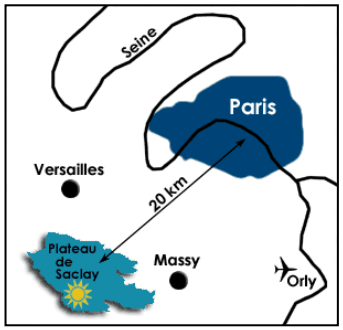
Combining Memprot & Dadimodo, programs for modeling the detergent belt in solubilized membrane protein complexes & re-orienting domains of multi-domain proteins

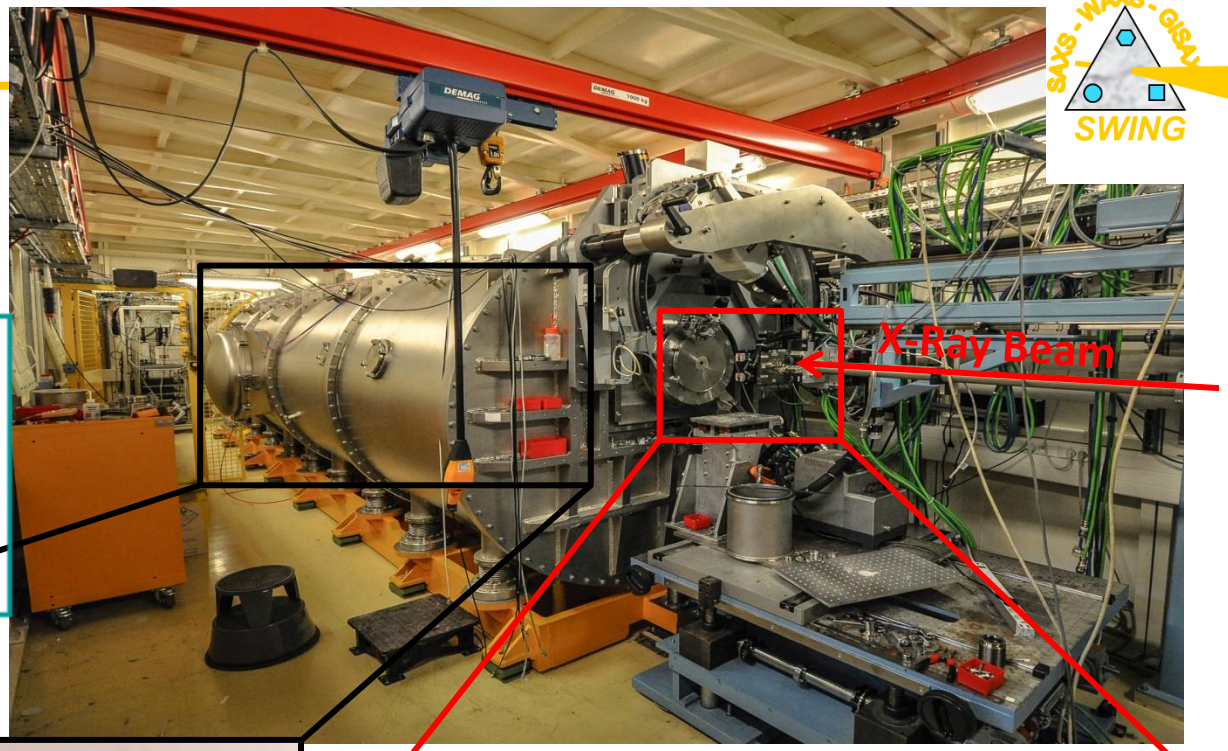


M. Baranowski, A. Thureau, O. Roudenko,
Javier Pérez

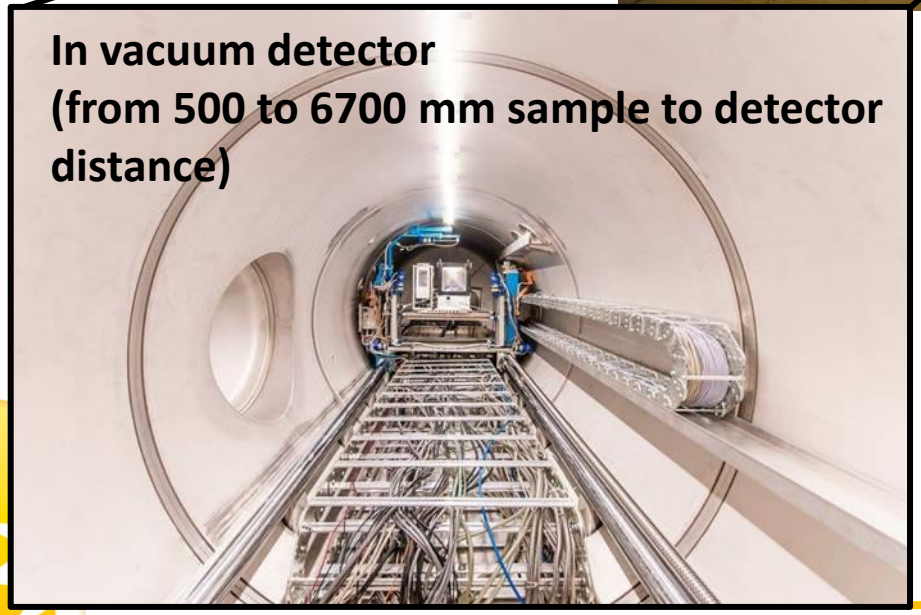
Synchrotron SOLEIL

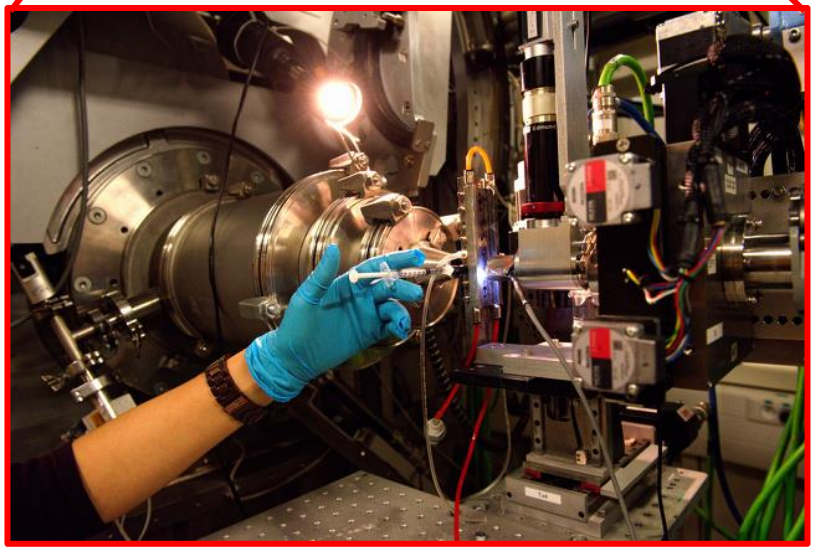
Full flux = 5.10^{12} ph/s @ 12 keV
Beam size (FWHM) =
400 (H) x 25-100 (V) µm^2

- **Structural Biology** (macromolecular shapes / low resolution structure)
- **Soft Condensed Matter** (crystal growth, colloids, polymers, liquid crystals, hierarchical systems, …)

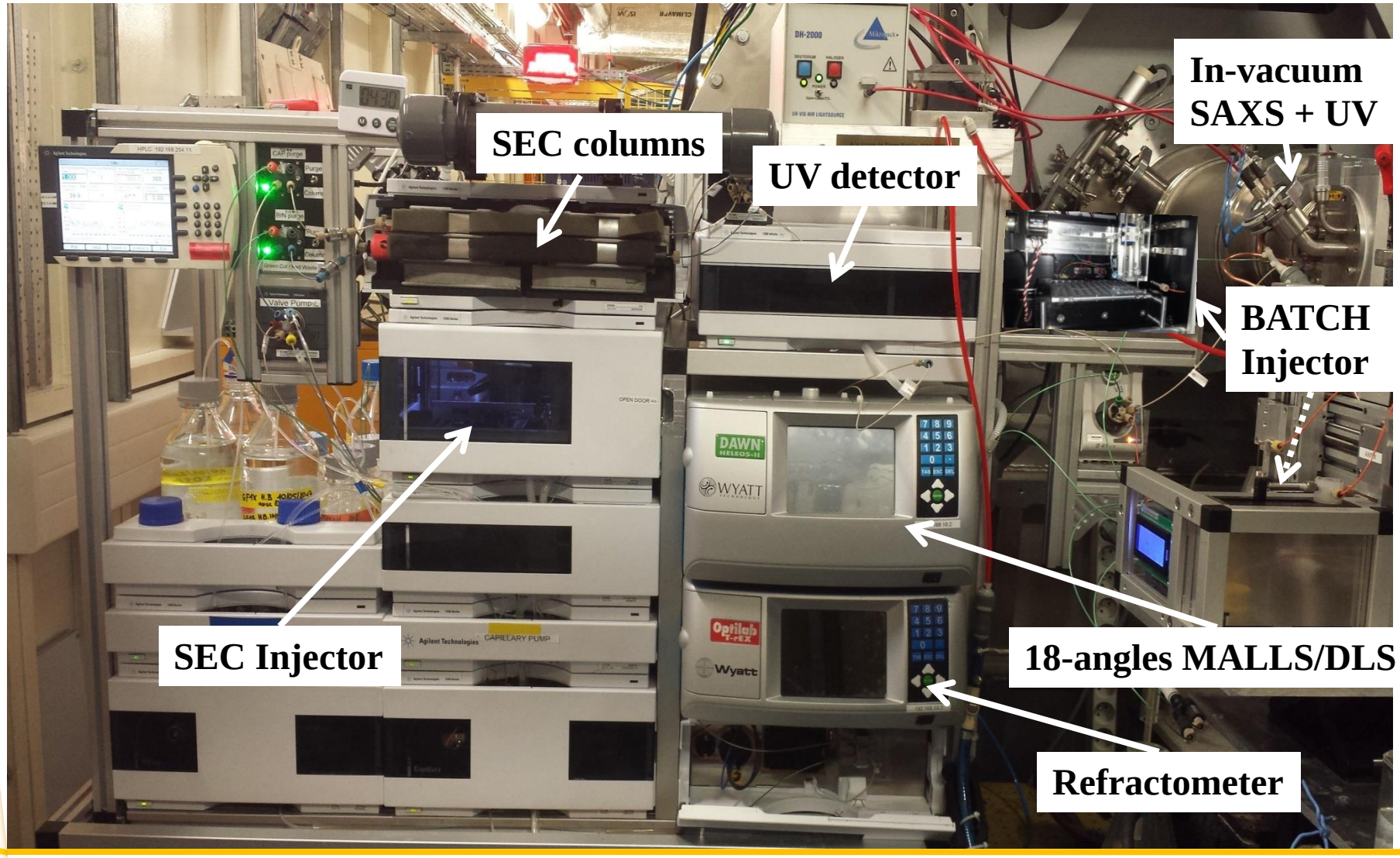


In vacuum detector
(from 500 to 6700 mm sample to detector distance)



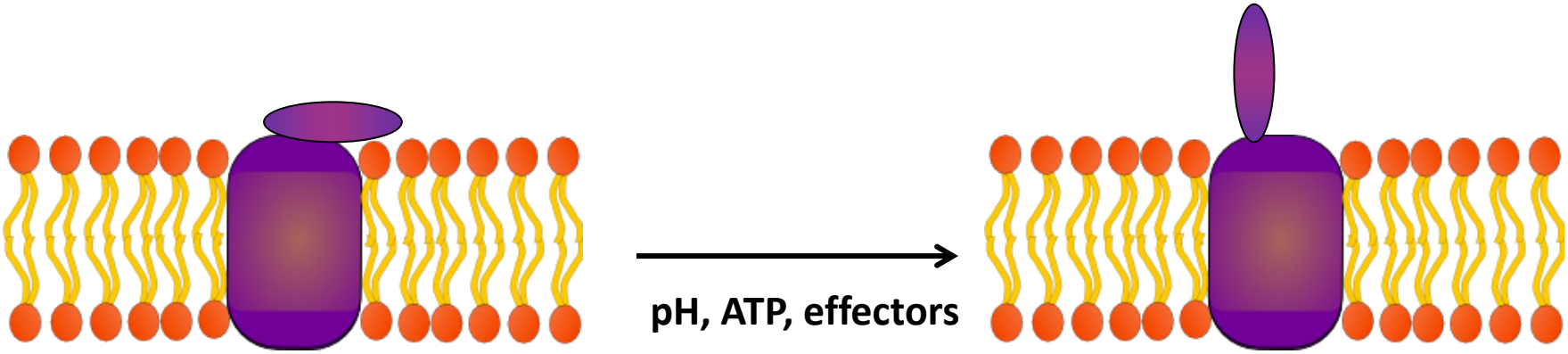


HPLC – MALLS/QELS – SAXS – RI online



Motivation: apply SAXS to Membrane proteins

- SAXS is good at monitoring conformation changes
- Membrane proteins undergo conformational changes

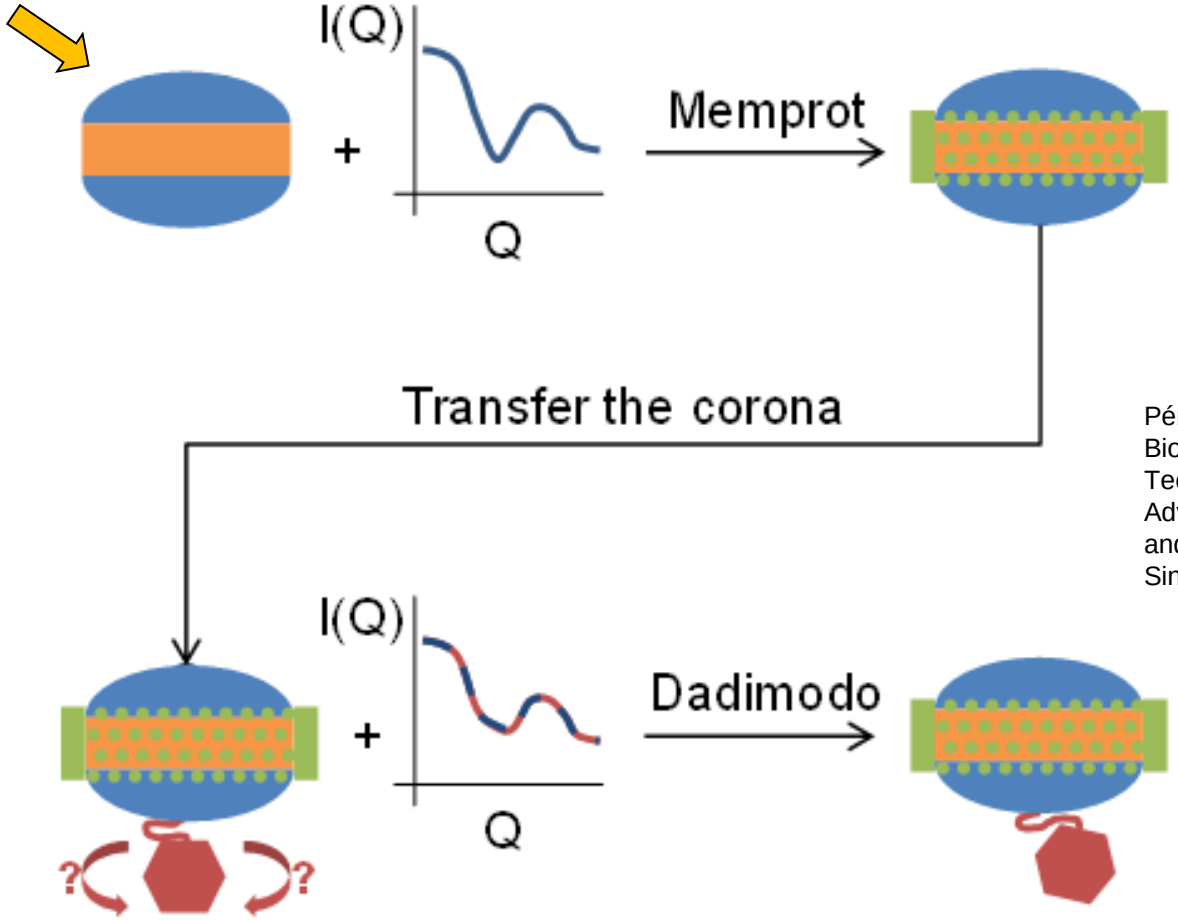


One conformation A **One conformation B**

- How can we use SAXS to monitor membrane proteins conformation changes ?
- How can we use SAXS with a membrane protein of known structure ?

