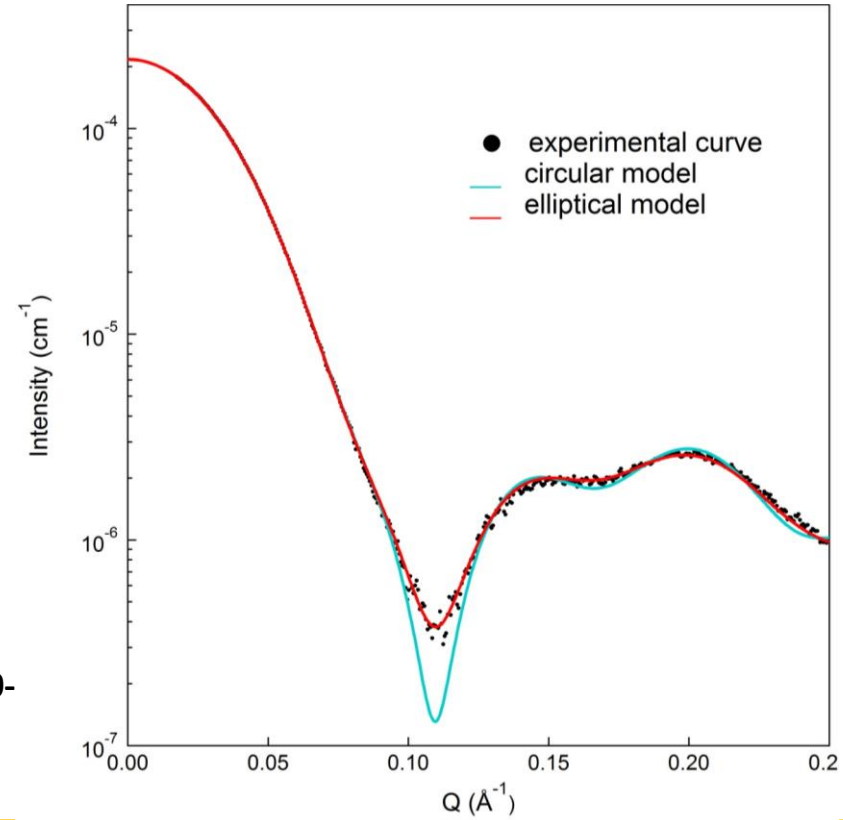
Mycobacterium tuberculosis DNA Gyrase

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

Pérez J. & Koutsioubas, A. (2014). Acta Cryst.D70
F. De Pol et al. (2024), J. Appl. Cryst 57

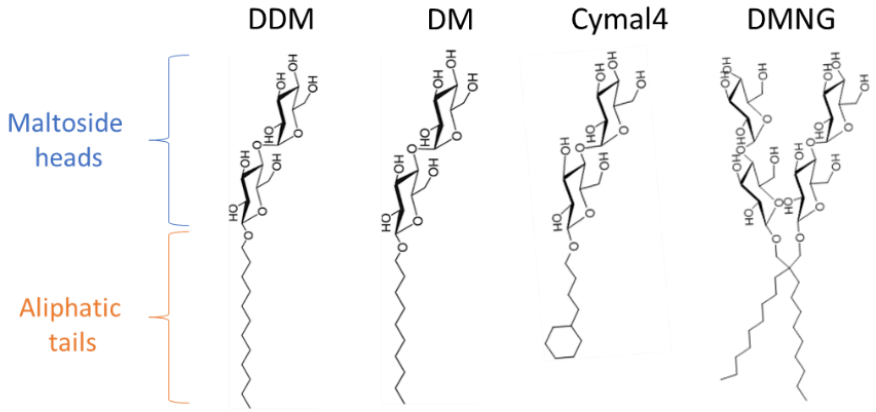


Algorithm of the *Memprot* program. The program essentially creates PDB files with the models made of the full-atom protein structure and the parameterized coarse-grained detergent corona, and *CRYSOL* is called to calculate the SAXS curves. An overall sorting on the χ value is performed to keep the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

