The Bio-SAXS beamline

Solution Scattering from Biological Macromolecules

What information can we obtain?

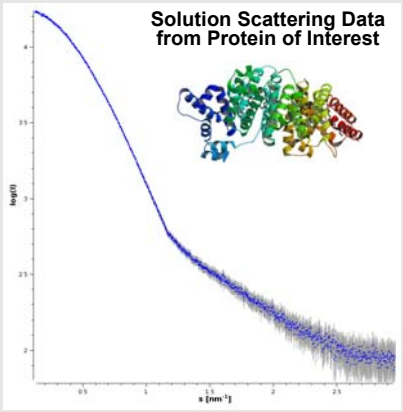
ESRF

EMBL

Contents

- What can be obtained from Bio-SAXS
  - Measurable parameters
  - Modelling strategies
- How to collect Solution SAXS data at ID14-3
  - Procedure
    - Data collection tests
    - Data Verification and quality control
    - Modelling and analysis
  - Practical demonstration at the Bio-SAXS beamline

Solution SAS Experiment



Black Box

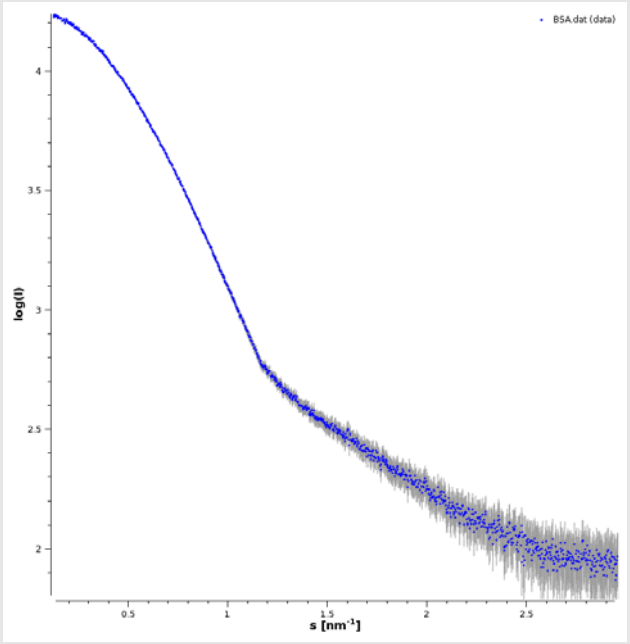
Neutron/X-ray beamline
Home source

?

ESRF EMBL

Radius of Gyration and Zero Angle Intensity

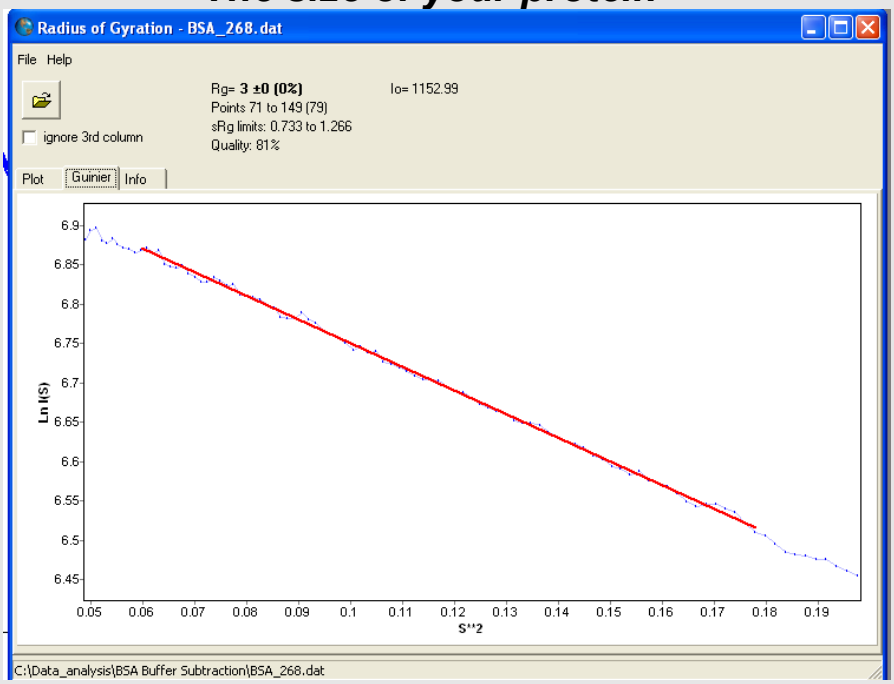
The size of your protein



ESRF EMBL

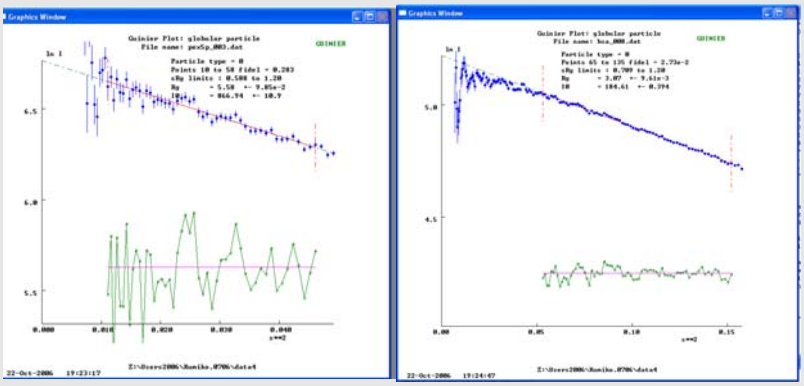
Radius of Gyration and Zero Angle Intensity

