TD Oléron SAXS 2016

Aurélien Thureau

23/05/2016

Instructions

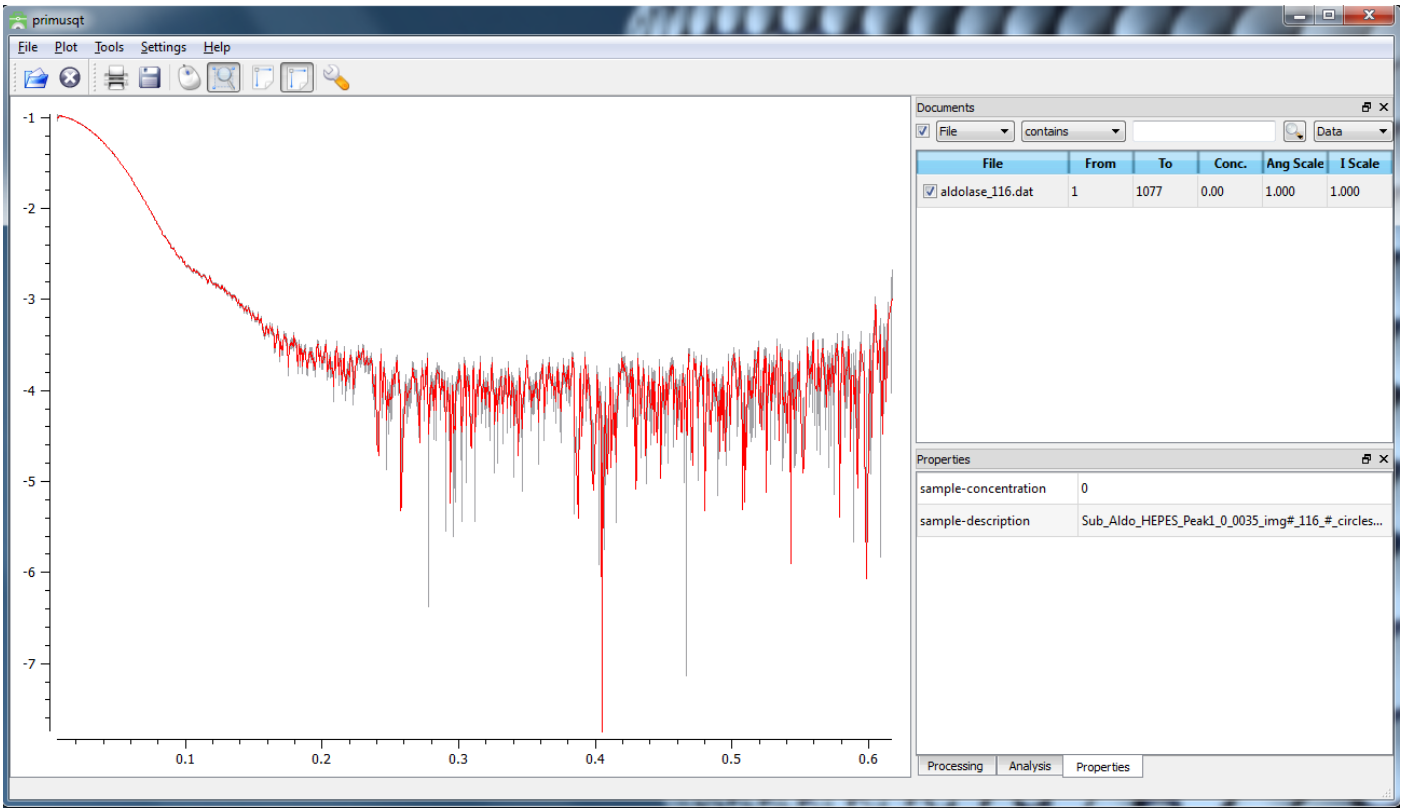
1. Aldolase (aldolase.dat)
    a. Is the protein is Folded ?
    b. What is the Rg value ? — primus
    c. What is the Dmax value ? — gnom
    d. Is the curve correspond to the monomer or the tetramere models (pdb files)? — crysol
    e. Try to generate a bead model. — dammif

