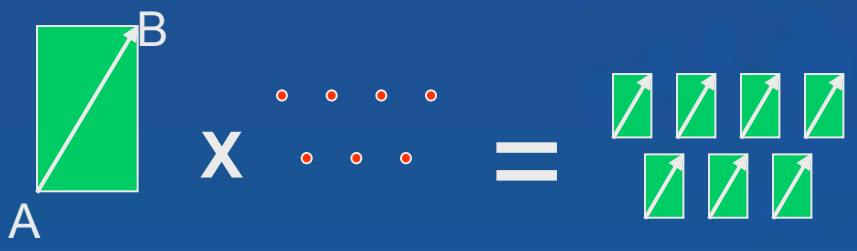
Bio-SAXS @ ID14-EH3

Adam Round

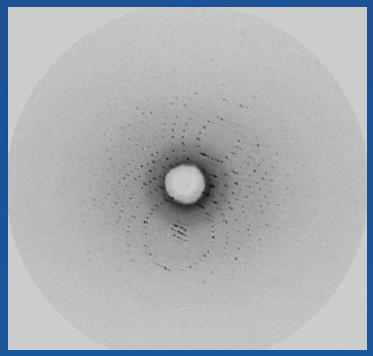
Contents

- What can be obtained from Bio-SAXS
  - Measurable parameters
  - Modelling strategies
- How to collect data at Bio-SAXS
  - Procedure
    - Data collection tests
    - Data Verification and quality control
    - Modelling and analysis
- Future Improvements for Bio-SAXS

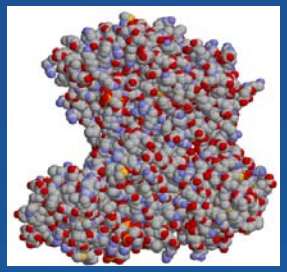
Crystals versus solution



Protein molecule — Lattice — Crystal



Fourier Transformation



3-dim information about:
Length of the vector **AB**
Orientation of vector **AB**

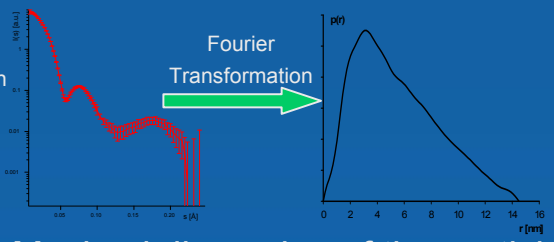
3-dim protein structure with atomic resolution

Protein molecule — Solution — Random

Radial integration

Scattering function

Pair distance distribution function



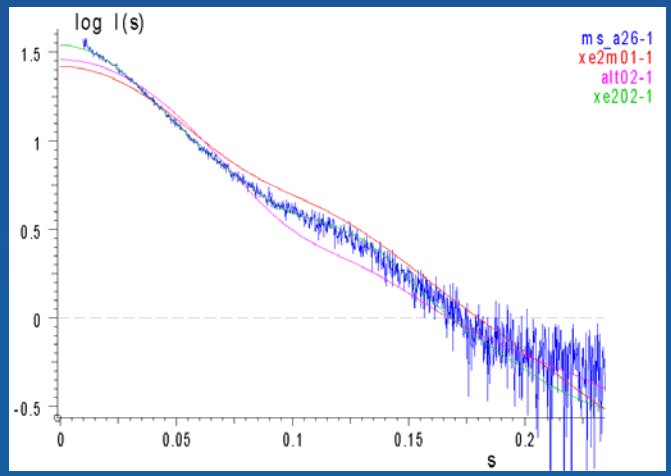
Maximal dimension of the particle
Mean value
Radius of gyration

Crystals Structure Validation

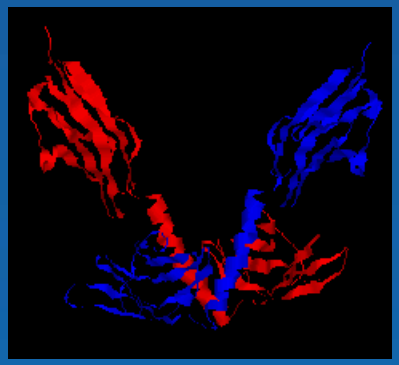
Based on existing high resolution protein structures from crystallography or NMR the corresponding SAXS profile can be calculated.

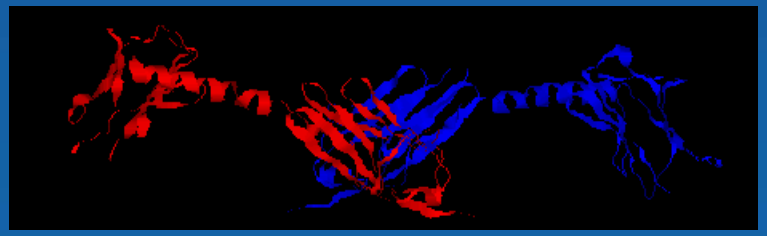
$$I(s) = \left\langle |A(\mathbf{s})|^2 \right\rangle_\Omega = \left\langle |A_a(\mathbf{s}) - \rho_s A_s(\mathbf{s}) + \delta\rho_b A_b(\mathbf{s})|^2 \right\rangle_\Omega$$

This allows one to verify high resolution structures especially in the case of multimeric protein and larger protein complexes.



Example: The muscle protein titin was found in different conformations in the crystal.





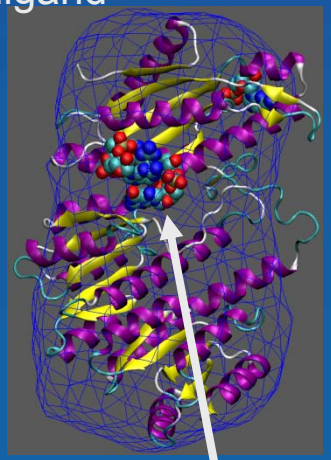
CRYSOL (X-rays): Svergun et al. (1995). *J. Appl. Cryst.* **28**, 768
***CRYSON* (*neutrons*):** Svergun et al. (1998) *P.N.A.S. USA*, **95**, 2267

Refinement of Rigid Domains

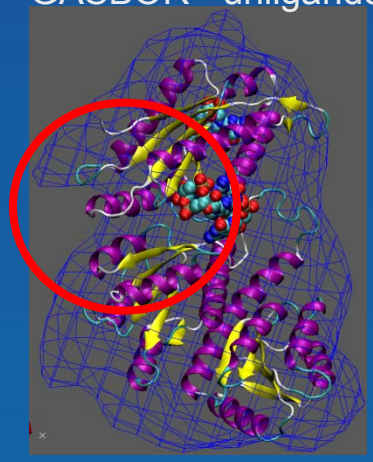
Rigid Body Refinement: Moving protein sub-parts (called domains) as rigid bodies to fit the scattering data

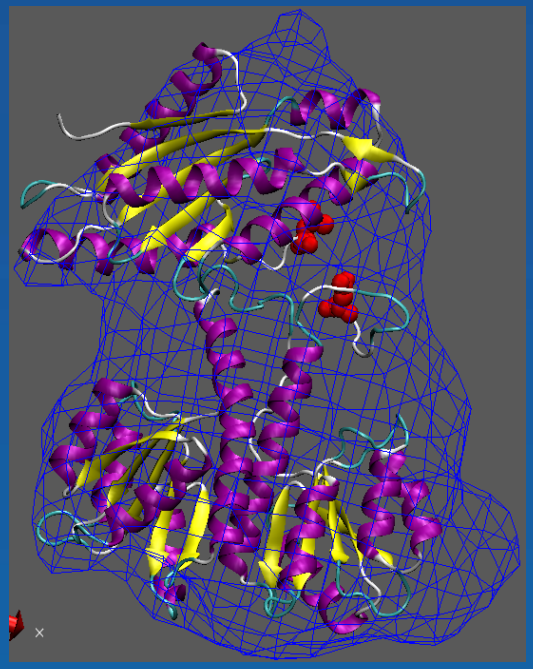
Example: Structural changes upon ligand binding

PX-structure with ligand



SAXS Shape obtained by GASBOR - unliganded state



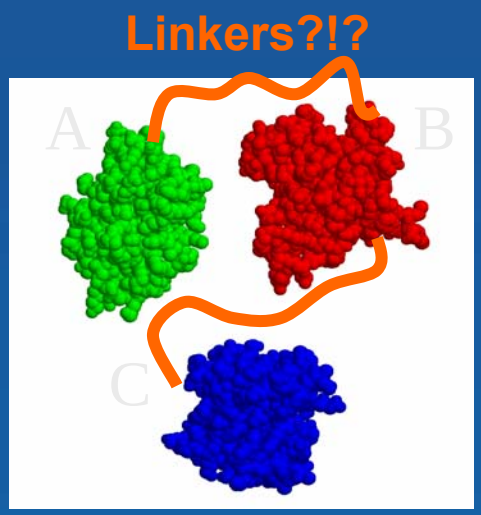


Rigid Body refinement using MASSHA

Adding Missing Linkers

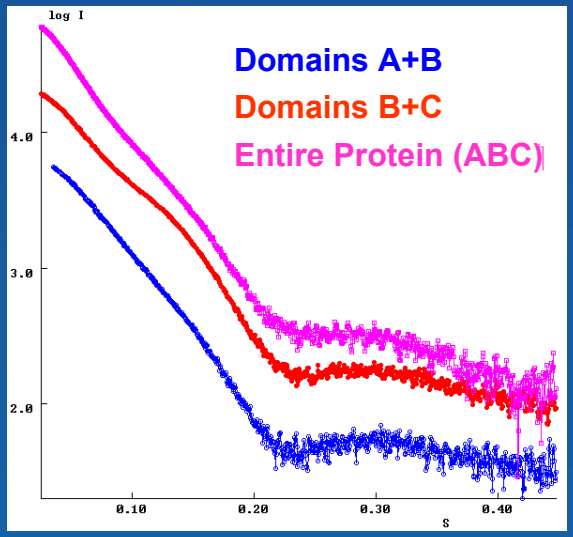
Remodeling of proteins from high resolution fragments/constructs

Program BUNCH (M. Petoukhov; Biophysical J.)



High resolution protein fragments from X-ray crystallography

+ sequence data (TrEMBL/Swissprot)

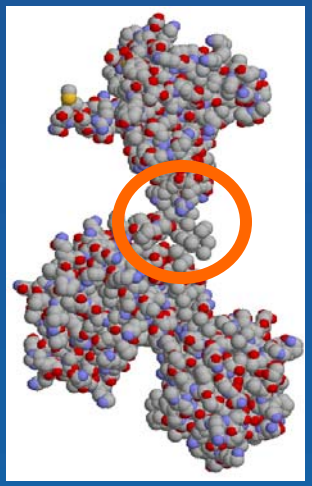


SAXS Data of constructs

AB

BC

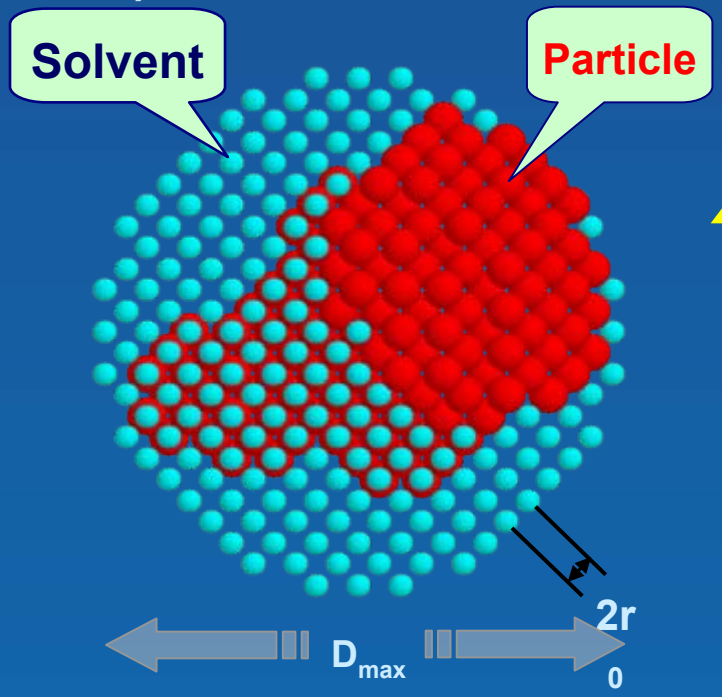
ABC



Model for the entire protein including not resolved linker components

Ab-initio Modelling

A sphere with diameter D_{max} is filled by densely packed beads of radius $r_0 << D_{max}$. A configuration vector X indicates whether the j-th atom belongs to the particle or to the solvent.



Vector of model parameters:

Position (j) = x(j) =

$$\begin{cases} 1 & \text{if particle} \\ 0 & \text{if solvent} \end{cases}$$

(phase assignments)

The number of model parameters $M \approx (D_{max} / r_0)^3 \approx 10^3$ is too large for conventional minimization methods.

A Monte-Carlo type search starting from a random X can be employed to find a configuration that yields the calculated scattering curve fitting the experimental data

Chacón, P. *et al.* (1998) *Biophys. J.* 74, 2760-2775.

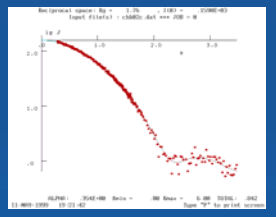
Svergun, D.I. (1999) *Biophys. J.* 76, 2879-2886

Ab-initio Modelling

DAMMIN modelling penalties

Using simulated annealing, finds a compact dummy atoms configuration X that fits the scattering data by minimizing

$$f(X) = \chi^2[I_{\exp}(s), I(s, X)] + \alpha P(X)$$



where χ is the discrepancy between the experimental and calculated curves, *P(X)* is the penalty to ensure compactness and connectivity, $\alpha>0$ its weight.