The size of your protein



Using calibration data

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



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

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

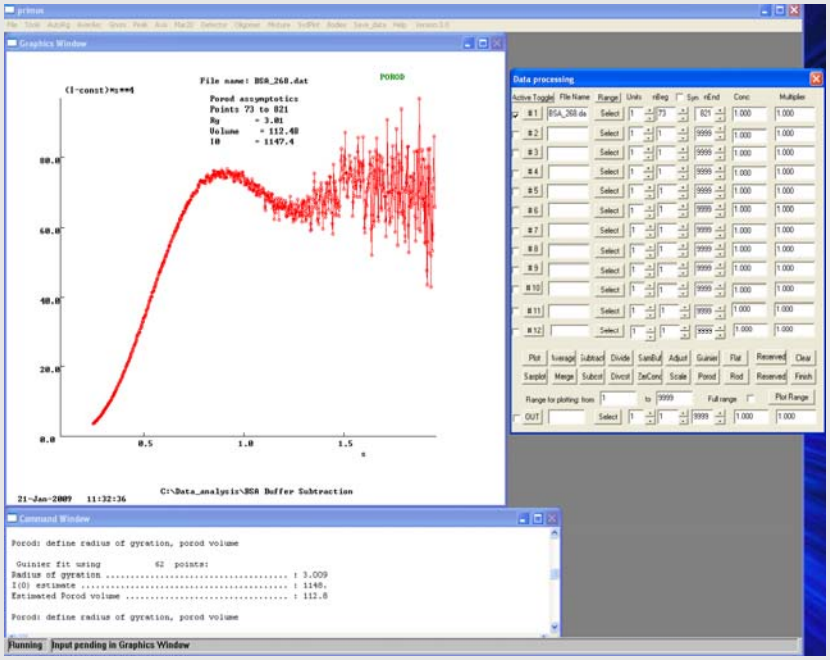
	Radius of Gyration	Forward scattering I_0
BSA standard	3.07 nm	185 units
Sample protein	5.58 nm	867 units

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

$$MW_{protein} = 307kDa$$

Thanks to M. Roessle for this slide

Porod Analysis in Primus gives excluded volume



ESRF EMBL

Ab-initio Modelling

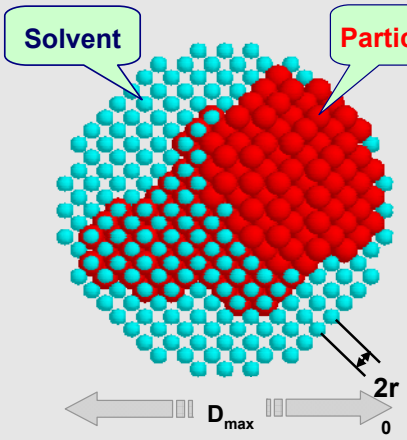
A sphere with diameter D_{max} is filled by densely packed beads of radius $r_0 << D_{max}$. A configuration vector X indicates whether the j-th atom belongs to the particle or to the solvent.

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

Thanks to D. Svergun for this slide

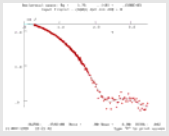
ESRF EMBL

Ab-initio Modelling

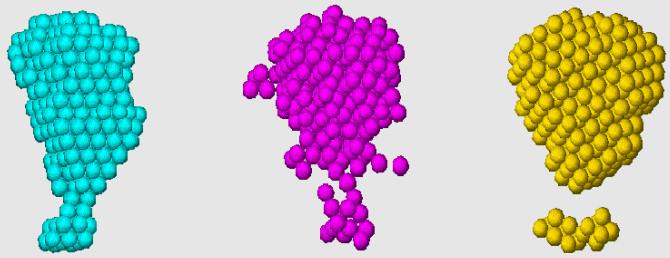
DAMMIN modelling penalties

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

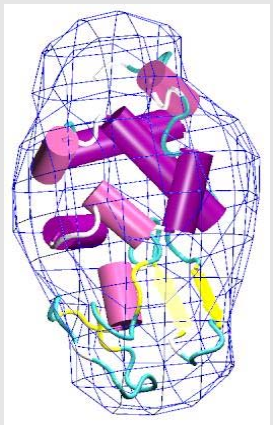
disconnected

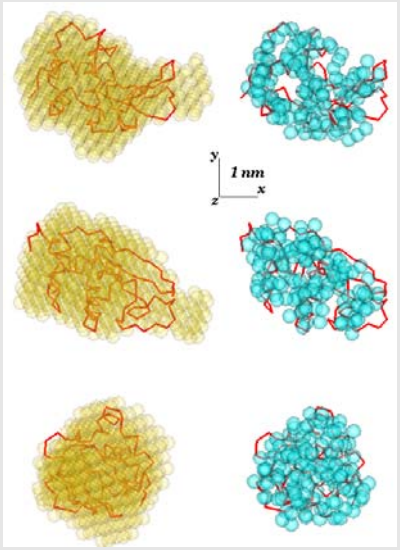
Thanks to D. Svergun for this slide

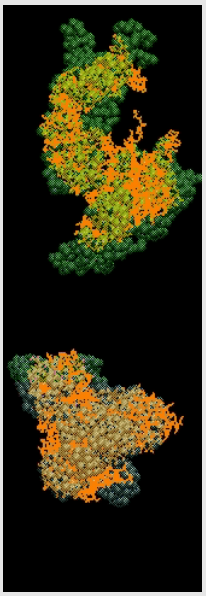
ESRF EMBL

Ab-initio Can it be Trusted?

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







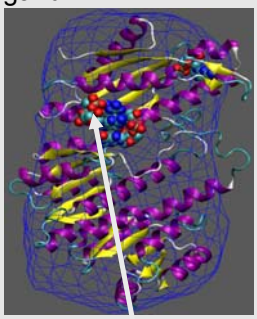
ESRF EMBL

Refinement of Rigid Domains

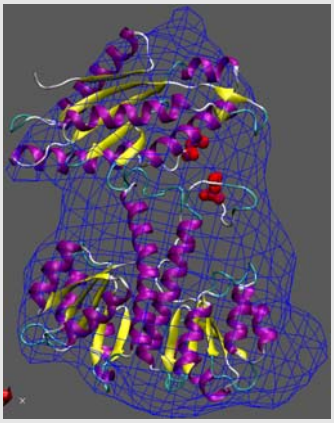
Rigid Body Refinement: Moving protein sub-parts (called domains) as rigid bodies to fit the scattering data

Example: Structural changes upon ligand binding

PX-structure with ligand



SAXS Shape obtained by GASBOR - unliganded state



Rigid Body refinement using MASSHA

Thanks to M. Roessle for this slide

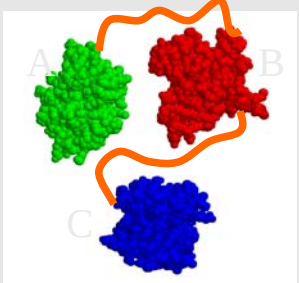
ESRF EMBL

Adding Missing Linkers

Remodeling of proteins from high resolution fragments/constructs

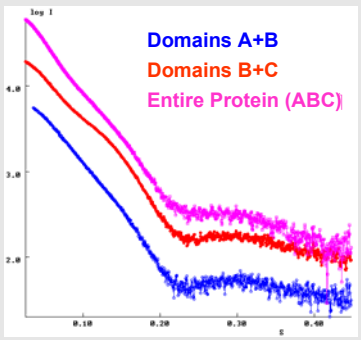
Program BUNCH (M. Petoukhov; Biophysical J.)