Our approach: a two-step strategy

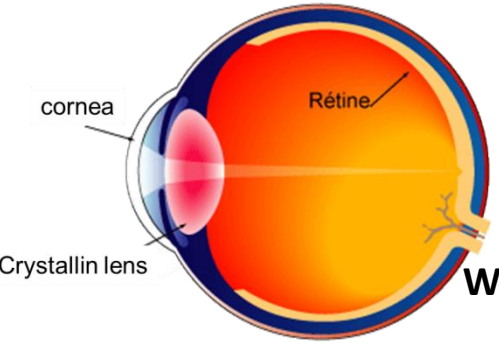
A construct of known structure is needed



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

Our starting model protein: Aquaporin-0

Crystalline lens (eye)

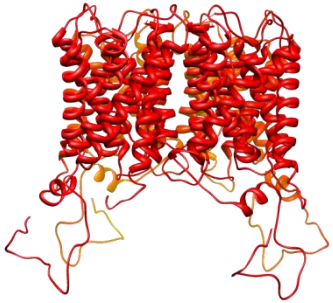


AQP0 (ex-MIP)
60 % of the membrane protein content

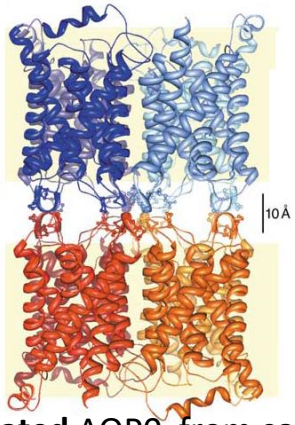
Natively tetramer

Water transport across cell membranes

✓ Two types of known existing states
✓ 3D already obtained



Full AQP0, from **cortex**
→ **Tetramer**

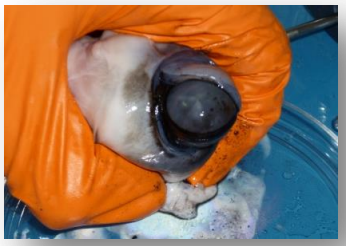


Truncated AQP0, from **core**
→ **Octamer**

Gonen et al., Nature 2004

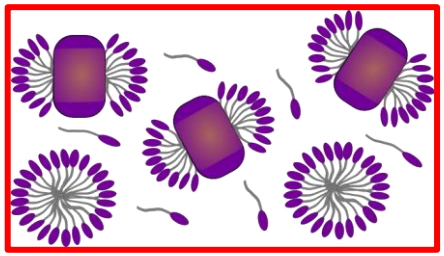
Purification of Full AQP0

- From bovine eye to lens membrane



- From lens membrane to AQP0 in solution
- Detergent:
  Dodecyl-β-D-maltopyranoside (DDM)

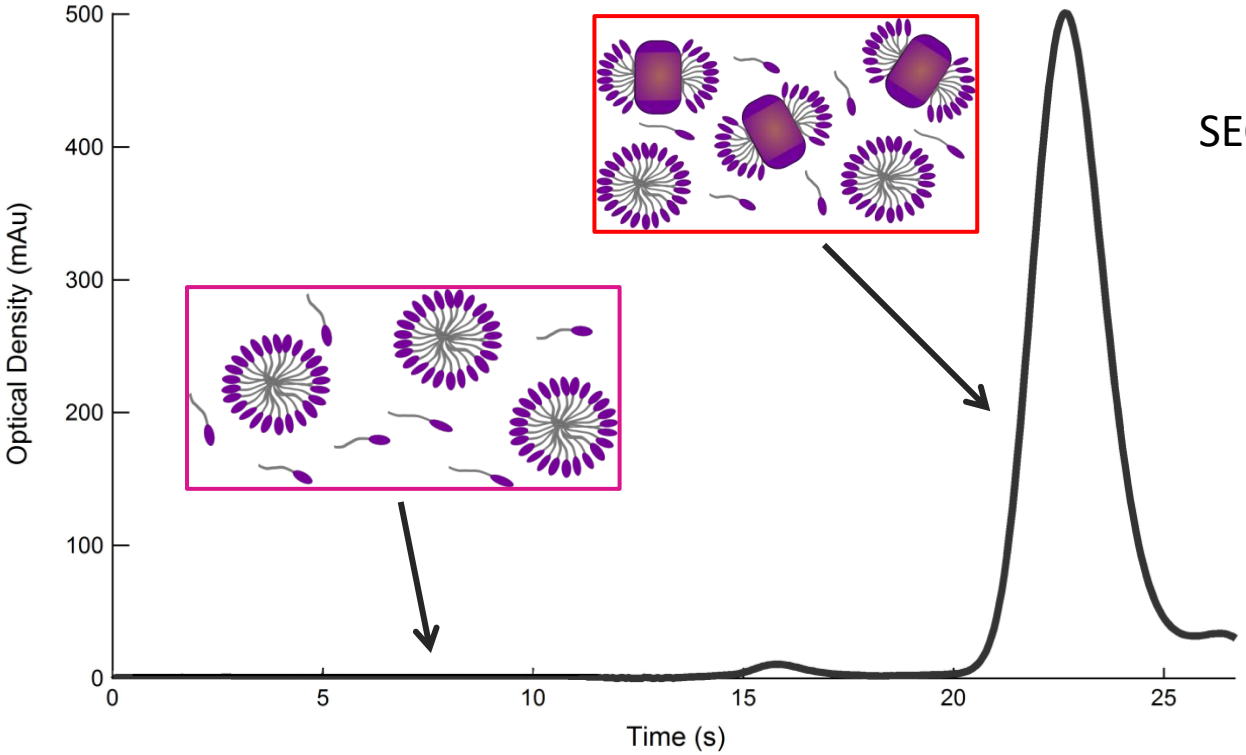
✓ Obtained concentration : **4 mg/ml** (2ml)



2 problems for SAXS:
- **Mixture**
- **Detergent belt**

SEC elution using a 0.2% DDM buffer

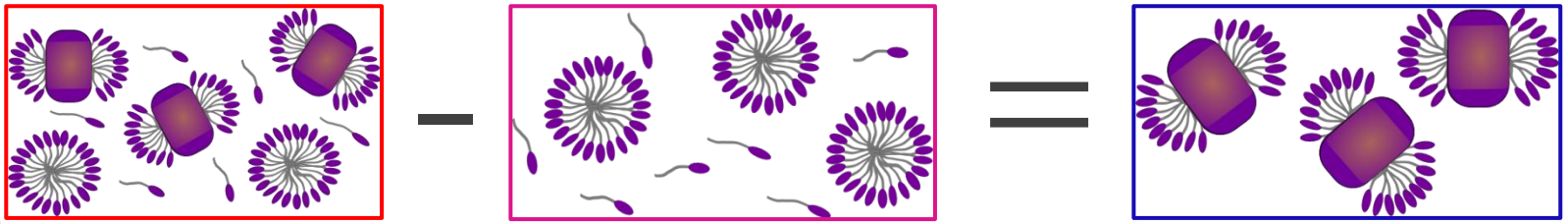
- **Mixture problem solved with the HPLC**



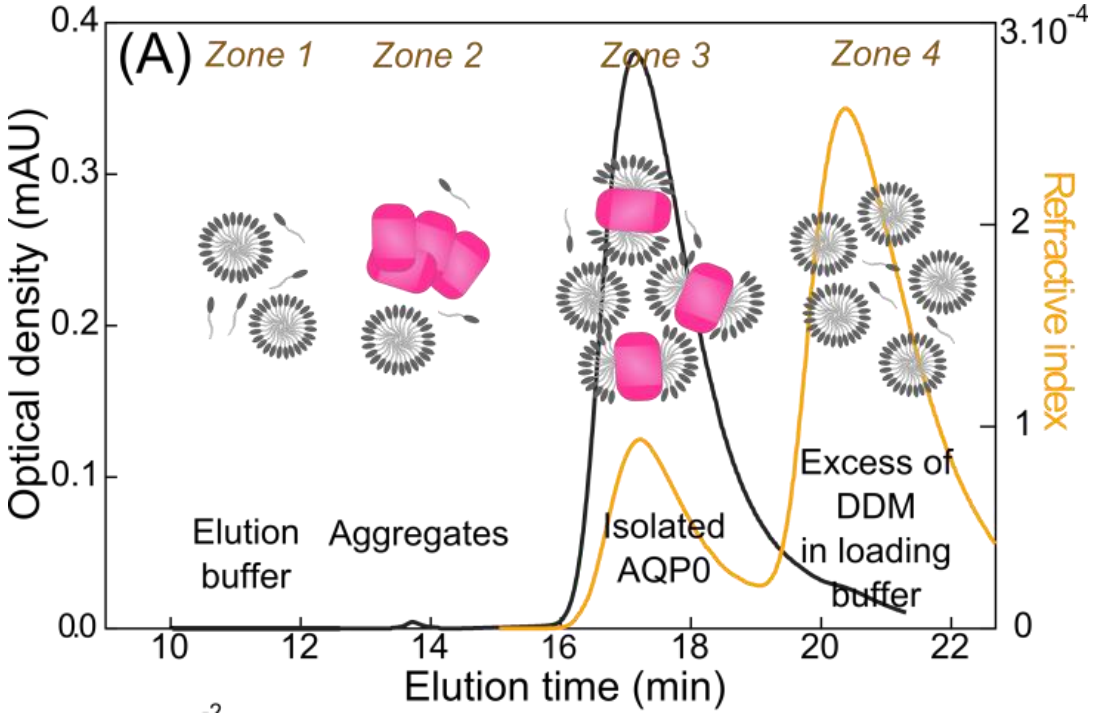
SEC-HPLC / SAXS combination

✓ Prevents the aggregates in the sample

✓ Subtraction of free micelles of detergent



SEC elution : O.D. (280nm) + R.I.

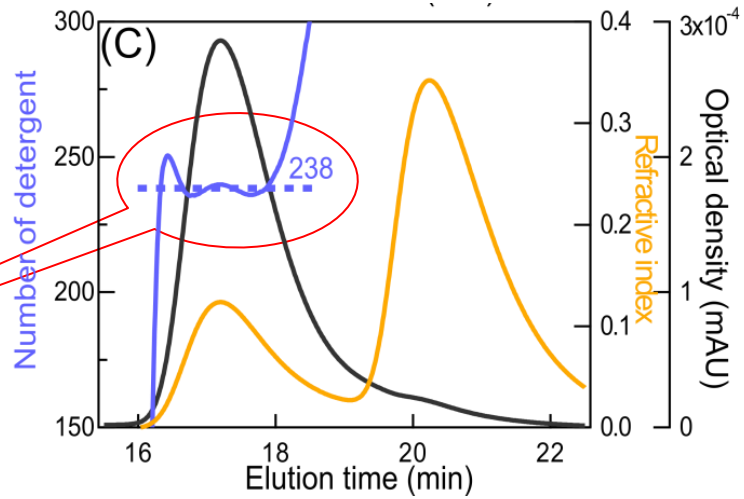


From Refractive Index & UV abs (280nm)

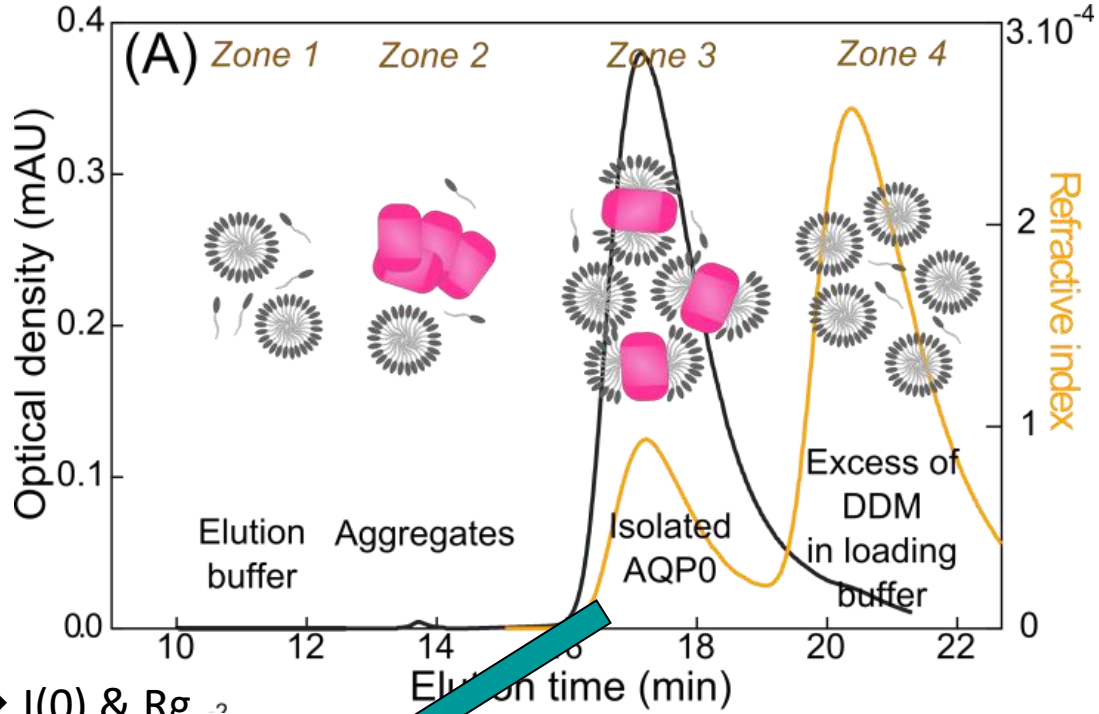
$$\varphi = \frac{OD \times (dn/dc)_{Det}}{\varepsilon_{AQP0}} \times \left[RI - \frac{OD \times [(dn/dc)_{Prot} - (dn/dc)_{Det}]}{\varepsilon_{AQP0}} \right]^{-1}$$

$$N_{Det} = \frac{1-\varphi}{\varphi} \times \frac{M_{Prot}}{M_{Det}}$$

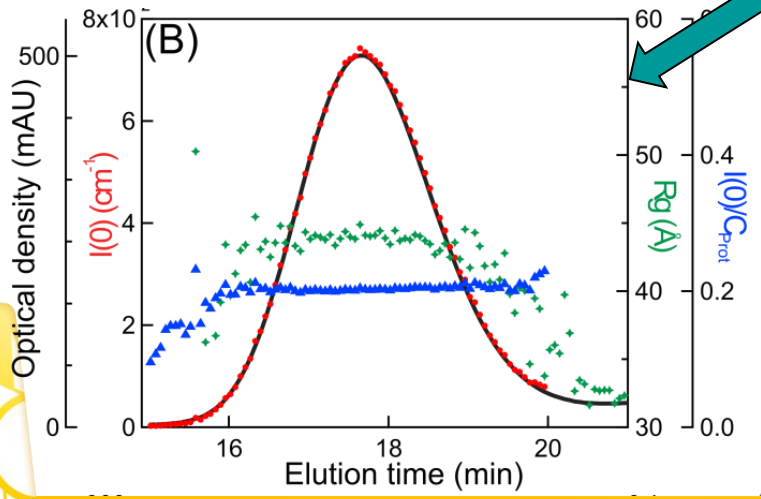
N_{Det} = 238 ± 15 molecules per protein