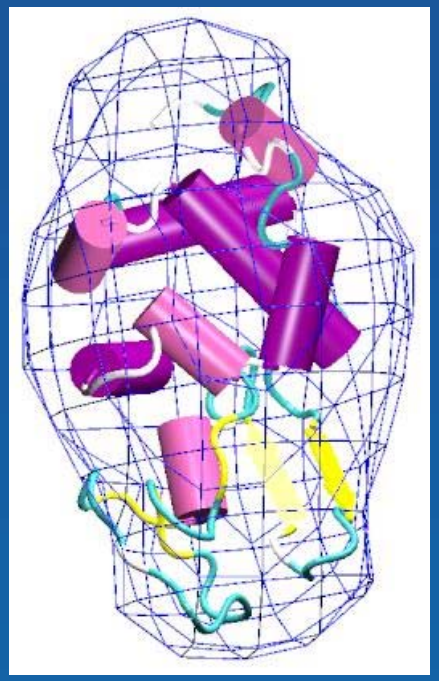
compact

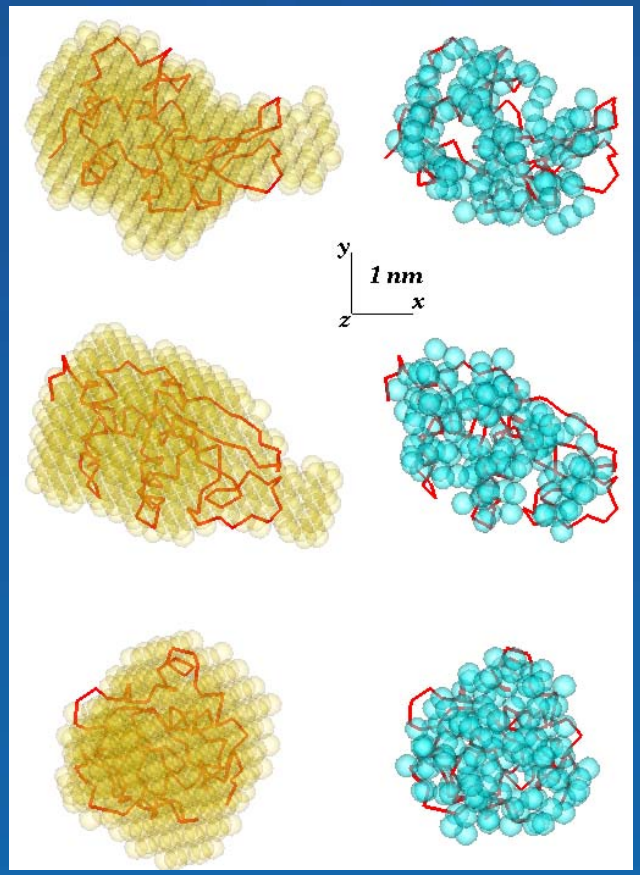
loose

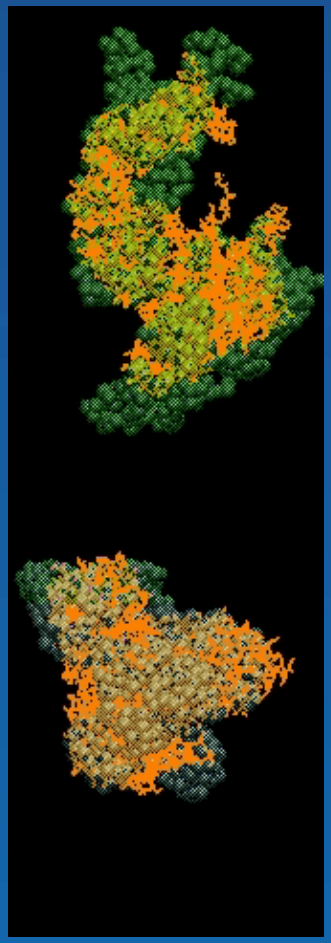
disconnected

Ab-initio Can it be Trusted?

Ab initio bead models compared to high resolution X-ray structures





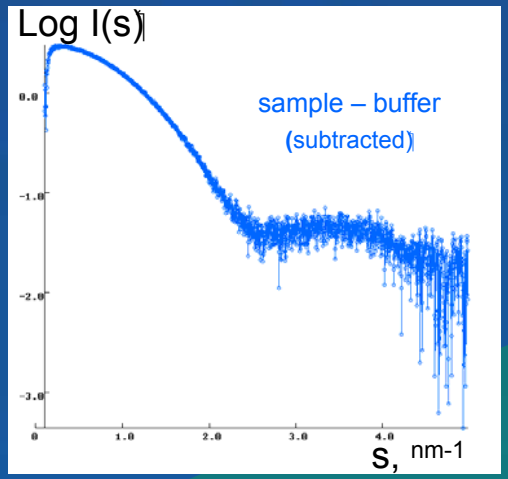


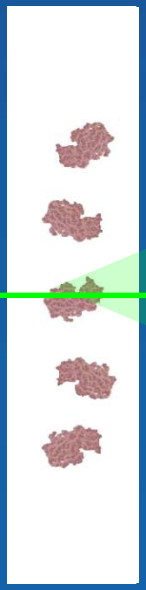
Summary

What can we learn from solution SAXS

- Complete high resolution structure known: validation of crystal structure in solution under physiological conditions
- Ligand binding reactions: Internal structure of multicomponent particles and large macromolecular complexes
- High resolution structure of domains/subunits known: quaternary structure using docking/rigid body refinement
- Incomplete high resolution structure known: probable configuration of missing portions
- Nothing known: *ab initio* low resolution structure

