

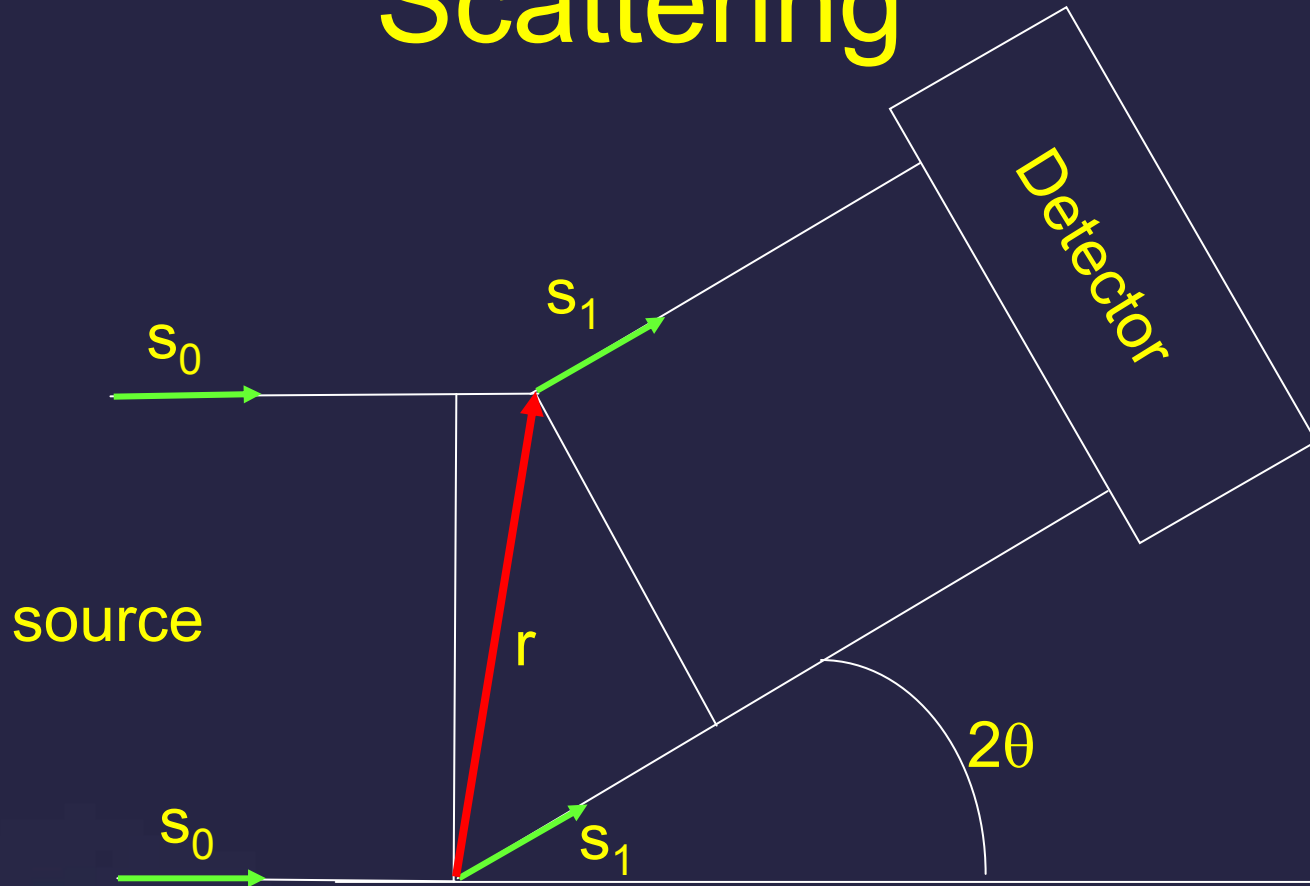
Data Analysis

Solution scattering

Marc Malfois – September 2010



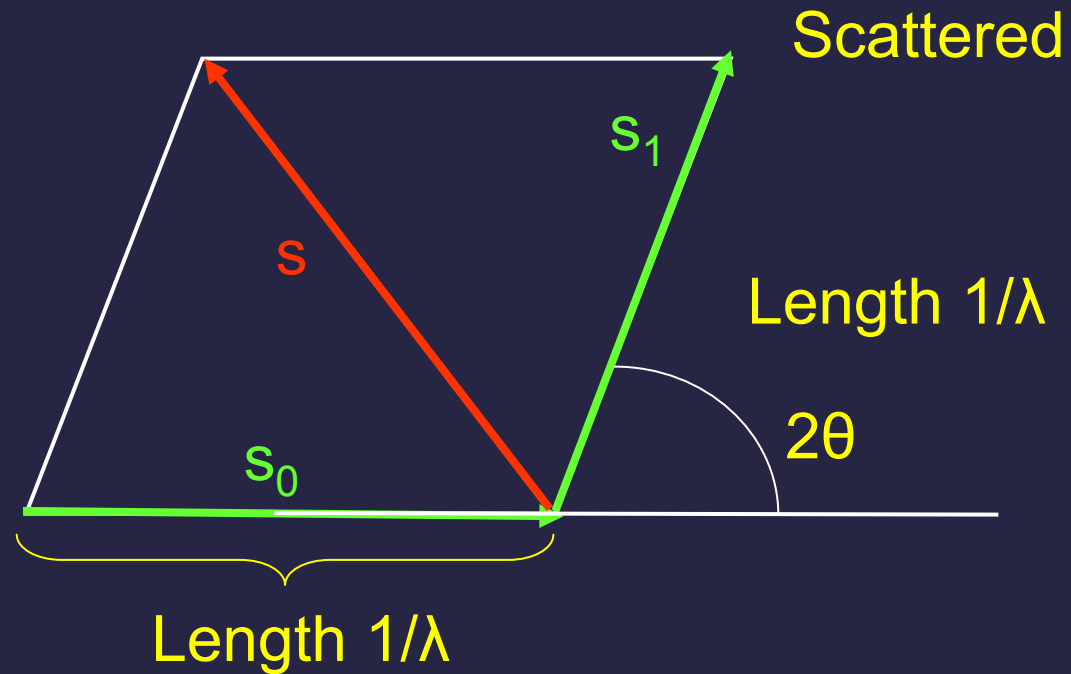
Scattering



$$\text{Path difference: } r \cdot s_1 - r \cdot s_0 = r (s_1 - s_0)$$

Scattering vector

$$|s_0| = |s_1| = 1/\lambda$$



$$s = |\mathbf{s}| = 2 \sin\theta / \lambda$$

s is called the scattering vector

Other notations: $q = h = "s" = 2\pi s$

SAS = SMALL ANGLE scattering $\Rightarrow s \sim 2\theta/\lambda$



Scattered amplitude

$$F(q) = \int_{V_r} \rho(r) e^{iq \cdot r} dV_r$$

Fourier Transform of the electron density $\rho(r)$

The scattered intensity is

$$I(q) = F(q) \cdot F^*(q)$$

At $s = 0$;

$$F(0) = \int_{V_r} \rho(r) dV_r$$

$$I(0) = (\text{number of electrons})^2$$

Scattering intensity

$$I(q) = F(q) \cdot F^*(q)$$

$$I(q) = \int_{V_r} \rho(r) e^{-i\vec{q}\vec{r}} dV_r \int_{V_{r'}} \rho(-r) e^{i\vec{q}\vec{r}} dV_{r'}$$

$$I(q) = \int_{V_r} \int_{V_{r'}} \rho(r) \rho(r') e^{-i\vec{q}(\vec{r}-\vec{r}')} dV_r dV_{r'}$$

$$p(R) = \int_{V_r} \rho(r) \rho(r+R) dV_r$$

Pair distribution function

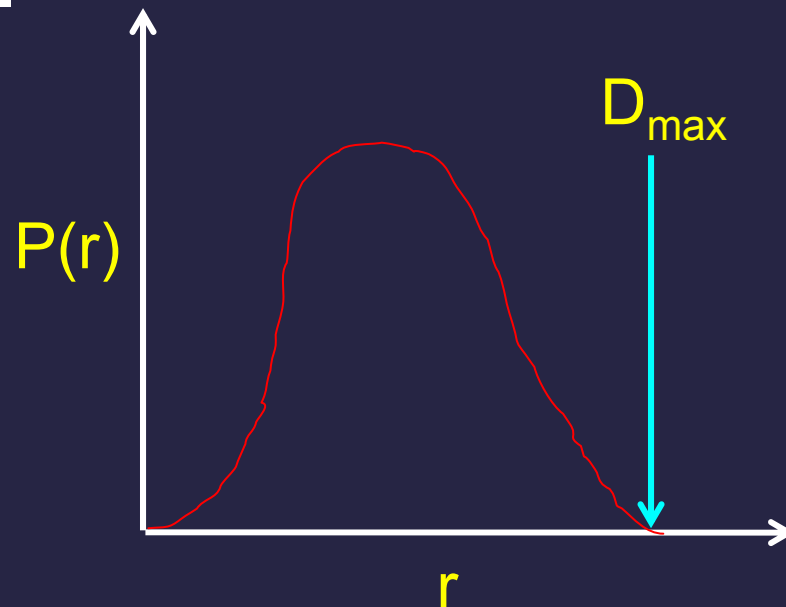
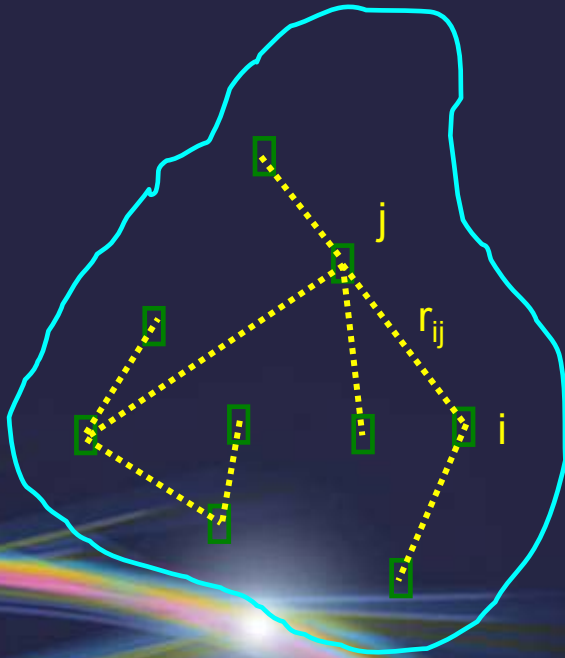
$$I(q) = \int_{V_r} p(R) e^{-i\vec{R}\vec{q}} dV_r$$