SEC elution : O.D. (280nm) + R.I. + SAXS



+ SAXS → I(0) & Rg



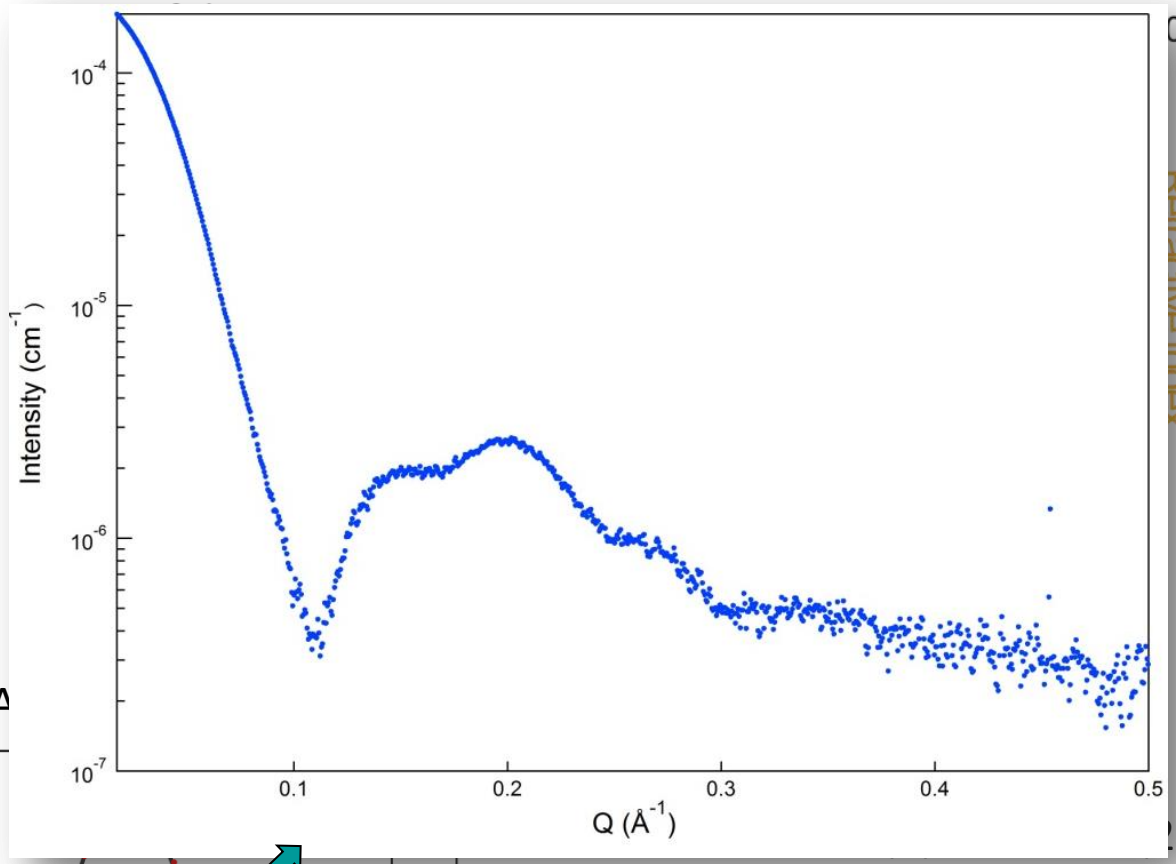
$$\frac{I(0)}{C} = M \cdot \frac{{r_e}^2}{N_A}\left(\frac{n_e N_A}{M} - \rho_0 \bar{v}\right)^2$$

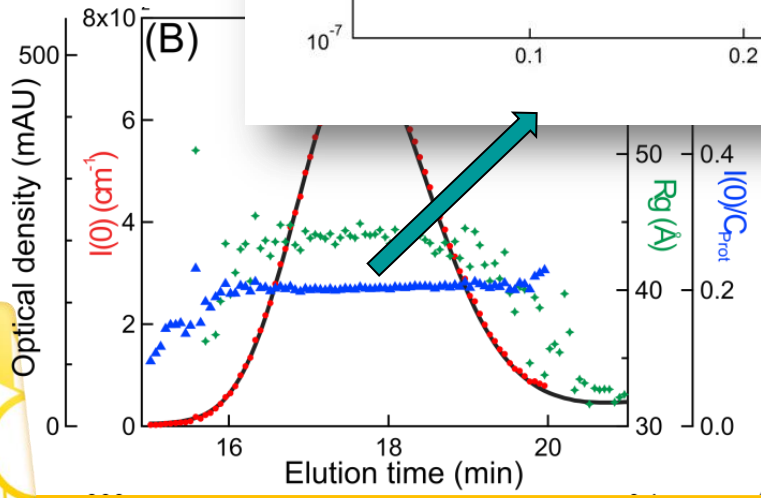
⇒ **M** is constant

⇒ No depletion of detergent

⇒ Monodisperse solution

SEC elution : O.D. (280nm) + R.I. + SAXS

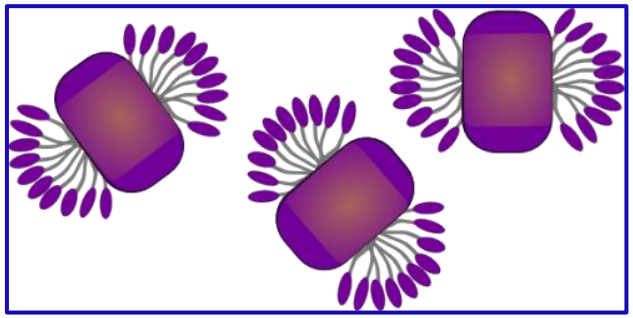




$$N_{Det} = \frac{\sqrt{\frac{I(0) M_{prot} N_A}{C_{prot} f^2}} - \left(n_{prot} N_A - \rho_S \bar{\nu}_{prot} M_{prot}\right)}{\left(n_{det} N_A - \rho_S \bar{\nu}_{det} M_{det}\right)}$$

N_{Det} = 225 ± 25 molecules per protein

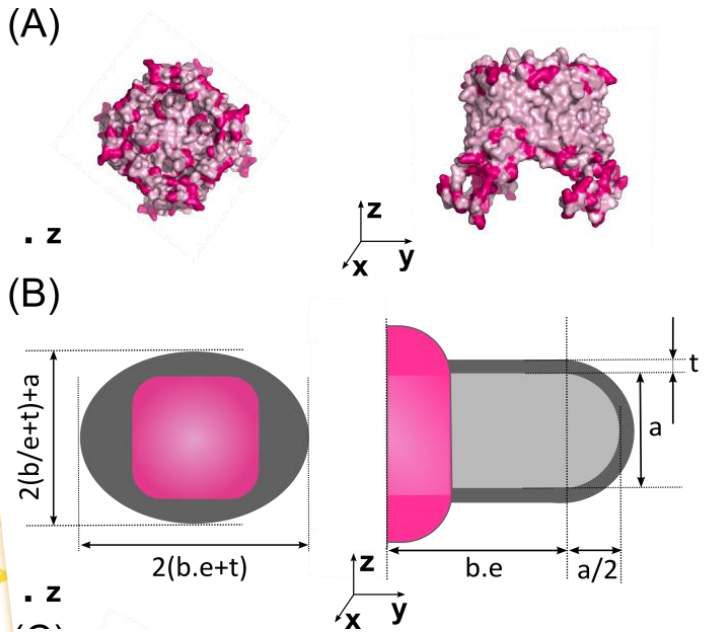
How to find a model which fits the curve ?



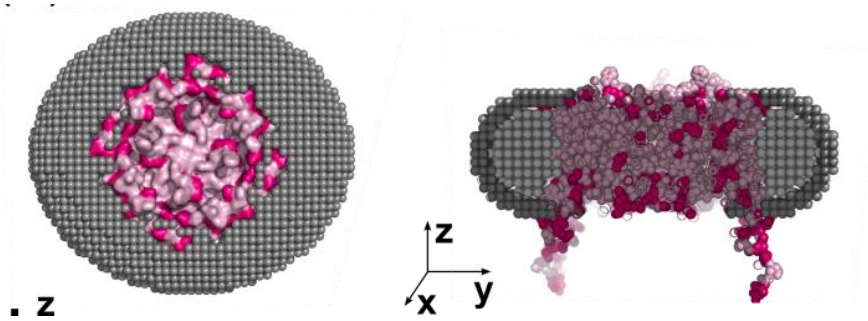
Several electronic densities : protein/detergent

→ Simple ~~Ab initio~~ method

A parametrized torus with two electronic densities



The torus volume is filled with beads.
The SAXS curve is calculated with CRYSOL



Beads « atoms » and grid parameters chosen for Crysol input :

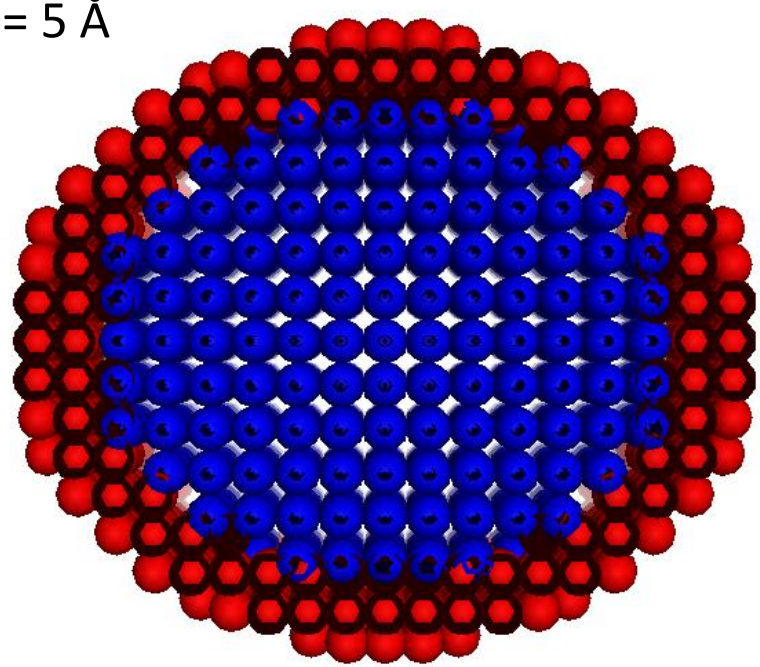
ρ_{tails} = 0.282 Å

ρ_{heads} = 0.520 Å

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

Validating the beads modeling

Core-shell ellipsoid
A = 22 Å
B = 18.2 Å
T = 5 Å



- Calculate with Crysol

ρ_0 = 0.334 Å
No hydration shell

- Fit using the analytical function from SASFit
- Check consistency

Beads « atoms » and grid parameters chosen for Crysol input :
ρ_{in} = 0.282 $e^-/Å^3$
ρ_{out} = 0.520 $e^-/Å^3$

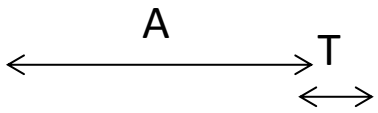
Validating the beads modeling

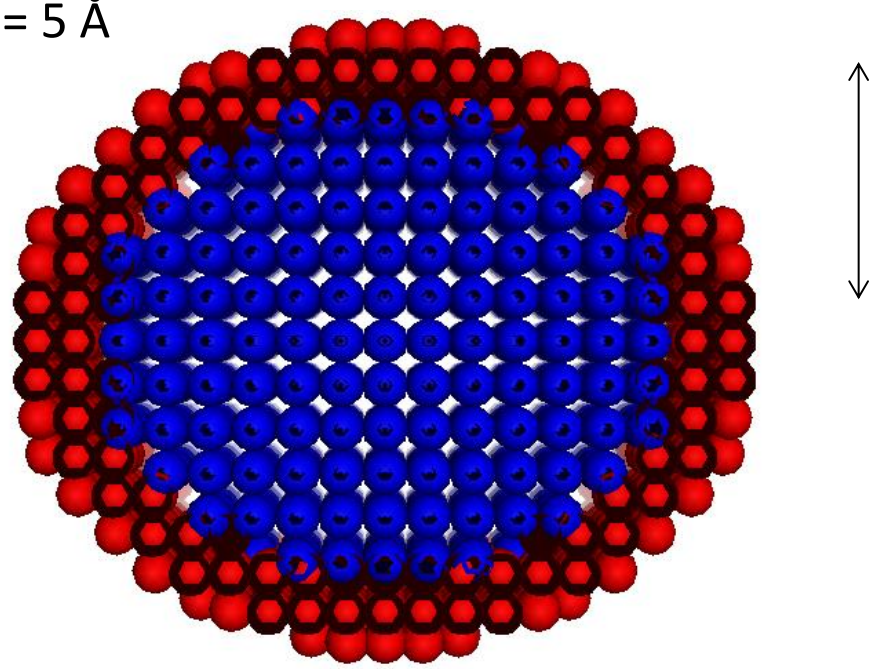
Core-shell ellipsoid

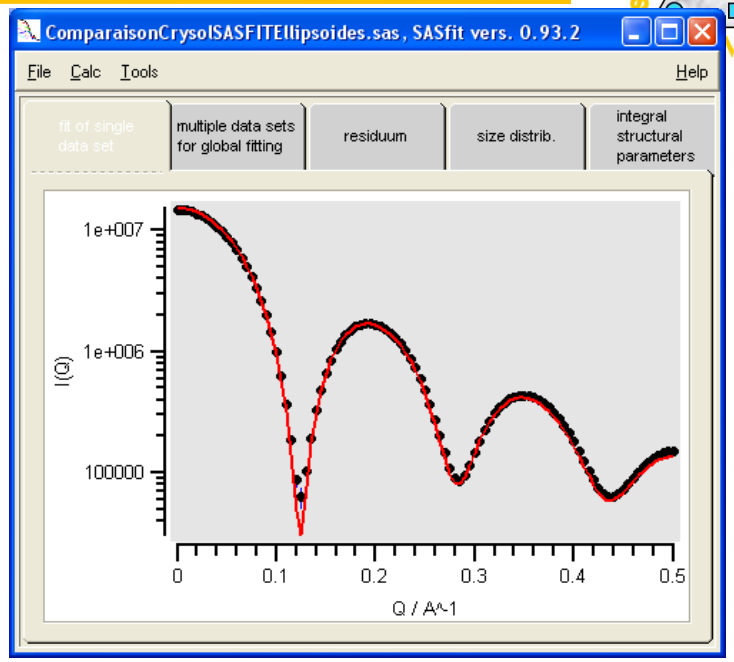
A = 22 Å

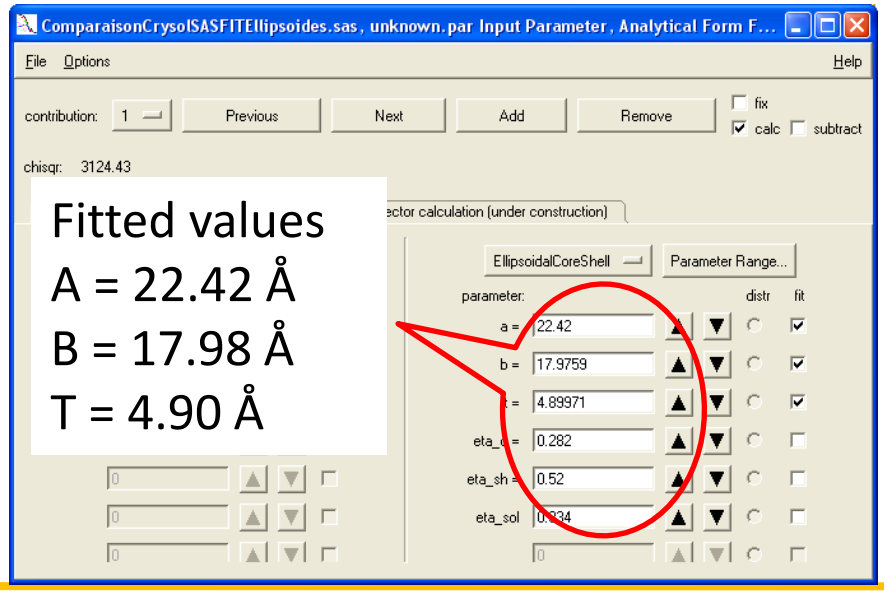
B = 18.2 Å

T = 5 Å









Fitted values

A = 22.42 Å

B = 17.98 Å

T = 4.90 Å

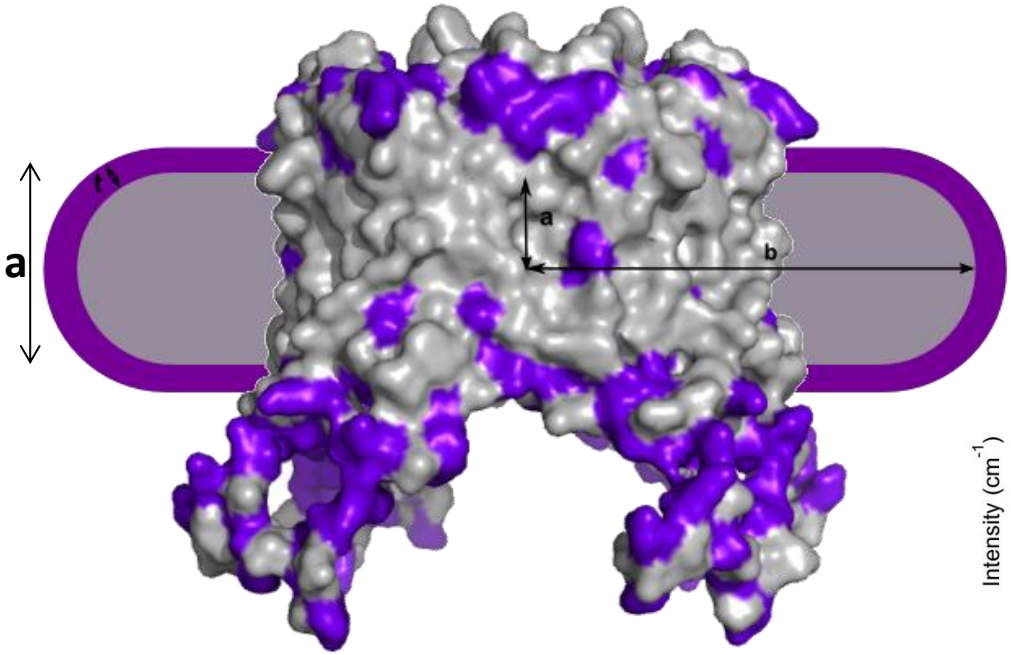
Beads « atoms » and grid parameters chosen for Crysol input :