Solution scattering data collection

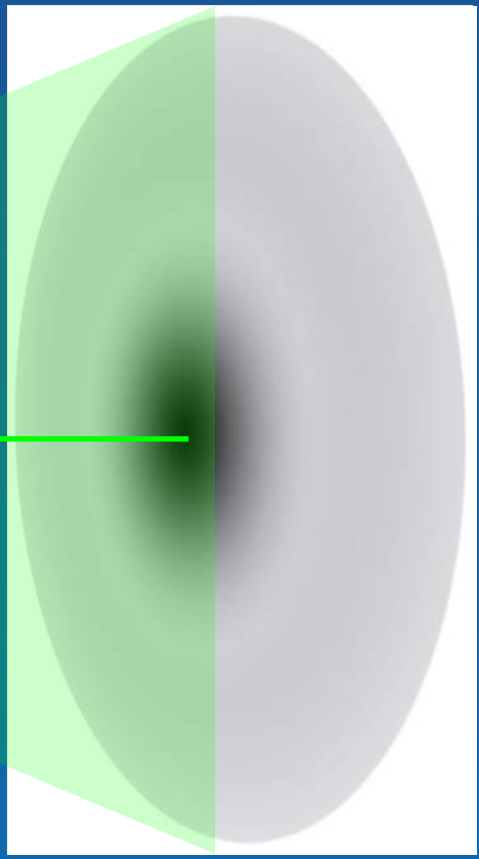




X-ray Beam

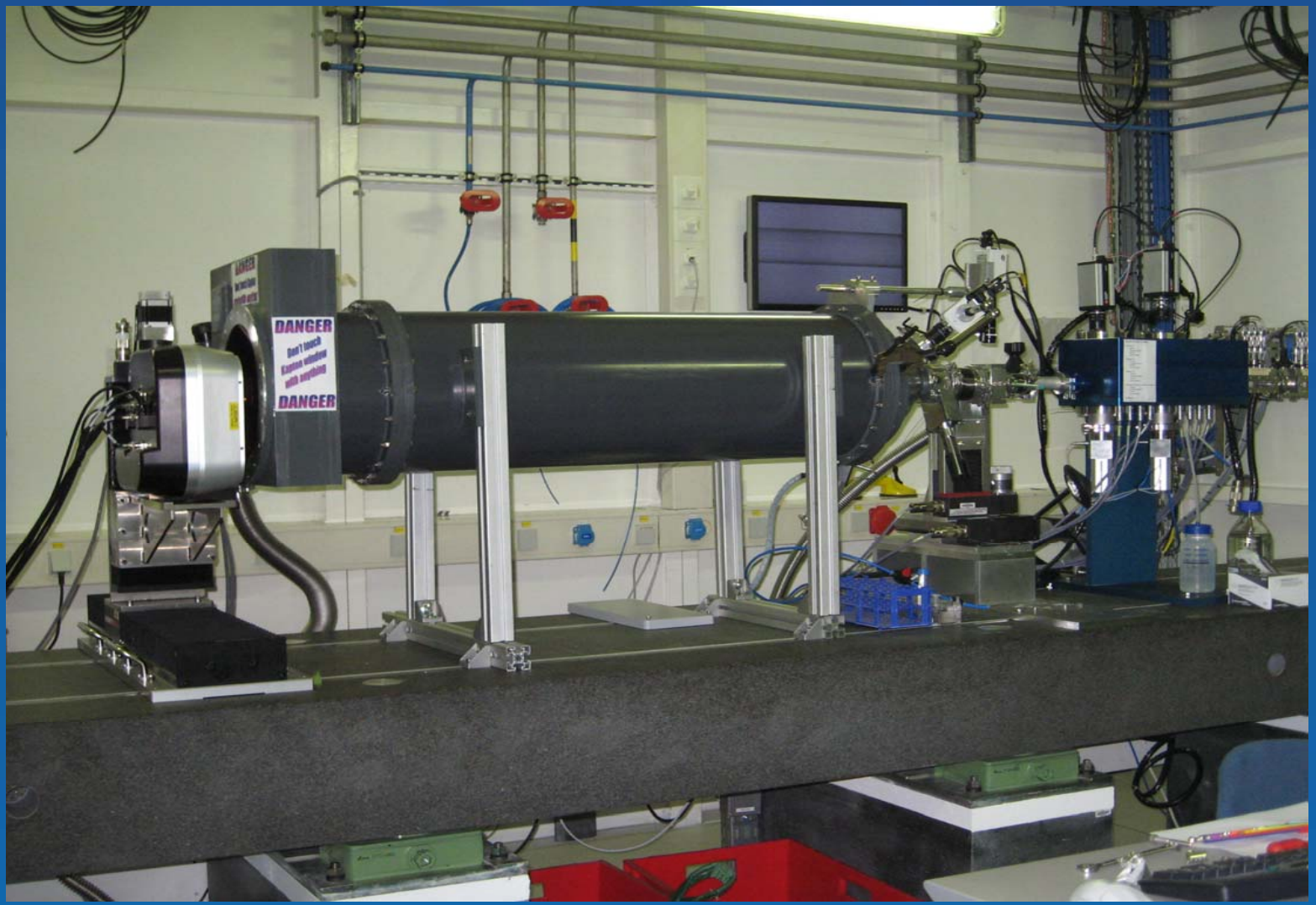
Sample

X-ray Scattering

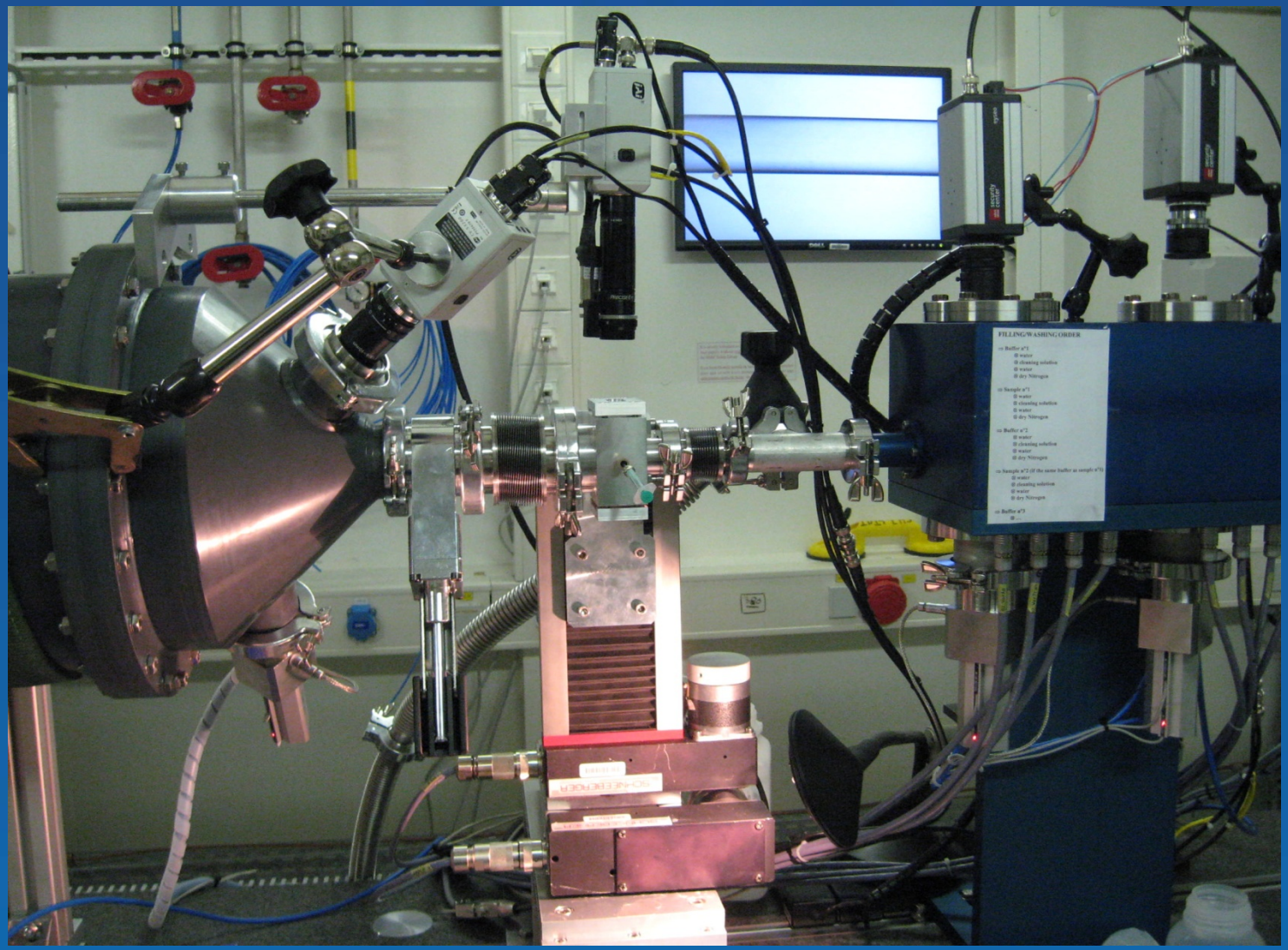


X-ray detector

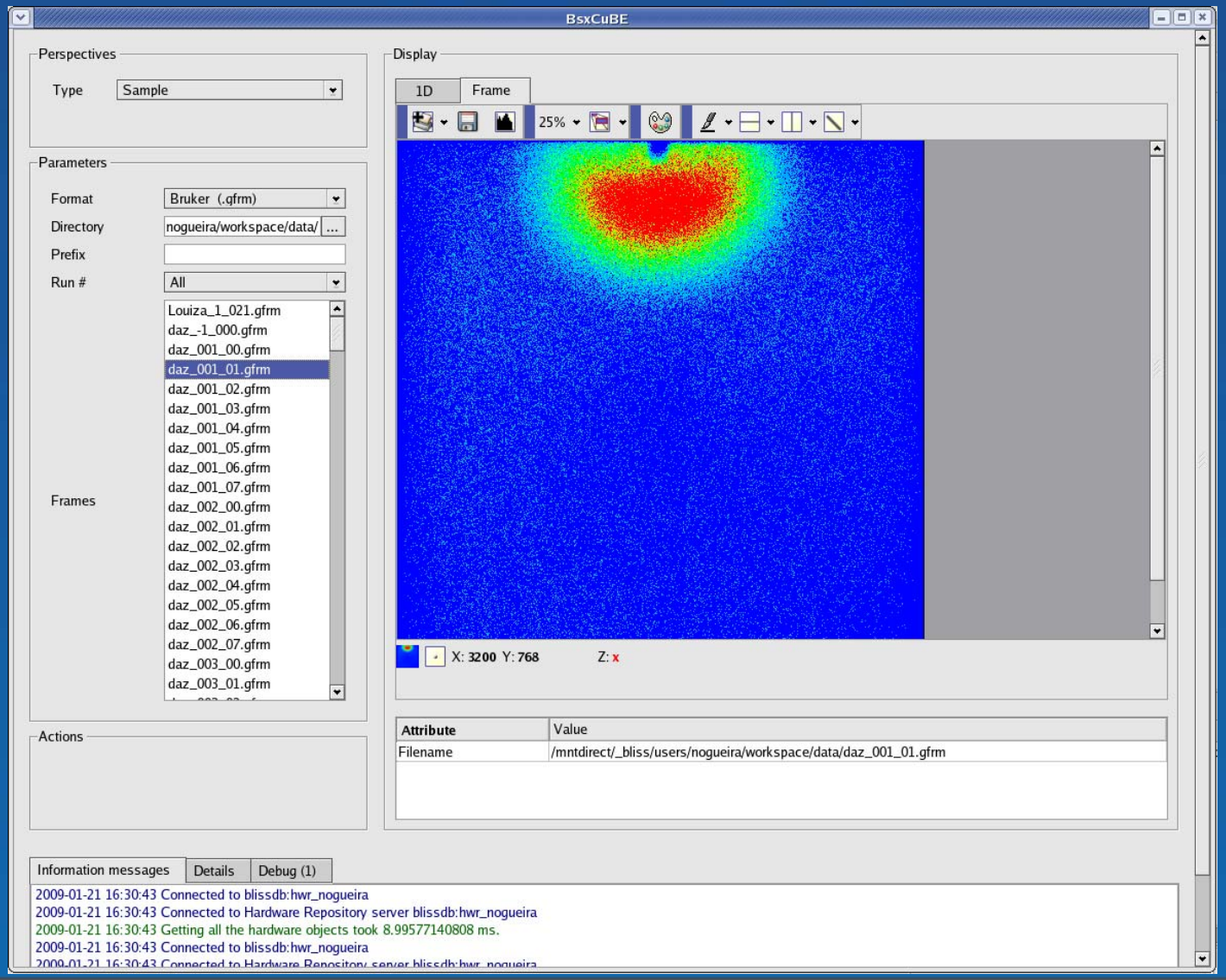
ID14-3 Experimental Hutch Setup



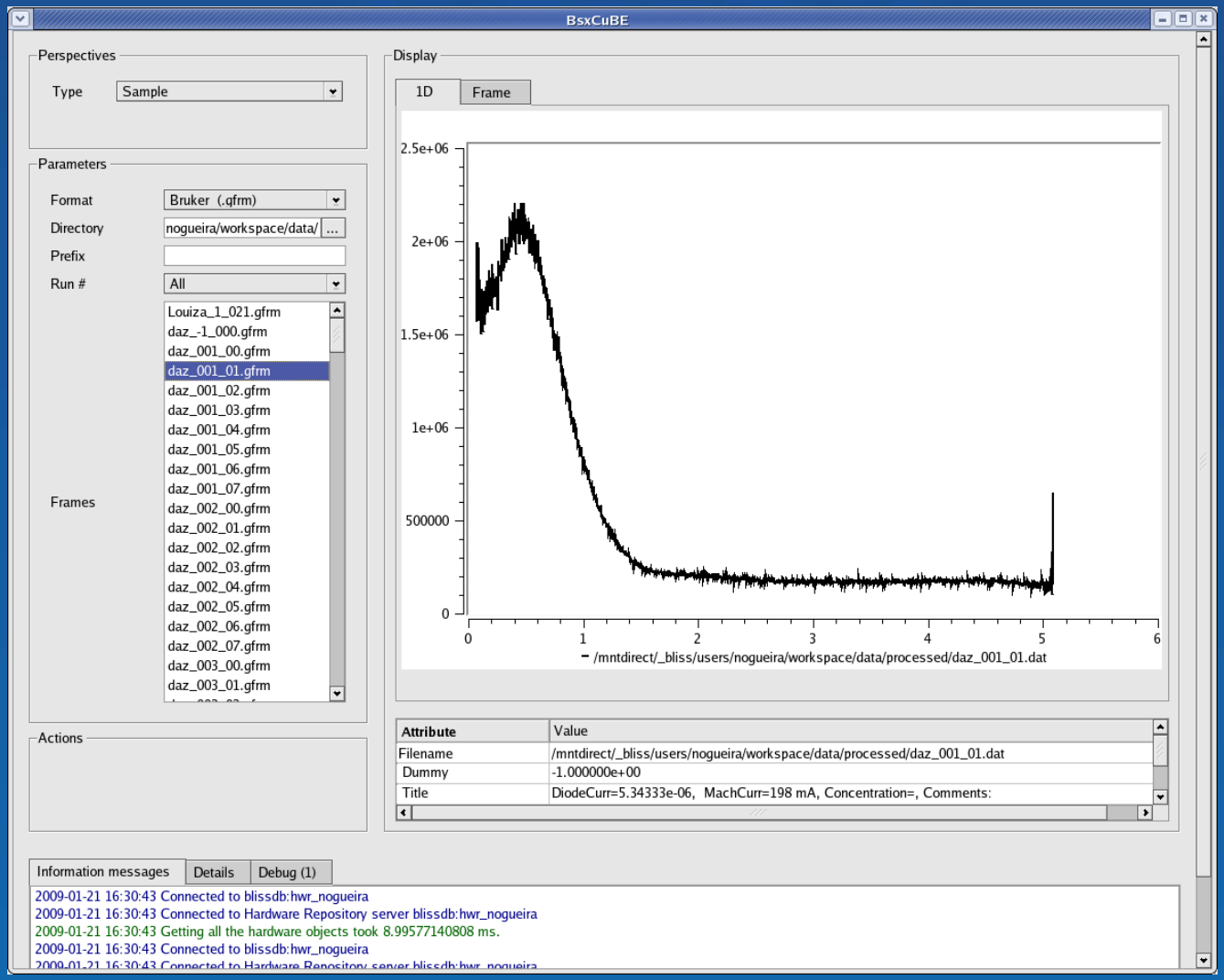