Linkers?!?

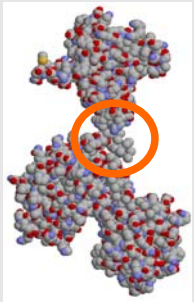


High resolution protein fragments from X-ray crystallography

+ sequence data (TrEMBL/Swissprot)



SAXS Data of constructs
AB
BC
ABC



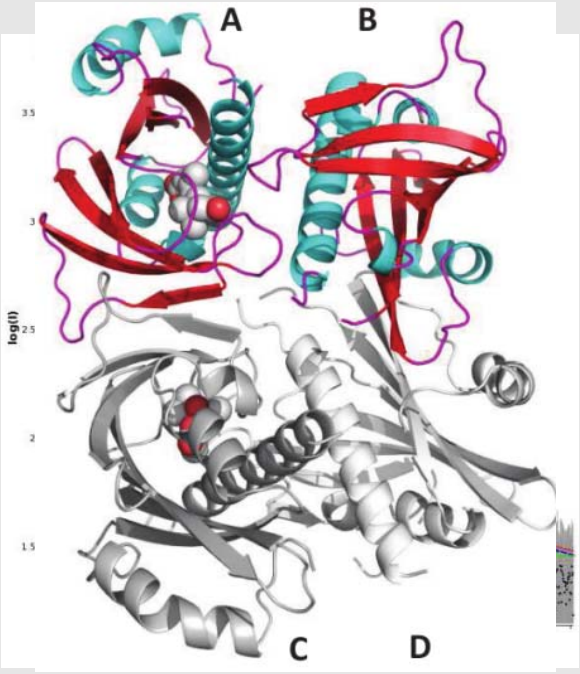
Model for the entire protein including not resolved linker components

Thanks to M. Roessle for this slide

ESRF EMBL

Most recent Published Results from ID14-3

08 November 2009



nature

Experimental X-ray scattering of the PYR1 protein in solution in the presence of 1mM (+) ABA.

Scattering curves for possible ensembles were calculated.

Only the curve for ensembles AB/CD produced a good fit to the experimental data (χ=0.72)

SAXS demonstrated that the AB ensemble corresponds to the biologically relevant form found under physiological conditions

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

J Santiago *et al. Nature* **000**, 1-4 (2009) doi:10.1038/nature08591

ESRF EMBL

Summary

What can we learn from solution SAXS

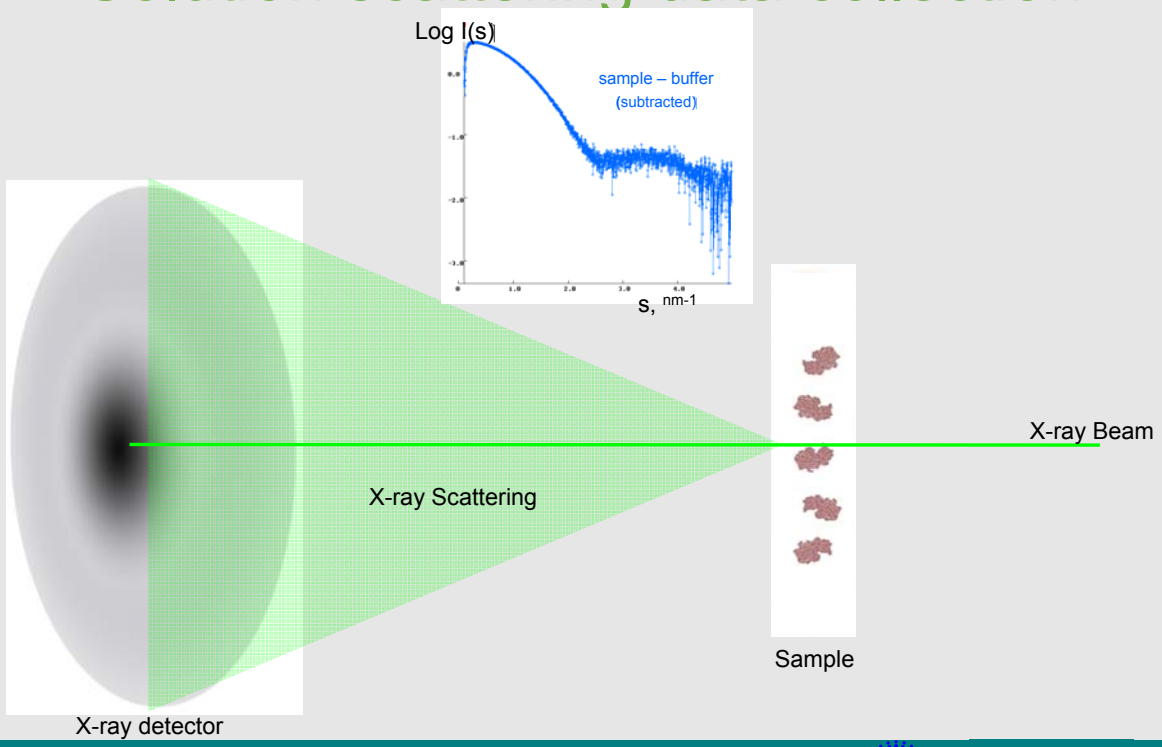
- Model independent parameters: Size! Rg, Dmax, Volume and MM estimates. Basic shape (Extended or Globular)
- Behaviour in different buffer conditions: To assess if interparticle effects or flexibility could be preventing crystallisation (find optimum conditions)
- Complete high resolution structure known: Validation of crystal structure in solution under physiological conditions
- 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 shape
- Dynamic investigations under physiological conditions: Conformational changes with temperature, pH, binding etc.