with $\vec{R} = \vec{r} - \vec{r}'$

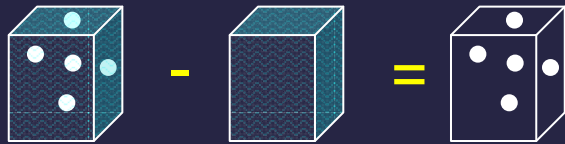
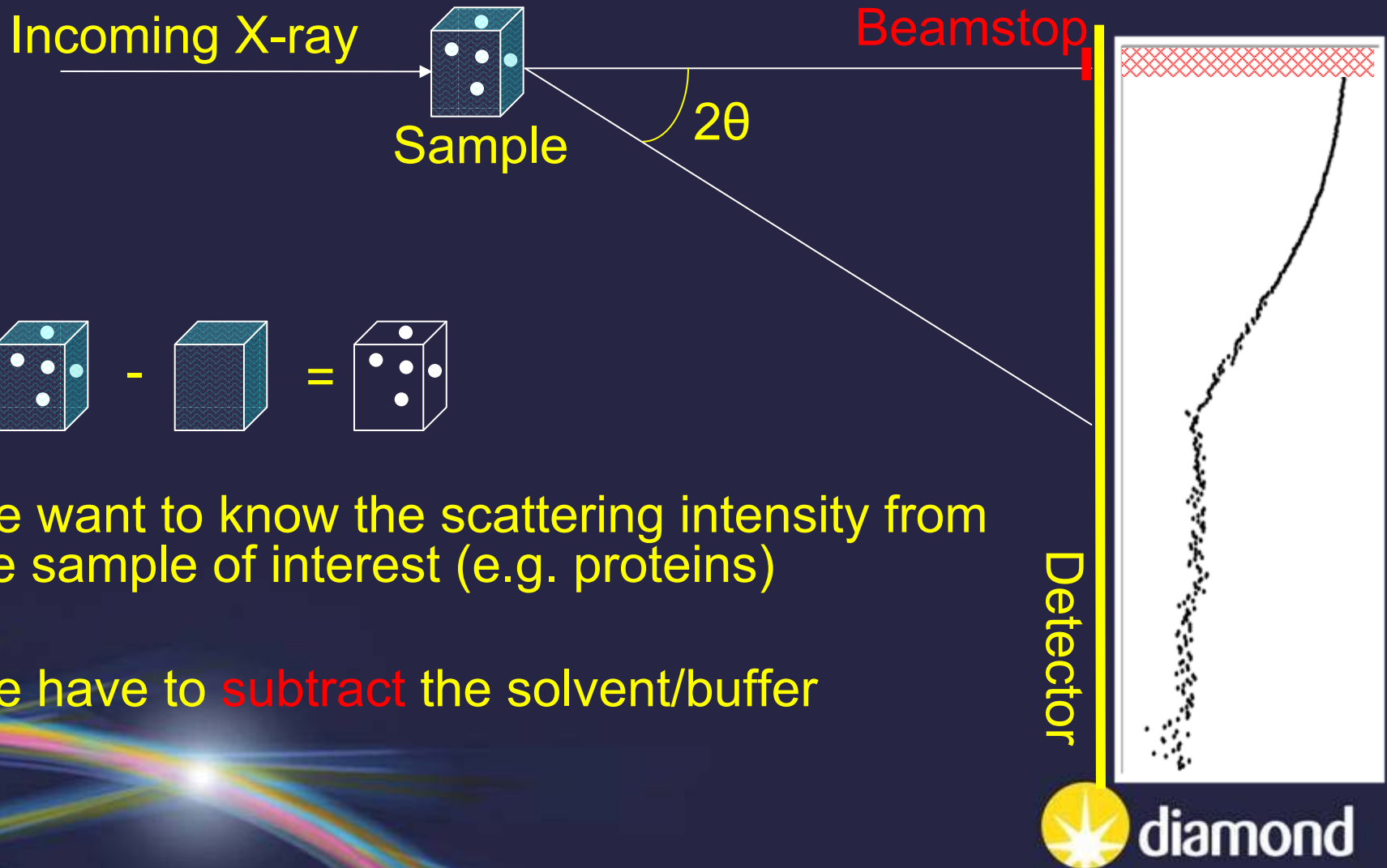
Pair distance distribution function

$$p(r) = \int_{V_r} \rho(r) \rho(r + R) dV_r$$

Probability of finding a point at r from a given point.



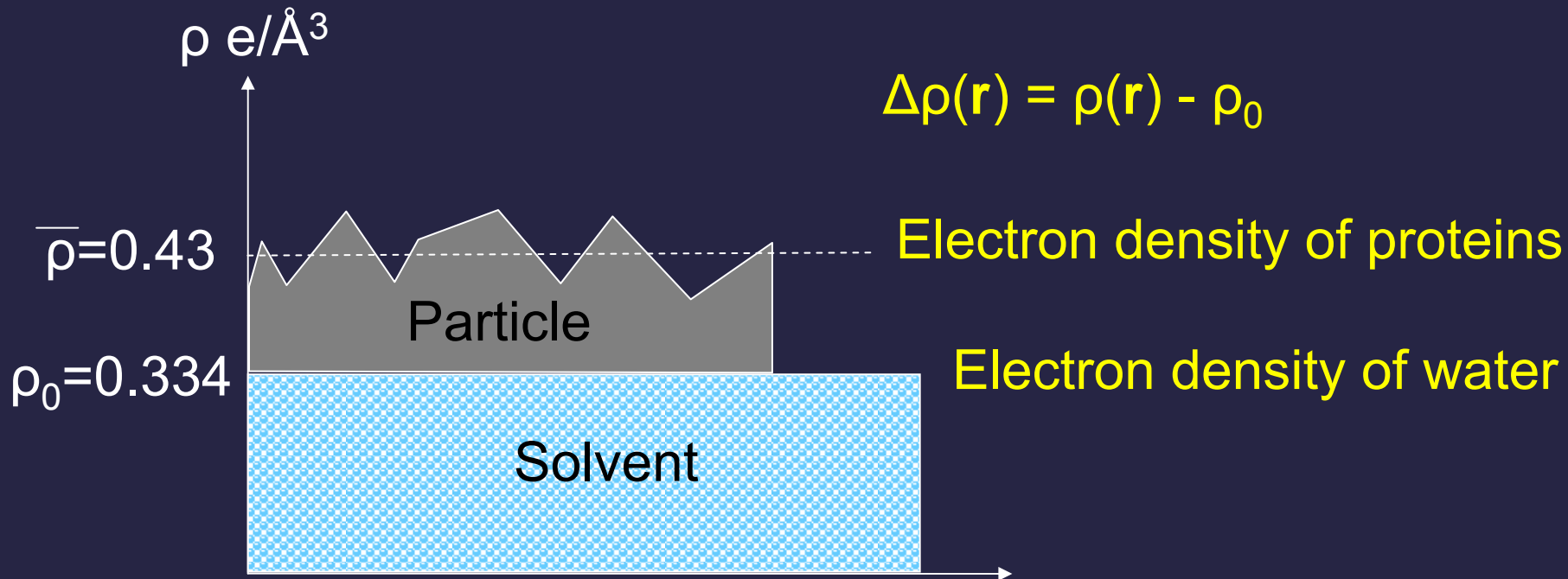
Experimental setup



We want to know the scattering intensity from the sample of interest (e.g. proteins)

We have to **subtract** the solvent/buffer

Contrast of electron density



The electron density for buffers with high salt quantity has to be calculated.

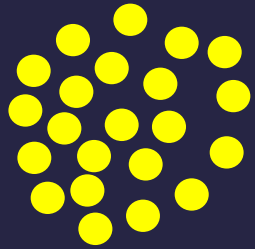
Very low contrast for proteins in solution

Contrast of electron density

$$I(q) = \int_{V_r} \int_{V_{r'}} \Delta\rho(r)\Delta\rho(r')e^{-i\vec{q}(\vec{r}-\vec{r}')}dV_r dV_{r'}$$

$$p(R) = \int_{V_r} \Delta\rho(r)\Delta\rho(r+R)dV_r$$

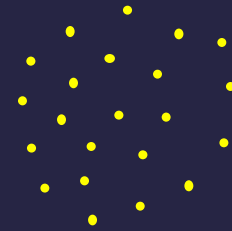
Solution of particles



=



*



Solution

=

Motif (Protein)

*

Lattice

$I(c,s)$

=

$F(0,s)$

.

$S(c,s)$

Form factor
of the particle

Structure factor
of the particle

Solution of particles

Assumption

Ideality : No intermolecular interactions

Monodispersity : Identical particles

Particle in solution => thermal motion => Particles have a random orientation/x-ray beam. The sample is **isotropic**. Only the **spherical average** of the scattered intensity is experimentally accessible.

$$F_1(q) = \int_{V_r} \Delta\rho(r) e^{irq} dV_r$$

Amplitude of 1 particle

$$I(q) = n \cdot \langle F_1(q) \cdot F_1(q)^* \rangle = n \cdot I_1(q)$$

Intensity of n particles
in solution

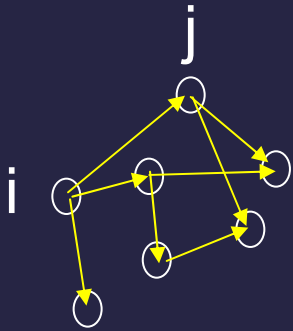
Debye formula

$$i_1(q) = \langle i_1(\vec{q}) \rangle = \int_{\Omega} i_1(\vec{q}) d\Omega$$

$$i_1(q) = \int_{V_1} \int_{V_2} \Delta\rho(r_1)\Delta\rho(r_2) \frac{\sin(r_{12}q)}{r_{12}q} dr_1 dr_2$$

$$i_1(q) = 4\pi \int_0^{\infty} p(r) \frac{\sin(qr)}{qr} dr$$

Debye formula



Particle : Discrete sum of elementary scatterers (Atoms).

$$i_1(q) = \sum_i^N \sum_j^N f_i(q) f_j(q) e^{i\pi q(r_i - r_j)}$$

With $f_i(q)$ the form factor of the elementary form factor

Debye 1915

$$i_1(q) = \sum_i^N \sum_j^N f_i(q) f_j(q) \frac{\sin qr_{ij}}{qr_{ij}}$$

Widely used for model calculations  diamond

Intensity at the origin

$$I(0) = \frac{I_0 r_e^2 c M d}{N_A a^2} \left[\overline{v_p} (\rho - \rho_0) \right]^2$$

r_e^2 : $7.95 \cdot 10^{-26}$ cm² scattering cross section of the e⁻

c : concentration of the sample (mg/ml)

d : Thickness of the sample

I_0 : Intensity of the direct beam

v_p : partial specific volume

The molecular weight can be determined from the intensity at origin

Intensity at the origin

Determination of I_0

- Well know reference sample (Glassy carbon on I22)
- Scattering by water(O. Glatter)

In case of **relative intensity**, it is possible to determine the molecular weight of the sample by measuring the scattering of a reference sample (Lysozyme or BSA)

$$Mw_s = \frac{I_r(0)}{I_s(0)} \cdot \frac{c_s}{c_r} Mw_r$$

Guinier Law

$$I(q) = 4\pi \int_0^\infty p(r) \frac{\sin(qr)}{qr} dr$$

Taylor series expansion of

$$\frac{\sin(qr)}{qr} = 1 - \frac{(qr)^2}{3!} + \frac{(qr)^4}{5!} - \dots$$

$$I(q) = 4\pi \int_0^\infty p(r) dr \left[1 - \left(\frac{q^2}{3!}\right) \frac{\int_0^\infty r^2 p(r) dr}{\int_0^\infty p(r) dr} \right] = I(0) \left[1 - \frac{q^2 R_g^2}{3} \right]$$

$I(0)$

$2R_g^2$

Taylor series expansion of

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} \quad \text{with} \quad x = -\frac{q^2 R_g^2}{3}$$