ID14-3 Sample Exposure System



ID14-3 Beamline Control Interface

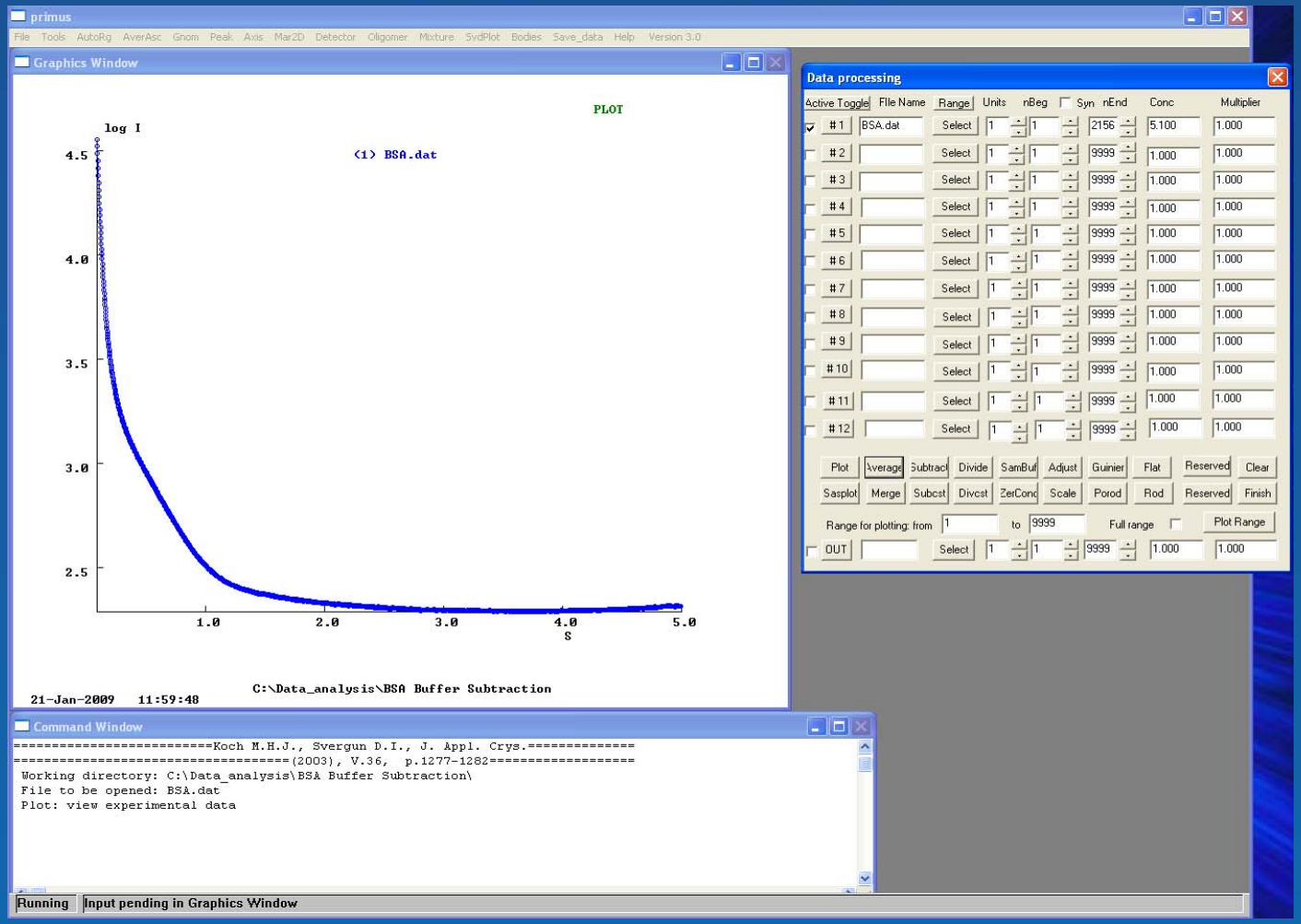


ID14-3 Beamline Control Interface



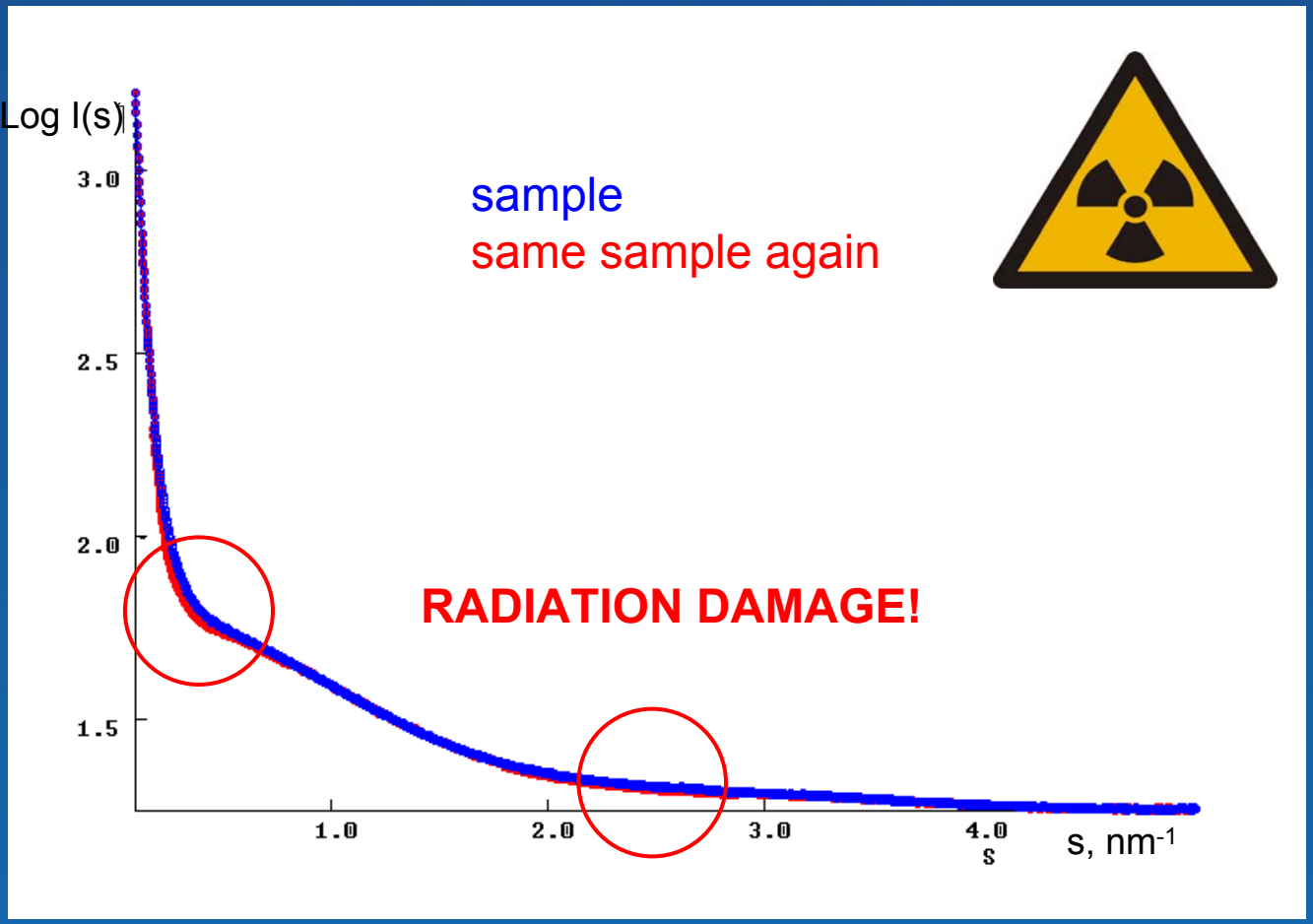
Primary Data Processing

PRIMUS



Quality Control Tests

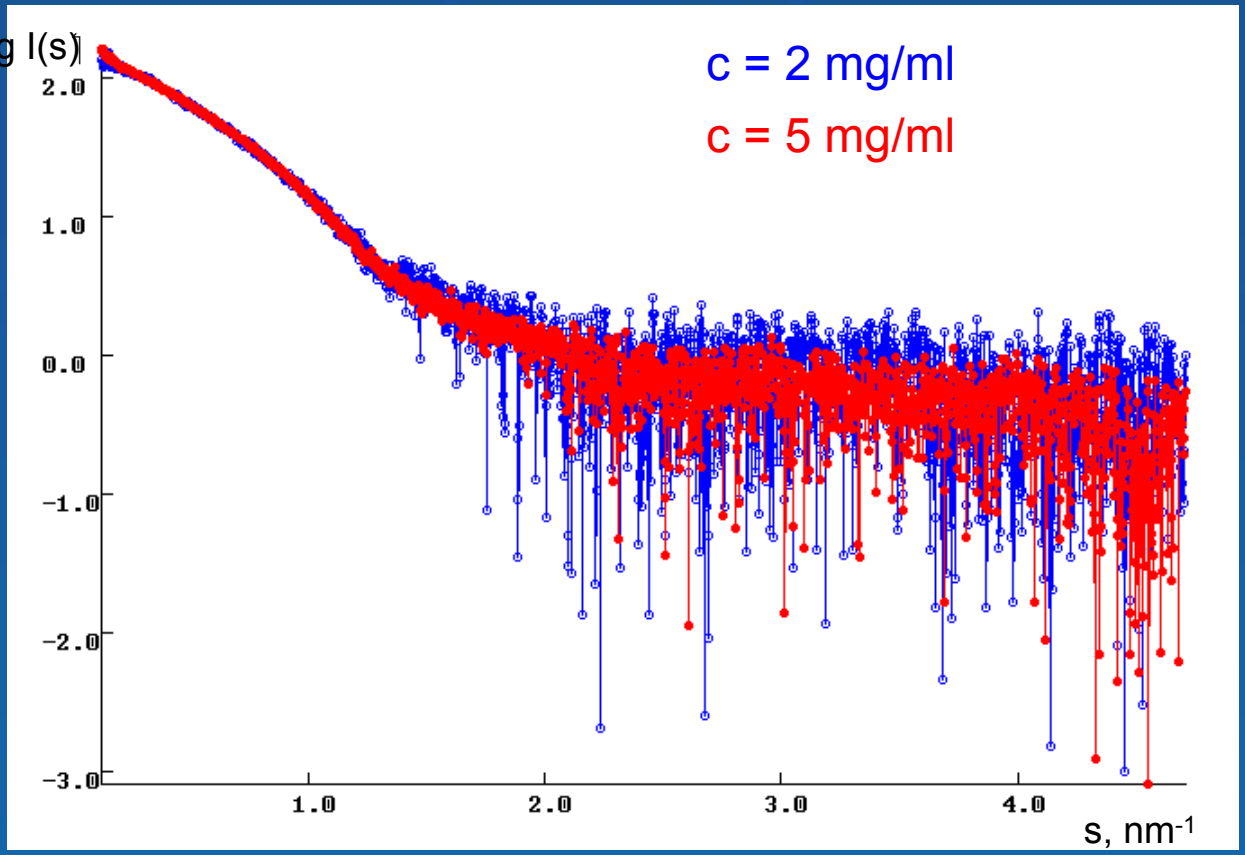
Multiple time frames used to check for radiation damage!