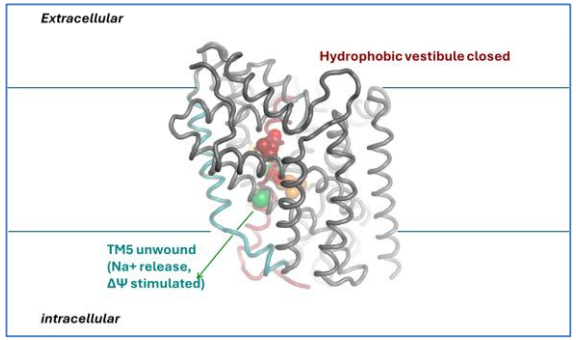


Berthaud et al. (2012), *JACS*, 134 (24), 10080-10088

MhsT in four different detergents

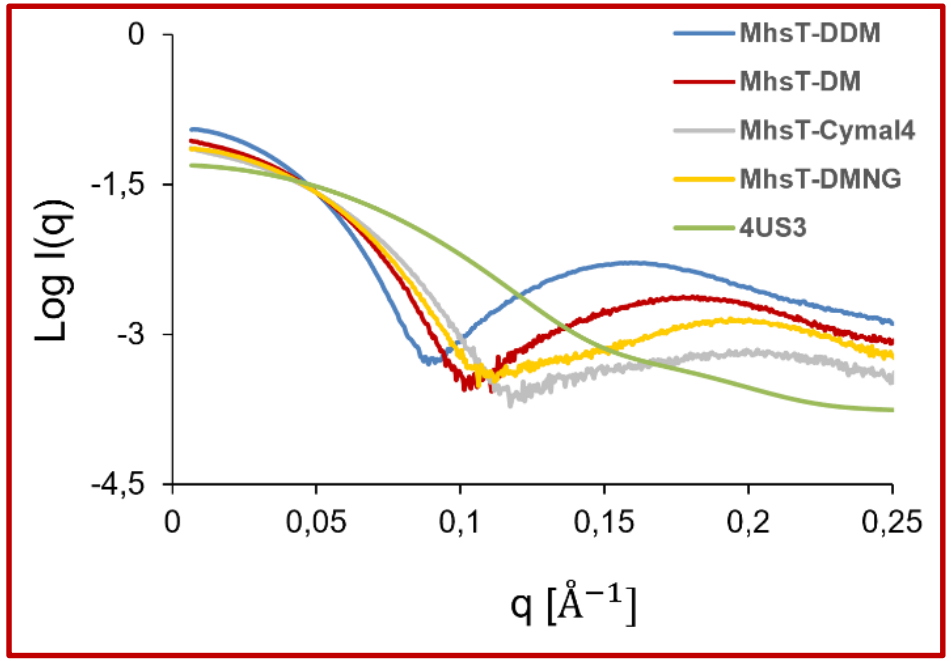


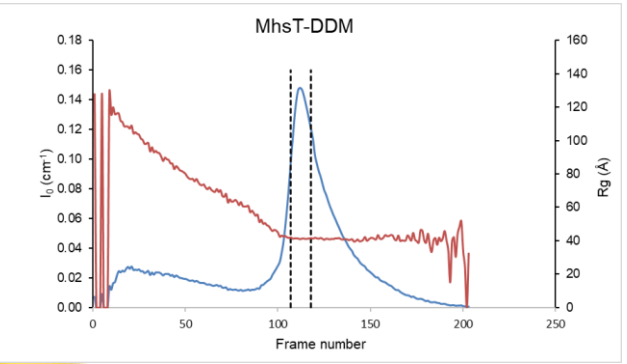
De Pol et al., J. Appl. Cryst (2024)

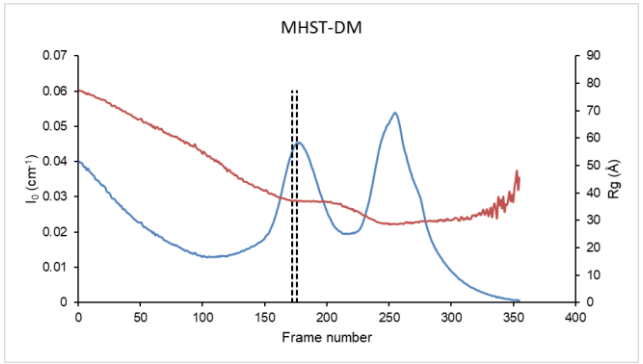


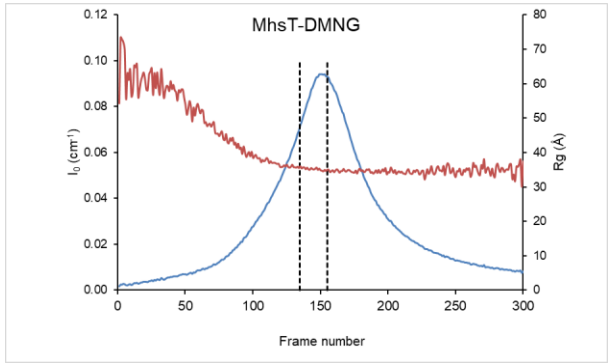
B. halodurans **multi-hydrophic amino acid transporter MhsT – bacterial SLC6 transporter**