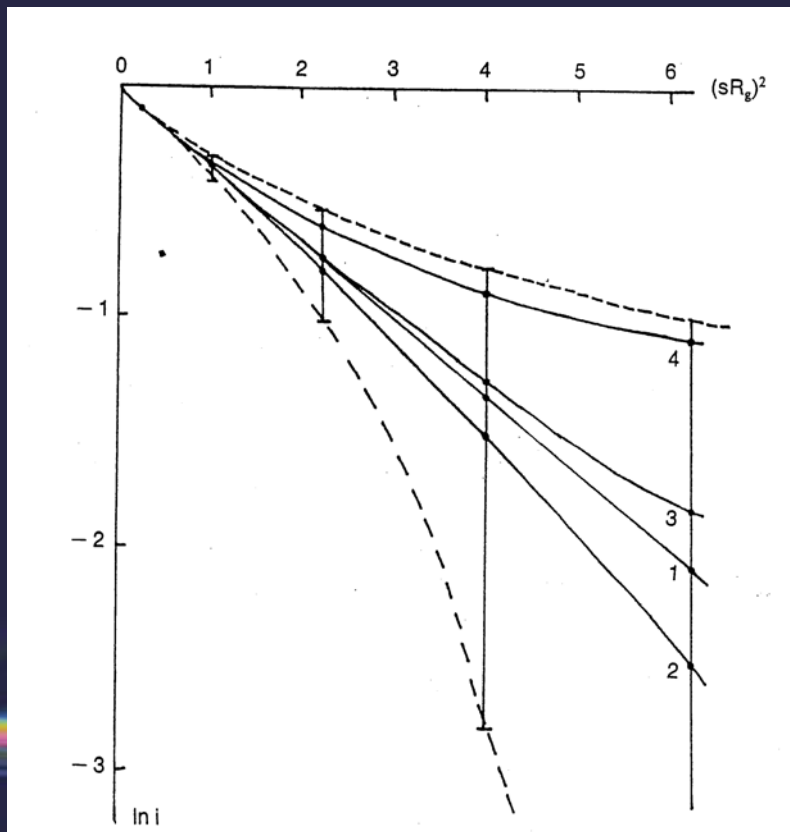
$qR_g \ll 1$

$$I(q) = I(0) e^{-\frac{q^2 R_g^2}{3}}$$



Accuracy of the Guinier law

Valide for $qR_g \ll 1$
but $qR_g < 1.3$ in the literature



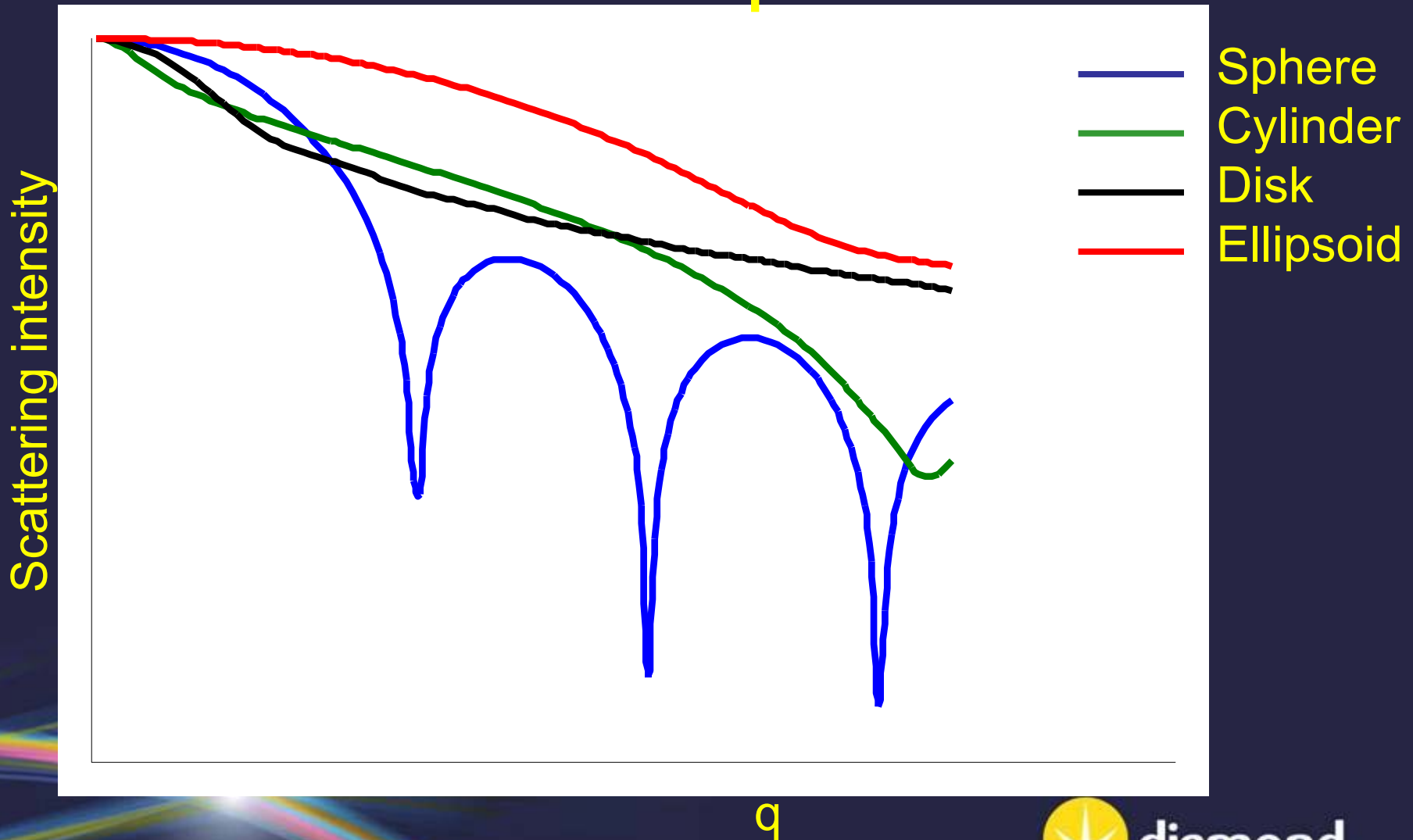
- 1 – Guinier law (exponential)
- 2 – Sphere
- 3 – Thin disk
- 4 – long rod

Depends on the shape of the particle

Structure Analysis by Small Angle X-ray and Neutron Scattering

L.A. Feigin and D.I. Svergun (1987),
Plenum Press.

Solution of particles



Radius of Gyration

$$R_g^2 = \frac{\int_{V_r} \Delta\rho(r) r^2 dV_r}{\int_{V_r} \Delta\rho(r) dV_r}$$

R_g is the quadratic mean of distances to the centre of mass weighted by the contrast of electron density

R_g is an index of non sphericity

Sphere: $R_g = \sqrt{\frac{3}{5}} R$

Cylinder (D,H)

$$R_g = \sqrt{\frac{D^2}{8} + \frac{H^2}{12}}$$

Smallest R_g for a given volume

Ellipsoid of revolution (a,b)

$$R_g = \sqrt{\frac{2a^2 + b^2}{5}}$$

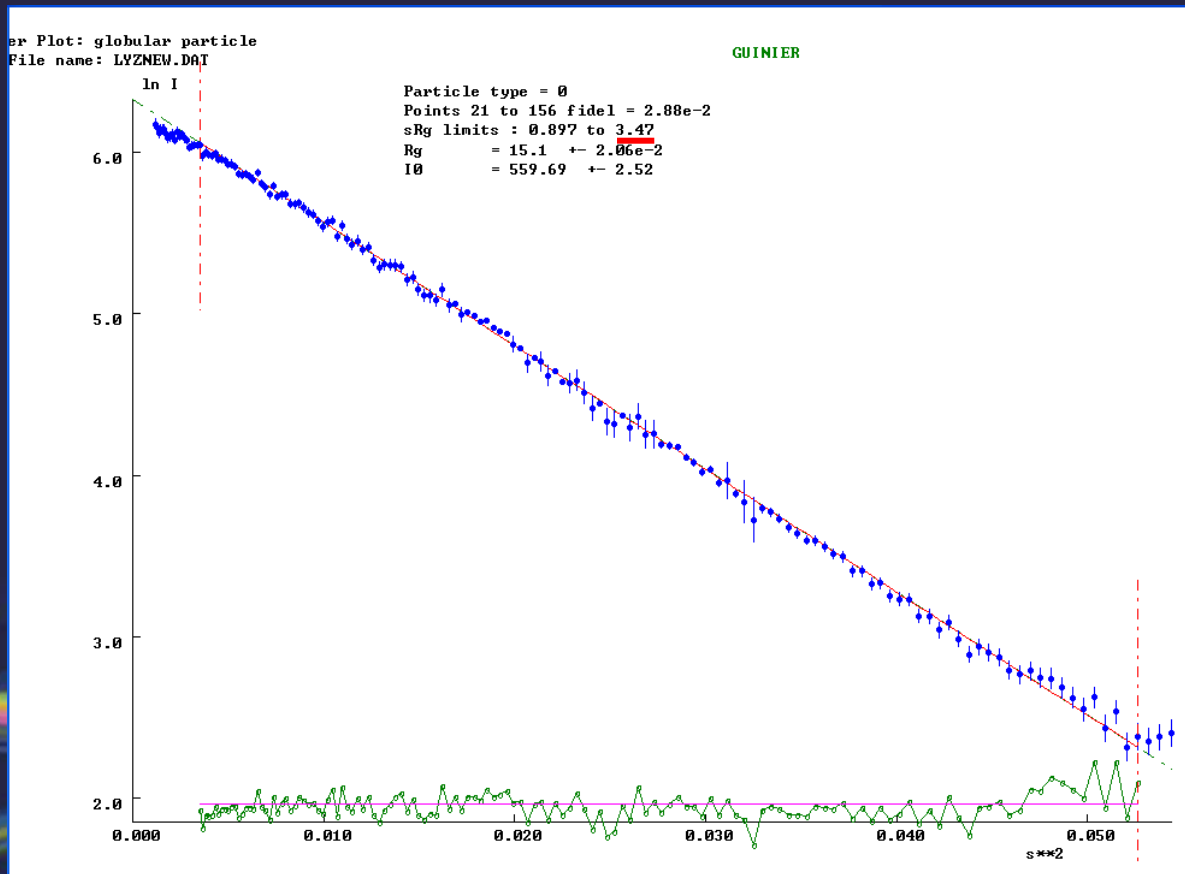


Guinier plot

$$I(q) = I(0)e^{-\frac{q^2 R_g^2}{3}}$$

$$\ln I(q) = \ln I(0) - \frac{q^2 R_g^2}{3}$$

- Linear regression
- ✓ Slope $\rightarrow R_g^2$
 - ✓ Intercept $\rightarrow I(0)$
 $\rightarrow M_w$



qR_g limit to 3.47

$$R_g = 15.1 \text{ \AA}$$

It is very unlikely that any considerable progress will be made in the future.

André Guinier on theory of small-angle scattering (Guinier 1969)

Pair distribution function

$$p(r) = \frac{1}{(2\pi)^2} \int_0^\infty I(q)qr \sin(qr) dq$$

In theory, very easy calculation

Problem:

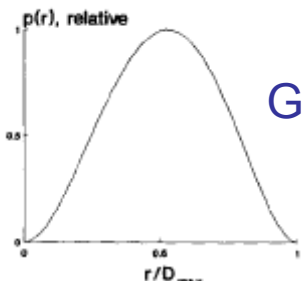
The intensity :

- ✓ only known over q_{\min} - q_{\max} (Detector size)
- ✓ affected by experimental errors

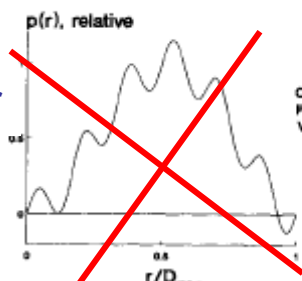
⇒ Fourier transform of incomplete and noisy data is a ill-posed problem

Solution: Indirect Fourier Transform

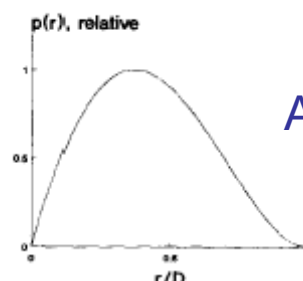
Pair distance distribution function



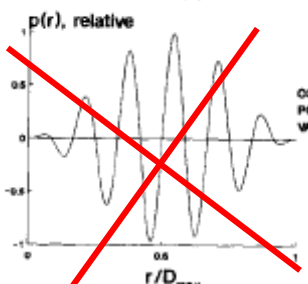
Globular



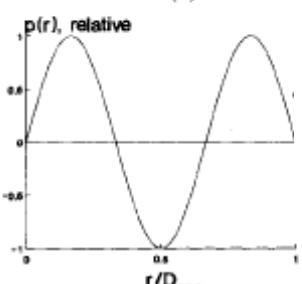
OSCILL 2.28
POSITV 0.99
VALCEN 0.96



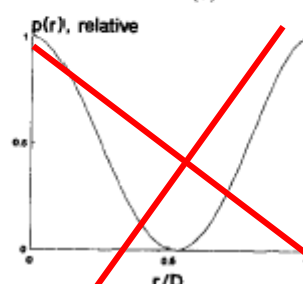
Anisotropic particle



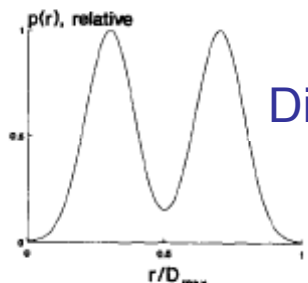
OSCILL 12.0
POSITV 0.71
VALCEN 0.46



Bilayer



OSCILL 1.20
POSITV 1.00
VALCEN 0.26



Dimers

J. Appl. Cryst. (1992). 25, 495–503

Determination of the Regularization Parameter in Indirect-Transform Methods Using Perceptual Criteria