Thanks to A. Kikhney for this slide

Merging Data

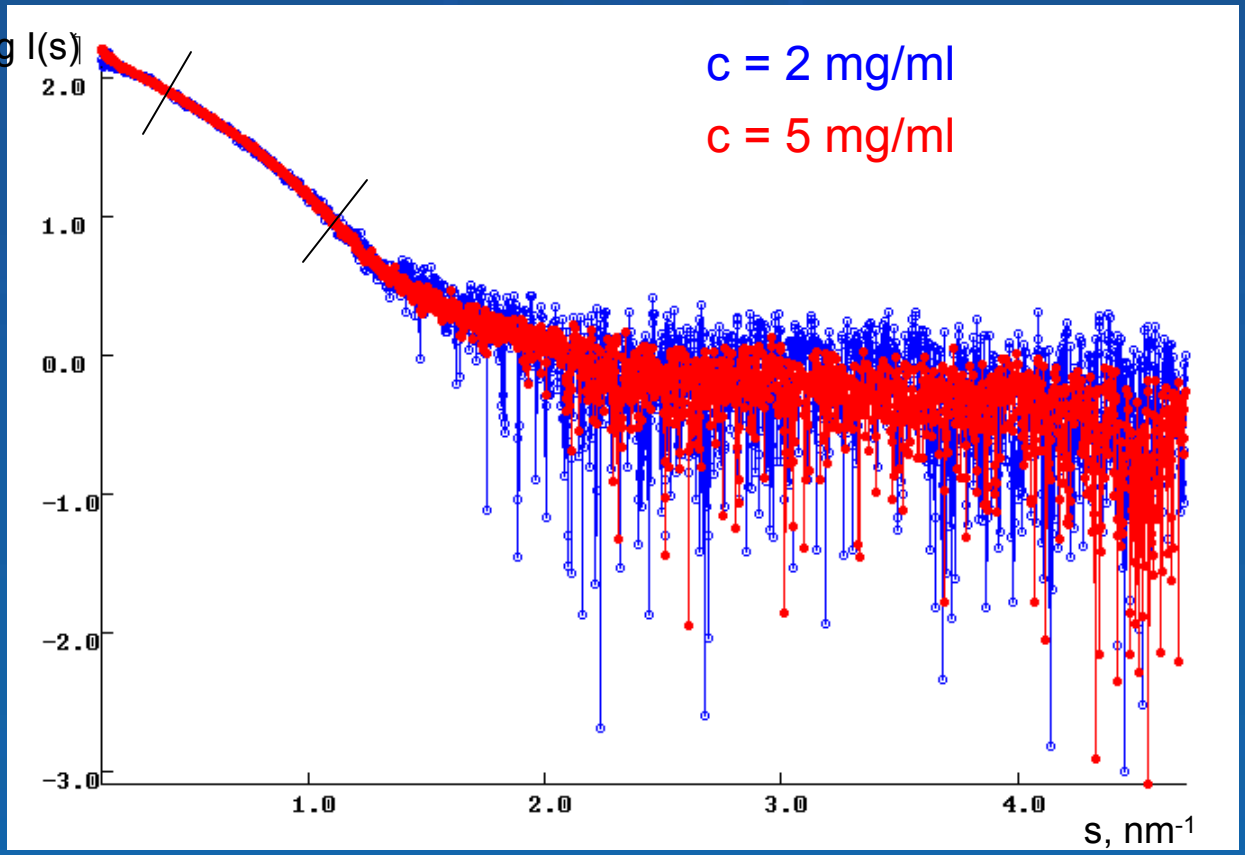
Low and High Concentration



Thanks to A. Kikhney for this slide

Merging Data

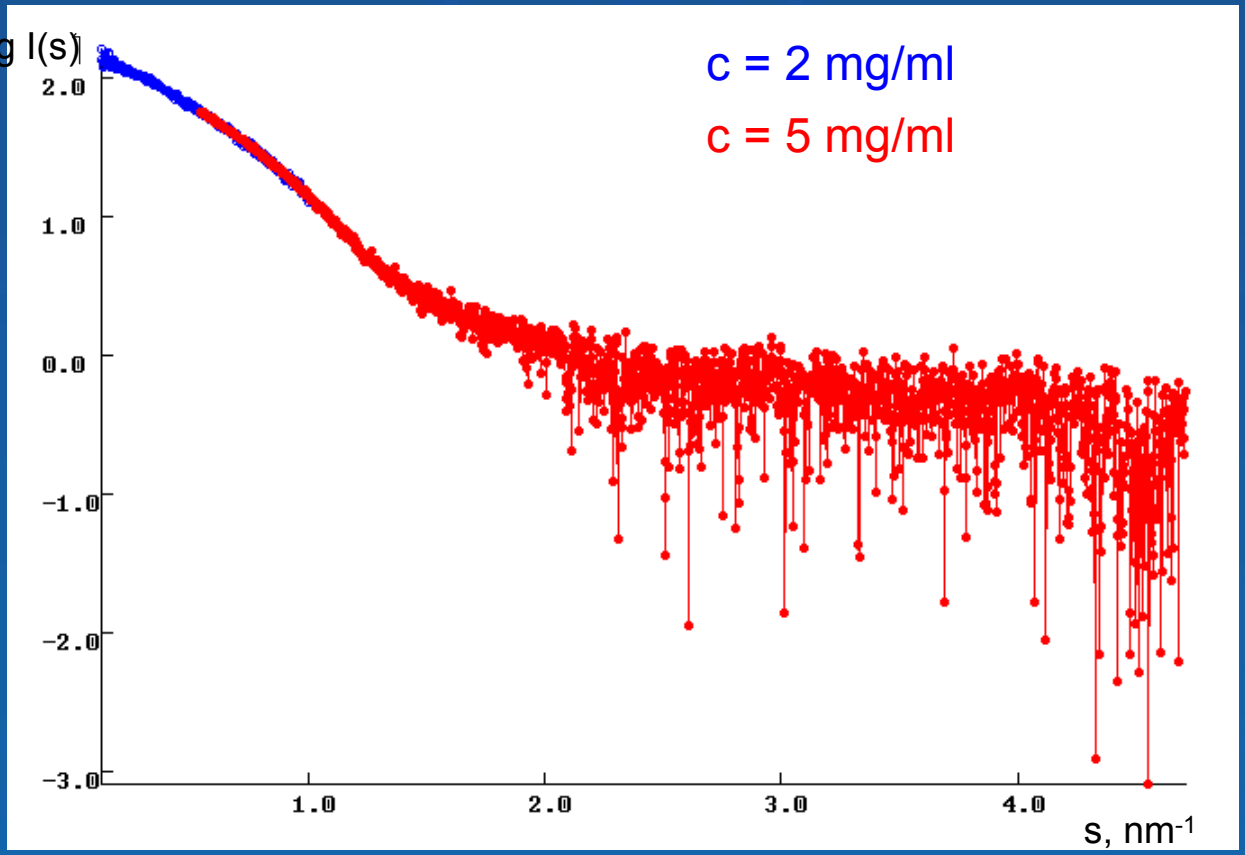
Low and High Concentration



Thanks to A. Kikhney for this slide

Merging Data

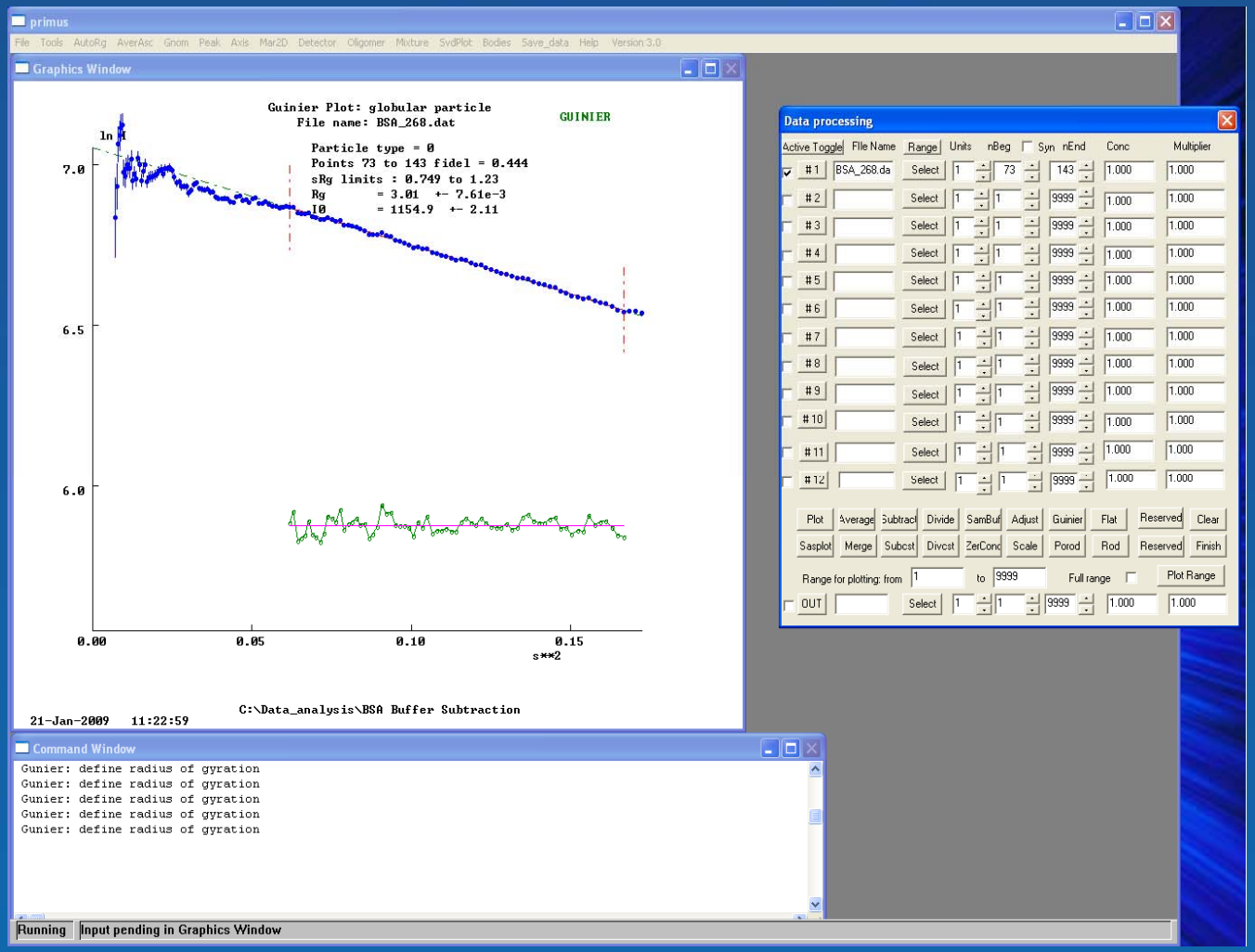
Low and High Concentration



Thanks to A. Kikhney for this slide

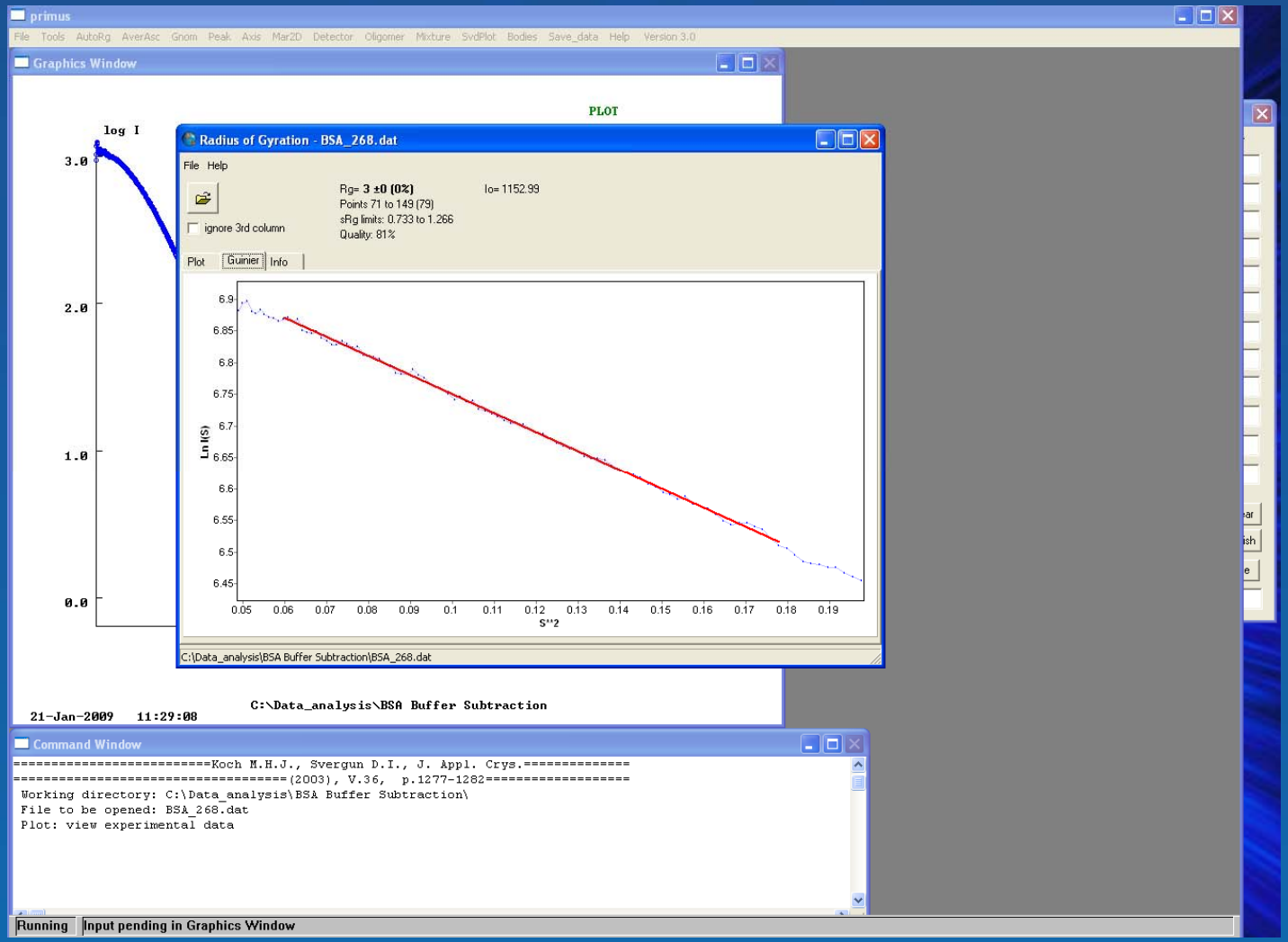
Rg and I(0)

Radius of Gyration and Zero Angle Intensity



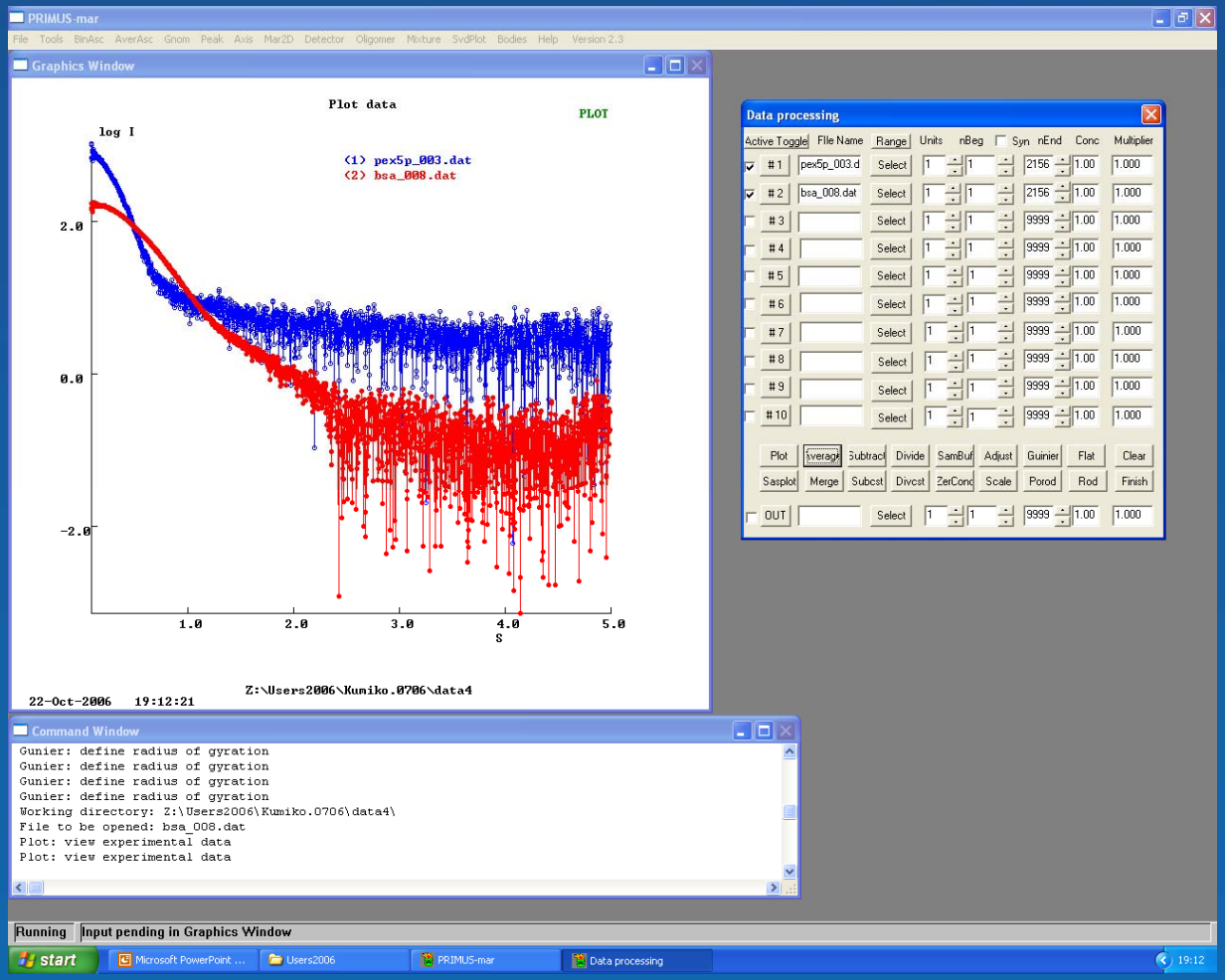
Rg and I(0)

Radius of Gyration and Zero Angle Intensity



Using calibration data

BSA as a standard.... Comparison with unknown proteins



The difference at low angles < 1 nm^{-1} is due to the different molecular weights (MW).

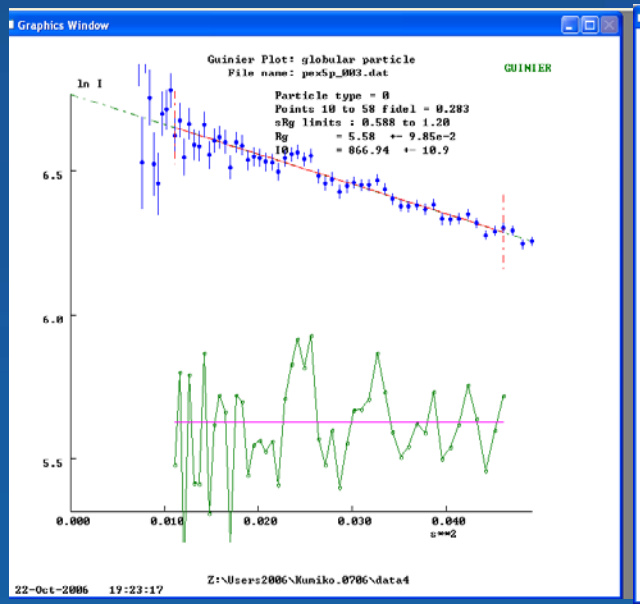
The scattering at low angles is proportional the product of $c_{protein} \cdot MW$.