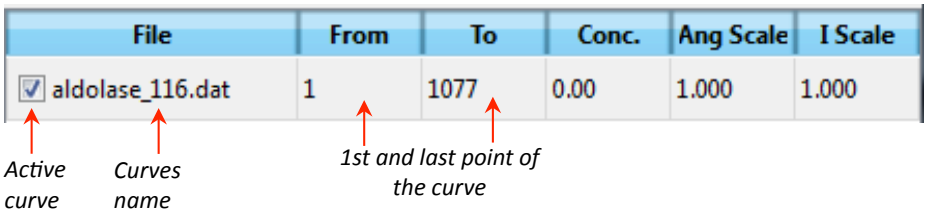
2. CD+Y (CDY.dat)
    a. Is the protein is Folded ?
    b. What is the Rg value ?
    c. What is the Dmax value ?
    d. Is the curve correspond to the model (CDy.pdb) ?
    e. Try to complete the model. — bunch

3. Others (BCDA3.dat & R4-9_SAXS.dat)
    a. Is the protein is Folded ?
    b. What is the Rg value ?
    c. What is the Dmax values ?

Primus

This program is used to display curves and apply some operation on the curves (scaling, substraction or merging…) and to determine the gyration radius Rg.

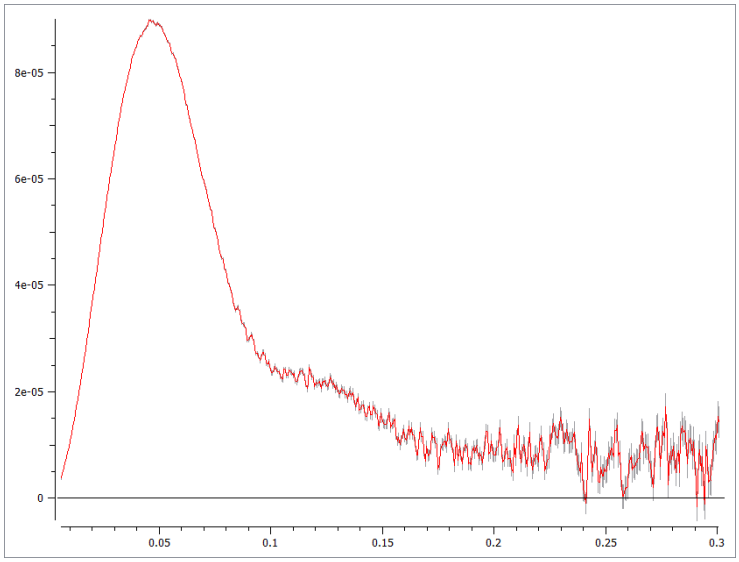




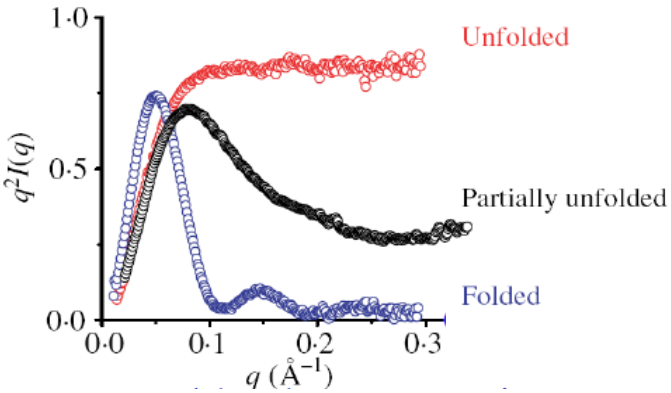
You can load the SAXS curves (click on “File” button). When the data are loaded in the table, the name appears in the “Documents” area.

With the buttons “From” and “To” you can remove respectively the first points or the last points of the curve.