By D. I. SVERGUN*†

GKSS Research Center, GKSS-WS, 2054 Geesthacht, Germany

Radius of Gyration and $I(0)$

$$I(0) = 4\pi \int_0^{\infty} p(r) dr$$

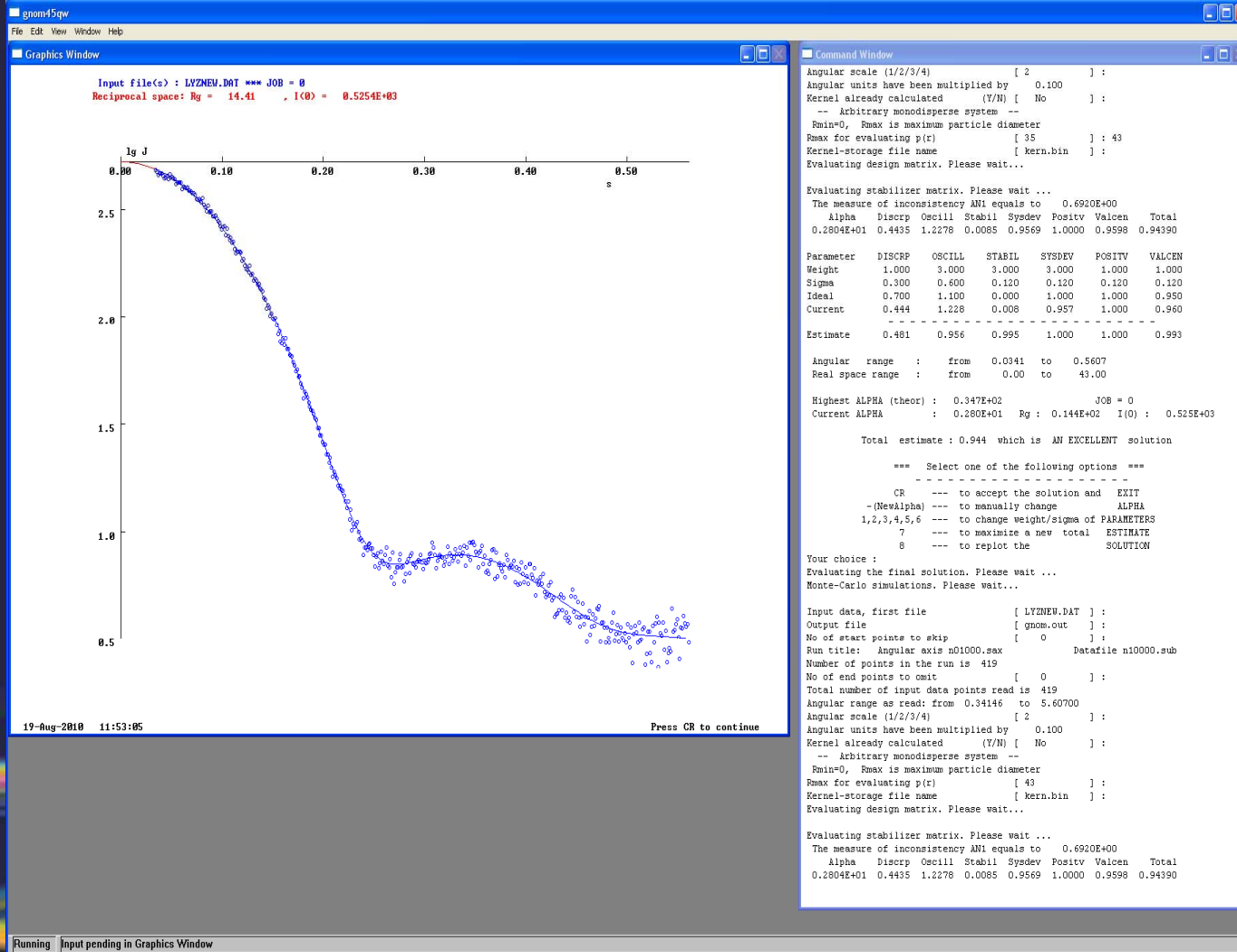
$$R_g^2 = \frac{1}{2} \frac{\int_0^{\infty} r^2 p(r) dr}{\int_0^{\infty} p(r) dr}$$

This estimate of R_g makes use of the whole scattering curve
And is less sensitive to interactions or to the presence
Of a small fraction of oligomers

Cross-check with the Guinier plot

GNOM

<http://www.embl-hamburg.de/ExternalInfo/Research/Sax/gnom.html>



Input file: ASCII
Output file: Ascii but results appended to the file

Number of point to omit:
Due to subtraction issues closed to the beamstop
Or to low concentration at higher q

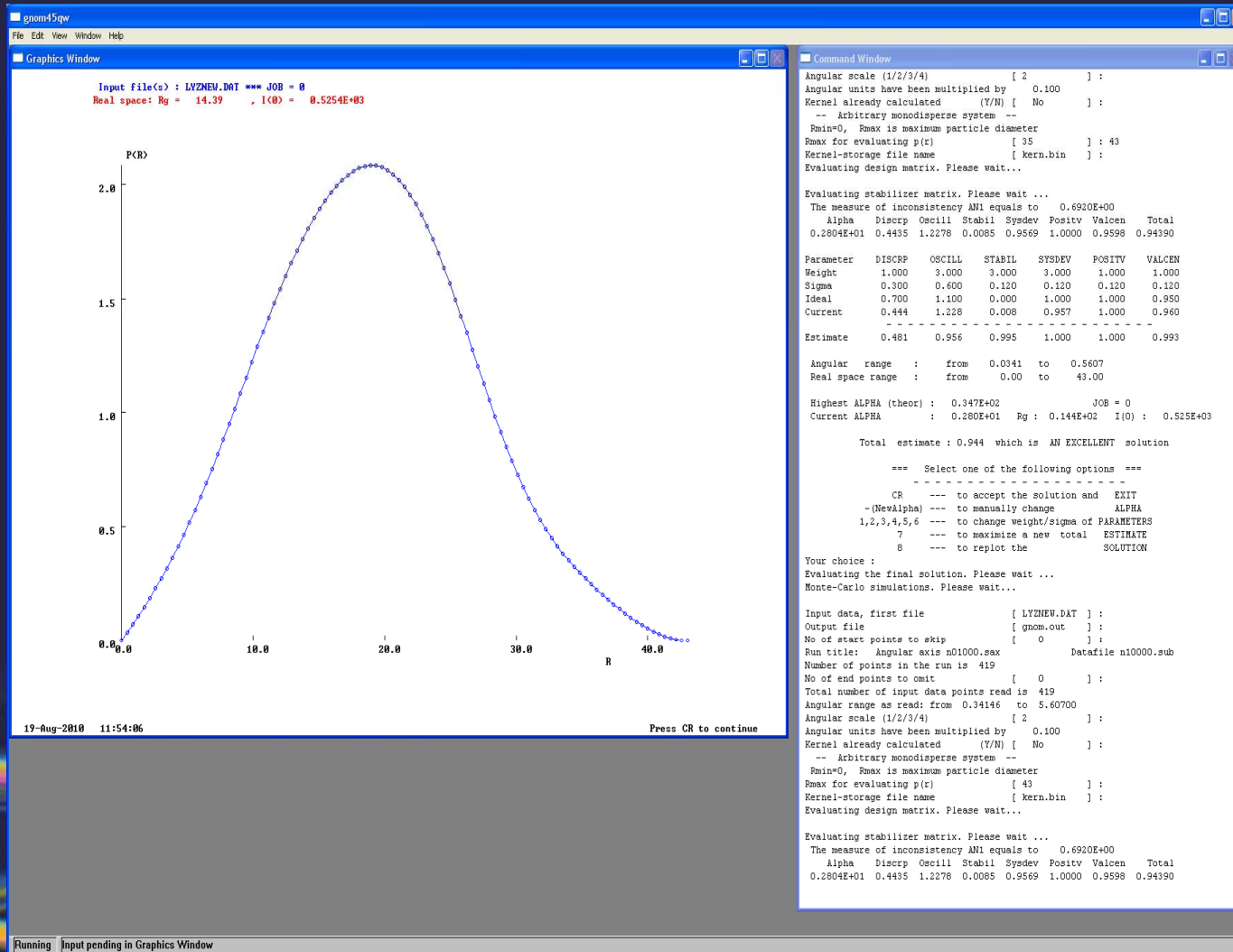
Rmax: Maximum distance of the particle
Here, Rmax=43Å

Other parameters, default values.