ESRF EMBL

Part 2

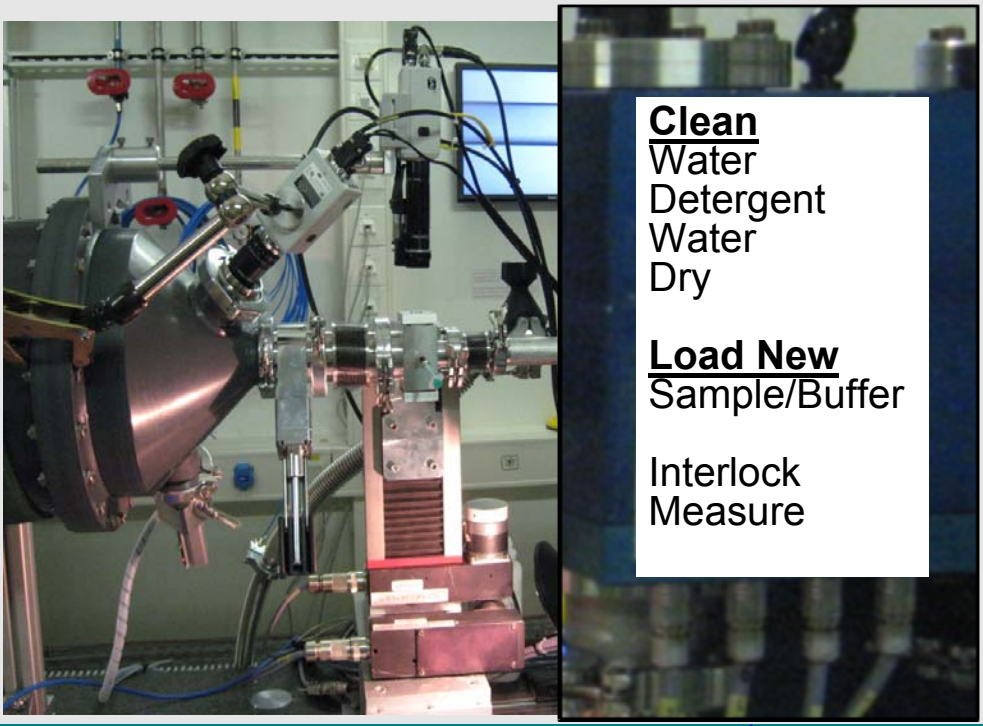
Data collection at the ESRF Bio-SAXS beamline ID14-3