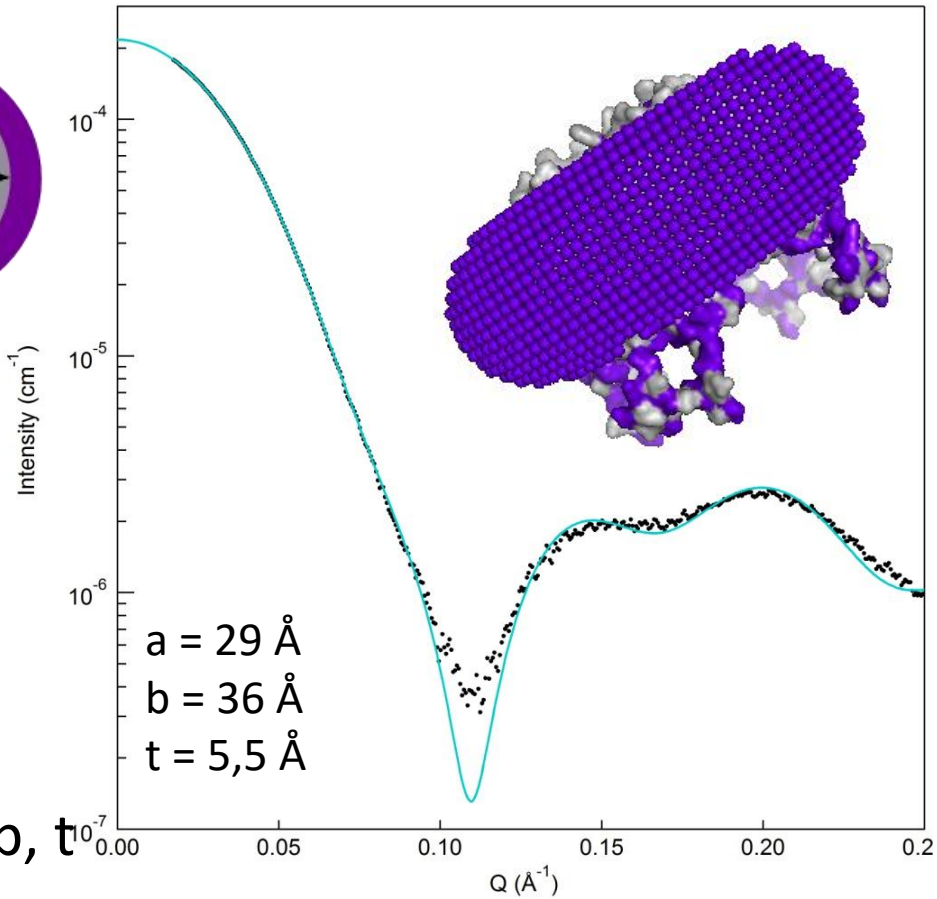
ρ_{in} = 0.282 $e^-/Å^3$

ρ_{out} = 0.520 $e^-/Å^3$

Validated

Circular detergent corona





Modelization of a circular torus of detergent

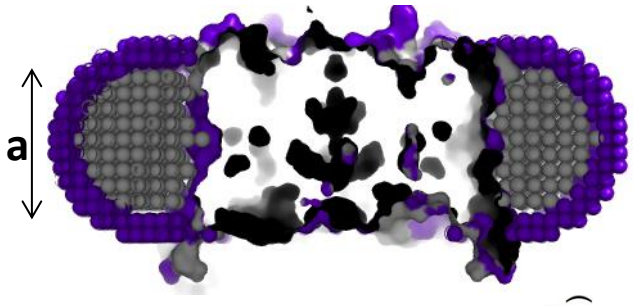
Three free geometric parameters, a, b, t

Elliptical detergent corona

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

Introduction of a parameter of ellipticity : e

a = 30 Å
b = 35 Å
t = 5.5 Å
e = 1.12

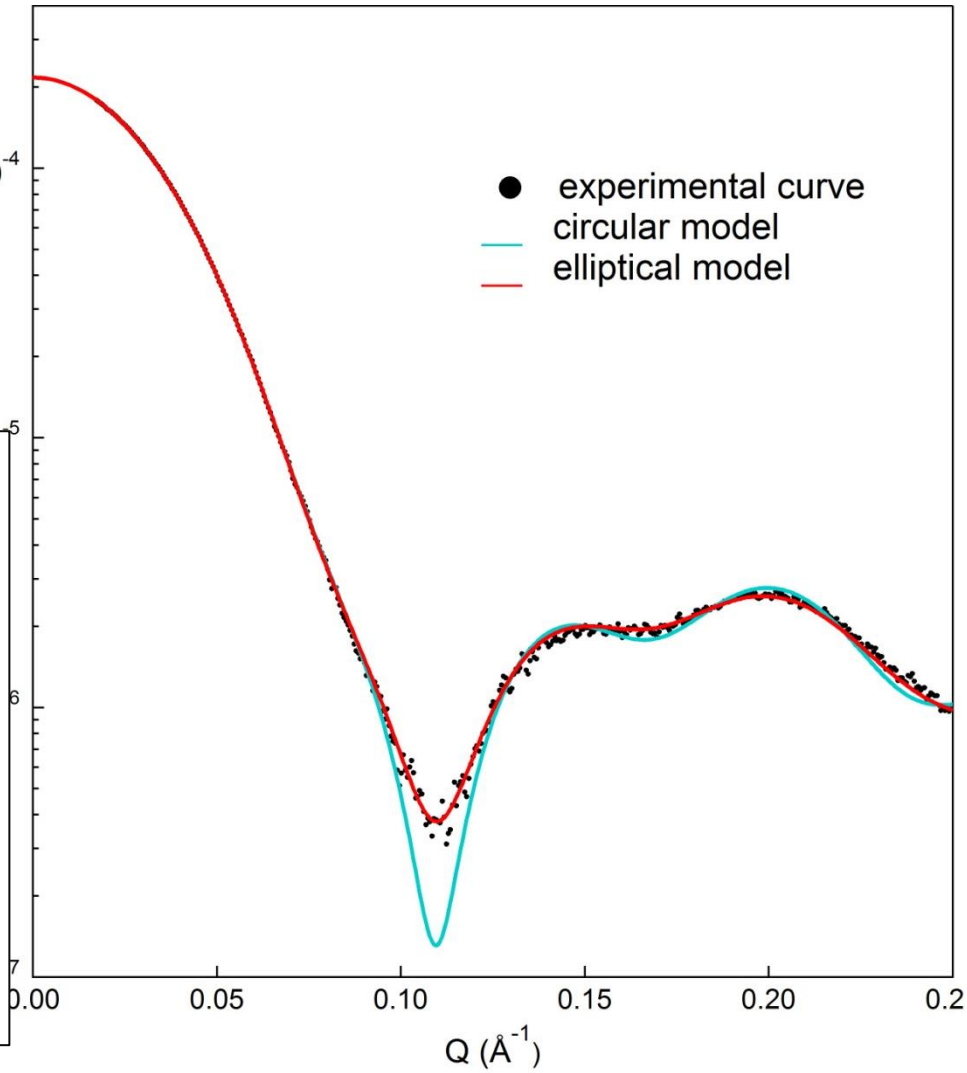


Relevancy of the model

Number of detergent molecules from the coarse grain model by volume's calculations:

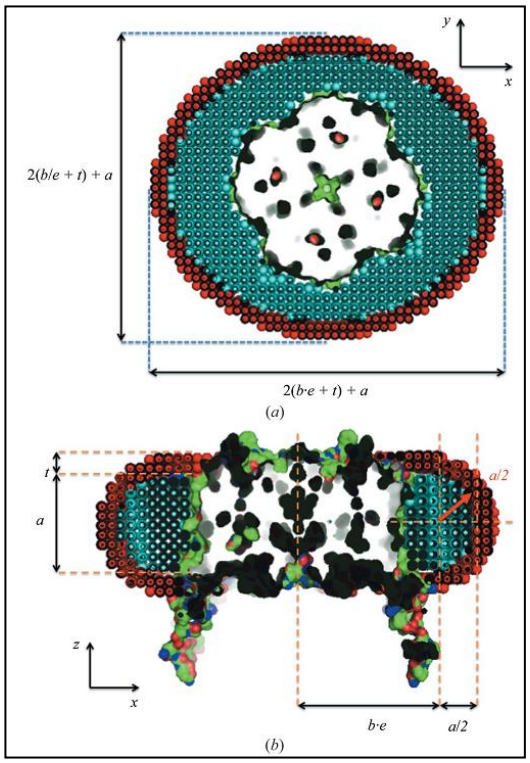
N = 270 ± 30 detergent molecules

⇒ Good agreement with previous values



Memprot : a program to generate/optimize the corona geometrical model

Pérez J. & Koutsioubas, A. (2015), *Acta Cryst., D71, 86-93*

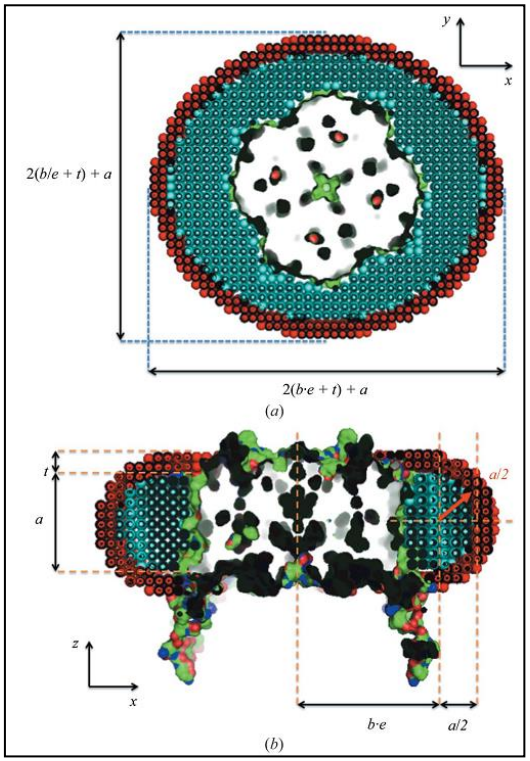



```
minimum_chi ← infinite
 for each a in the range [a_min,a_max], do
  for each b in the range [b_min,b_max], do
   for each t in the range [t_min,t_max], do
    for each e in the range [e_min,e_max], do
     for each phi in the range [phi_min,phi_max], do
      generate corona_model(a,b,t,e,phi)
      calculate chi (corona, protein pdb, experimental data) calling CRYSOL
      if minimum_chi > chi, then
       minimum_chi ← chi
 return chi
```

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.

Memprot : a program to generate/optimize the corona geometrical model

Pérez J. & Koutsioubas, A. (2015), *Acta Cryst., D71, 86-93*



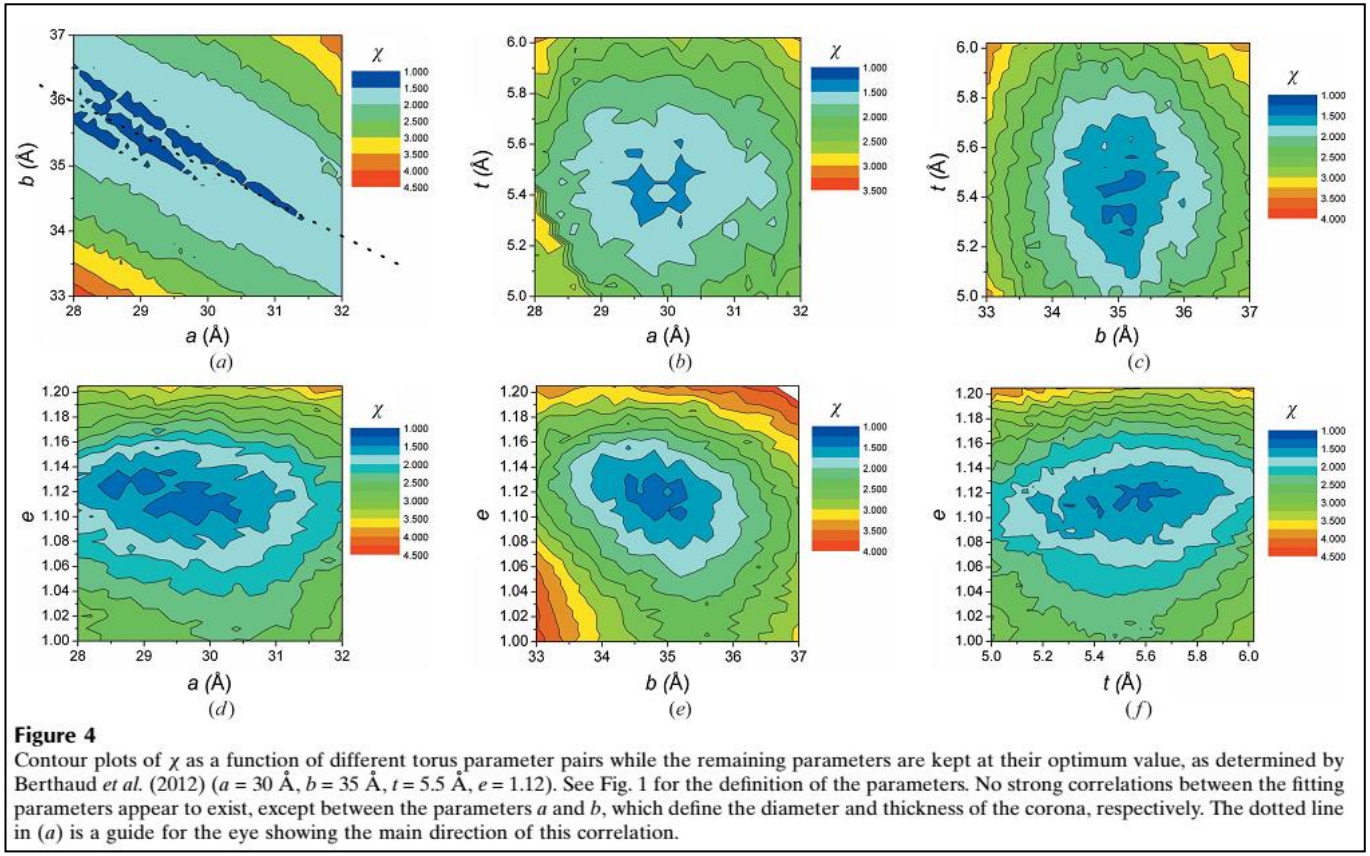
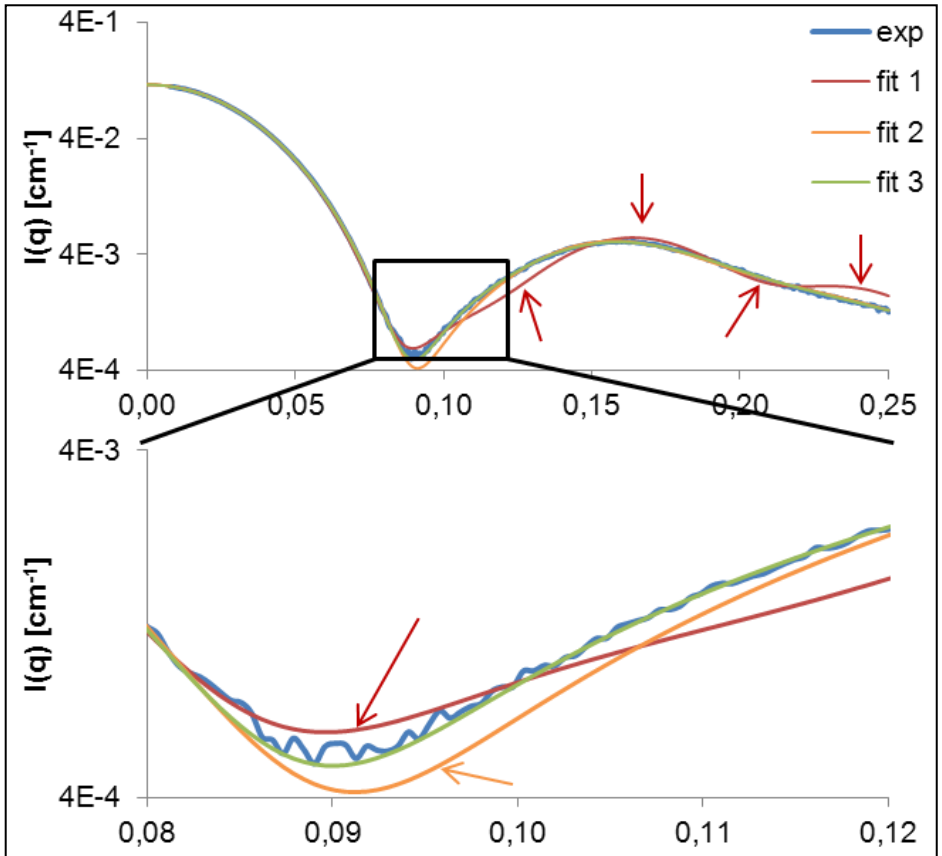