GNOM

<http://www.embl-hamburg.de/ExternalInfo/Research/Sax/gnom.html>



$R_g = 14.39 \text{ \AA}$
 $I(0) = 525.0$

Guinier plot:
 $R_g = 15.1 \text{ \AA}$
 $I(0) = 559.0$

- P(r):
- Smooth at Rmax
 - Only positive
 - Peak maximum centered



GNOM

<http://www.embl-hamburg.de/ExternalInfo/Research/Sax/gnom.html>

```
gnom45qw
File Edit View Window Help

Command Window
Real space range : from 0.00 to 43.00
Highest ALPHA (theor) : 0.347E+02 JOB = 0
Current ALPHA : 0.280E+01 Pg : 0.144E+02 I(O) : 0.525E+03
Total estimate : 0.944 which is AN EXCELLENT 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 ...
Monte-Carlo simulations. Please wait...

Input data, first file [ LYZNEW.DAT ] :
Output file [ gnom.out ] :
No of start points to skip [ 0 ] :
Run title: Angular axis n01000.sax Datafile n10000.sub
Number of points in the run is 419
No of end points to omit [ 0 ] :
Total number of input data points read is 419
Angular range as read: from 0.34146 to 5.60700
Angular scale (1/2/3/4) [ 2 ] :
Angular units have been multiplied by 0.100
Kernel already calculated (Y/N) [ No ] :
-- Arbitrary monodisperse system --
Rmin=0, Rmax is maximum particle diameter
Rmax for evaluating p(r) [ 43 ] :
Kernel-storage file name [ kern.bin ] :
Evaluating design matrix. Please wait...

Evaluating stabilizer matrix. Please wait ...
The measure of inconsistency AMI equals to 0.6920E+00
Alpha Discrp Oscill Stabll SysDev Positr Valcen Total
0.2804E+01 0.4435 1.2278 0.0085 0.9569 1.0000 0.9598 0.94190

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 0.960
Current 0.444 1.228 0.008 0.957 1.000 0.960

Estimate 0.481 0.956 0.995 1.000 1.000 0.993

Angular Range : from 0.0341 to 0.5607
Real space range : from 0.00 to 43.00

Highest ALPHA (theor) : 0.347E+02 JOB = 0
Current ALPHA : 0.280E+01 Pg : 0.144E+02 I(O) : 0.525E+03
Total estimate : 0.944 which is AN EXCELLENT 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

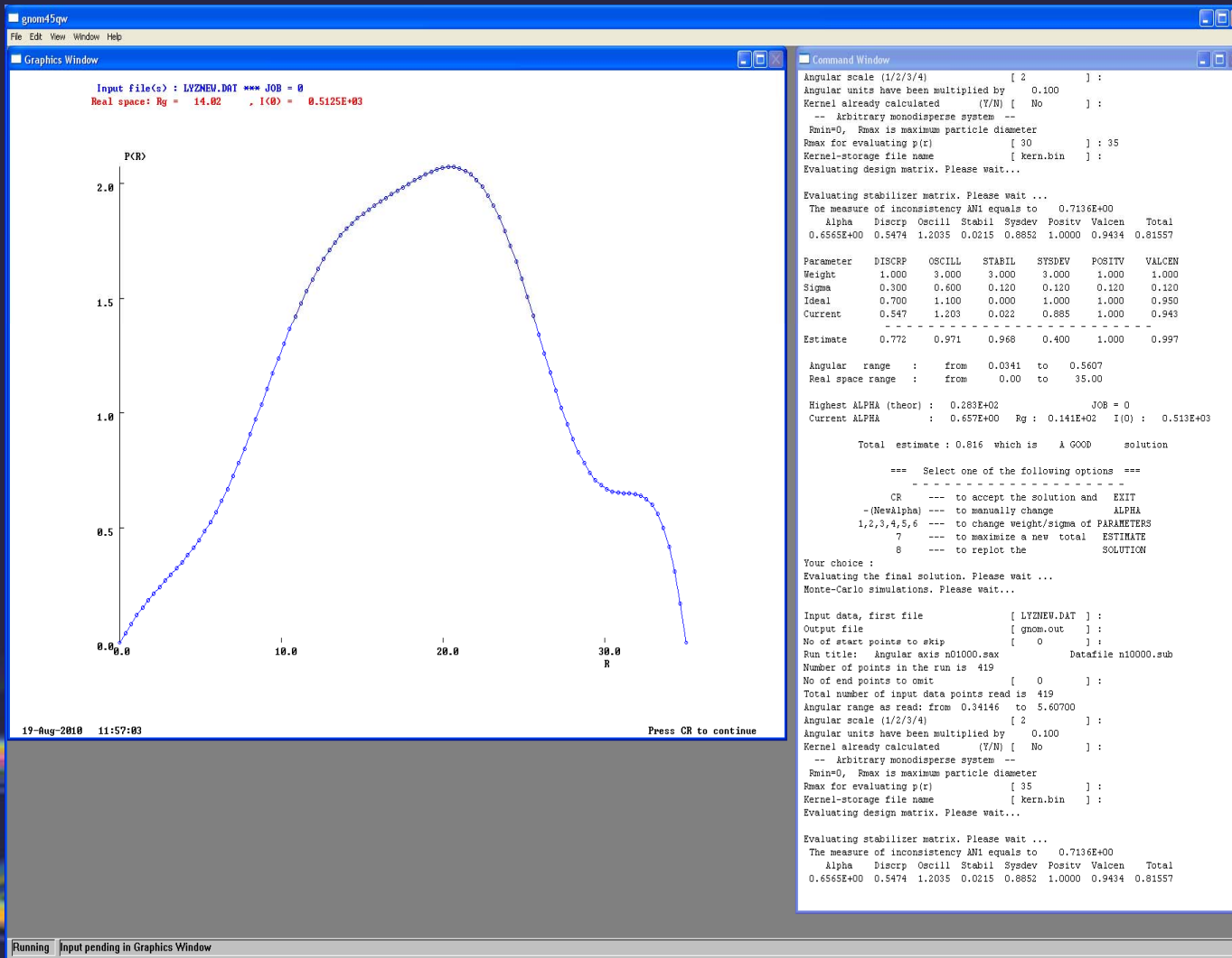
Your choice :
```

Total estimate:
Between 0 and 1
1 is the best
solution



GNOM

<http://www.embl-hamburg.de/ExternalInfo/Research/Sax/gnom.html>



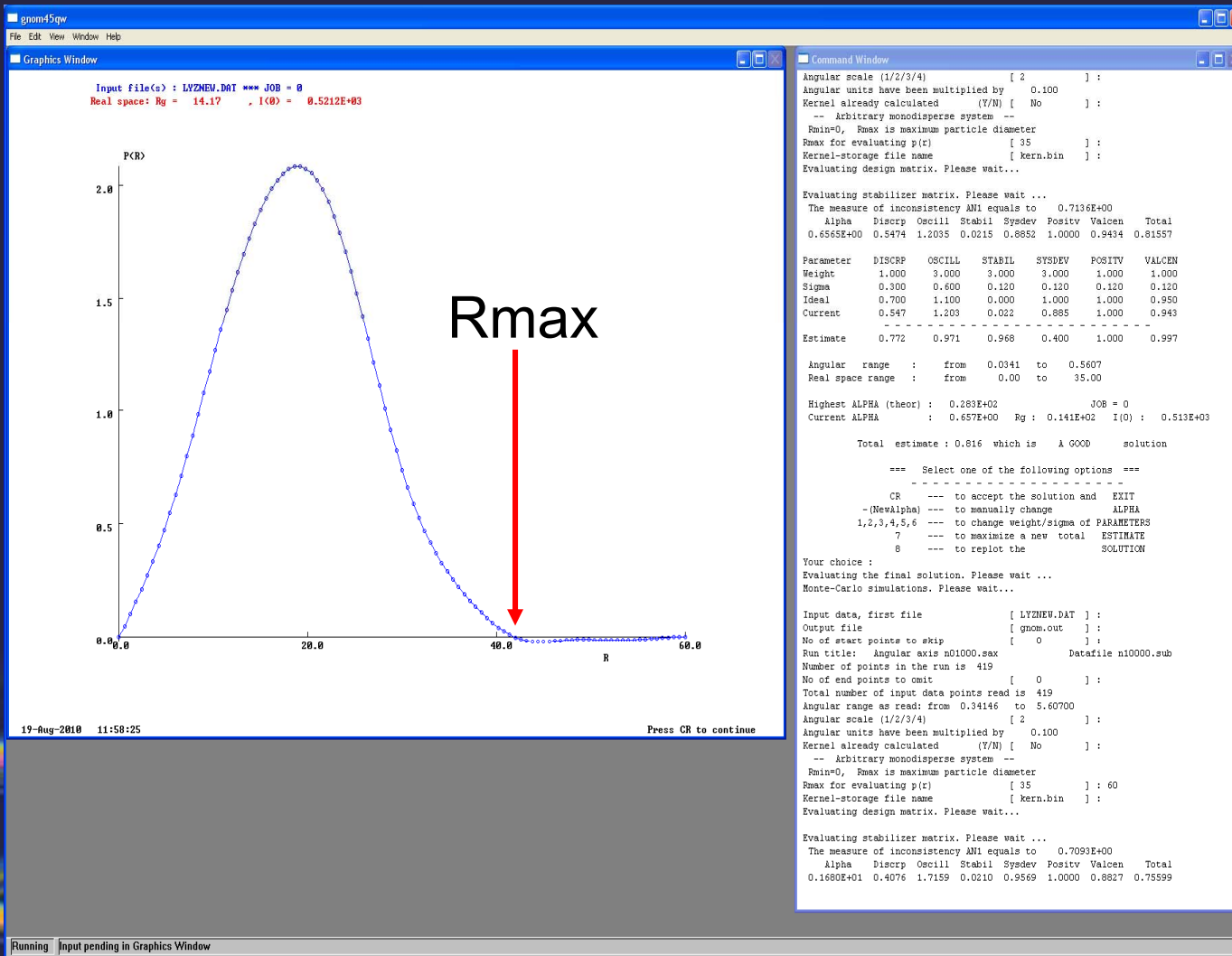
$R_{max} = 35\text{\AA}$

$P(r)$ not smooth
at R_{max}
 $\Rightarrow R_{max}$ must
be bigger



GNOM

<http://www.embl-hamburg.de/ExternalInfo/Research/Sax/gnom.html>



$P(r)$ negative a high r
 \Rightarrow The particle must be smaller. R_{max} is likely to be at the intercept of the $p(r)$ with the x-axis



Using crystallographic structure

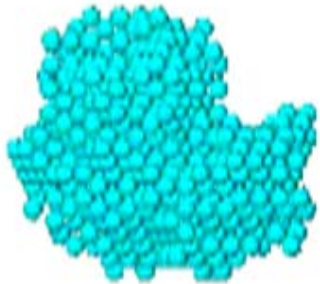
- To validate theoretical model
- To analyse similarities/difference in solution and in the crystal
- To build a complex from subunits

CRYSOL

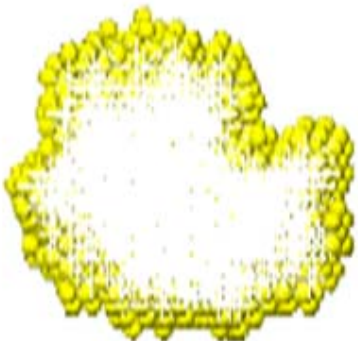
$$I(q) = \left\langle |F(q)|^2 \right\rangle_{\Omega} = \left\langle |A(q) - \rho_s E(q) + \partial \rho_b B(q)|^2 \right\rangle_{\Omega}$$



A(s): Atomic scattering in vacuum



E(s): Scattering from the excluded volume

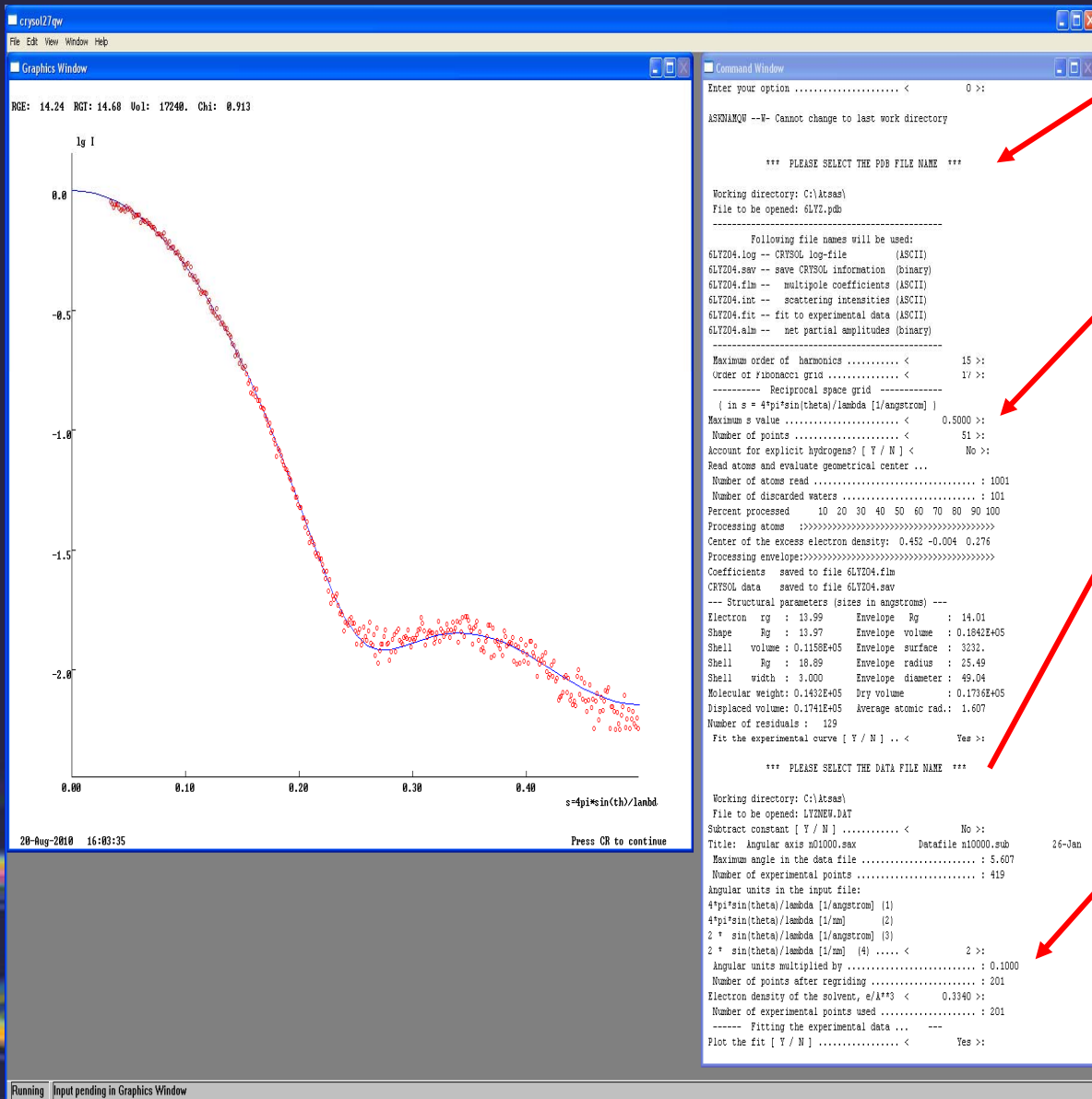


B(s): scattering from the hydration shell

Svergun D.I., Barberato C. & Koch M.H.J.
(1995) *J. Appl. Cryst.*, **28**, 768-773.