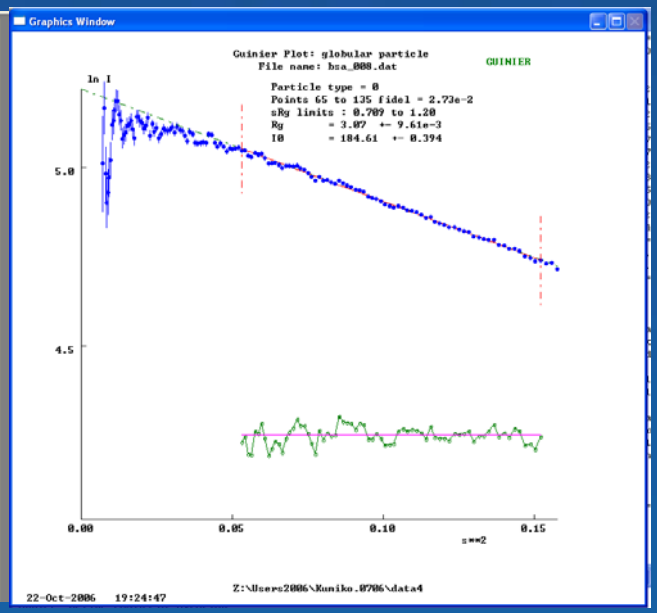
With known protein concentration the molecular weight of the sample can be estimated using the parameters estimated from the Guinier-plot.

Thanks to M. Roessle for this slide

Using calibration data

BSA as a standard.... Comparison with unknown proteins





$$I(s) \cong I_0 e^{-\frac{Rg^2}{3}s^2}$$

$$\ln I(s) \cong \ln I_0 - \frac{Rg^2}{3}s^2$$

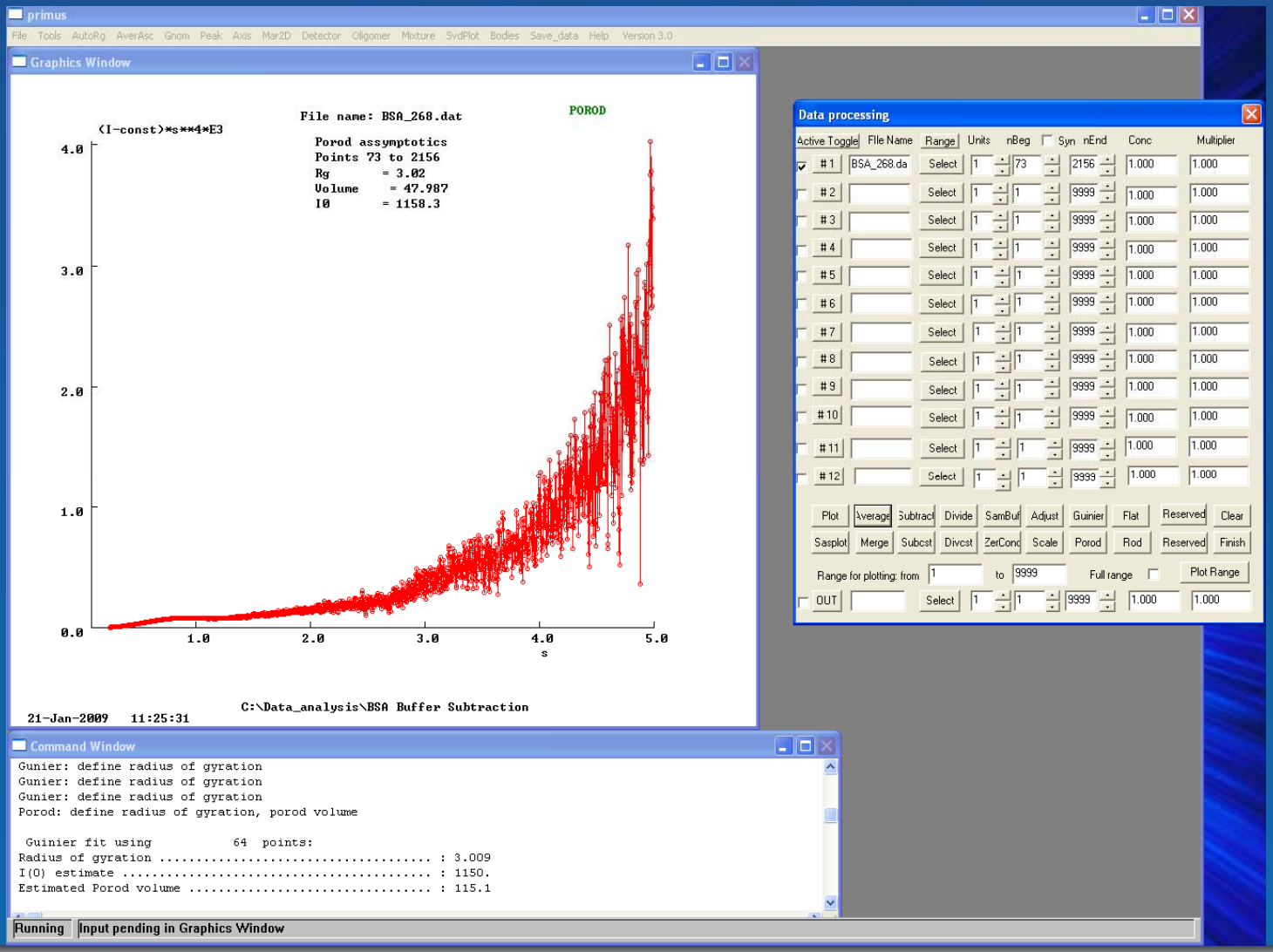
	Radius of Gyration	Forward scattering I_0
BSA standard	3.07 nm	185 units
Sample protein	5.58 nm	867 units