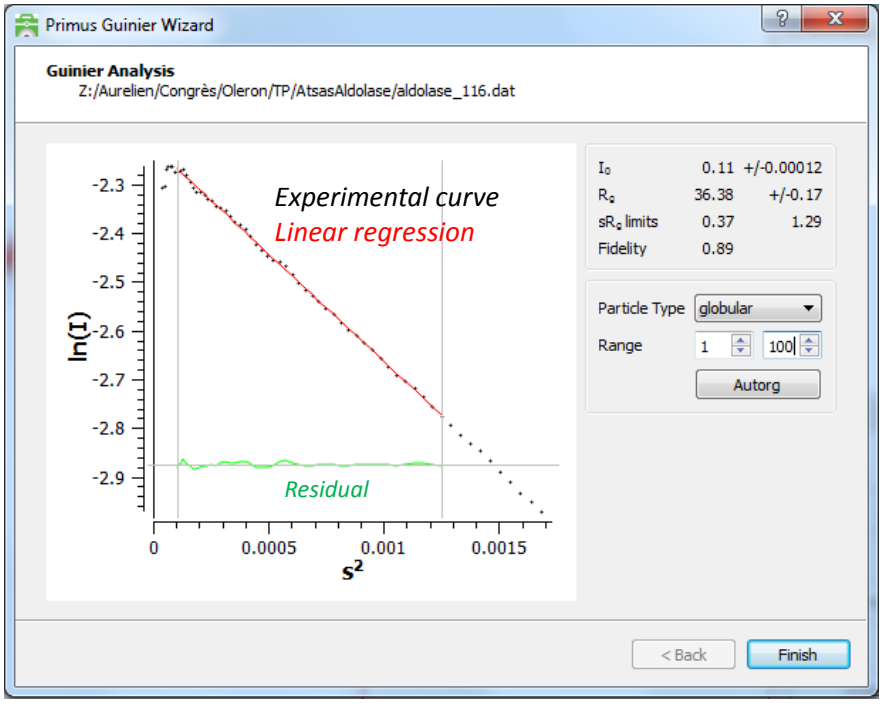
Primus



For Kratky representation, you can go to **Plot -> I*s² vs. s (Kratky plot).**
Only the first points are necessary (q between 0 and 0.3)



For Rg calculation, you can go to **Tools -> Analysis -> Radius of Gyration** and modify manually the parameters or use directly the program “AutoRg”.



You can evaluate the quality of the fit with the residual plot which represents the distribution of the experimental points around the regression line. You can control also the limit s*Rg max which should be inferior to 1. This limit depends of the geometry of the object : For globular protein, the limit can be increased up to 1,4 whereas for elongated or unfolded protein, the Guinier region is more restricted (less than 0,8). The first points which correspond to the lowest values of q can be also removed due to large error measurement (data close to the beam stop).

Gnom

The program GNOM is used to determine the autocorrelation function p(r) from the SAXS data.

```
Command Window
            ***  PLEASE SELECT THE FIRST DATA FILE NAME  ***

Working directory: Z:\Aurelien\Congrès\Oleron\TP\AtsasAldolase\
File to be opened: aldolase_116.dat
Output file                          [ gnom.out    ] : aldolase.out
No of start points to skip           [    0        ] : 8
Run title:  Sample description: Sub_Aldo_HEPES_Peak1_0_0035_img#_116_#_circle
Number of points in the run is 1069
Input data, second file              [ none        ] :
No of end points to omit             [    0        ] : 370
Total number of input data points read is  699
Angular range as read: from  0.01083   to  0.40806
 *** Input data points joined to optimize the performance
   2 successive data points joined
Angular scale (1/2/3/4)              [    1        ] :
Plot input data                (Y/N) [   Yes       ] :
File containing expert parameters    [ none        ] :
Kernel already calculated      (Y/N) [   No        ] :
Type of system        (0/1/2/3/4/5/6) [    0        ] :
Zero condition at r=rmin       (Y/N) [   Yes       ] :
Zero condition at r=rmax       (Y/N) [   Yes       ] :
  --  Arbitrary monodisperse system  --
 Rmin=0,  Rmax is maximum particle diameter
Rmax for evaluating p(r)                             : 180
Number of points in real space       [    141      ] :
Kernel-storage file name             [ kern.bin    ] :
Experimental setup           (0/1/2) [    0        ] :
Evaluating design matrix. Please wait...

Evaluating stabilizer matrix. Please wait ...
 The measure of inconsistency AN1 equals to    0.1566E+01

         Warning: using the chosen range in the real space
            it is not possible to fit the data set within the
            given error band.
Initial ALPHA                        [   0.0       ] :
Plot alpha distribution        (Y/N) [   Yes       ] :
    Alpha    Discrp  Oscill  Stabil  Sysdev  Positv  Valcen    Total
 0.4200E+01  1.1058  2.0085  0.0032  0.8539  1.0000  0.7434  0.43281

Plot results                   (Y/N) [   Yes       ] :
Parameter    DISCRP    OSCILL    STABIL    SYSDEV    POSITV    VALCEN
Weight        1.000     3.000     3.000     3.000     1.000     1.000
Sigma         0.300     0.600     0.120     0.120     0.120     0.120
Ideal         0.700     1.100     0.000     1.000     1.000     0.950
Current       1.106     2.009     0.003     0.854     1.000     0.743
               - - - - - - - - - - - - - - - - - - - - - - - - -
Estimate      0.160     0.101     0.999     0.227     1.000     0.052

 Angular   range    :     from    0.0111   to    0.4081
 Real space range    :     from      0.00   to    180.00

 Highest ALPHA (theor) :   0.289E+04                JOB = 0
 Current ALPHA         :   0.420E+01   Rg :  0.361E+02   I(0) :   0.108E+00

          Total  estimate : 0.433  which is  A SUSPICIOUS  solution

               ===   Select one of the following options  ===
                   - - - - - - - - - - - - - - - - - - -
               CR     ---  to accept the solution and   EXIT
            -(NewAlpha) ---  to manually change          ALPHA
            1,2,3,4,5,6  ---  to change weight/sigma of PARAMETERS
                  7     ---  to maximize a new  total   ESTIMATE
                  8     ---  to replot the              SOLUTION
Your choice :
Evaluating the final solution. Please wait ...
Evaluate errors                (Y/N) [   Yes       ] : N
Next data set           (Yes/No/Same) [   No        ] : Y
Input data, first file               [ aldolase 116] :
```