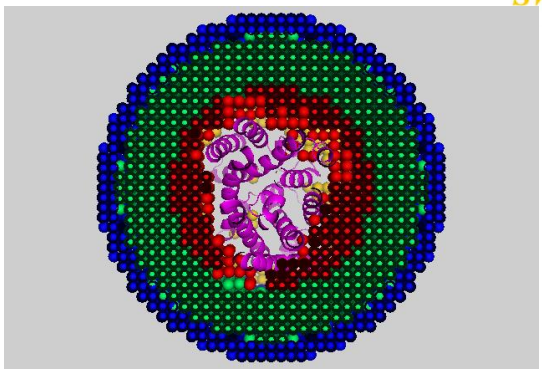
Figure 4
Contour plots of χ as a function of different torus parameter pairs while the remaining parameters are kept at their optimum value, as determined by Berthaud *et al.* (2012) (a = 30 Å, b = 35 Å, t = 5.5 Å, e = 1.12). See Fig. 1 for the definition of the parameters. No strong correlations between the fitting parameters appear to exist, except between the parameters a and b, which define the diameter and thickness of the corona, respectively. The dotted line in (a) is a guide for the eye showing the main direction of this correlation.

Example of fitting improvement steps on MHST protein

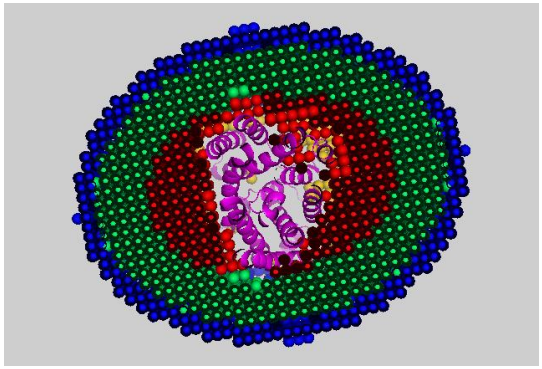
(coll: Poul Nissen, Aarhus University, Denmark)



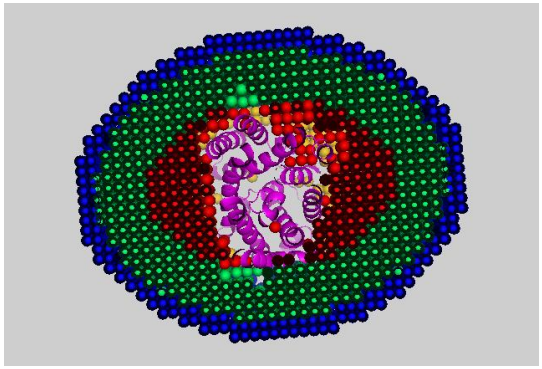
	a[Å]	b[Å]	t[Å]	e	rotation[deg]	χ^2	N_t	N_h
fit 1	28.0	28.0	5.0	1.00	0.0	83.807	230	215
fit 2	28.0	29.0	5.0	1.20	-10.0	6.398	255	220
fit 3	28.3	28.1	5.0	1.22	-5.0	1.937	237	219



Fit 1

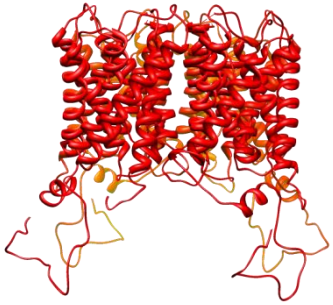


Fit 2

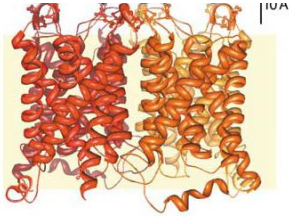


Fit 3

How far can we discriminate different protein constructs ?



Full Aqp-0
(2b6p)



Truncated Aqp-0
(2b6o)

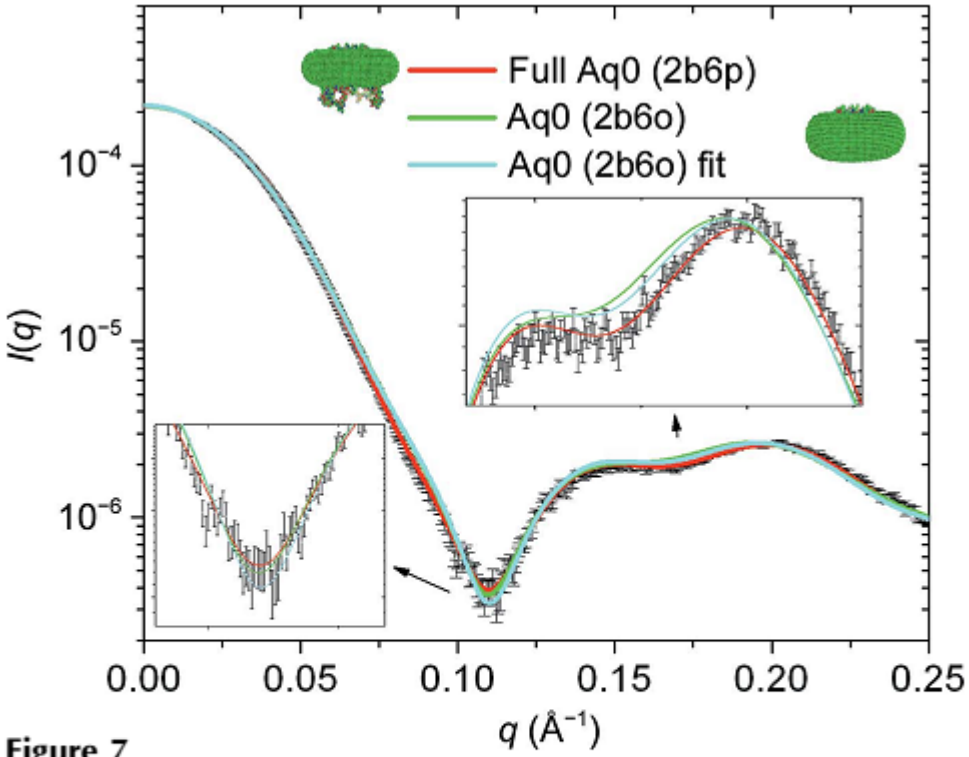
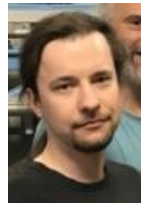


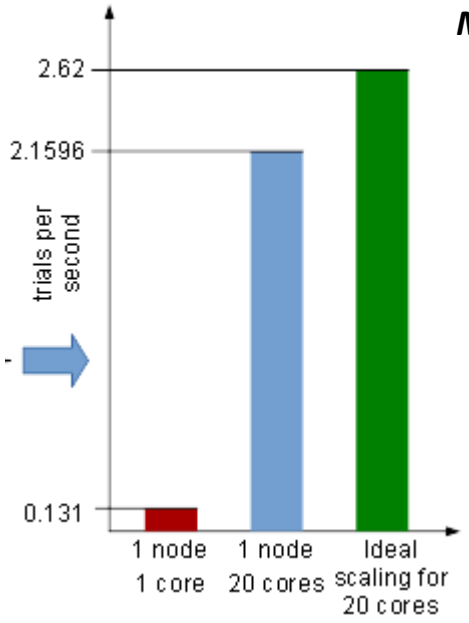
Figure 7
Scattering curves corresponding to corona parameters a = 29.6 Å, b = 35.4 Å, t = 5.6 Å, e = 1.12, e'_{heads} = 0.512 e Å^{-3}, e'_{tails} = 0.270 e Å^{-3}) for the full (2b6p) and truncated (2b6o) structures of aquaporin-0. The respective χ values are 1.31 and 3.79. The curve corresponding to an artificial optimized corona using the truncated form of aquaporin-0 is also plotted. The associated χ value is 3.47, which is still much higher than that for the complex based on the actual 2b6p structure.

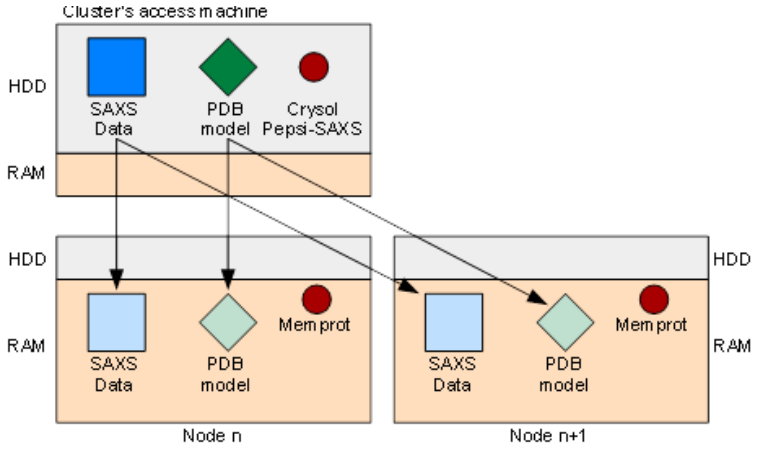
Memprot : Speed Improvements

- Typical Memprot runs range from thousands to hundreds of thousands of trials – the speed of calculations and scalability is an important issue
- We have implemented MPI-based, data-driven parallelization in Memprot to benefit from HPC clusters (here SOLEIL HPC)
- 449 residue protein MHST (PDB id 4us3) was used as a test case (sample provided by collaborators and measured at SWING)
- To prevent saturation of the cluster's network, Memprot stores everything (experimental data, protein's PDB model, intermediate files) locally in the node's RAM, utilising /dev/shm partition

M. Baranowski







Example of log file after a Memprot run

```
7053 ----------------+-------------------------------------------------------------------------------
7054 Best model #   6272
7055  a =  28.000
7056  b =  28.300
7057  t =   5.000
7058  e =   1.200
7059  r =   0.000
7060  d =   1.000
7061  chi^2 =     1.952
7062  Total pseudo atoms =    7436
7063  Total hydrophobic pseudo atoms =    3564
7064  Total hydrophylic pseudo atoms =    3872
7065  Vexcl{fit}/Vexcl{calc} (alpha) =  1.045
7066  Final ratio of tails to heads (TOH) = 1.307
7067  Initial electron density of hydrophobic part    =  0.270
7068  Calculated electron density of hydrophobic part =  0.245
7069  Initial electron density of hydrophilic part    =  0.540
7070  Calculated electron density of hydrophilic part =  0.508
7071  Number of detergent hydrophobic tails calculated =     285
7072  Number of detergent hydrophilic heads calculated =     218
7073
7074 No model found below cutoffs.
7075
7076 -------------------------------------------------------------------------------------------
7077  TIMINGS:
7078  Memprot took                                2174.44(   100.0%) seconds to complete, out of which:
7079  |-Building models took                       106.65(     4.9%) seconds to complete, out of which:
7080  | |-Adaptive Shape Algorithm took              0.00(     0.0%) seconds to complete, and:
7081  |-Crysol calls took                         1998.71(    91.9%) seconds to complete.
7082  seconds per trial   :          0.3117
7083  sec. per 1k trials  :        311.7472
7084  min per 1k trials   :          4.9880
7085  hours per 1k trials :          0.0000
7086  trials per second   :          3.2077
7087  trials per hour     :      11547.8167
7088  trials per day      :     277147.6004
7089 -------------------------------------------------------------------------------------------
7090  End of calculation!
```

DADIMODO : <u>All-atom</u> modeling of flexible parts *vs.* SAXS / NMR data

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

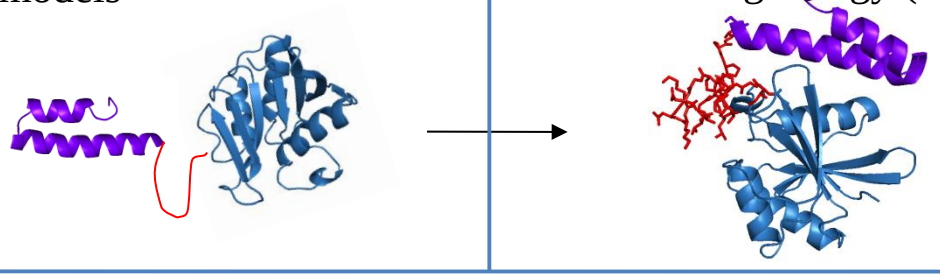
F. Mareuil, et al. (2007) *Eur Biophys J.*

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

Modelling approach : complete atomic model

Full structure initiated with :

- Crystal or NMR domain structures
- Homology models



Prior knowledge:

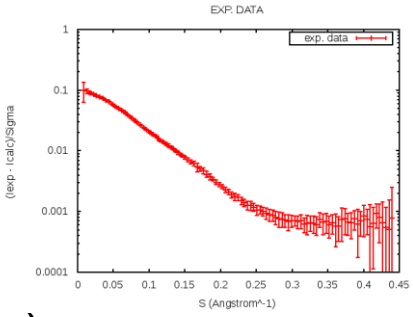
- Sequence
- Sub-parts moved as rigid-bodies (user-defined)
- A correct stereochemistry is maintained at all steps by minimizing energy (Amber 99 Force Field)

Experimental data:

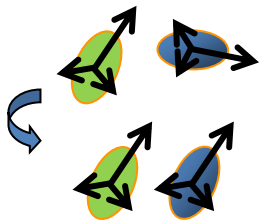
- SAXS
- NMR

  RDC

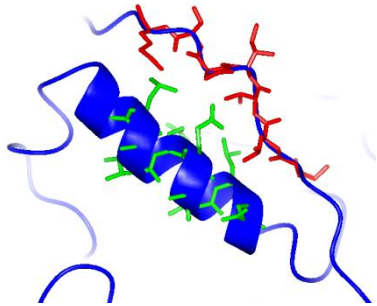
  ADR (chem. shift map.)



SAXS score



RDC score



ADR score

Optimisation of the structure via a genetic algorithm

Dadimodo: refining domain orientations with an all-atom description

- **Initial (slow) version** : **Evrard et al. (2011),** *J. Appl. Cryst.*, **44**:1264-1271.