CRY SOL



Input file: PDB file

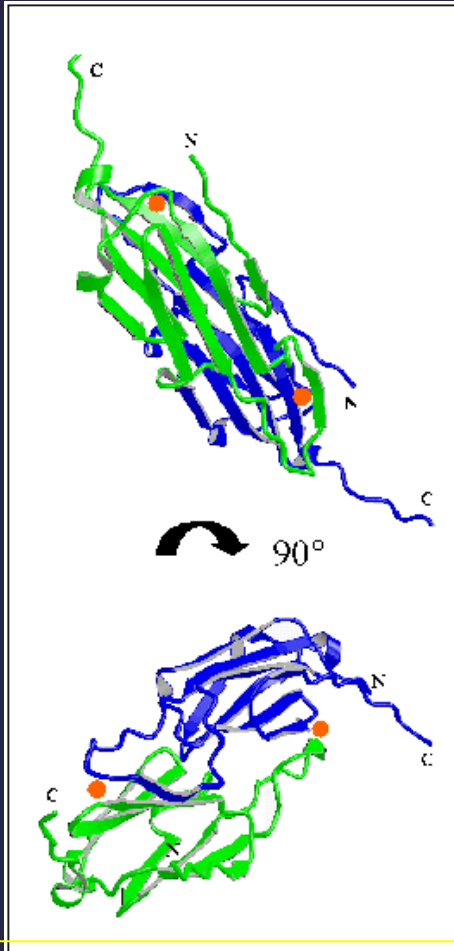
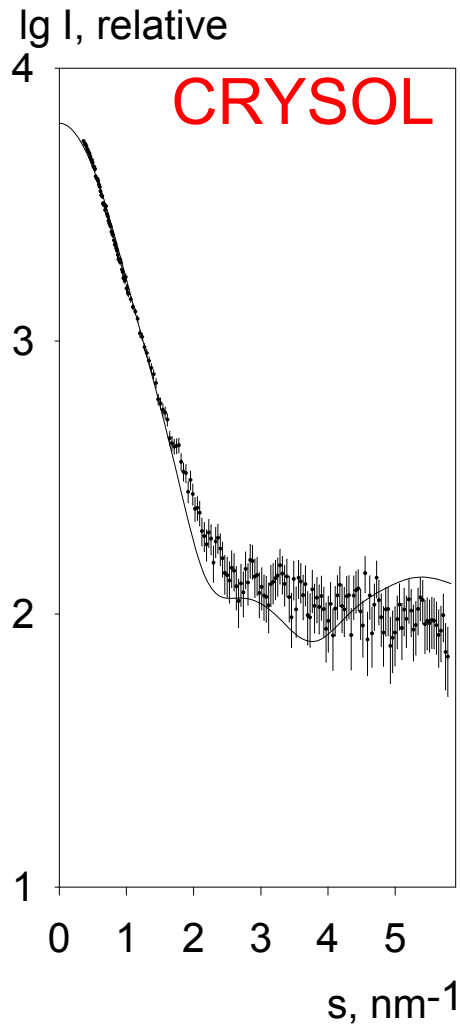
q-range: must be checked in the data file.

Input file: Ascii file

Electron density: to be calculated if highly salted buffer, By default, electronic density of water.



α B-crystallin domain

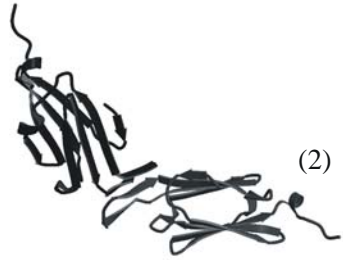


Crystallographic
structure of
MjHSP 16.5

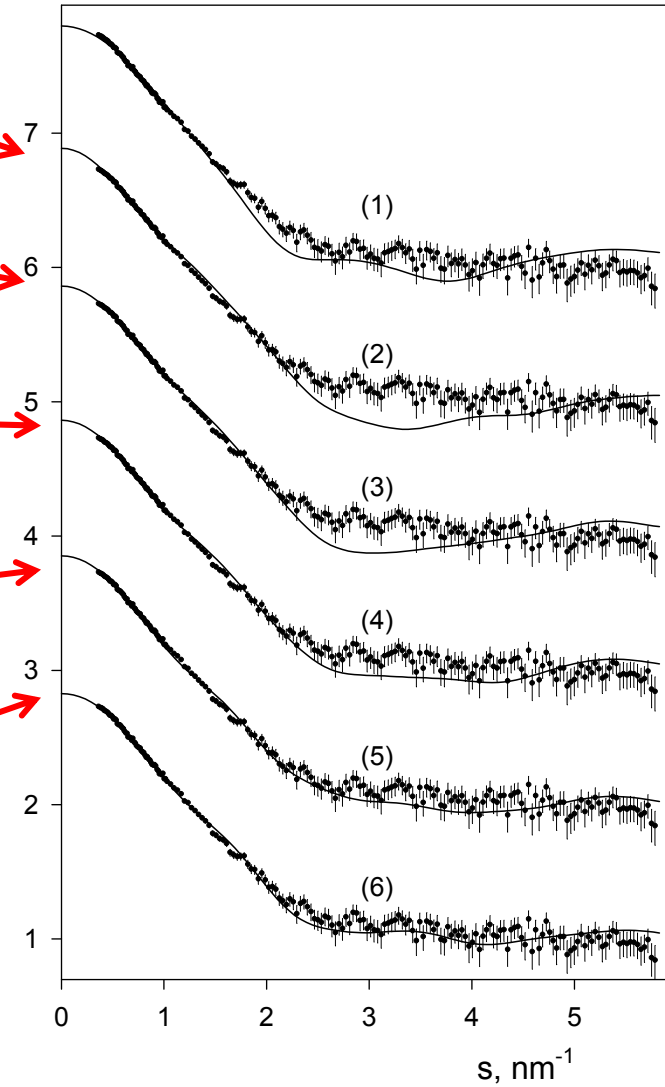
Feil I.K., Malfois M., Hendle J., van der Zandt H., Svergun D.I.
(2001)
J. Biol. Chem., **276** (15), 12024-12029



Dimer modelling

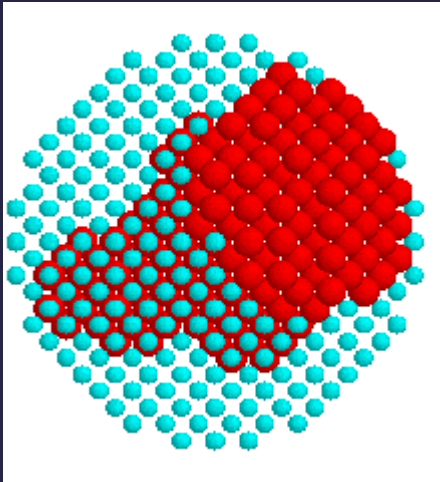


Ig I, relative



DAMMIN/DAMMIF

<http://www.embl-hamburg.de/ExternalInfo/Research/Sax/dammin.html>



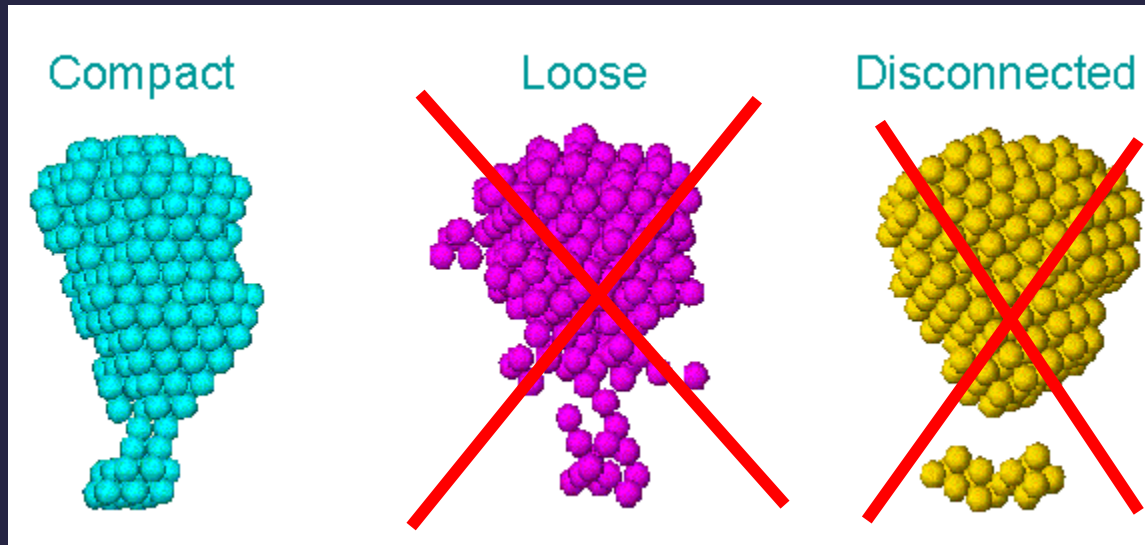
- From D_{max} determined in GNOM, a sphere is created.
- The sphere is filled by dummy atoms.
- The dummy atoms belong to the protein or belong to the solvent.
- The number of dummy atoms is about 1000
- Can describe complex shape
- No unique shape restoration unless constraints

Svergun, D.I. (1999)
Biophys. J. **76**, 2879-2886



DAMMIN CONSTRAINTS

<http://www.embl-hamburg.de/ExternalInfo/Research/Sax/dammin.html>



DAMMIN/DAMMIF

<http://www.embl-hamburg.de/ExternalInfo/Research/Sax/dammin.html>

```
Command Window

***  Ab initio shape determination by simulated  ***
***  annealing using a single phase dummy atoms model  ***
***  Win 9x/NT, UNIX/Linux/Mac release version 5.3  ***
***  Last modified --- 06/05/09 19:00  ***
***  Please reference: D.Svergun (1999). Biophys. J.  ***
***  76, 2879-2886.  ***
***  Copyright (c) ATSAS Team  ***
***  EMBL, Hamburg Outstation, 1999 - 2009  ***

===== DAMMIN53 started on 20-Aug-2010 16:20:44

Mode: <[F]>ast, [S]low, [J]ag, [K]eep, [E]xpert < Fast >:
Log file name ..... < .log >: t1

*** PLEASE SELECT THE INPUT FILE NAME ***

Working directory: C:\atsas\
File to be opened: gnom.out
Project identifier ..... : t1
Enter project description ..... :
Random sequence initialized from ..... : 162052
** Information read from the GNOM file **
Data set title: Angular axis n01000.sax Datafile n10000.sub

Raw data file name: LYZNEW.DAT
Maximum diameter of the particle ..... : 43.00
Solution at Alpha = 0.280E+01 Rg : 0.144E+02 I(0) : 0.525E+03
Radius of gyration read ..... : 14.40
Number of GNOM data points ..... : 446
Angular units in the input file:
4*pi*sin(theta)/lambda [1/angstrom] (1)
4*pi*sin(theta)/lambda [1/nm ] (2) < 1 >:
Maximum s value [1/angstrom] ..... : 0.5607
Number of Shannon channels ..... : 7.675
Portion of the curve to be fitted ..... < 1.000 >:
Number of knots in the curve to fit ..... : 20
*** Warning: constant reduced to avoid oversubtraction
A constant was subtracted ..... : 2.537
Maximum order of harmonics ..... : 10
Initial DAM: type S for sphere [default],
E for ellipsoid, C for cylinder, P for parallelepiped
or start file name ..... < .pdb >:
Symmetry: P1...19 or Pn2 (n=1,...,12)
or P23 or P432 or PICO ..... < P1 >:
Sphere diameter [Angstrom] ..... : 43.00
Packing radius of dummy atoms ..... : 1.500
Radius of the sphere generated ..... : 21.50
Number of dummy atoms ..... : 2171
Number of equivalent positions ..... : 1
Expected particle shape: <P>rolate, <O>blate,
or <U>nknown ..... < Unknown >:
```