Solution scattering data collection



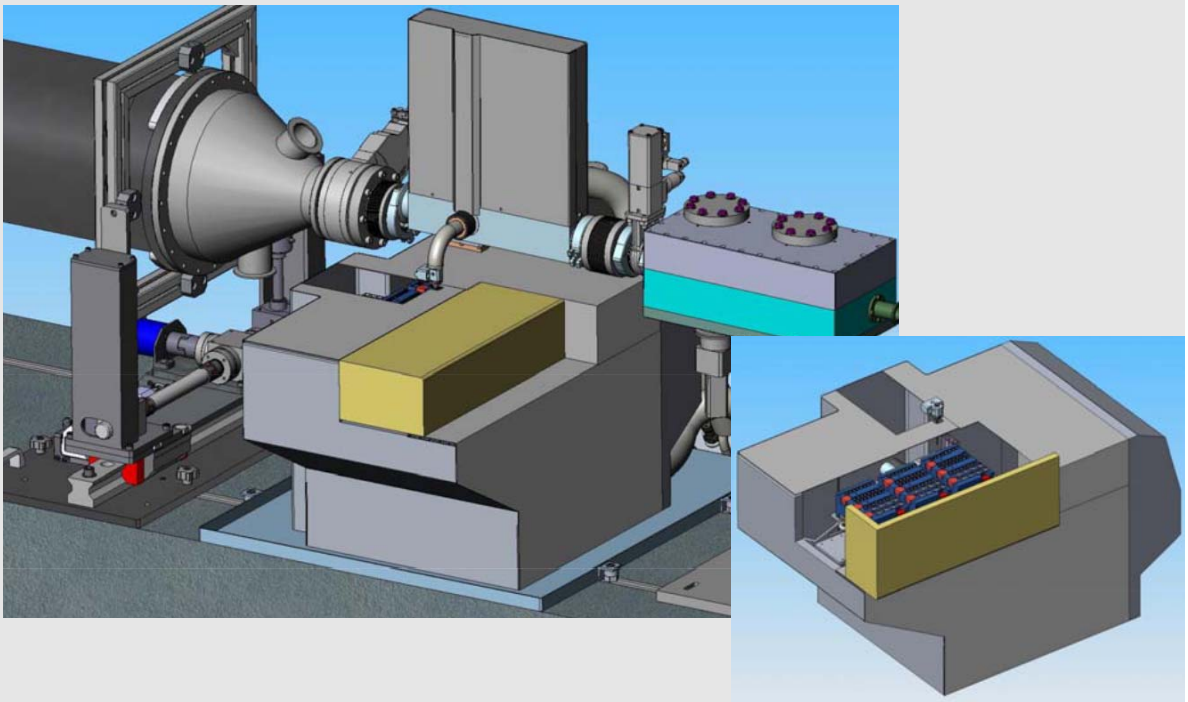
Thanks to A. Kikhney for this slide

Experimental Procedure



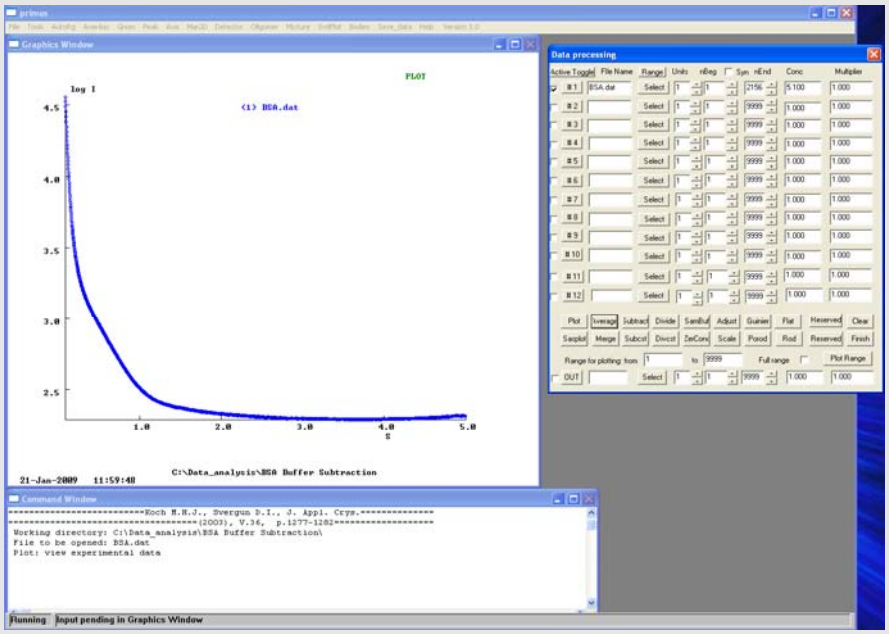
Automation: Sample loading and cleaning

Automation: Sample loading and cleaning



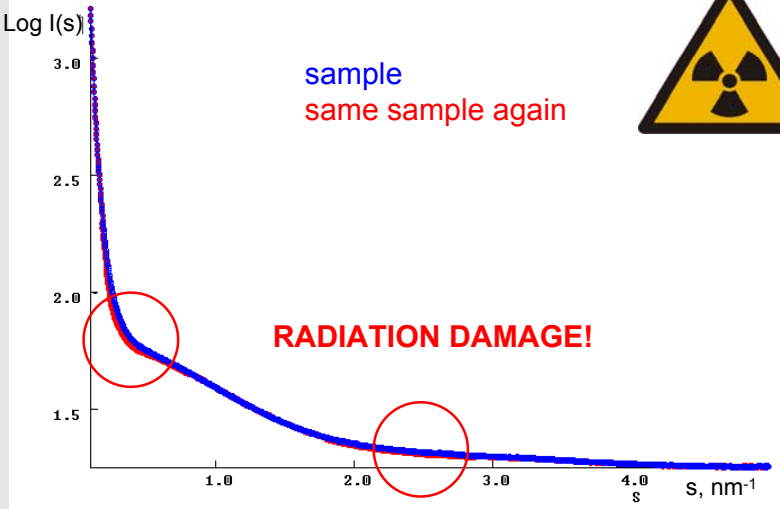
Primary Data Processing

PRIMUS



Quality Control Tests

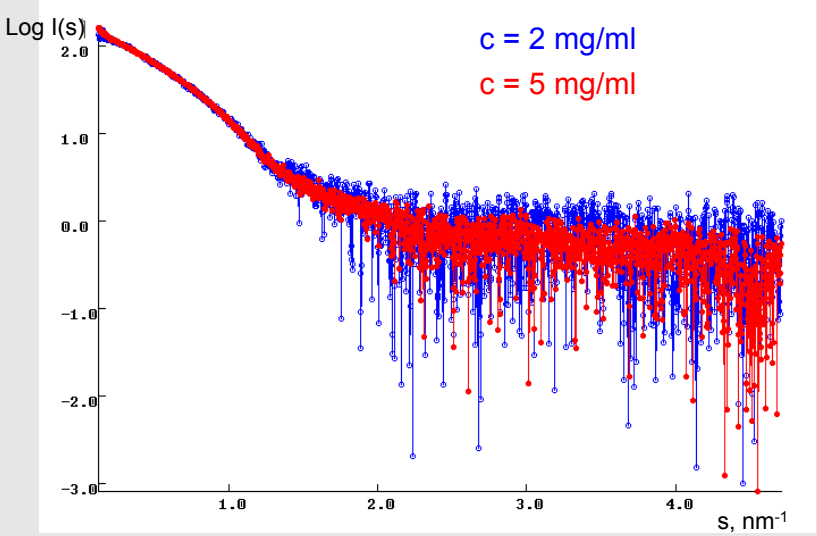
Multiple time frames used to check for radiation damage!