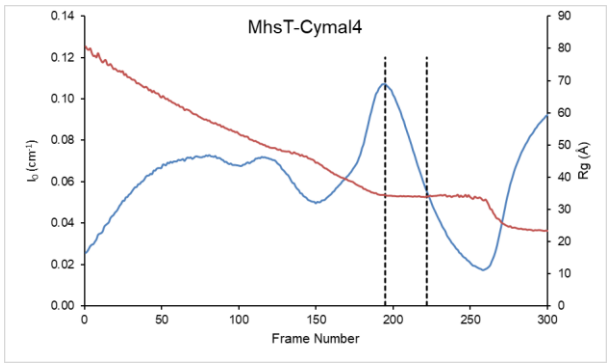
Small, rigid, known structure: 4US3











50 µl, c=6.7 mg/ml

50 µl, c=14.2 mg/ml

50 µl, c=10.5 mg/ml

50 µl, c=11.0 mg/ml

Buffer (pH=7): Tris-HCl: 10 mM, NaCl: 100 mM**, glycerol: 10% (v/v)**, L-tryptophan: 0.5 mM

+ DDM, c = 2 cmc
cmc = 0.17 mM

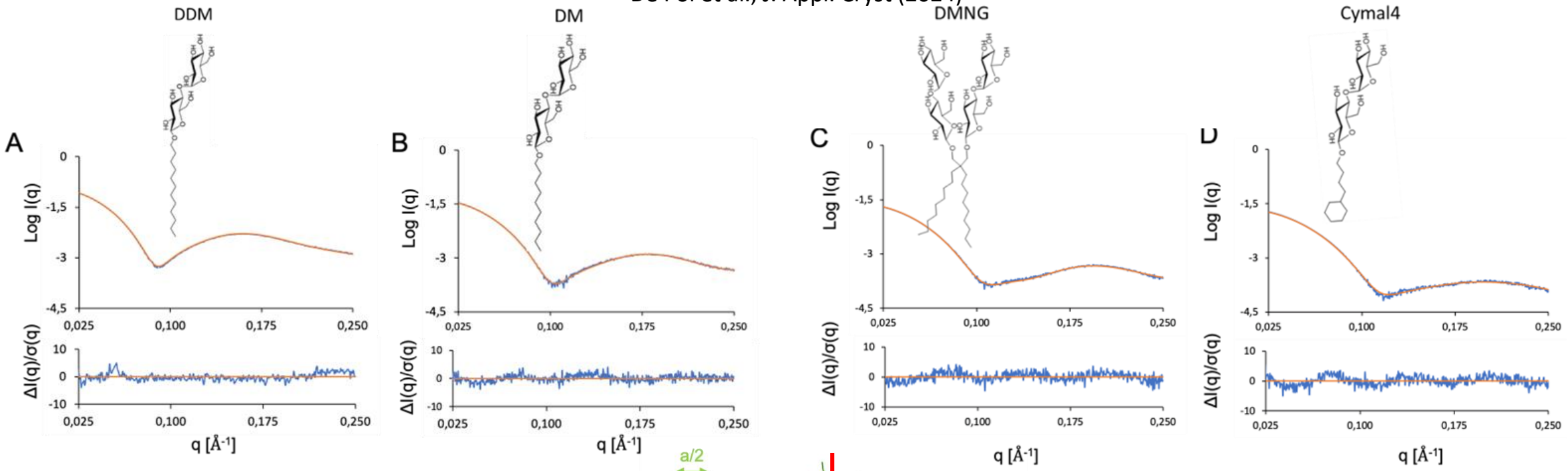
+ DM, c = 2 cmc
cmc = 1.66 mM

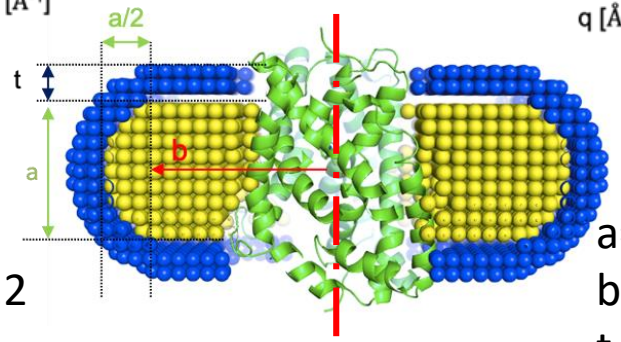
+ DMNG, c = 4 cmc
cmc = 0.036 mM

+ Cymal4, c = 2 cmc
cmc = 7.6 mM

MhsT in four different detergents