$$MW_{protein} = \frac{66kDa}{185} \cdot 867$$

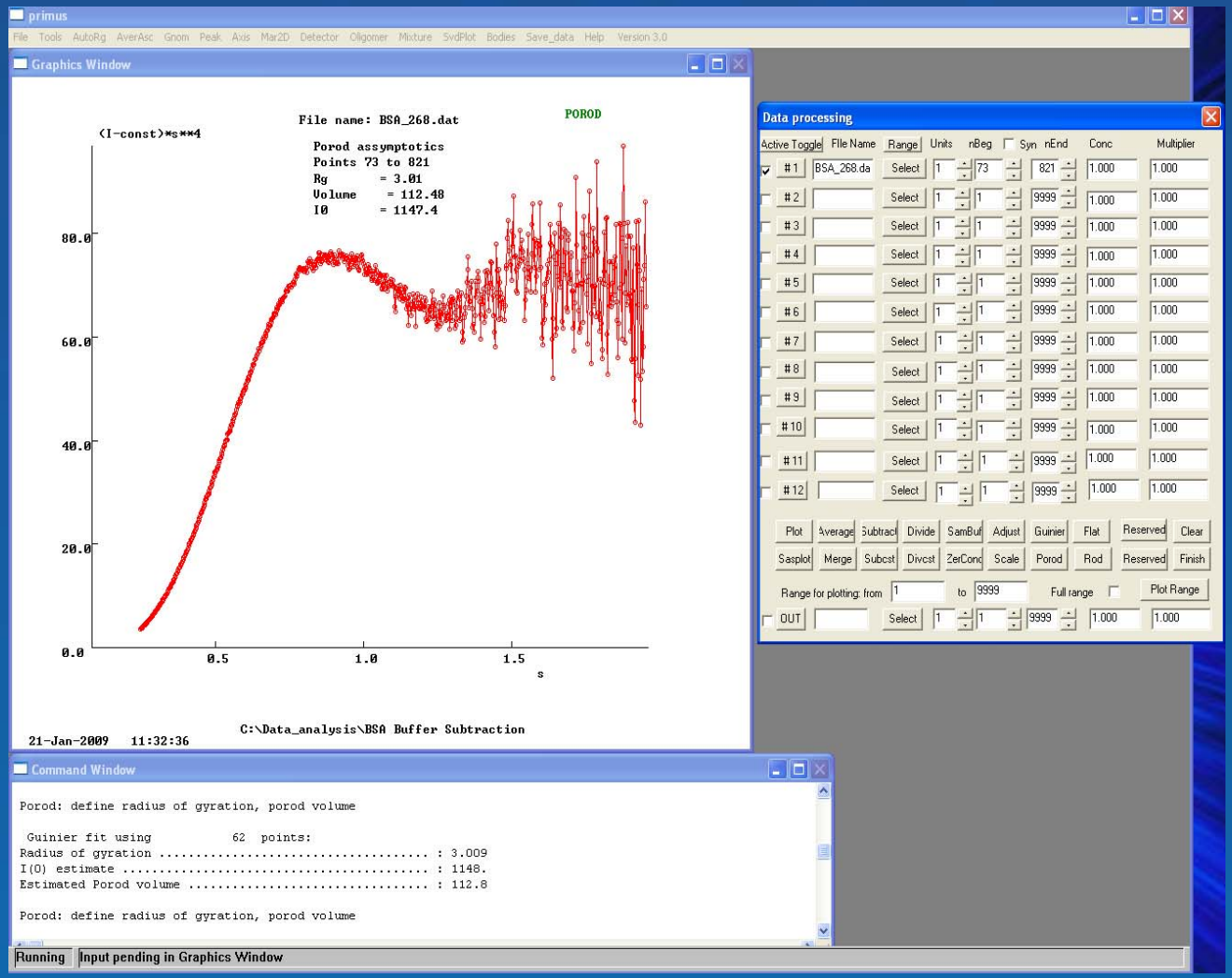
$$MW_{protein} = 307kDa$$

Thanks to M. Roessle for this slide

Porod Analysis in Primus gives excluded volume

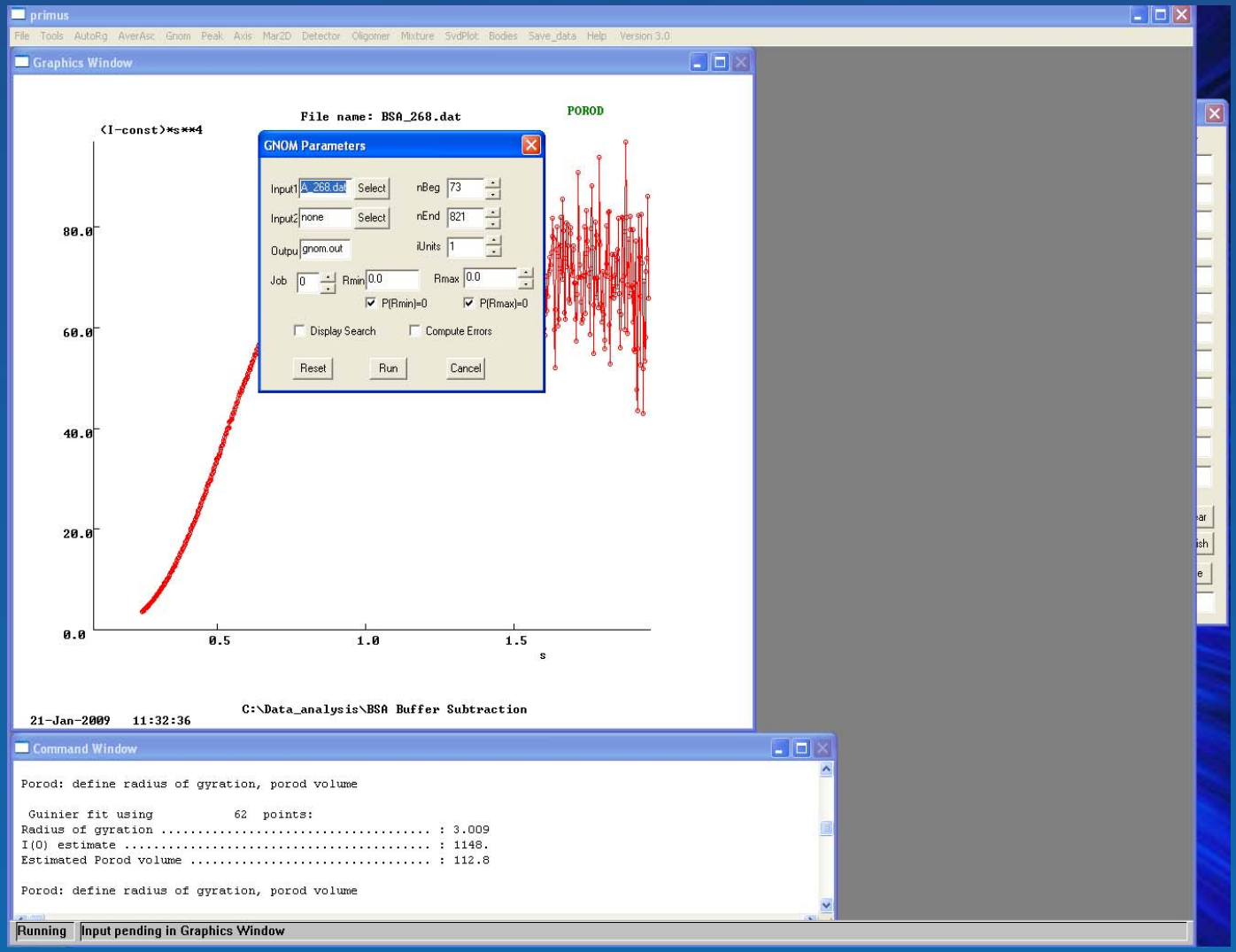


Porod Analysis in Primus gives excluded volume



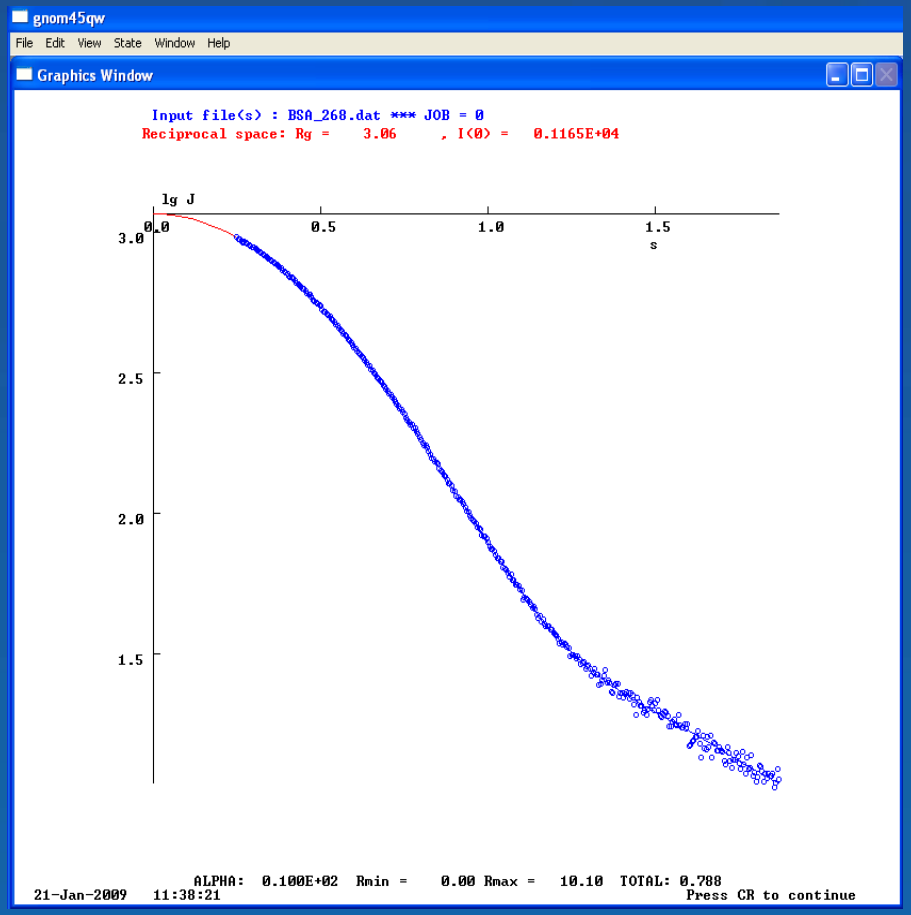
Distance Distribution P(r) Function

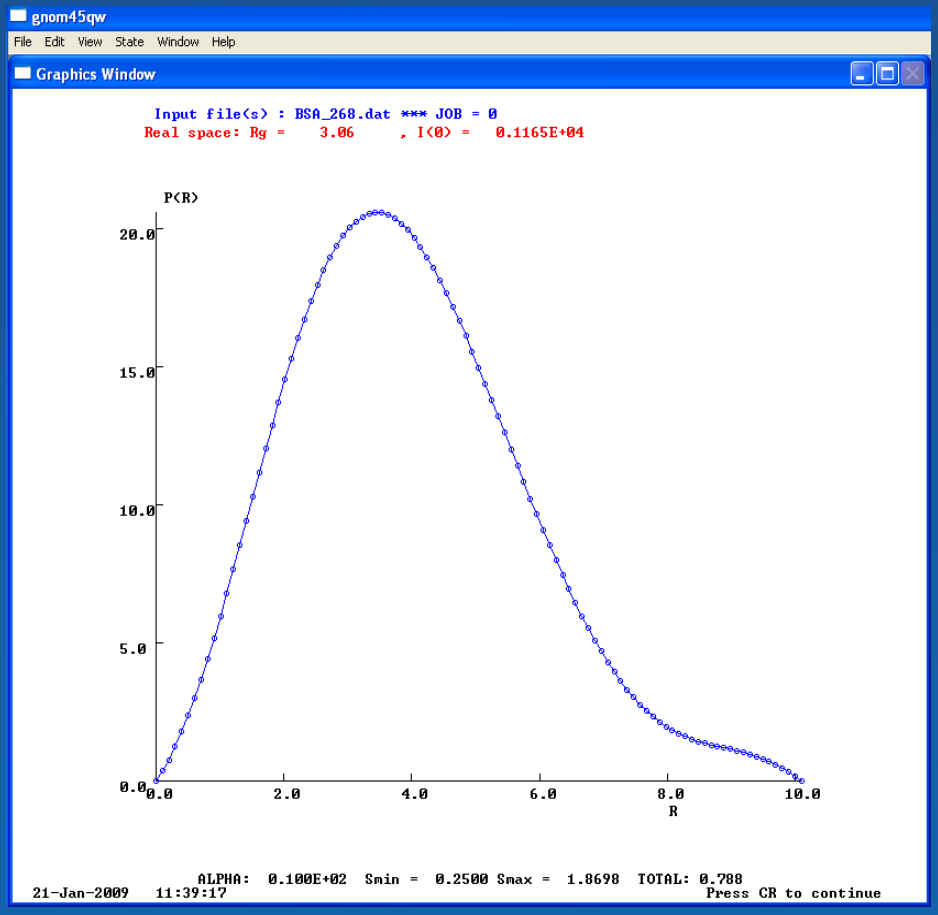
Calculated by GNOM give the Dmax and the input file required for *Ab-initio* modeling



Distance Distribution P(r) Function

Calculated by GNOM give the Dmax and the input file required for *Ab-initio* modeling

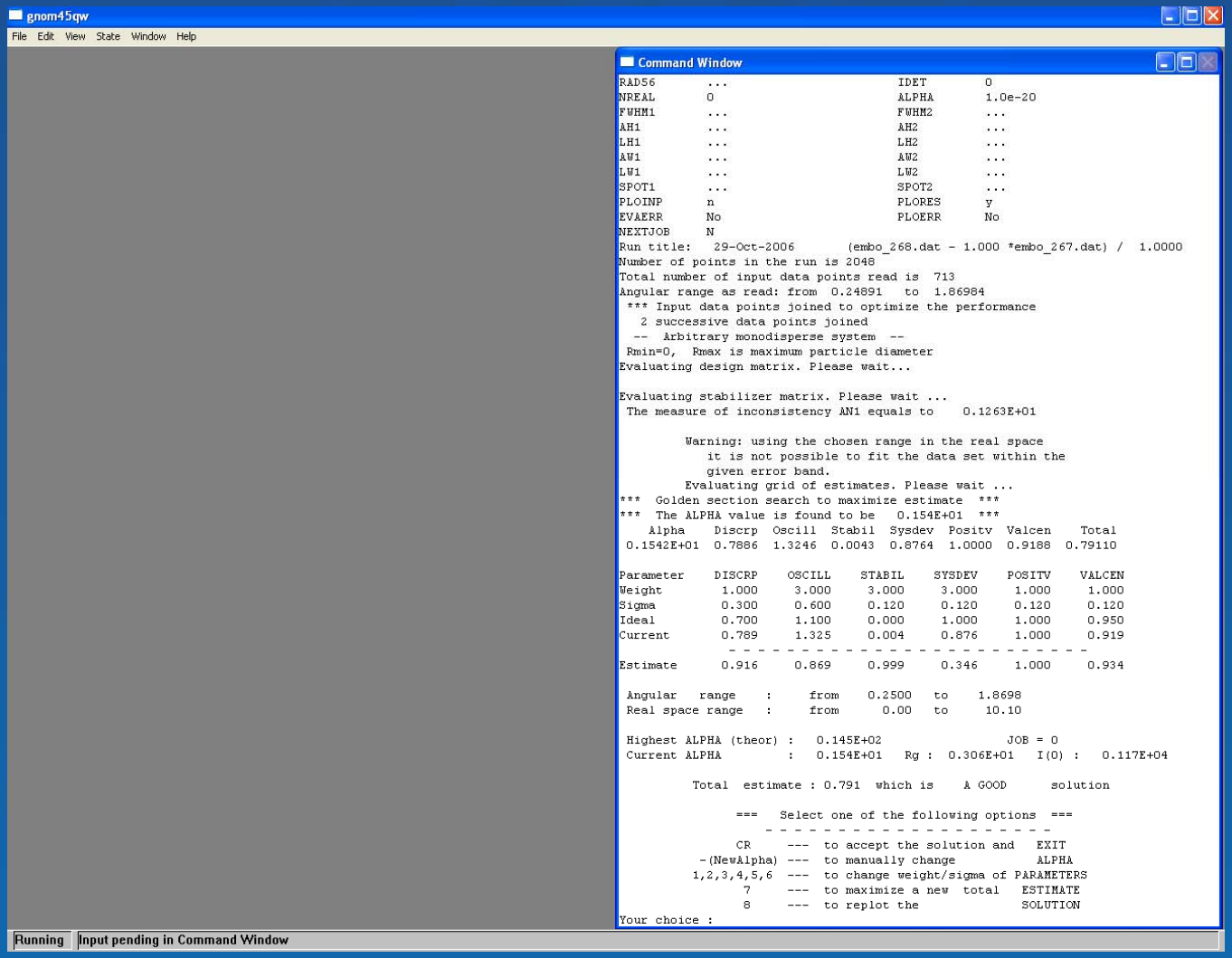




Indirect Fourier Transform!

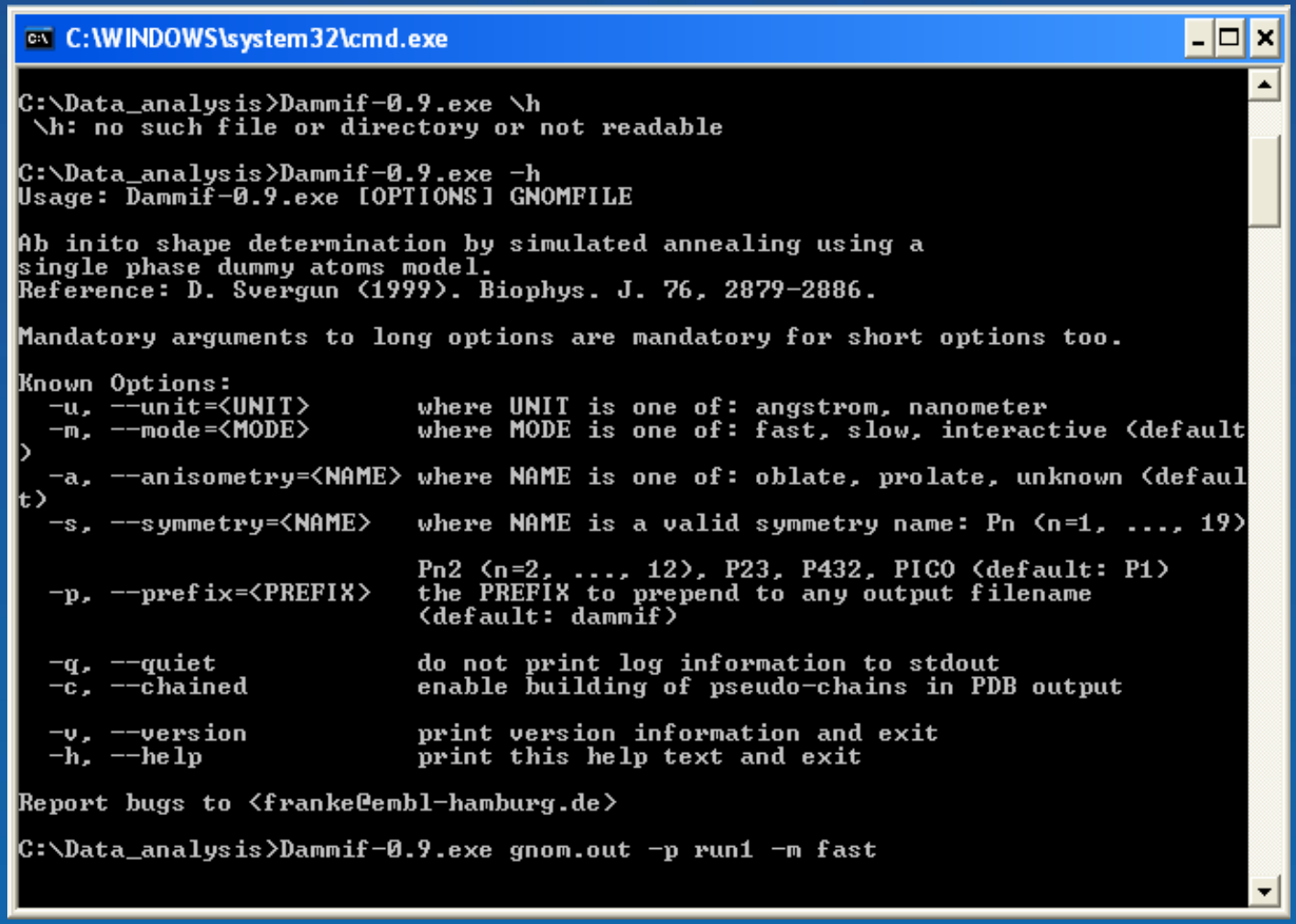
Distance Distribution P(r) Function

Calculated by GNOM give the Dmax and the input file required for *Ab-initio* modeling



Distance Distribution P(r) Function

Calculated by GNOM give the Dmax and the input file required for *Ab-initio* modeling