- **Current (faster) version : O. Roudenko , A. Thureau, J. Pérez**

  - **Parallel implementation of the genetic algorithm**
    - 7300 Atoms → 7 hours on a 20 processor node (200 generations)

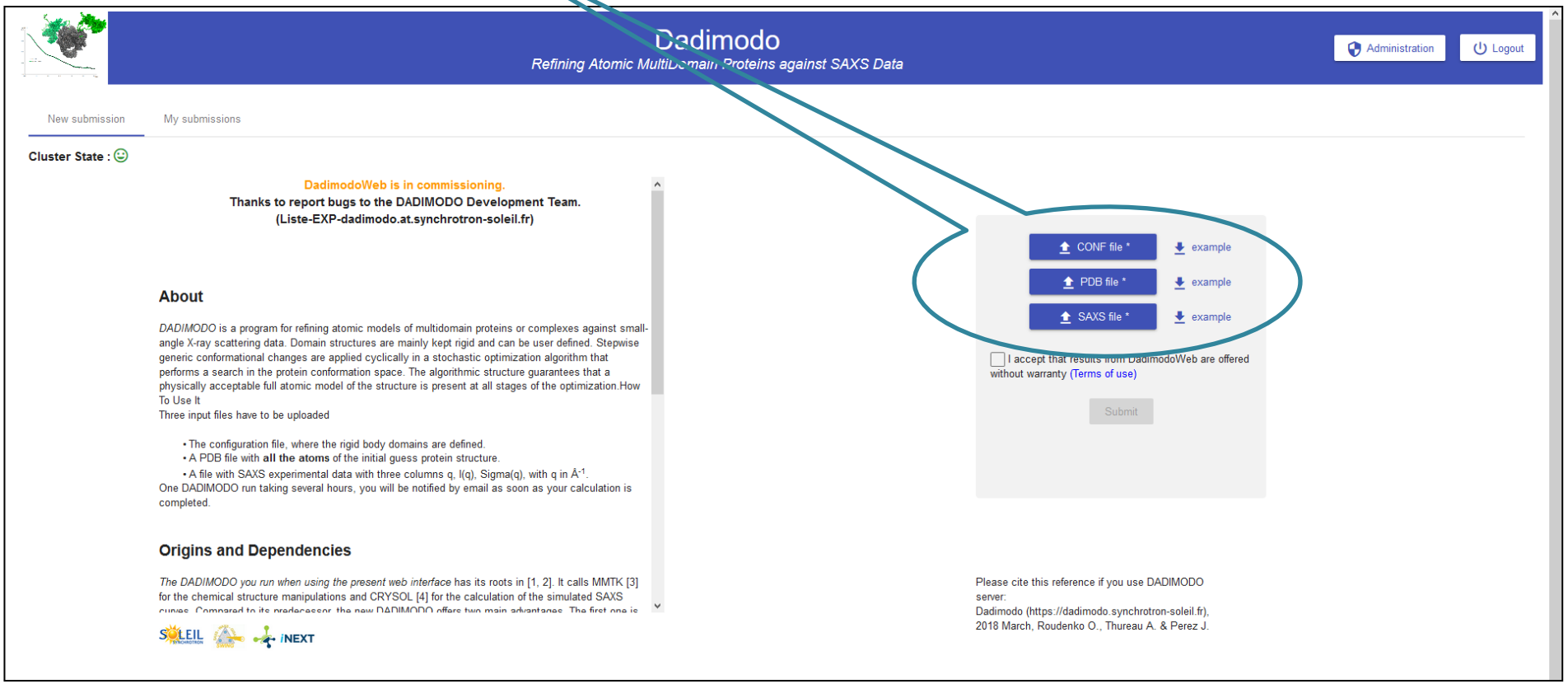
  - **User-friendly input**
    - Tools for completion of pdb input files (if needed)
    - User-defined topology : Pdb file + rigid bodies definitions

  - **Web server since end 2018**
    - Accessible to external users (after login in Soleil DB)
    - Five independent runs launched in parallel

Dadimodo Web Server

https://dadimodo.synchrotron-soleil.fr

3 input files needed to launch Dadimodo on the Web Server

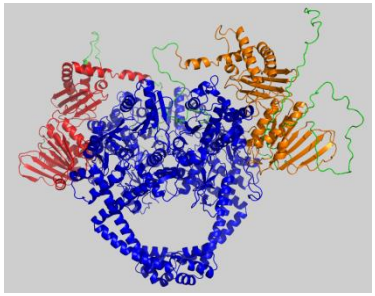


Dadimodo Web Server

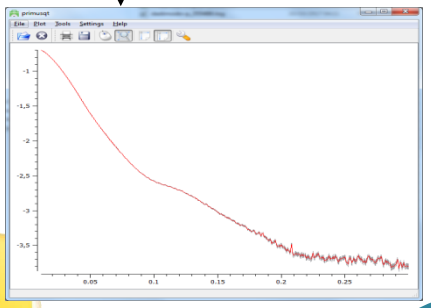
https://dadimodo.synchrotron-soleil.fr

3 input files needed to launch Dadimodo on the Web Server

Complete PDB file



SAXS data



```
########################## USER DEFINED PARAMS ##########################

[structure] ; ------------------ STRUCTURE ------------------

# Define here the rigid bodies in the same nomenclature as your pdb
# Syntax: rigid_body =  list of expressions 'chain: first_residue - last_residue' separated by comma
#
body1 = A: 25-104, A: 123-330, A: 345-421
body2 = A: 453-1179, B: 455-1179
body3 = B: 35-82, B: 125-231, B: 249-426

[experimental data] ; ------------- EXPERIMENTAL DATA -------------


# The saxs curve will be fitted between the following Q values:
new_q_min =  0.0112 ;  default = 0.001
new_q_max = 0.30 ;    default = 0.40
#
# This interval will be automatically truncated if its bounds
# lay outside your experimental Q-range

# Weights for experimental data.
saxs_weight = 1. ;              # weight for SAXS data in chi squared
adr_weight = 0. ;               # weight for ADR data in chi squared
rdc_weight = 0. ;               # weight for RDC data in chi squared


########################## EXPERT USER  ##########################


[optimization] ; --------------- OPTIMIZATION PARAMS ------------

mut_sigma = 40 ;              # mutation radius (degrees)
```

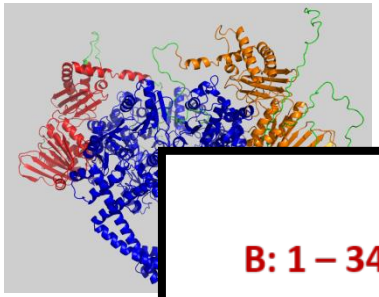
Configuration file

Dadimodo Web Server

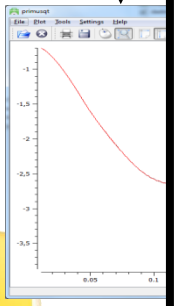
https://dadimodo.synchrotron-soleil.fr

3 input files needed to launch Dadimodo on the Web Server

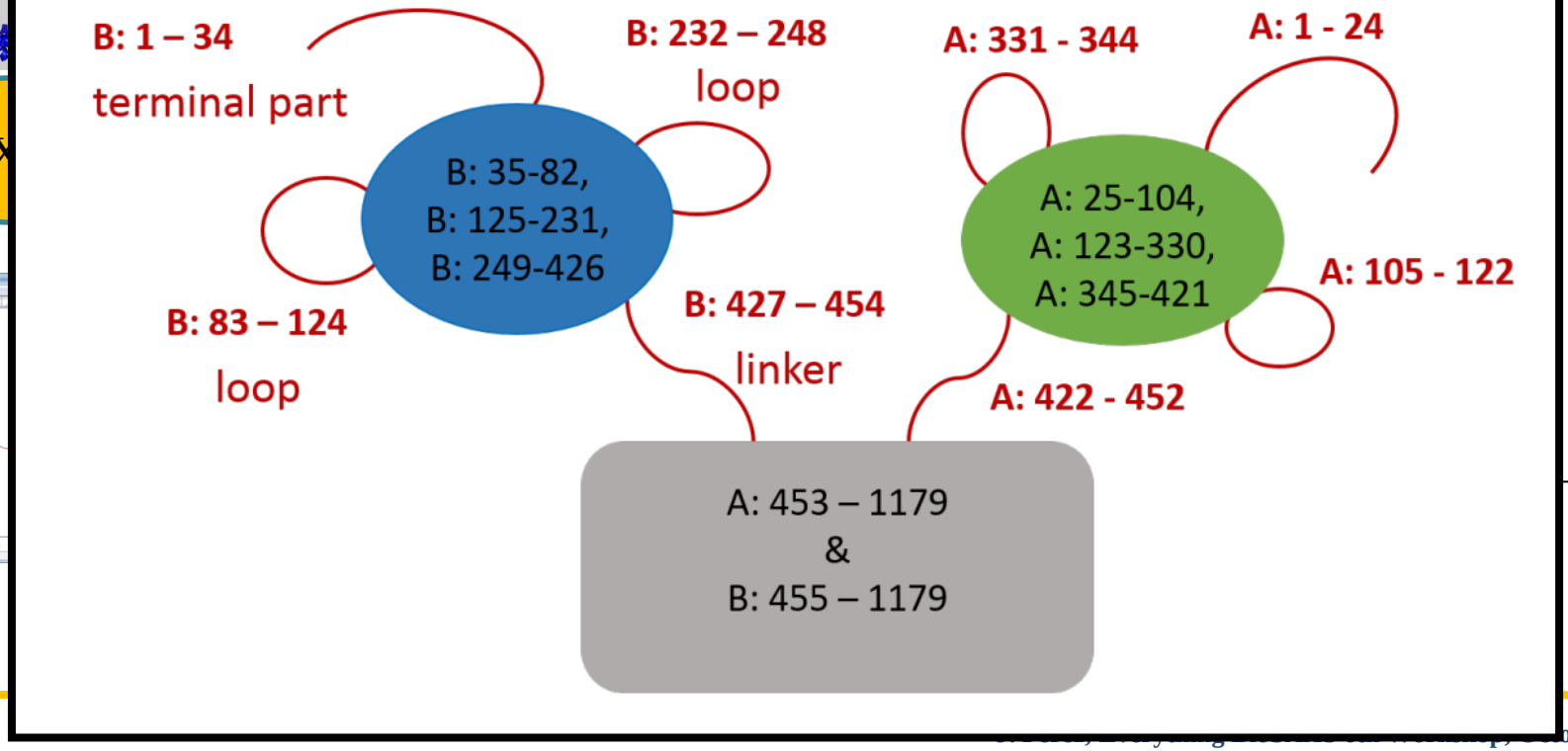
Complete PDB file



SAX



```
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
```



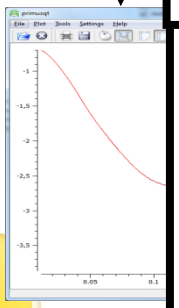
Dadimodo Web Server

https://dadimodo.synchrotron-soleil.fr

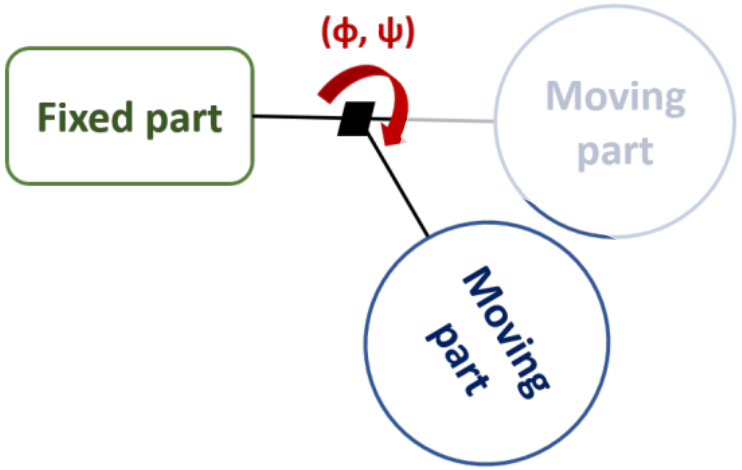
3 input files needed to launch Dadimodo on the Web Server

Complete

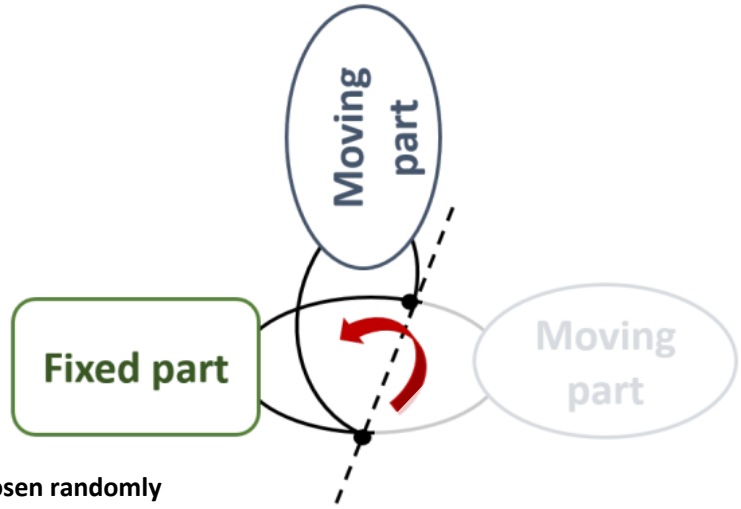
SAX



Linker (ϕ, ψ) mutation



Double linker (axis) mutation



Mutated residue is chosen randomly

B: 83 – 124

loop

B: 427 – 454

linker

A: 422 - 452

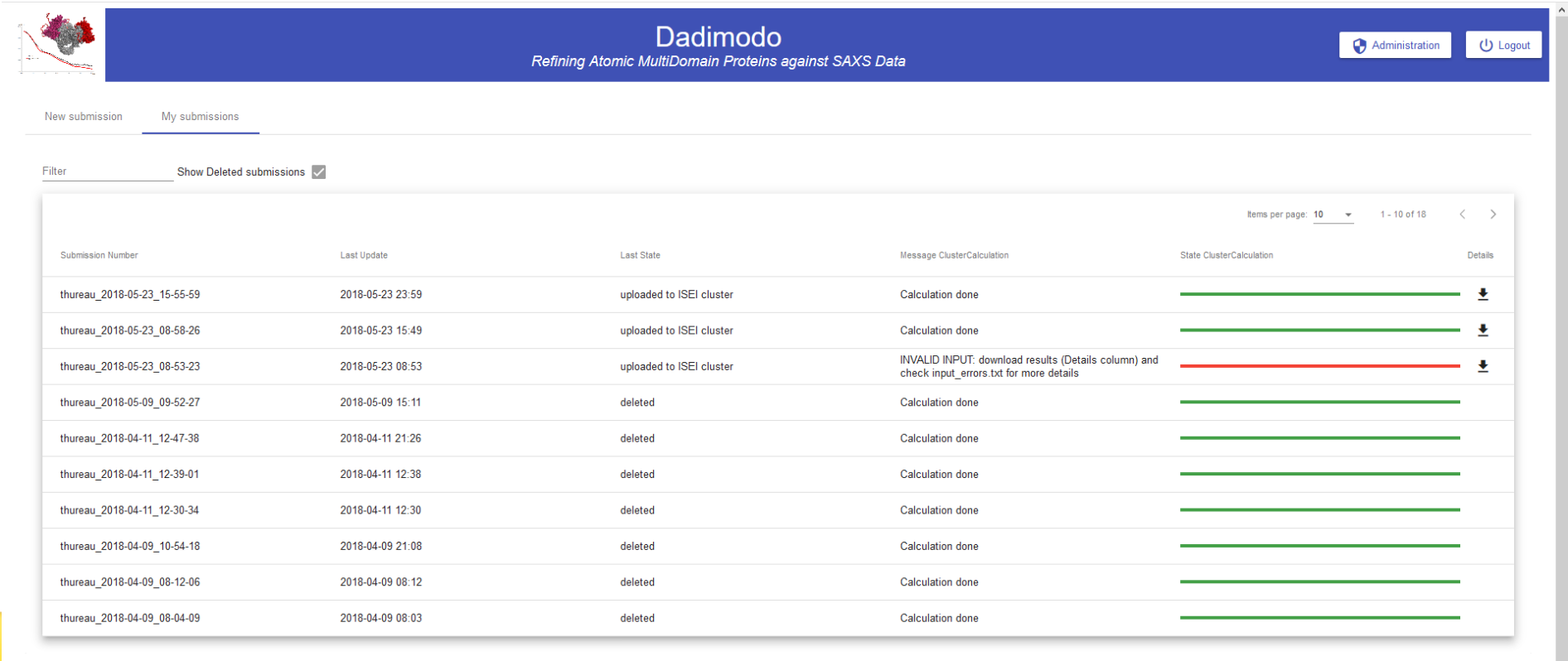
A: 453 – 1179
&
B: 455 – 1179

Dadimodo Web Server

https://dadimodo.synchrotron-soleil.fr

« My submissions » tab:

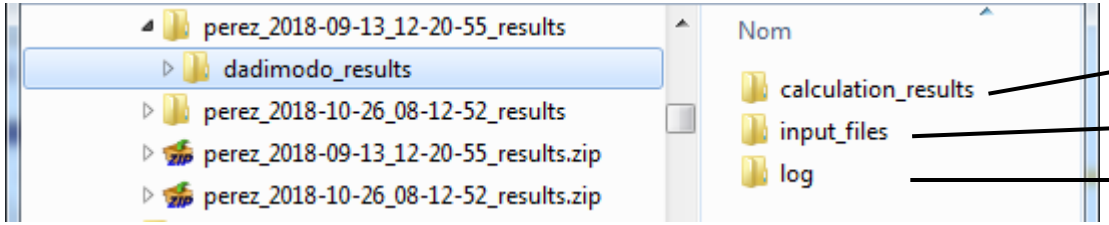
- ***Status of current submission and history of past jobs***
- ***Results download (zip file)***



Submission Number	Last Update	Last State	Message ClusterCalculation	State ClusterCalculation	Details
thureau_2018-05-23_15-55-59	2018-05-23 23:59	uploaded to ISEI cluster	Calculation done		
thureau_2018-05-23_08-58-26	2018-05-23 15:49	uploaded to ISEI cluster	Calculation done		
thureau_2018-05-23_08-53-23	2018-05-23 08:53	uploaded to ISEI cluster	INVALID INPUT: download results (Details column) and check input_errors.txt for more details		
thureau_2018-05-09_09-52-27	2018-05-09 15:11	deleted	Calculation done		
thureau_2018-04-11_12-47-38	2018-04-11 21:26	deleted	Calculation done		
thureau_2018-04-11_12-39-01	2018-04-11 12:38	deleted	Calculation done		
thureau_2018-04-11_12-30-34	2018-04-11 12:30	deleted	Calculation done		
thureau_2018-04-09_10-54-18	2018-04-09 21:08	deleted	Calculation done		
thureau_2018-04-09_08-12-06	2018-04-09 08:12	deleted	Calculation done		
thureau_2018-04-09_08-04-09	2018-04-09 08:03	deleted	Calculation done		

Results

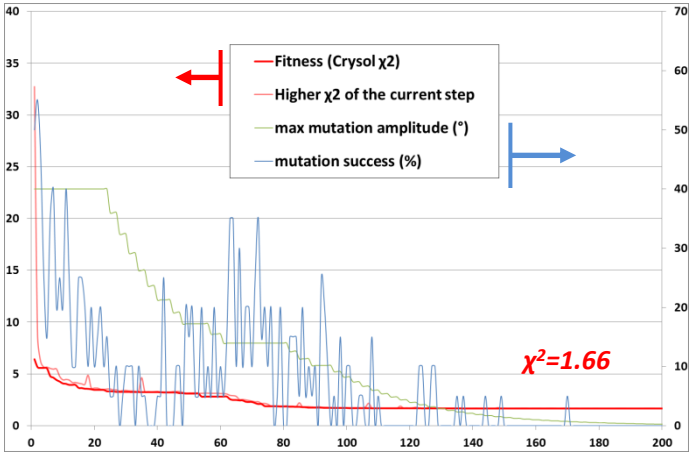
https://dadimodo.synchrotron-soleil.fr

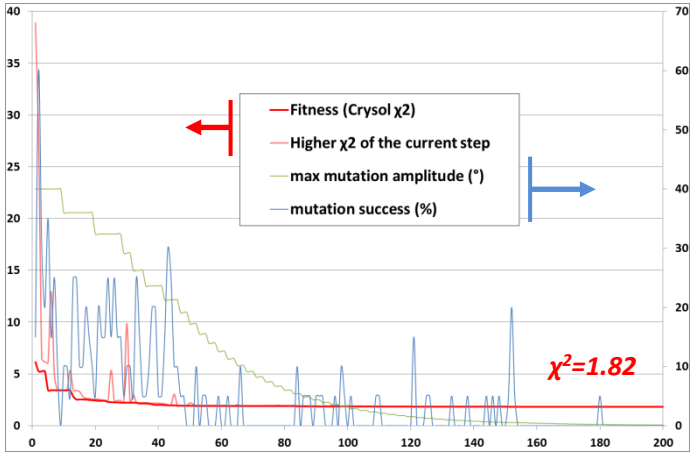


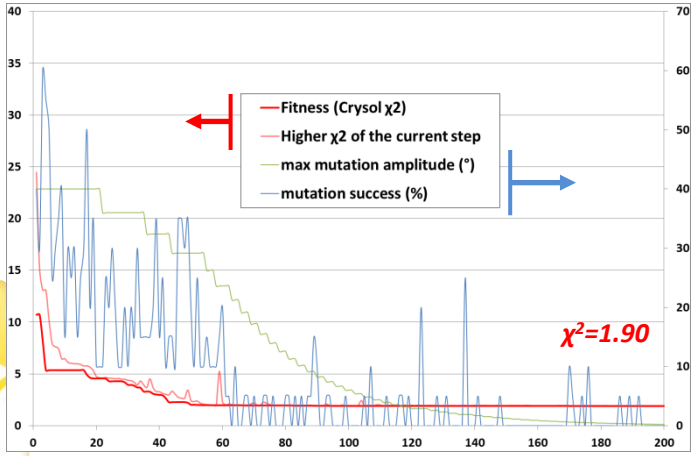
The final pdb files from the several runs