De Pol et al., J. Appl. Cryst (2024)





a=30.3±0.2
b=26.2±0.1
t=5.8±0.1
ρ_H=0.539±0.003
ρ_T=0.280±0.002

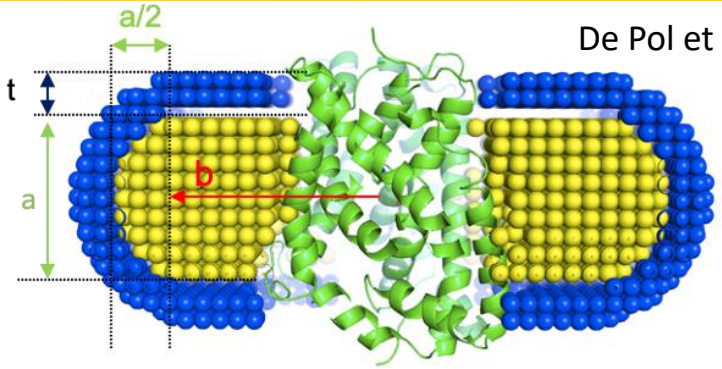
a=32±2.5
b=20.5±1.2
t=6.3±0.2
ρ_H=0.518±0.004
ρ_T=0.290±0.002

a=31.1±1.0
b=19.8±0.5
t=6.2±0.15
ρ_H=0.541±0.003
ρ_T=0.271±0.002

a=31±2
b=19±1
t=6.0±0.2
ρ_H=0.512±0.012
ρ_T=0.264±0.004

MhsT in four different detergents

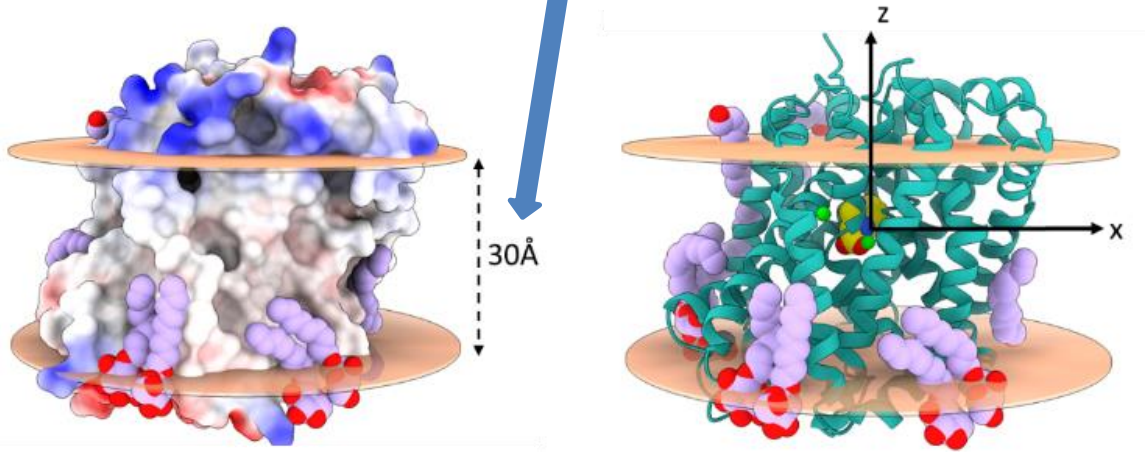
Comparison with predicted membrane thickness



De Pol et al., J. Appl. Cryst (2024)