Select the dat file

Choose the output file

Remove the first points (noisy)

Remove the last points (noisy)

To see the saxs curve

Limits condition (try to change)

Dmax (to be evaluated) : start with 5*Rg

Not necessary

Will plot the Regularized curve and the Autocorrelation function p(r)

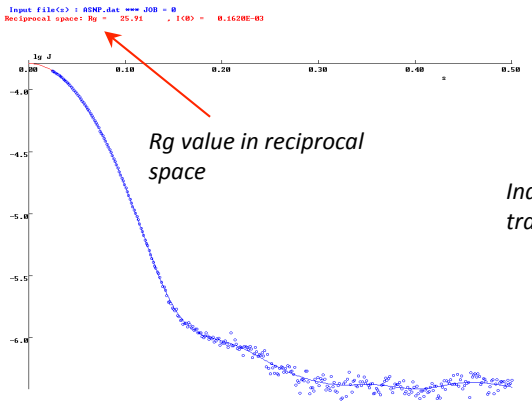
Not necessary for this practical

Let's start again with an other Dmax value !!!

Gnom

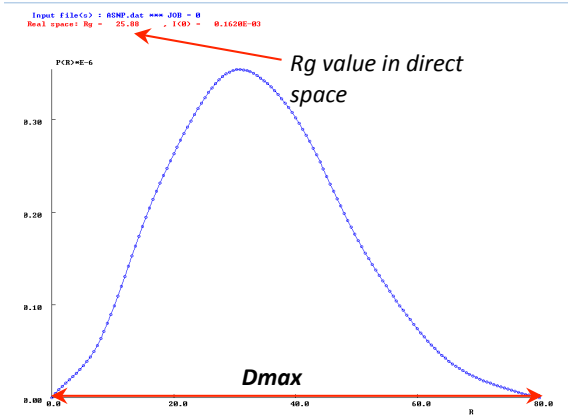
To determine a correct value of Dmax, we must proceed by trial and error to find a Rg calculated with GNOM similar to that found with the calculation of Guinier. We start in general with a value of Dmax equal to 4 or 5 times the value of Rg, and decrease gradually the value of the Dmax in order to obtain a smoothed p(r) that cuts the axis of the distance r.