Thanks to A. Kikhney for this slide

Merging Data

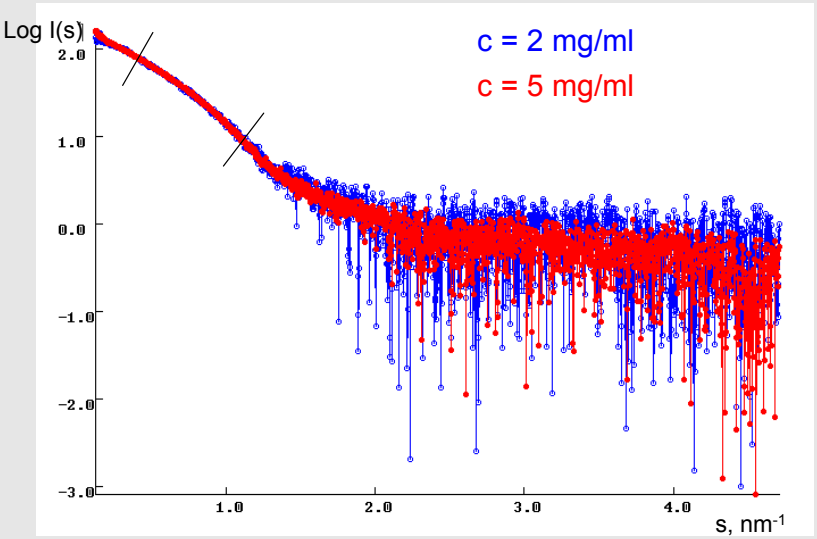
***Low* and *High* Concentration**



Thanks to A. Kikhney for this slide

Merging Data

Low and *High* Concentration

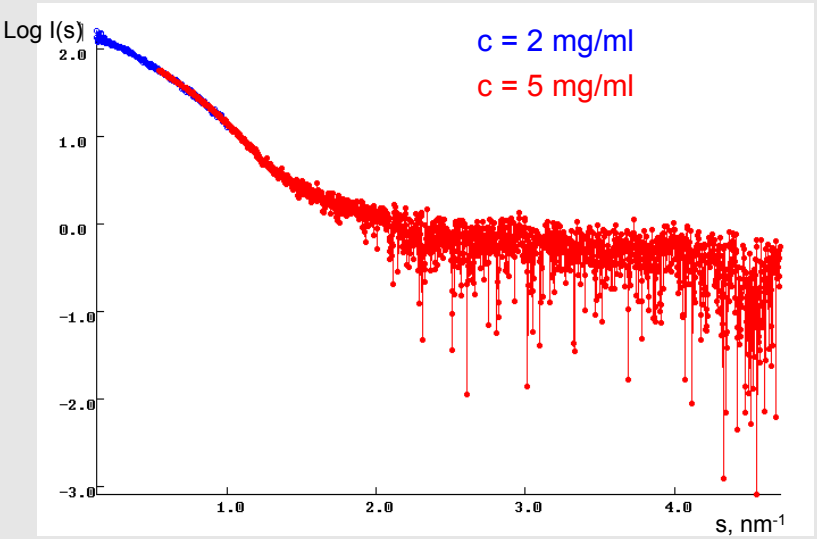


Thanks to A. Kikhney for this slide

ESRF EMBL

Merging Data

Low and *High* Concentration

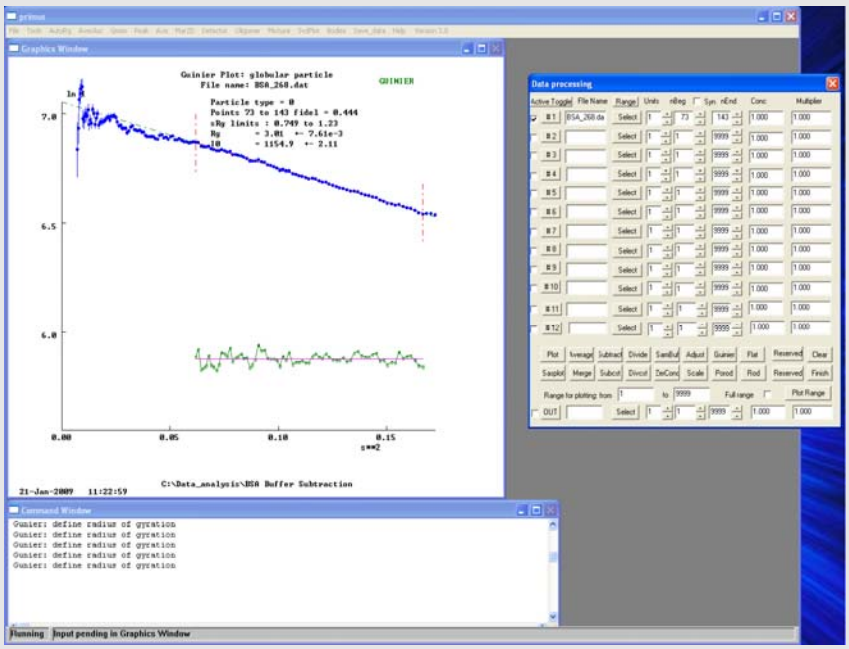


Thanks to A. Kikhney for this slide

ESRF EMBL

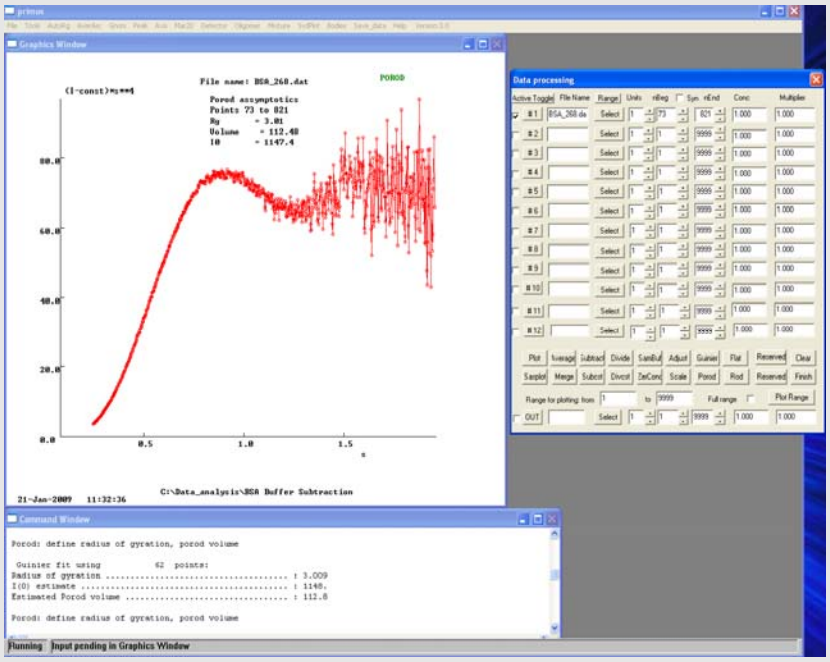
Rg and I(0)