Input name: GNOM file
Dmax has to be determined
before running dammin/dammif

Initial shape

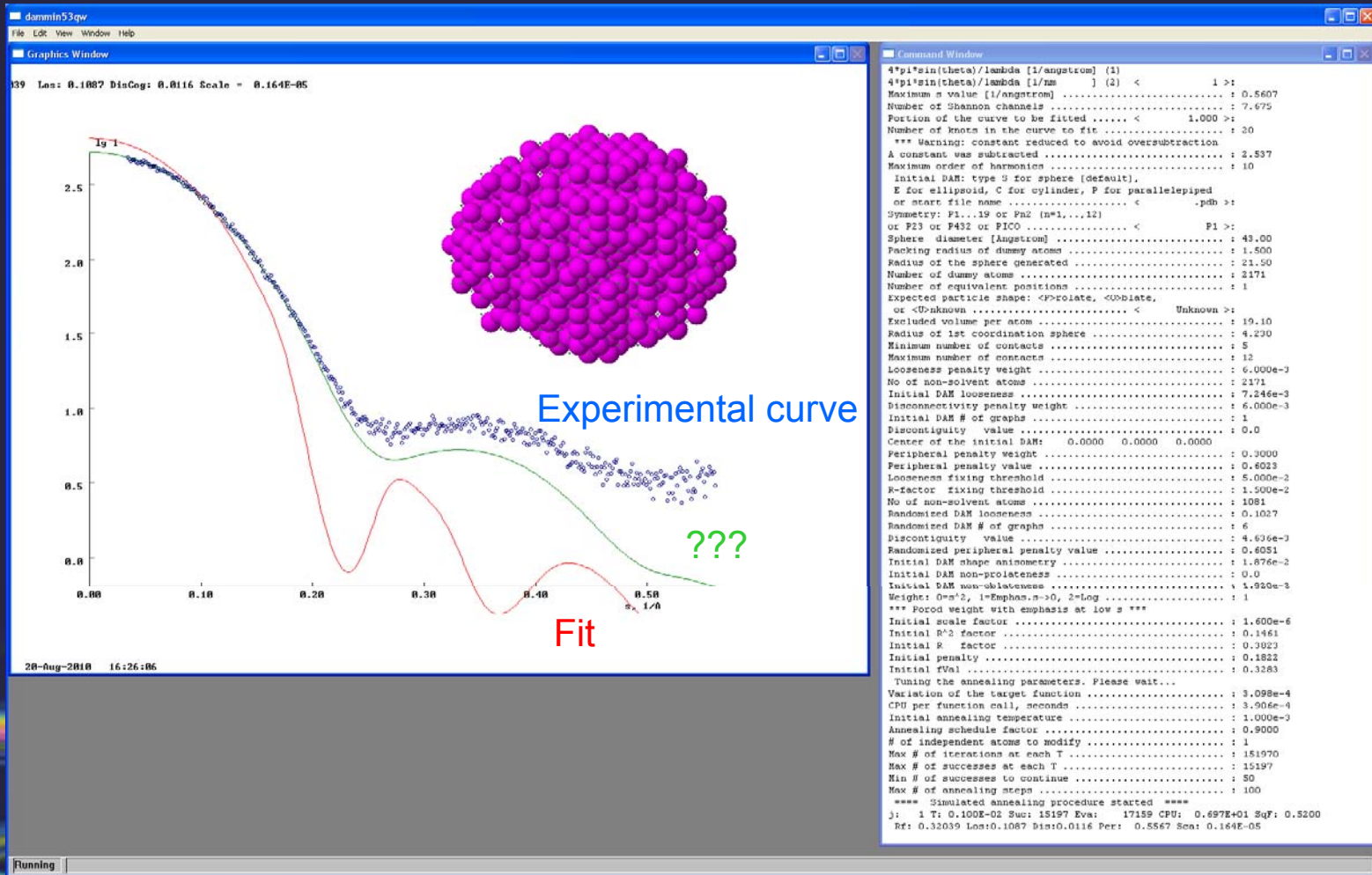
Symetry

Particle shape



DAMMIN/DAMMIF

<http://www.embl-hamburg.de/ExternalInfo/Research/Sax/dammin.html>



nond

DAMMIN/DAMMIF

<http://www.embl-hamburg.de/ExternalInfo/Research/Sax/dammin.html>

Shape restoration: Equivalent to the excluded volume in CRY SOL

$$I(q) = \left\langle |F(q)|^2 \right\rangle_{\Omega} = \left\langle |A(q) - \rho_s E(q) + \partial \rho_b B(q)|^2 \right\rangle_{\Omega}$$

DAMMIN/DAMMIF **DO NOT** take into account the hydration shell

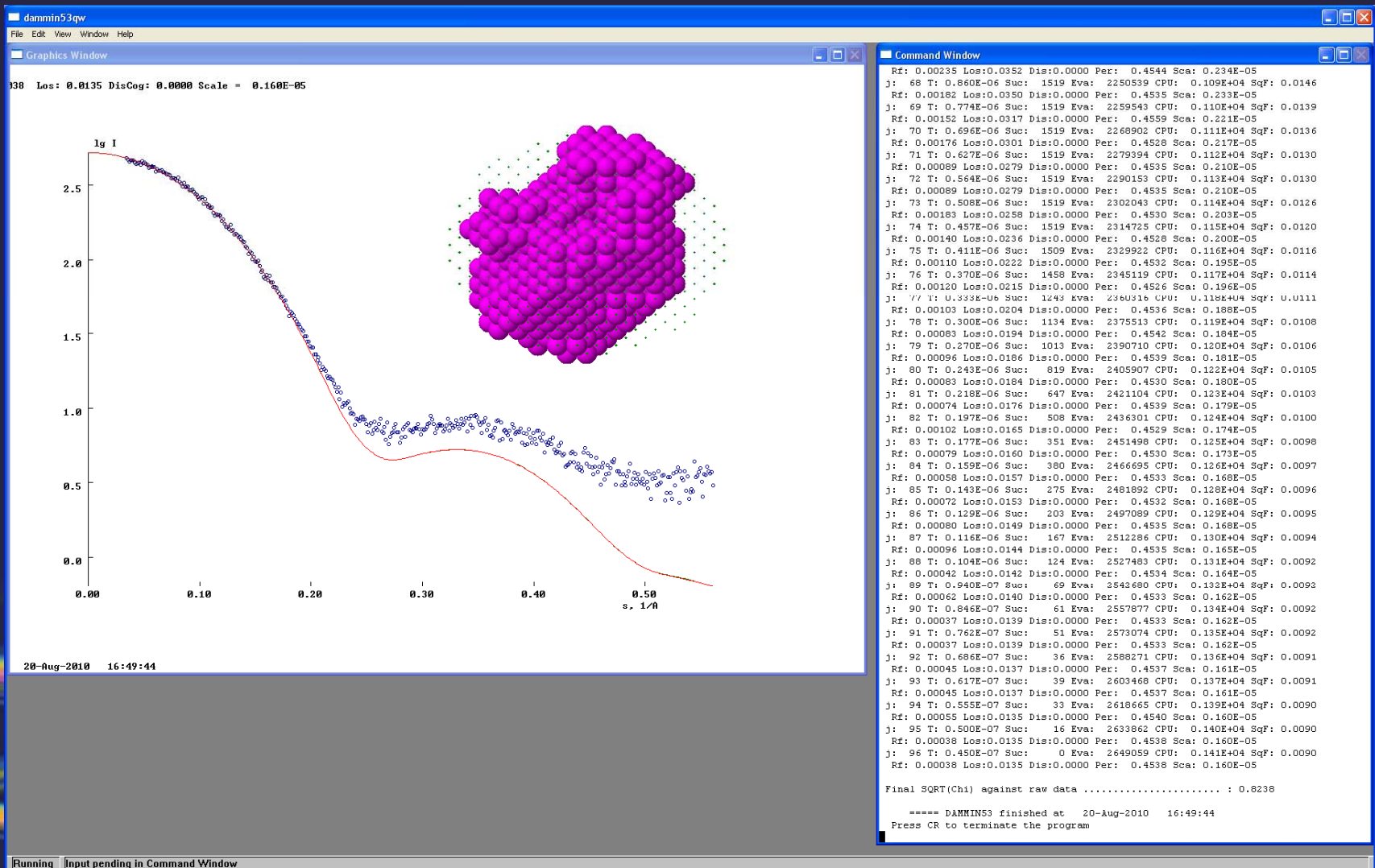
The atomic scattering in vacuum is subtracted from the experimental data

A constant is subtracted from the data to force the s^{-4} decay of the intensity at higher angles.

Green curve is the curve to be fitted.