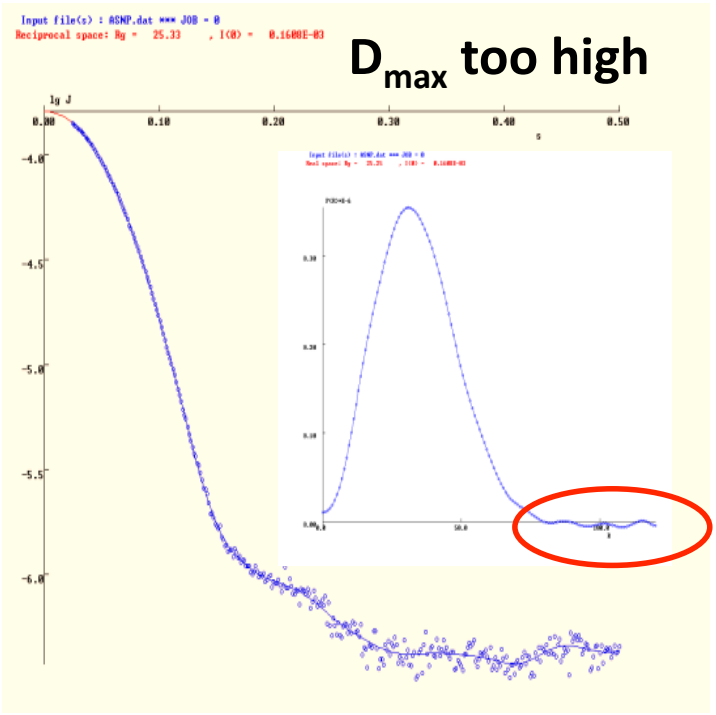
Regularized curve

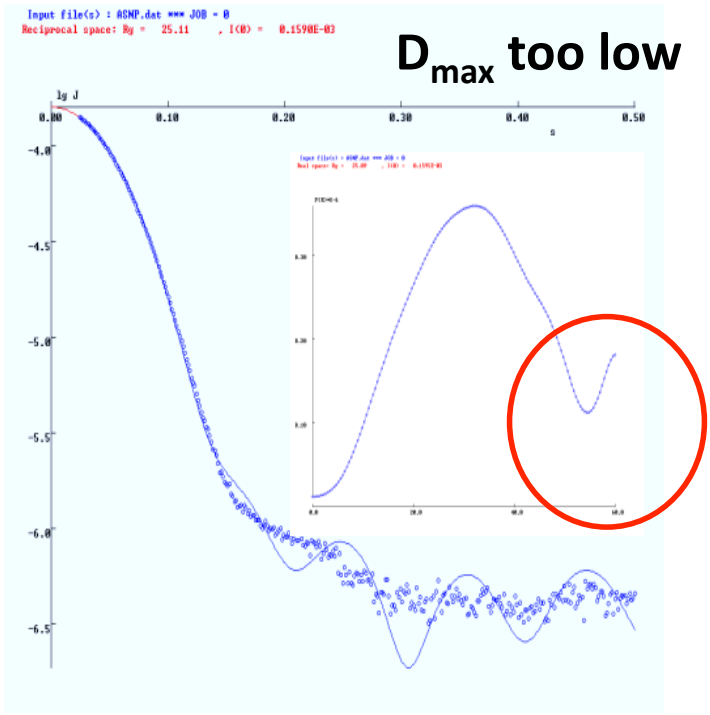


Indirect Fourier transform

Autocorrelation function p(r)







During the process, you will press return to validate each step of the p(r) calculation and the end, the program offers a summary table grouping parameters to appreciate the quality of the fit.

At the end, you will obtain a new file called <*.out> containing all informations about the p(r) determination such as parameters defined preliminary in GNOM (nBeg, nEnd, …), biophysical parameters (Rg, I(0) and Dmax). You will find also in ASCII format, the SAXS curve with the corresponding regularization curve and the p(r) function.

The file <*.out> will be used by ab initio program such as DAMMI, DAMMIN or GASBOR.