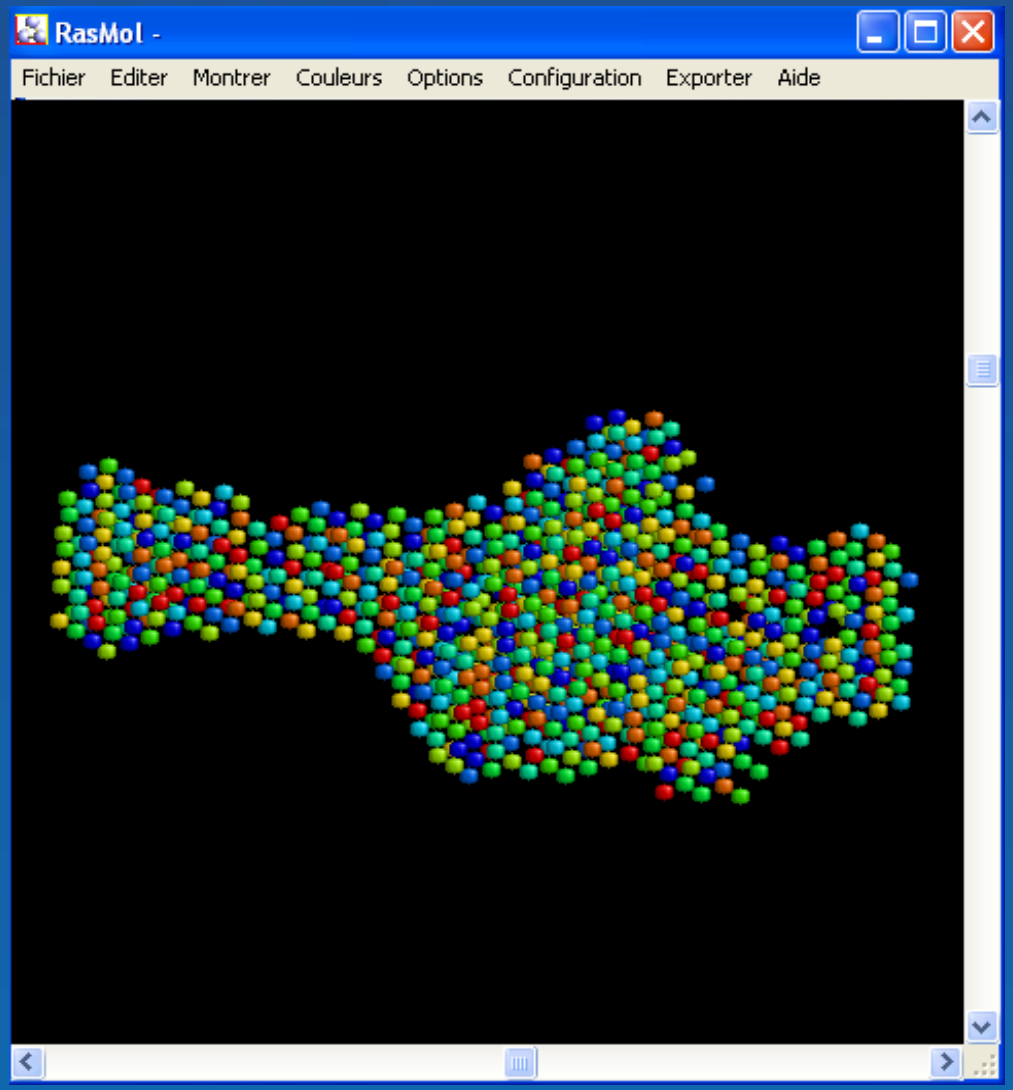


Distance Distribution P(r) Function

Calculated by GNOM give the Dmax and the input file required for *Ab-initio* modeling

C:\WINDOWS\system32\cmd.exe

```
C:\Data_analysis>Dammif-0.9.exe gnom.out -p run1 -m fast
dammif ..... : 0.9 (beta)
Log opened ..... : 2009-01-21, 11:48:06
Run as:
Full command line ..... : Dammif-0.9.exe gnom.out -p
run1 -m fast
Configuration mode ..... : fast
Angular units ..... : undefined
Prefix ..... : run1
Threads ..... : 1
GNOM input file ..... : gnom.out
Expected particle anisometry ..... : unknown
Enforced particle symmetry ..... : P1
Pseudo-chains in PDB output ..... : no
GNOM file:
Title ..... : 15-Dec-2008       (MJ_192.d
at - 0.883 *Bicine.dat) /  6.0000
Angular units ..... : nanometer
Maximum particle diameter [Angstrom] ..... : 160.
Radius of gyration [Angstrom] ..... : 41.2
Minimum s [1/Angstrom] ..... : 0.00
Maximum s [1/Angstrom] ..... : 0.176
Configured as:
Output filename prefix ..... : run1
Search volume shape ..... : sphere
Radius [Angstrom] ..... : 80.0
Approximate number of dummy atoms ..... : 2000
Expected particle anisometry ..... : unknown
Number of spherical harmonics ..... : 15
Proportion of the curve to be fitted ..... : 1.00
Maximum s [1/Angstrom] ..... : 0.176
Number of shannon channels ..... : 9
Number of supporting points ..... : 20
Weighting function ..... : emphasised porod
Maximum number of steps ..... : 200
Maximum number of iterations ..... : 20000
Minumum number of successes ..... : 20
Maximum number of successes ..... : 2000
Temperature scheduling factor ..... : 0.900
Peripheal penalty weight ..... : 0.200
Looseness penalty weight ..... : 0.500E-02
Disconnectivity penalty weight ..... : 0.00
Anisometry penalty weight ..... : 0.00
Constant subtraction ..... : skipped (negative value)
Dummy atom model status ..... : initializing
Number of dummy atoms ..... : 2093
Atom radius ..... : 5.70
Overall volume ..... : 0.214E+07
Annealing procedure status ..... : warming up
Initial looseness penalty ..... : 0.160
Initial disconnectivity penalty ..... : 0.00
initial peripheal penalty ..... : 0.546
Annealing procedure status ..... : started
Step:   1, T: 0.657E-03, Succ:   2001, Eval:     5150, CPU: 00:00:00
 RF: 0.701    , Los: 0.124, Dis: 0.000, Per: 0.561, Ani: 0.000, Fit: 0.77770
Step:   2, T: 0.592E-03, Succ:   2001, Eval:    10958, CPU: 00:00:00
 RF: 0.658    , Los: 0.137, Dis: 0.000, Per: 0.573, Ani: 0.000, Fit: 0.73266
Step:   3, T: 0.532E-03, Succ:   2001, Eval:    21251, CPU: 00:00:00
 RF: 0.399    , Los: 0.136, Dis: 0.000, Per: 0.543, Ani: 0.000, Fit: 0.49769
Step:   4, T: 0.479E-03, Succ:   2001, Eval:    40758, CPU: 00:00:01
 RF: 0.172    , Los: 0.143, Dis: 0.000, Per: 0.360, Ani: 0.000, Fit: 0.28803
Step:   5, T: 0.431E-03, Succ:   1734, Eval:    60759, CPU: 00:00:01
 RF: 0.119    , Los: 0.142, Dis: 0.000, Per: 0.322, Ani: 0.000, Fit: 0.23918
Step:   6, T: 0.388E-03, Succ:   1671, Eval:    80764, CPU: 00:00:01
 RF: 0.919E-01, Los: 0.158, Dis: 0.000, Per: 0.284, Ani: 0.000, Fit: 0.20672
Step:   7, T: 0.349E-03, Succ:   1532, Eval:   100766, CPU: 00:00:02
 RF: 0.102    , Los: 0.143, Dis: 0.000, Per: 0.289, Ani: 0.000, Fit: 0.20468
Step:   8, T: 0.314E-03, Succ:   1539, Eval:   120767, CPU: 00:00:02
 RF: 0.723E-01, Los: 0.134, Dis: 0.000, Per: 0.254, Ani: 0.000, Fit: 0.17364
```



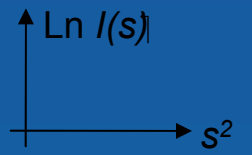
Summary

what should be done while at the beamline

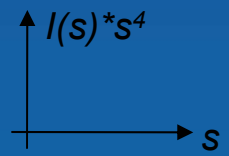
- Data collection
- Radial averaging → 1D
- Normalization
- Background subtraction
- Checks for effects of
  - Radiation
  - Concentration

- Log plot

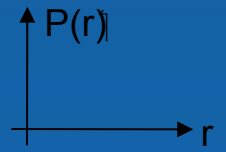
- Guinier plot (R_g, MM)



- Porod plot (volume)



- P(r) plot

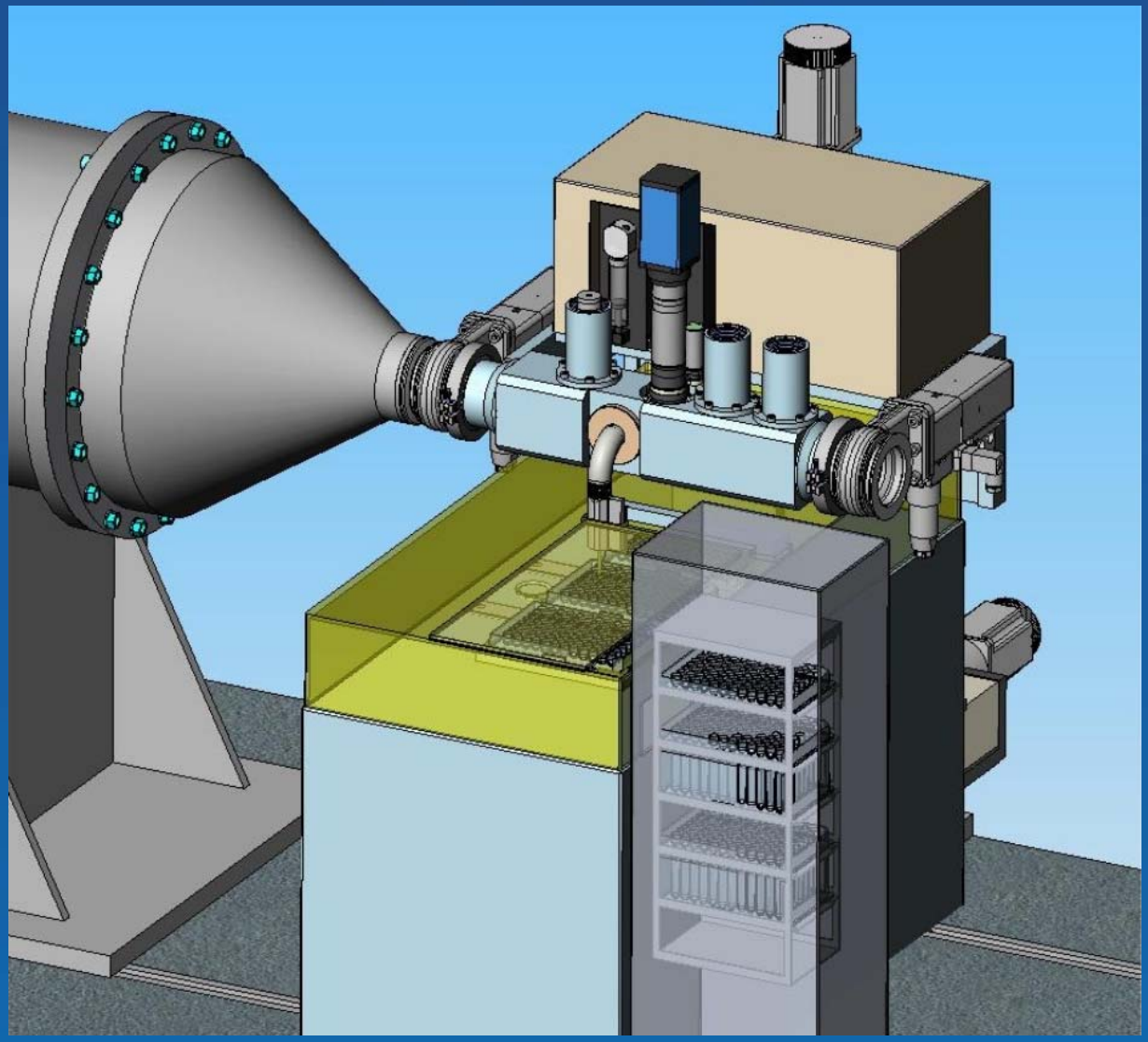


- Lo-res 3D model

Automation!

Making life easier!
The future of ID14-3

ID14-3 BioSAXS Data collection to be Automated



Automated Data Processing Pipeline

Data

AUTO*Pilatus*

Automatic data transformation of 2D raw data into 1D scattering curves

AUTOSUB

Automatic sample-buffer subtraction with automatic estimation of radius of gyration (AUTORG)

AUTOGNOM

Automatic estimation of distance distribution function

DAMMIF

Automaticly start the *ab-initio* shape determination program and fitting with simple geometrical bodies

Results:

Sample checks for radiation damage and concentration effects

Sample Size: Rg, Dmax, Volume and Mw

Low Resolution particle shape

Closing Remarks

- SAXS is a powerful tool for structural biology
  - Structure validation
  - Structure of multi domain proteins
    - With addition of missing linkers if needed
  - *Ab-initio* modeling

- Optimized data collection at ID14-3
  - Standard experimental setup greatly improves efficiency
  - Easy access using rolling application procedure

- The future is Automation
  - Data collection through collaboration between ESRF and the EMBL Grenoble and Hamburg outstations
  - Data Processing based on the Hamburg SAXS data processing pipeline

Acknowledgments

- EMBL-Hamburg
  - SAXS Group
  - D. Svergun
  - M. Roessle
  - M. Petoukhov
  - D. Franke
  - A. Kikhney
  - P. Konarev
- EMBL-Grenoble
  - S. Cusack
  - A. Mc Carthy
  - F. Cipriani
  - Synchrotron Instrumentation Team
- ESRF
  - P. Pernot
  - S. McSweeney
  - G. Leonard
  - D. Nurizzo
  - D. Spruce
  - R. Fernandez
  - P. Theveneau
  - J. Surr
  - T. Giraud
  - D. Davison
  - D. Flot
  - S. Larson