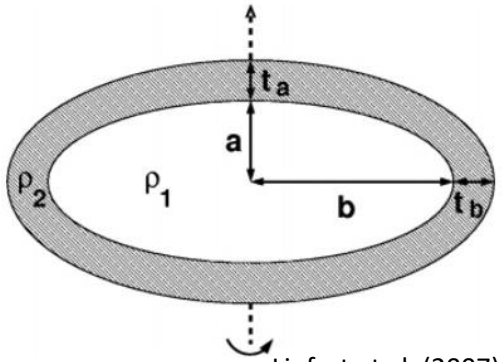
30.3 < « a » < 32 Å

Very close to predicted value



From « Orientations of Proteins in Membranes" database (Lomize et al., 2006)

Comparison with DM/DDM free micelles models



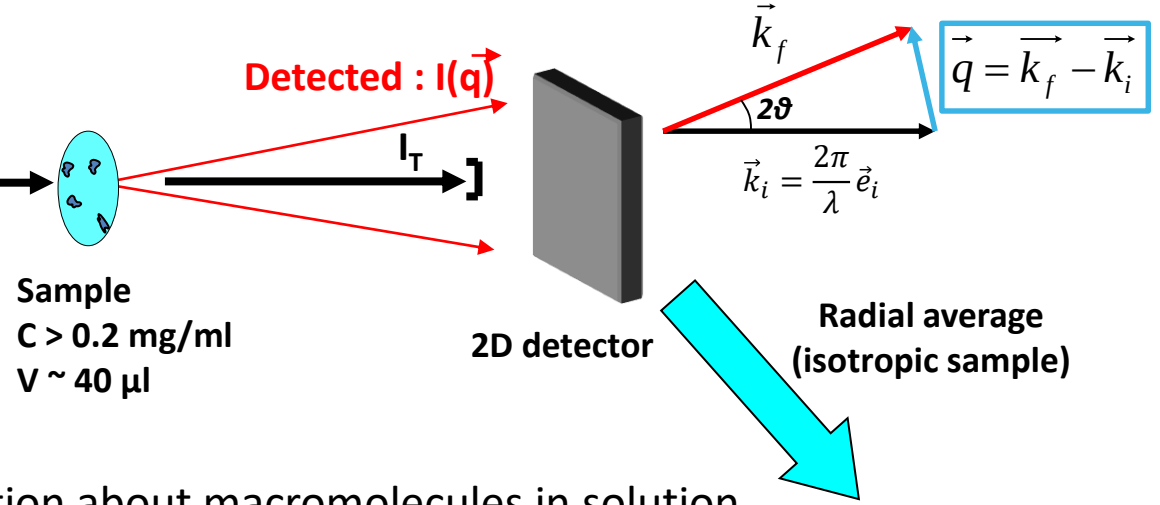
Lipfert et al. (2007), Phys.Chem.B, 111, 12427−12438

30.3 < « a » < 32 Å	*vs*	24 -28 Å (small ell. axis)
0.52 < « ρ_{Heads} » < 0.54 $e^-/Å^3$	*vs*	0.52 $e^-/Å^3$
0.26 < « ρ_{Tails} » < 0.29 $e^-/Å^3$	*vs*	0.28 $e^-/Å^3$
5.8 < « t » < 6.3 Å	*vs*	6.2 Å

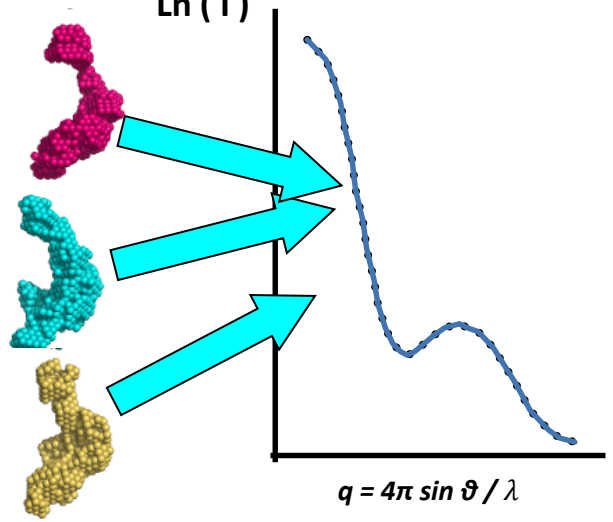
Very compatible values

The parameters resuting from the hybrid models seem to make sense for all four detergents

Summary



BioSAXS provides structural information about macromolecules in solution



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

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