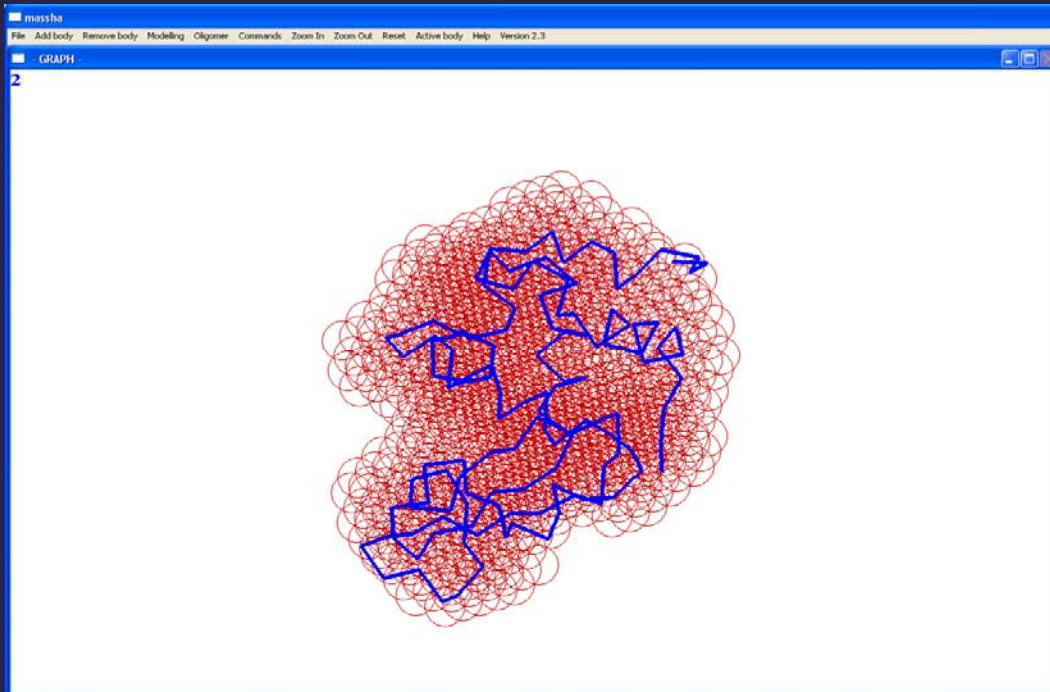
DAMMIN/DAMMIF

<http://www.embl-hamburg.de/ExternalInfo/Research/Sax/dammin.html>



DAMMIN/DAMMIF

<http://www.embl-hamburg.de/ExternalInfo/Research/Sax/dammin.html>

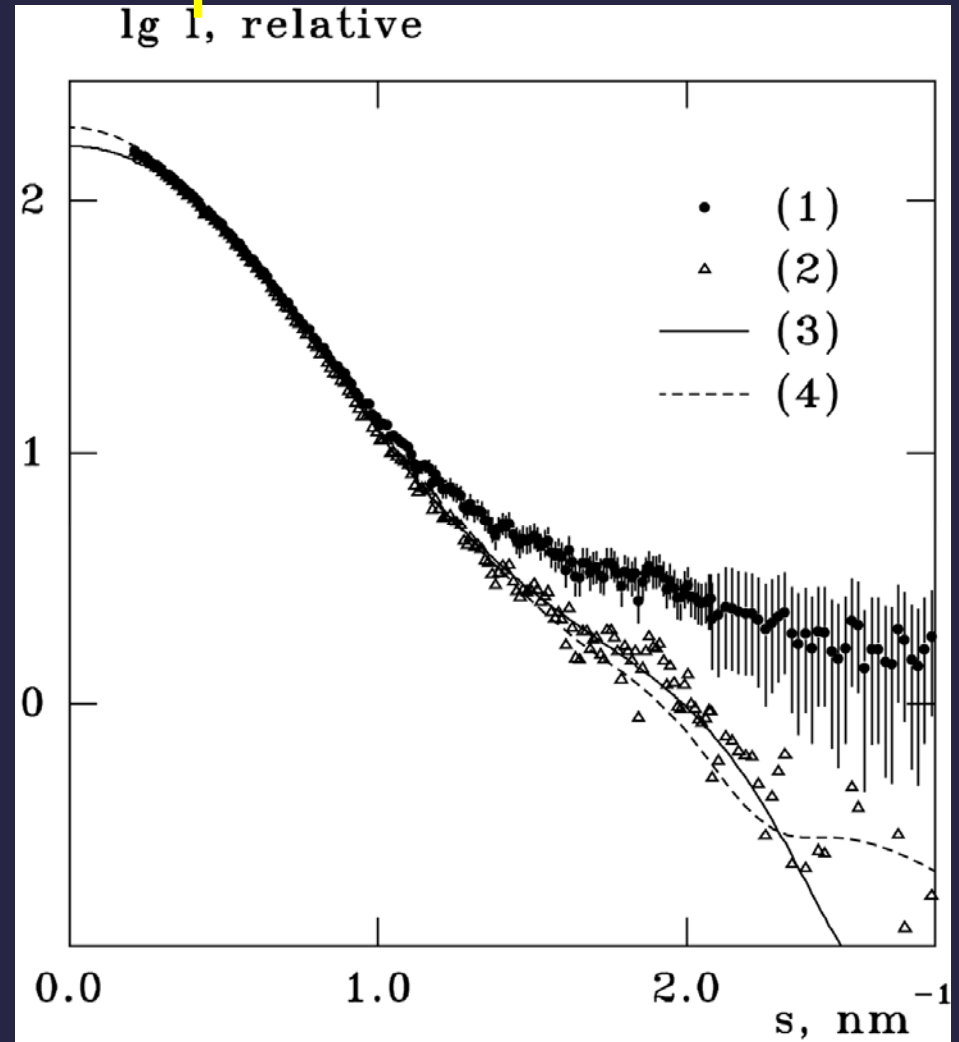
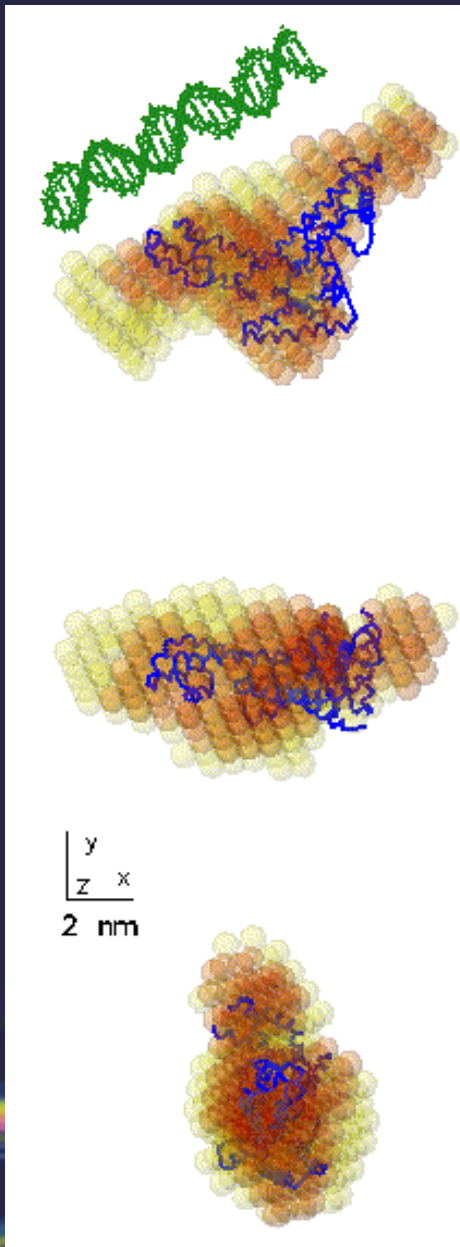


Output file: PDB file

Results can be
visualise with MASHA

$\sigma 54$ Transcription Factor

DAMMIN



1.Svergun D.I., Malfois M., Koch M.H.J., Wigneshweraraj S.R., Buck M.(2000), *J. Biol. Chem.*, **275** (6), 4210-4214

Experimental Considerations

$$I_{particle}(q) = \frac{1}{c} \left[\frac{I_{solution}(q)}{It_{solution}} - \frac{I_{background}(q)}{It_{background}} \right] \frac{1}{D(q)}$$

c: concentration (mg/ml)

D(q): Detector response

We measure

- The solution containing the particles of interest
- The background (Cell with the solvent/buffer)

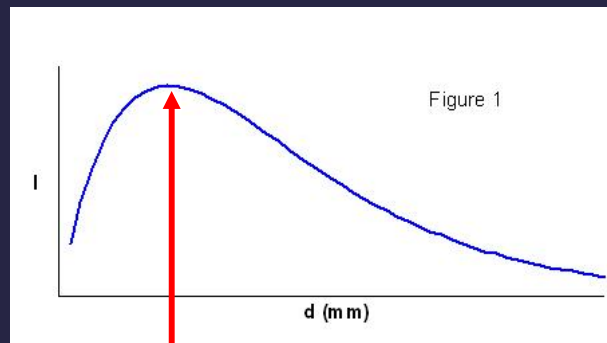
Eliminate contributions from parasitic background

Thickness of the sample

The maximum scattering intensity is achieved by selecting the optimal thickness of the sample.

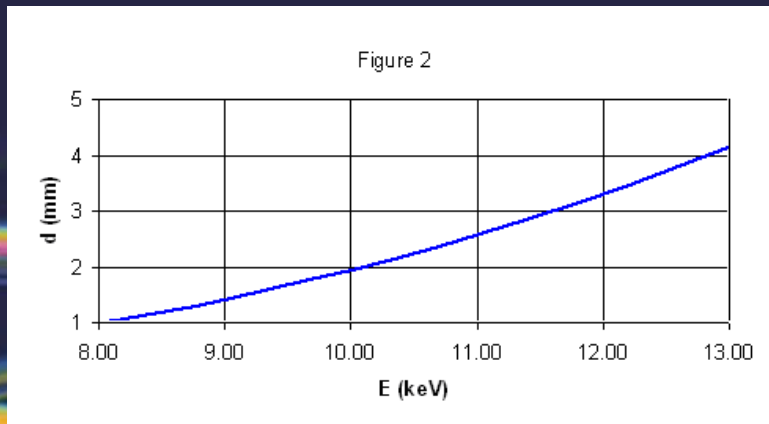
The scattering intensity can be formulated as: $I(q) \propto de^{-\mu d}$

d : the thickness of the sample
 μ : linear absorption coefficient



Optimal thickness

$$\frac{\partial I}{\partial d} = k(1 - \mu d)e^{-\mu d} \Rightarrow d_{optimal} = \frac{1}{\mu}$$



The linear absorption coefficient can be calculated or measured.

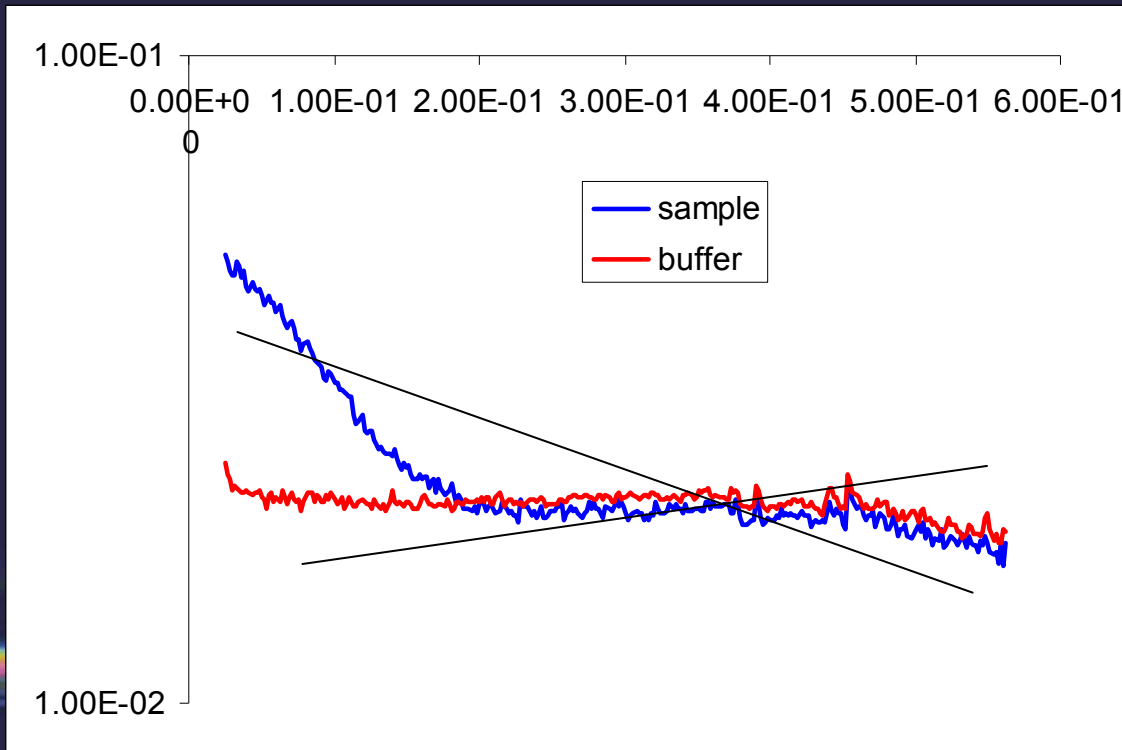
Optimal thickness of the sample (water in this case) as a function of the energy.

Experiment

Liquid cell or capillaries?

Liquid cell (Same windows and same thickness)

Capillaries (Different thickness)



Capillary thickness smaller for the sample than for the buffer

The scattering of the sample must be higher or equal to the scattering of the buffer

Liquid cell or capillaries?

$$I(q) \propto de^{-\mu d}$$

Can be measured by a diode located
in the beamstop

Unknown

For the liquid cell, the thickness is the same for the sample or for the buffer so there is no need to measure an empty cell.

The empty capillary for the sample and capillary for the buffer must be measured => 2 extra measurements

Experimental Considerations

Sample volume: 0.120ml with liquid cell

Sample volume: 0.05ml with HPLC