Crysol

The program CRYSOL is employed to calculate a theoretical SAXS curve starting from a pdb file and to compare the result to the experimental SAXS data.

```
 ------------------------------------------------
              Program options :
 0 - evaluate scattering amplitudes and envelope
 1 - evaluate only envelope and Flms
 2 - read CRYSOL information from a .sav file
 ------------------------------------------------

Enter your option ..... <           0 >:

           ***  PLEASE SELECT THE PDB FILE NAME  ***

Working directory: Z:\roblin\back-up_2014\Formation_SAXS\TP\CRYSOL-ASDG\
File to be opened: ASDG-dimere-PISA.pdb
 2 (AB) chains were found in ASDG-dimere-PISA.pdb
Process chain (0: all chains) ..............................: 0
 ------------------------------------------------
        Following file names will be used:
ASDG-dimere-PISA01.log -- CRYSOL log-file          (ASCII)
ASDG-dimere-PISA01.sav -- save CRYSOL information  (binary)
ASDG-dimere-PISA01.flm -- multipole coefficients   (ASCII)
ASDG-dimere-PISA01.int -- scattering intensities   (ASCII)
ASDG-dimere-PISA01.fit -- fit to experimental data (ASCII)
ASDG-dimere-PISA01.alm -- net partial amplitudes   (binary)
 ------------------------------------------------
 Maximum order of  harmonics ..... <          15 >: 50
 Order of Fibonacci grid ......... <          18 >:
 ----------  Reciprocal space grid  -------------
  ( in s = 4*pi*sin(theta)/lambda [1/angstrom] )
Maximum s value ................. <       1.000 >: 0.5
Number of points ................ <          51 >: 256
Account for explicit hydrogens? [ Y / N ] <          No >:
Read atoms and evaluate geometrical center ...
 Number of atoms read ........................................ : 10234
Percent processed      10  20  30  40  50  60  70  80  90 100
Processing atoms   :>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
Center of the excess electron density:  0.000  0.075  0.000
Processing envelope:>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
Coefficients   saved to file ASDG-dimere-PISA01.flm
CRYSOL data    saved to file ASDG-dimere-PISA01.sav
--- Structural parameters (sizes in angstroms) ---
Electron   rg   :  33.04       Envelope   Rg      :  32.80
Shape      Rg   :  33.08       Envelope  volume   : 0.2160E+06
Shell    volume : 0.5276E+05   Envelope  surface  : 0.1625E+05
Shell      Rg   :  42.11       Envelope  radius   :  58.85
Shell    width  :  3.000       Envelope  diameter :  115.5
Molecular weight: 0.1445E+06   Dry volume         : 0.1752E+06
Displaced volume: 0.1804E+06   Average atomic rad.:  1.614
Number of residuals :  1302
 Fit the experimental curve [ Y / N ] .. <         Yes >:

           ***  PLEASE SELECT THE DATA FILE NAME  ***

Working directory: Z:\roblin\back-up_2014\Formation_SAXS\TP\CRYSOL-ASDG\
File to be opened: ASDG.dat
Subtract constant ..................... <          no >: y
Title: ? Data from ASDG.dat
 Maximum angle in the data file ......................... : 0.5004
 Number of experimental points .......................... : 872
Angular units in the input file:
4*pi*sin(theta)/lambda [1/angstrom] (1)
4*pi*sin(theta)/lambda [1/nm]       (2)
2 *  sin(theta)/lambda [1/angstrom] (3)
2 *  sin(theta)/lambda [1/nm]       (4)  <           1 >:
 Angular units multiplied by ............................ : 1.000
 Number of points after regriding ....................... : 201
Electron density of the solvent, e/A**3  <       0.3340 >:
 Number of experimental points used ..................... : 200
 ------  Fitting the experimental data ...    ---
Plot the fit [ Y / N ] ................. <         Yes >:
Another set of parameters [ Y / N ] .... <          No >:
Rg from the slope of net intensity ..................... : 35.07
Average electron density ............................... : 0.4624
Data fit       saved to file ASDG-dimere-PISA01.fit
Intensities    saved to file ASDG-dimere-PISA01.int
Net amplitudes saved to file ASDG-dimere-PISA01.alm
 Press CR to terminate the program
```

Select option 0 (default)

Select the pdb file

Select the totality of the pdb contains or just some chain

*Enter the number of spherical harmonic determined by the relationship : **Lmax – 15 = Qmax*Dmax / 2***

For the Fibonacci grid, less the default value

Give the Qmax value for the fit

Enter number of point for the fitting curve (default to 256)

default proposition

After the processing of atoms and envelope, CRYSOL generate two files, the .flm file containing the enveloppe structure and function parameters , the .sav file containing all information to restart CRYSOL with already made calculation