Radius of Gyration and Zero Angle Intensity



EMBL

Porod Analysis in Primus gives excluded volume

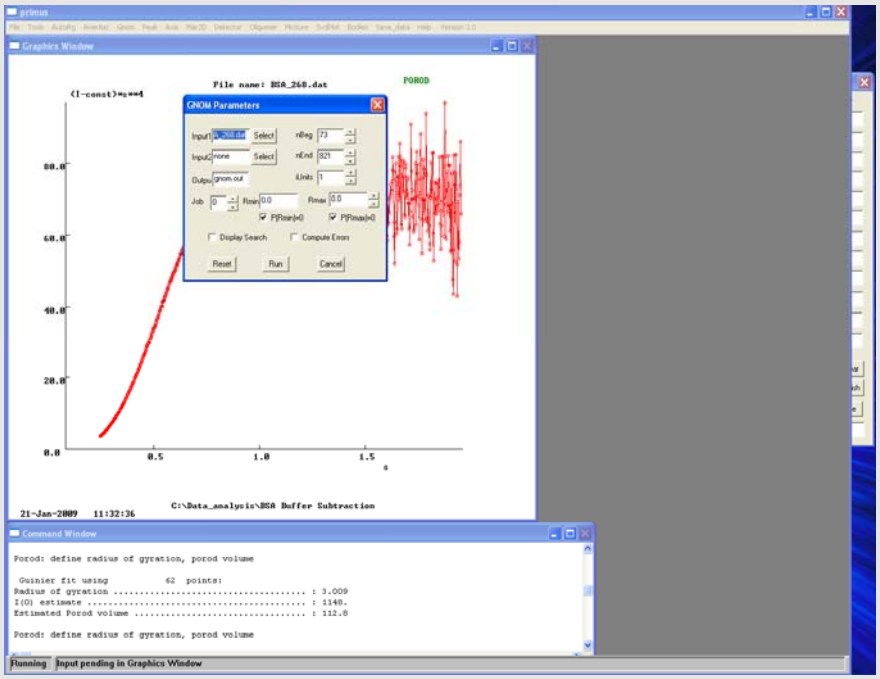


ESRF

EMBL

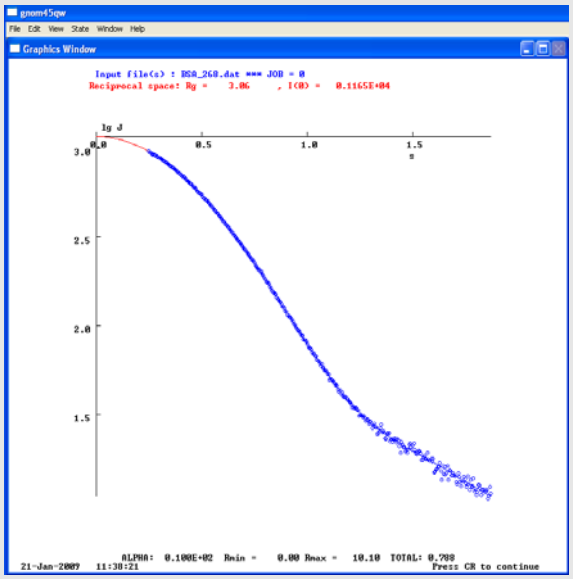
Distance Distribution P(r) Function

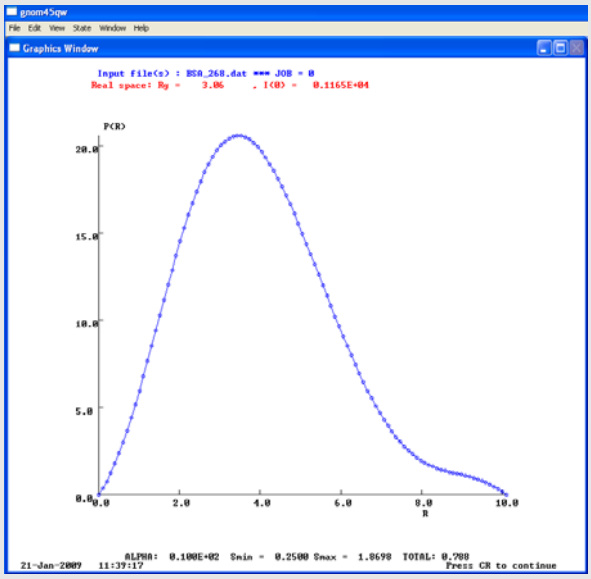
Calculated by GNOM give the Dmax and the input file required for *Ab-initio* modeling



Distance Distribution P(r) Function

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



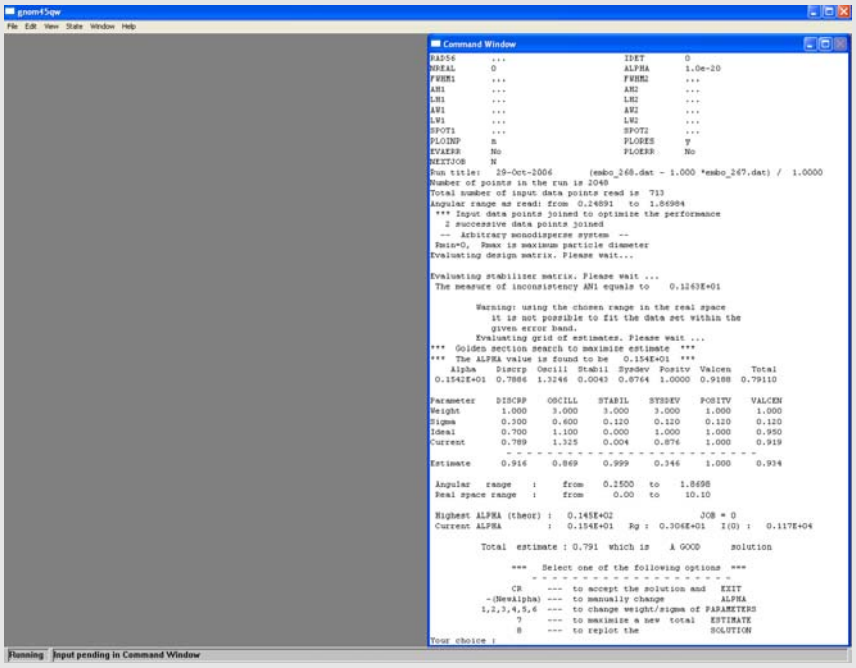


Indirect Fourier Transform!