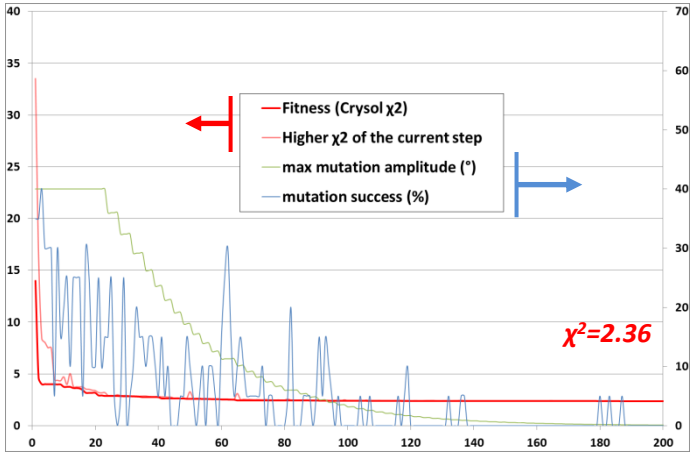
The 3 input files

The exhaustive log files + the statistics files









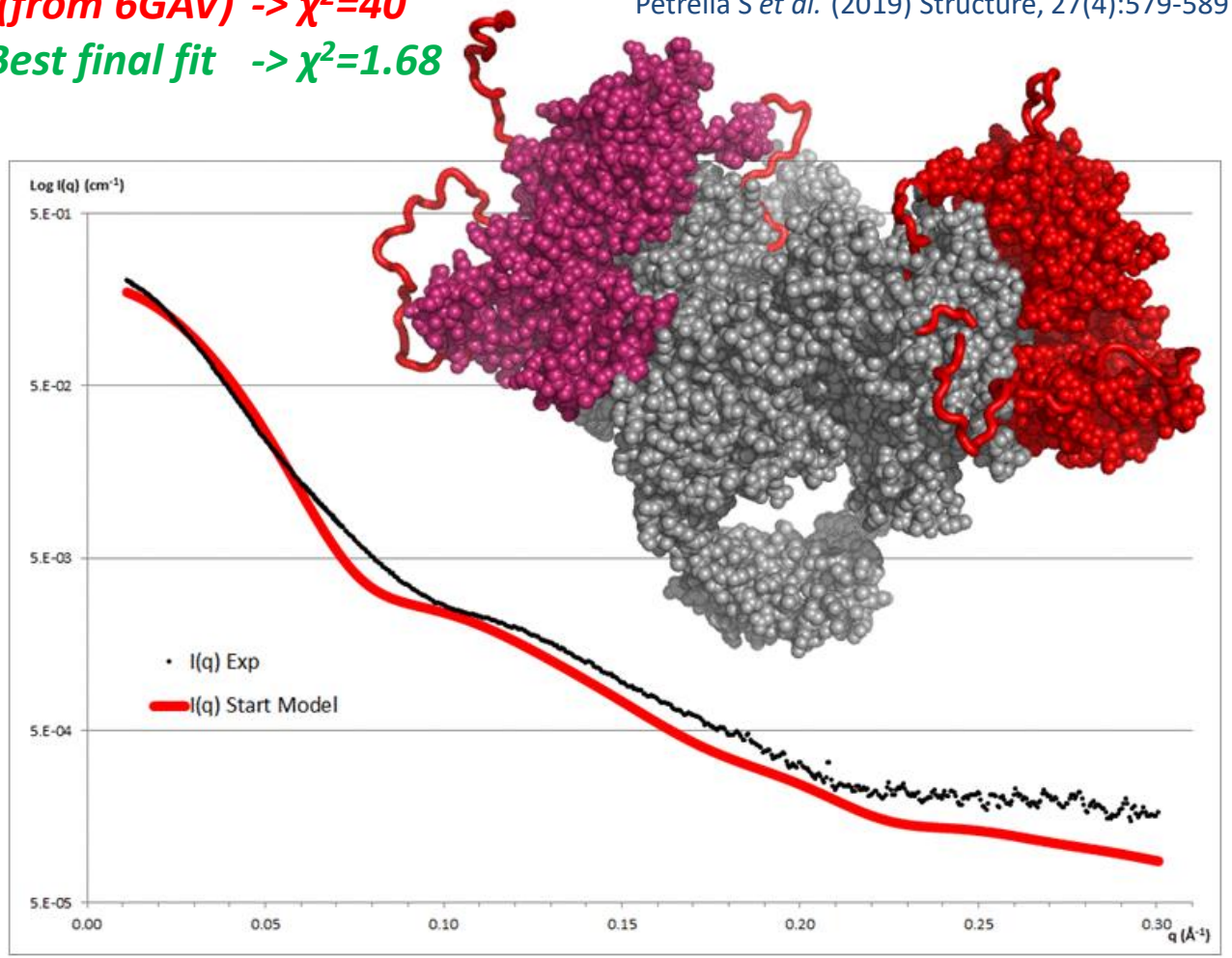
Results

Mycobacterium tuberculosis DNA Gyrase

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

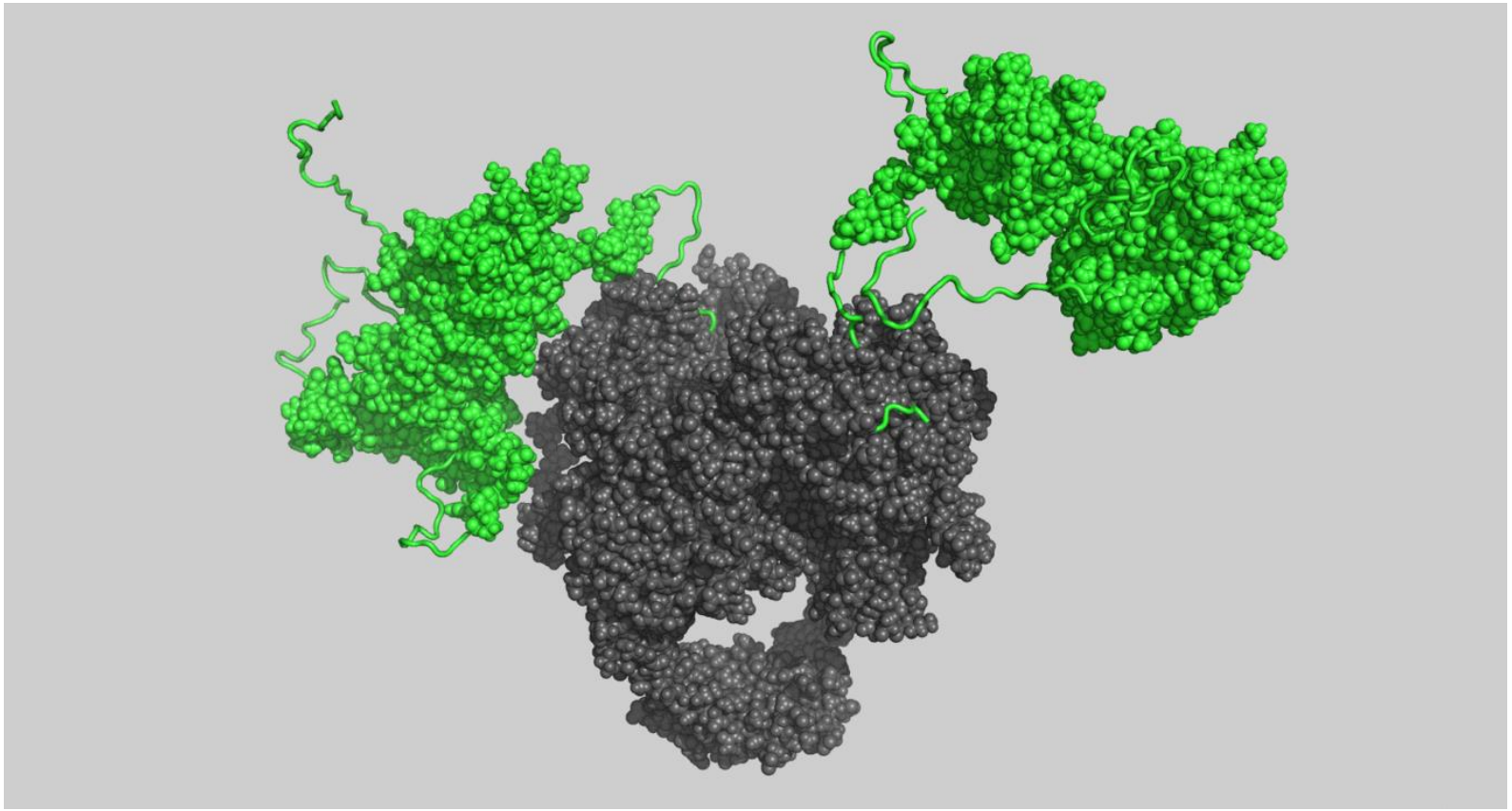
Start model (from 6GAV) -> χ^2=40

Best final fit -> χ^2=1.68



Results

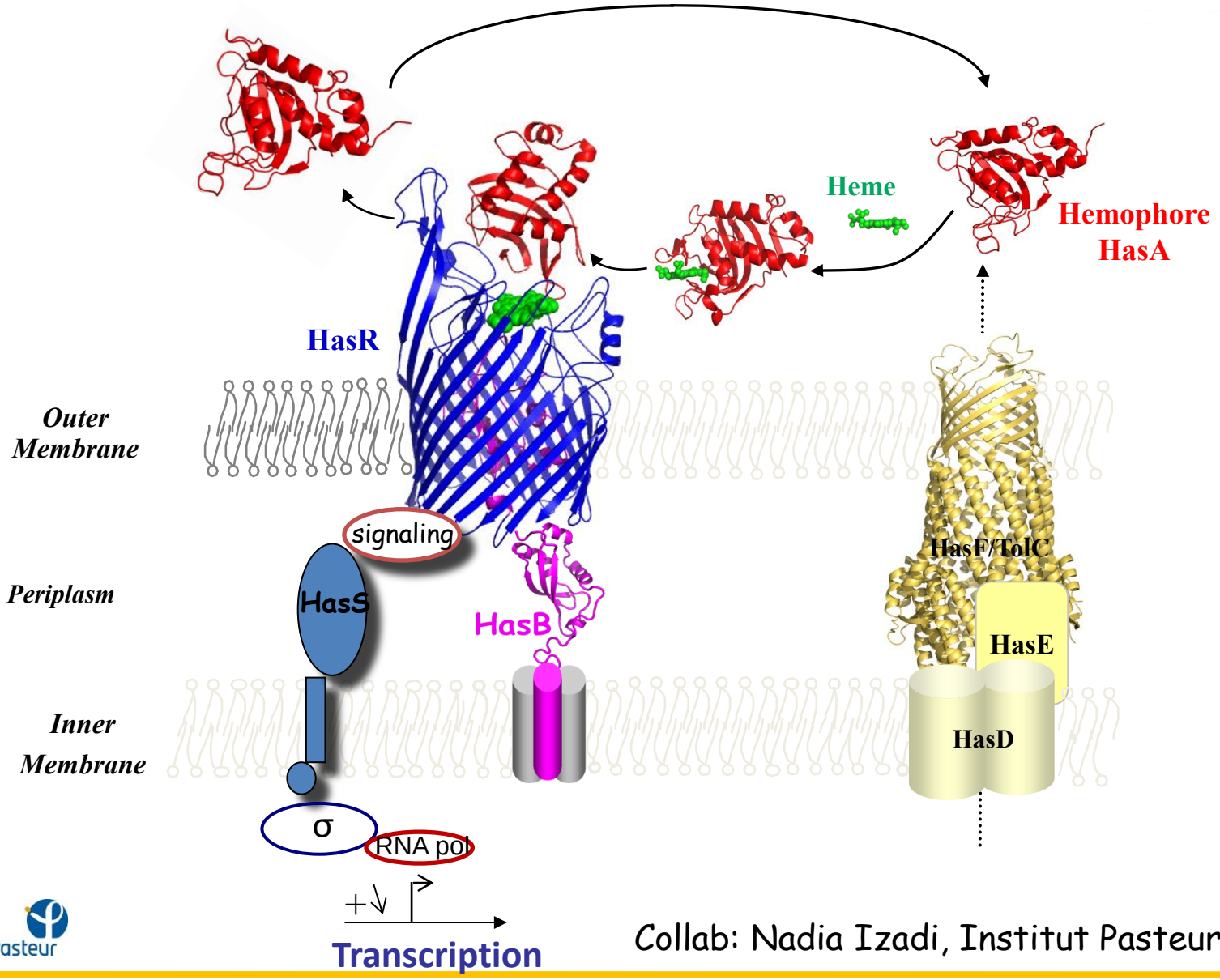
5 best final fits : 1.68 < χ^2 < 1.76



Mycobacterium tuberculosis DNA Gyrase

Combining Memprot & Dadimodo on Has: Heme acquisition system

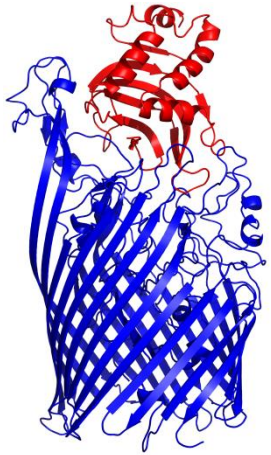
Gram- Bacteria, e.g. Serracia marcescens, opportunist pathogen



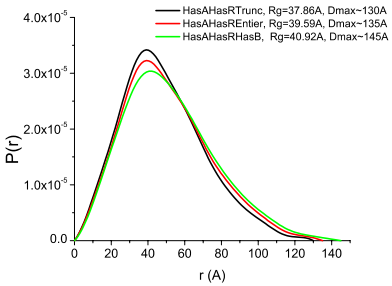
Collab: Nadia Izadi, Institut Pasteur, Paris

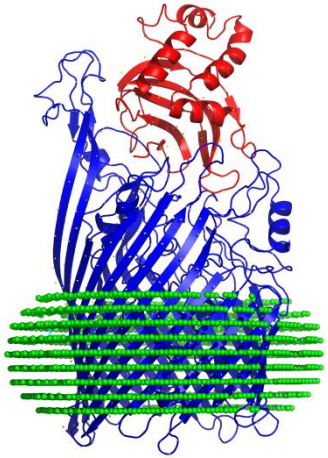
HasA-HAsR: Modeling strategy

1/

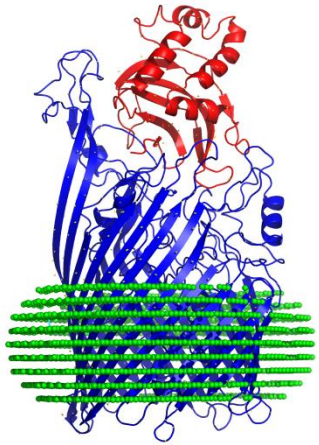


SAXS measurements
(distance constraints)

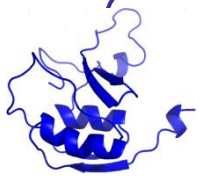


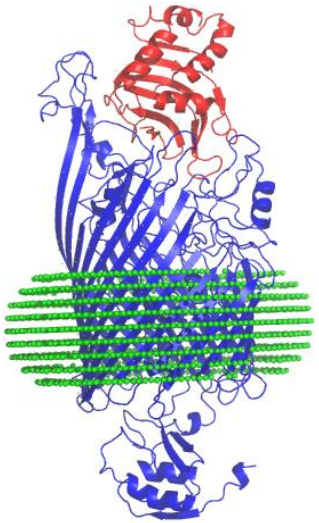


2/



NMR structure of
HasR signaling domain
+ SAXS data

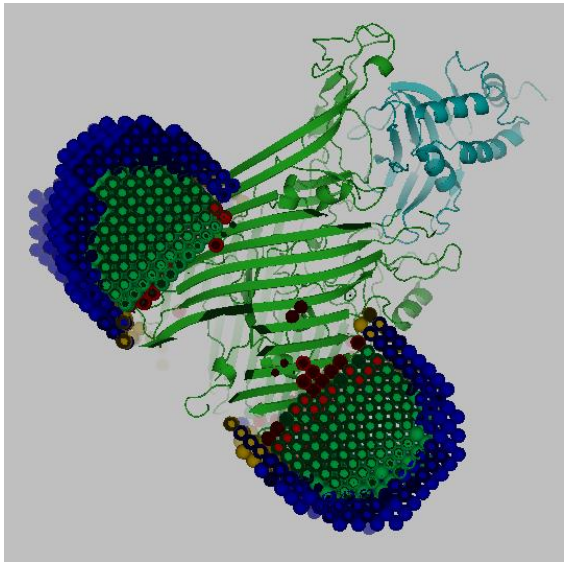


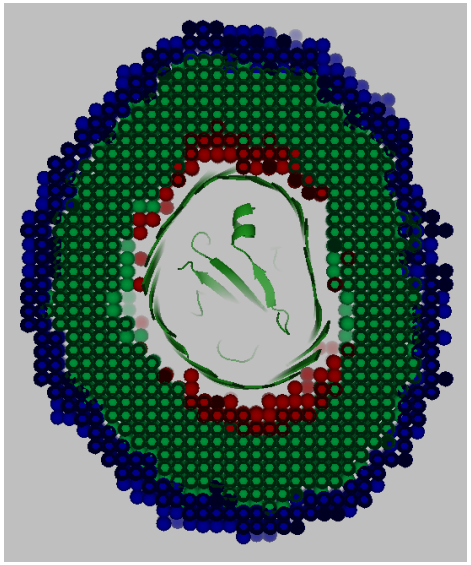


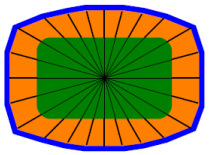
Entire HasR complex
including the signaling
domain

Corona geometrically adapted to protein shape

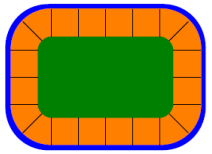
Wojtowicz *et al.*, Biochem. J. (2016) **473**, 2239–2248





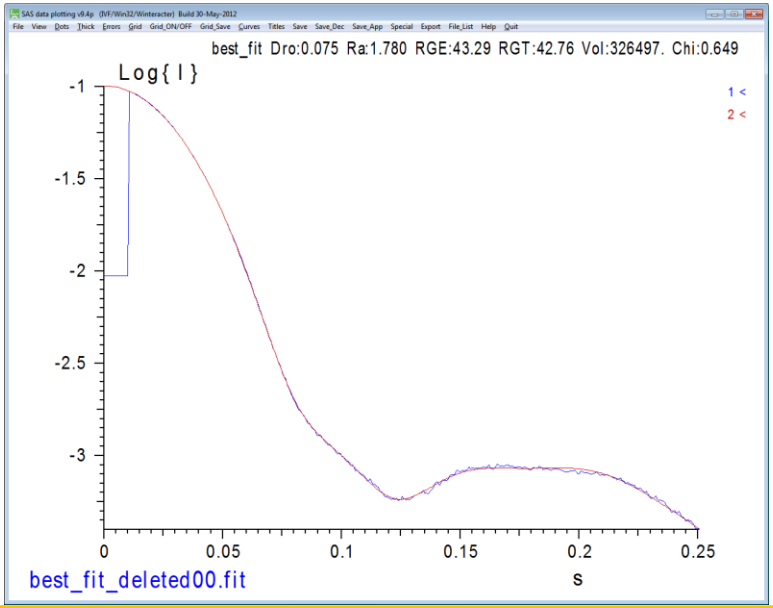


ASA1: Constant thickness on the **line between a given pseudoatom** of the corona **and the center** of the corona.



ASA2: Thickness defined as **the shortest distance to the protein's surface.**

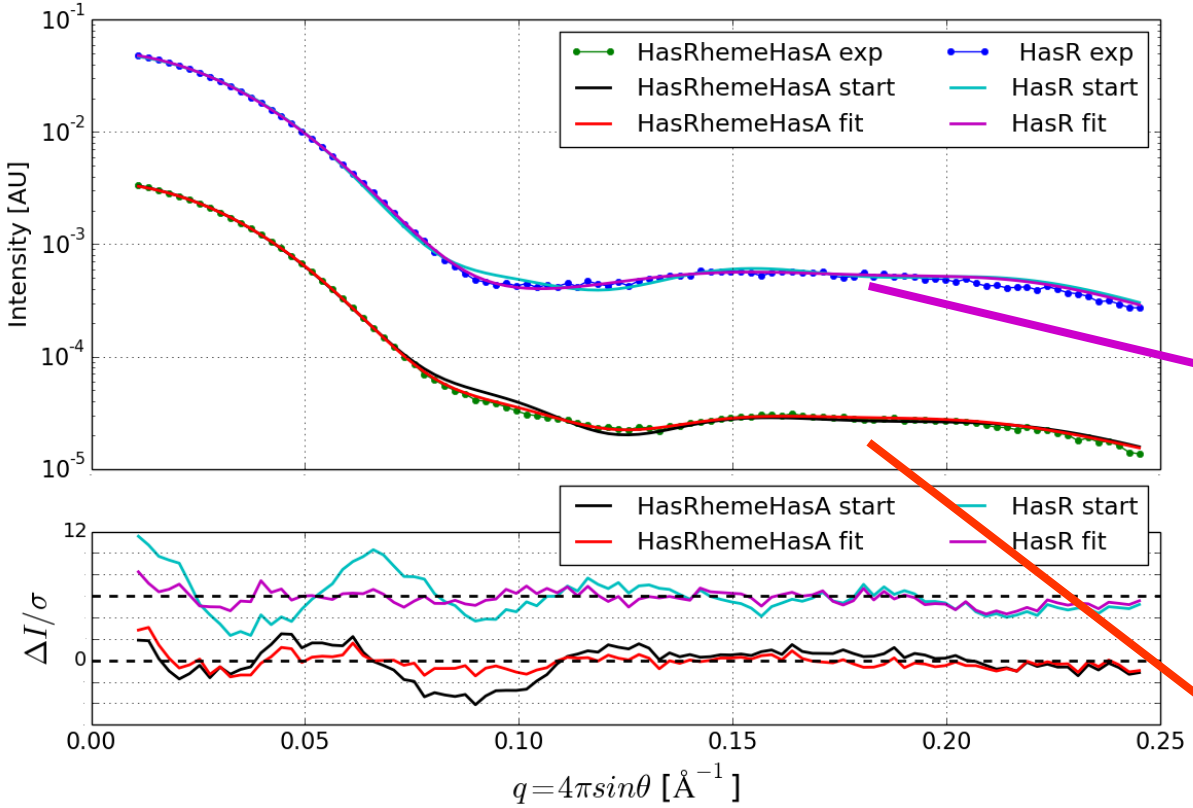
a = 33.500
b = 2.600
t = 5.400
e = 1.110
chi^2 = 2.005
electron density of hydrophobic part= 0.272
electron density of hydrophilic part= 0.506
Number of detergents (tails calc) = 285
Number of detergents (heads calc) = 240

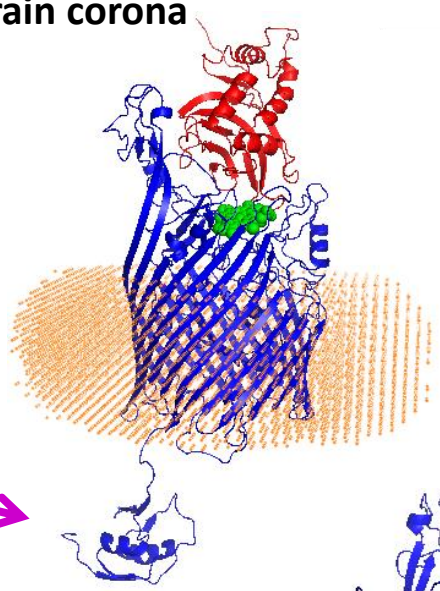


HasR-HasA, now including the signal domain

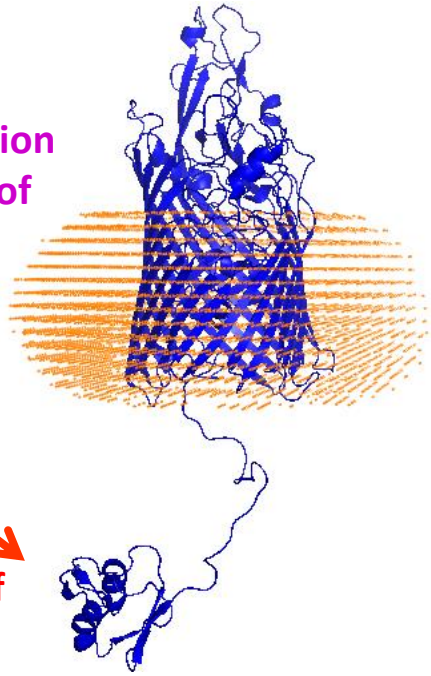
Dadimodo → specially adapted for coarse grain corona

Wojtowicz *et al.*, Biochem. J. (2016) **473**, 2239–2248





Average position in presence of HasA



Average position in the absence of HasA

The interaction of HasA with HasR seems to bring the signaling domain closer to the membrane

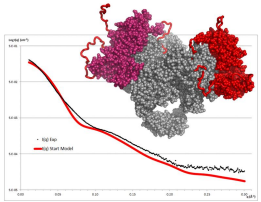
Roadmap

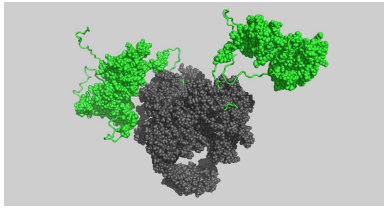
Dadimodo

- Better stop criteria (different from number of generations)
- More friendly output for each run (plot figures,…)
- Summary file for all runs (classification of individual results)
- ADR constraints available on WebServer version (currently only available on local version)

Memprot

- Commissioning of other geometries (bicelles & nanodiscs)
- Commissioning of PepsiSAXS implementation (collab. S Grudinin)
- Web server for direct access by users (currently only staff can use the HPC)





Collaborations

- AQP-0
  - **Alice Berthaud**, Institut Curie
  - Stéphanie Mangenot, Institut Curie
  - **Alexandros Koutsioumpas**, Swing + Julich ForschungsZentrum
- MHST
  - Poul Nissen team, Aarhus University
  - **Maciej Baranowski**, Swing
- HasA-HasR
  - Nadia Izadi
  - **Alexandros Koutsioumpas**, Julich ForschungsZentrum
- DNA Gyrase
  - Stéphanie Petrella, Unité de Microbiologie Structurale
- Memprot
  - **Maciej Baranowski**, Swing Post-Doc
  - **Alexandros Koutsioumpas**
- Dadimodo on the Web
  - **Olga Roudenko**, SOLEIL
  - Aurélien Thureau, Swing
  - Alejandro Diaz, SOLEIL

Beamline SWING
- Maciej Baranowski
- Javier Pérez
- Thomas Bizien
- Youssef Liatimi
- Aurélien Thureau