CRYSOL give also biophysical parameters such as the Rg of enveloppe or the volume (in A^3)

Select the experiment data file

You can add a supplementary parameter of the fitting process to offset the backgroud (linked to the substraction of solvent in major case)

Put the correct value of the angular units

For the classical buffer without excess of salt of organic molecule, the default value is correct

You can display the experimental curve fitting

A the end CRYSOL generate the .alm file containing the amplitudes which are used by SASREF or BUNCH, the .fit file in ascii format usable with another spreadsheet, and the .int file containing the intensity of the different component (electron, border shape and excluded volume)

Dammif

The program DAMMIF is dedicated to low resolution shape modeling using a sphere containing beads (with a defined value of electronic density) as initial model.

```
GNOM output file to read? ..................................................... :
e.out
Angular unit? Select one of: (u) unknown, (a) angstrom, (nm)
nanometer (default: unknown) .................................................. :
Output file prefix? (default: dammif) ......................................... :
Omit output of solvent in PREFIX-0.pdb? (default: yes) ......................... :
Create pseudo chains in PDB output? (default: no) .............................. :
Expected particle symmetry? Select one of: (1) p1, (2) p2, (3) p3,
(4) p4, (5) p5, (6) p6, (7) p7, (8) p8, (9) p9, (10) p10, (11) p11,
(12) p12, (13) p13, (14) p14, (15) p15, (16) p16, (17) p17, (18) p18,
(19) p19, (20) p22, (20) p222, (21) p32, (22) p42, (23) p52, (24) p62,
(25) p72, (26) p82, (27) p92, (28) p102, (29) p112, (30) p122
(default: p1) ................................................................. :
Expected particle anisometry? Select one of: (u) unknown, (p)
prolate, (o) oblate (default: unknown) ........................................ :
Constant to subtract? 0 to disable constant subtraction, undefined
for automatic (default)? ...................................................... :
Simulated annealing setup? Select one of: (i) interactive, (f) fast,
(s) slow (default: slow) ...................................................... :
Maximum bead count (default: unlimited)? ...................................... :
```

Load the GNOM file

Precise the units

outputfile

Generate an output file in pdb format of the solvent

No

Precise the symmetry of the protein

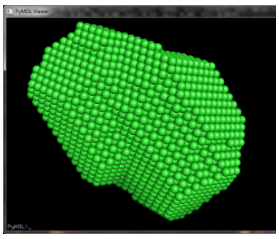
If known, precise the global form of the protein

Constant substraction

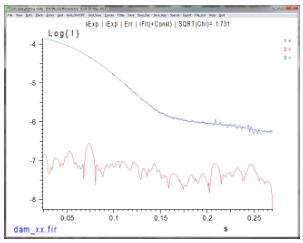
Speed of the simulated annealing (slow mode better)

??? No documentation

By following the instruction described below, you will obtain at the end, a pdb file containing the shape composed of dummy residues, a fitting curve (smoothed curve) dam_xx.fit, a fitting curve corresponding to the experimental data dam_xx.fir and a log file with initial parameters and the process of minimization steps.



Pymol
(sphere representation)



Sasplot
<outputname>.fit

Additional information

** This procedure has to be repeated 10 to 100 times. On linux (bash syntax):

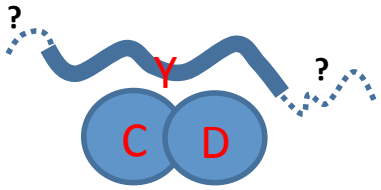
```
for i in `seq 1 10` ;
   do dammif --prefix=prot-$i --mode=slow prot.out;
done
```

** All envelope can be compared with each other with the DAMAVER package suite. This package allows to superimpose and to compare the shapes between them. A table contained in the file damsel.log presents a matrix giving a non-deviation standard score (NSD) for each pair of shapes, and gives a classification of the shape. The most representative shape presents the lowest mean value of NSD. Be careful with the damfilt.pdb which represents a filtered shape corresponding to the common part off all shape. This model, in general case, does not fit the SAXS data !

On linux:

```
damaver -a *-1.pdb
```

Bunch



BUNCH performs modeling of multidomain proteins against SAXS data using a combined rigid body and *ab initio* modeling approach. The program allows determination of three-dimensional domain structure of proteins based on multiple scattering data sets from deletion mutants when the structure(s) of individual domains are available.

pre_bunch

Initial approximation is made by a tool called **pre_bunch** which generates a PDB file containing a single CA-chain (even if there are several symmetry related polypeptide chains) with the length equal to the full-length sequence.

```
Initial random seed? (default: use current time) ..............
 Initialised random seed as    1463129901
 Input sequence file name ..... <         .seq >: CD
 Number of residues read ...................... : 798
 Number of domains ............ <           0 >: 1
 Input pdb file name .......... <         .pdb >: CD
 Shift the structure to the origin ? [ Y
 / N ] ........................ <         Yes >:
 Output pdb file name ......... <         .pdb >: test
```

default

complete sequence of the sample (CDY.seq)

number of pdb files (1 in our case)

uncomplete pdb file

Yes

Out prebunch pdb file => input bunch pdb file

Bunch need also the partial scattering amplitude file (.alm) of each domain (computed by **CRYSOL**). You need to run **crysol** without fitting an experimental curve.
After say *No* for fitting an experimental curve, all answers are default.

```
 Computation mode (User or Expert) ..... <         User >:
 Log file name ................... <         .log >:
 Log file name ................... <         .log >: test1
 Project identificator ..................... : test1
 Enter project description ..... : test oleron
 Random sequence initialized from .......... : 1463131476
 Initial structure ............... <         .pdb >: pre_bunch
  LOADAM --W- : rAtom not assigned
 Number of atoms read ...................... : 884
 Center of the initial structure :     0.0003   -0.0002   -0.0000
 Maximum radius ............................ : 65.22
Averaged formfactors of DRs used
 DR formfactor multiplier .................. : 1.000
 Symmetry: P1...19 or Pn2 (n=1,..,12) ... <          P1 >:
 Angles penalty ............................ : 5.661
 Dihedrals penalty ......................... : 1.142
 Angles penalty weight ..................... : 10.00
 Dihedrals penalty weight .................. : 1.000
 Cross penalty ............................. : 0.2644
 Cross penalty weight ...................... : 100.0
 Extended loops penalty .................... : 3.410
 Extended loops penalty weight ............. : 1.000
 Distances penalty ......................... : 1.316
 Distances penalty weight .................. : 10.00
 Shift penalty ............................. : 1.146e-5
 Shift penalty weight ...................... : 1.000
 File name, contacts conditions, CR for none <         .cnd >:
 Input total number of scattering curves  <           1 >:
 Input first & last residues in 1-st
 construct ................. <          1,         884 >:
 Enter file name, 1-st experimental data  <         .dat >: CDA2_Mrg
 Number of experimental points found ....... : 703
 Angular units in the input file :
 4*pi*sin(theta)/lambda [1/angstrom] (1)
 4*pi*sin(theta)/lambda [1/nm      ] (2)
 2*   sin(theta)/lambda [1/angstrom] (3)
 2*   sin(theta)/lambda [1/nm      ] (4)  <           1 >:
 Fitting range in fractions of Smax ..... <       1.000 >:
 Experimental radius of gyration ........... : 37.94
 Number of points in the Guinier Plot ...... : 0
 Amplitudes, 1-st subunit ............... <         .alm >: CD00.alm
  LOALMS ---W-: Max. order of harmonics reduced to          20
 Maximum order of harmonics ................ : 20
 Number of points in partial ampliudes ..... : 101
 Current subunit:   6278 atoms read, center at     0.24    25.98     1.80
 Residues in the full-length protein ....... : 22     -      819
 Fix the subunit at this position? [ Y /
 N ] ....................................... <          No >: Y
```

May be one day you will change the default…

Output log file name

Text information

Output prebunch pdb file as input

Symmetry of your system

You can add contacts condition (not in our case)

Number of saxs curve (1 in our case)

The curve correspond to all the sequence (1-884)

Saxs curve scattering amplitude file

units

We want to take into account all the curve

scattering amplitude file generated by crysol

Yes

Angular step : 20.00 is correct