ESRF EMBL

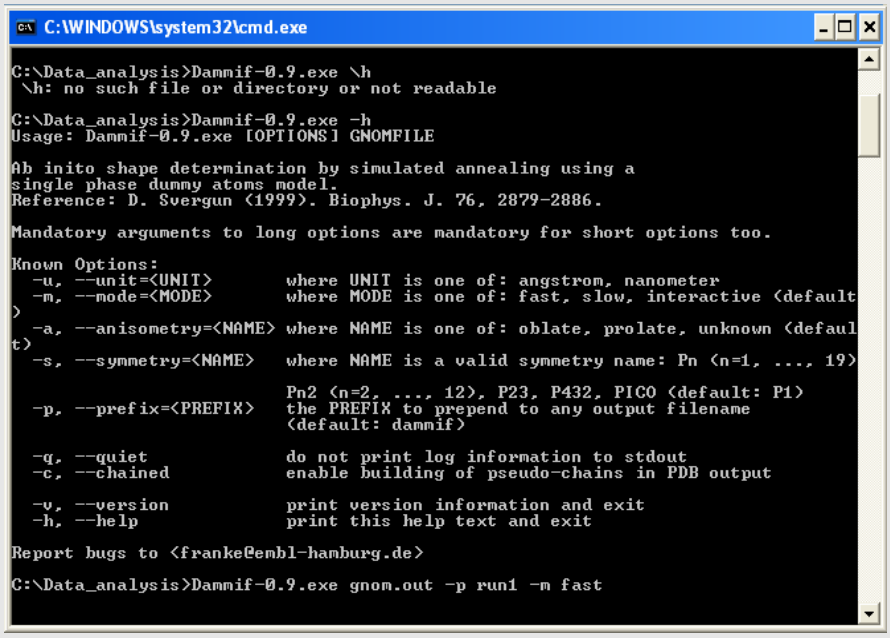
Distance Distribution P(r) Function

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



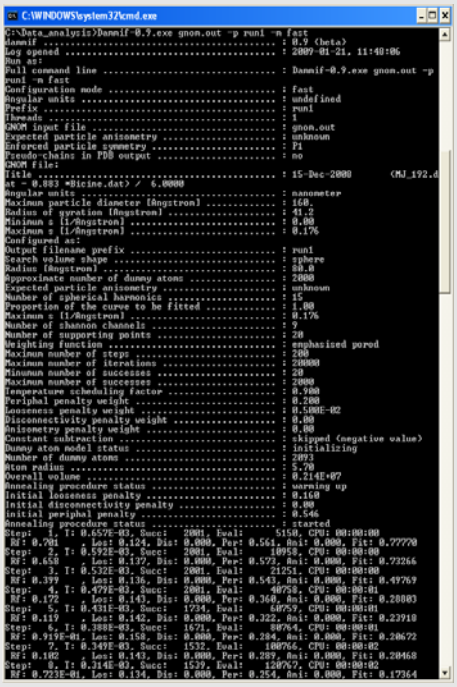
Distance Distribution P(r) Function

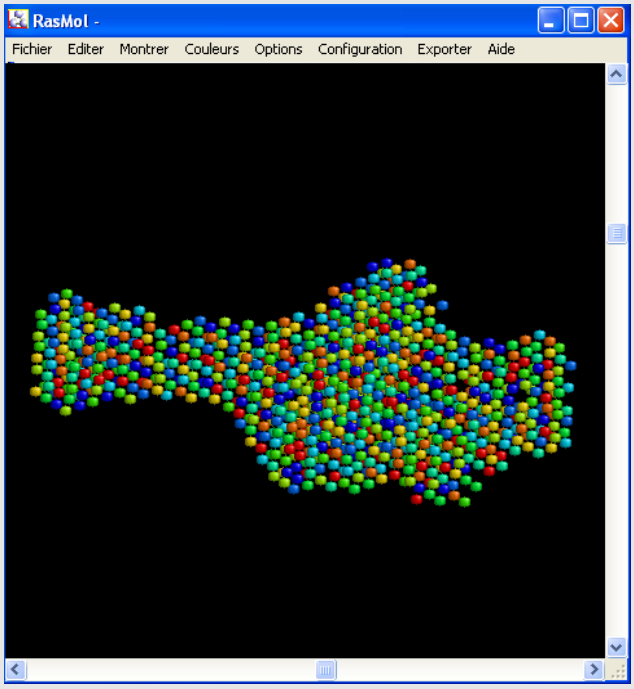
Calculated by GNOM give the Dmax and the input file required for *Ab-initio* modeling



Distance Distribution P(r) Function

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



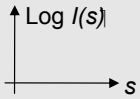


Summary

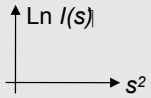
what should be done while at the beamline

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

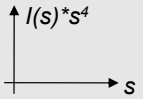
- Log plot



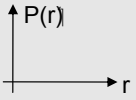
- Guinier plot (R_g, MM)



- Porod plot (volume)

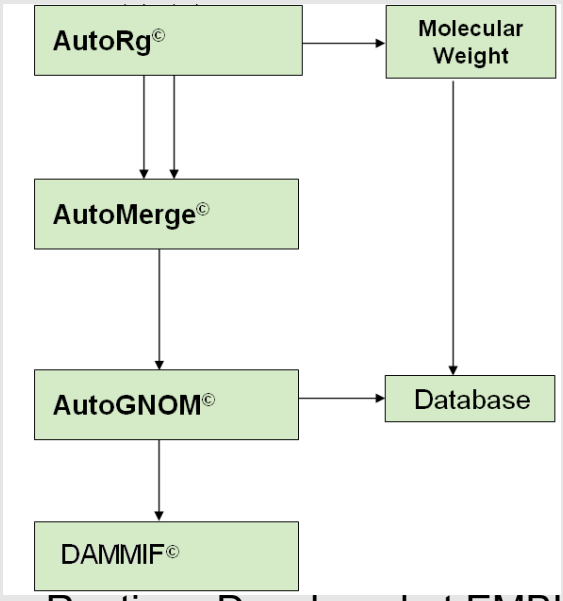


- P(r) plot



- Lo-res 3D model

Automation: Data Processing



Processing Routines Developed at EMBL Hamburg
Results To be Stored in ESRF IspyB Databse
(Currently being modified for bio-SAXS)

Automation: Data Processing

Log started: 2009-05-23 10:47

Run #	File	Conc., mg/ml	Description	R_g, nm	I(0)	Guinier points	D_{max}, nm	MM, kDa	Volume, nm^3	Quality, %	Comments
3	bsa_003.dat	2.0	bsa	3.05±0%	111.50	71 - 134 (64)	10.5	66	143	81	11:23
6	cnq_006.dat	1.0	cnq 2-1 1	5.01±1%	87.87	27 - 71 (45)	16.0	46	225	88	11:33
8	cnq_008.dat	2.0	cnq 2-1 2	5.40±0%	132.10	21 - 63 (43)	18.9	70	388	91	11:48
10	cnq_010.dat	6.3	cnq 2-1 6.3	6.64±0%	265.80	15 - 45 (31)	21.2	140	544	91	12:03
12	cnq_012.dat	2.4	cnq 4-3 2.4	5.41±0%	75.43	38 - 64 (27)	18	P(r) function	135	84	12:17
14	cnq_014.dat	1.0	cnq 6-1 1	1.98±0%	21.06	63 - 218 (156)	6.5	11	22	86	12:37
16	cnq_016.dat	4.0	cnq 6-1 4	2.01±0%	23.28	66 - 217 (152)	7.0	12	19	86	12:51
18	cnq_018.dat	10.0	cnq 6-1 10	2.08±0%	23.44	88 - 207 (120)	7.3	12	47	79	13:11 Aggregated.
21	mig1_021.dat	1.0	mig1	2.61±0%	202.40	53 - 161 (109)	8.5	107	85	87	13:44
23	mig2_023.dat	1.0	mig2	2.57±0%	197.20	83 - 164 (82)	8.4	104	45	78	14:01
25	mgp1_025.dat	1.0	mgp1	2.48±0%	110.50	69 - 163 (95)	8.1	58	32	83	14:15
27	mgp2_027.dat	4.0	mgp2	2.39±0%	95.07	75 - 177 (103)	7.6	50	34	81	14:30
29	ind1_029.dat	1.0	indica	2.73±0%	334.80	58 - 151 (94)	9.3	177	105	85	14:45
31	ind2_031.dat	1.0	indica	3.24±3%	132.80	74 - 124 (51)	10.9	70	106	56	14:59
34	zipkwt_034.dat	1.0	zip kinase wild	3.37±2%	139.30	59 - 118 (60)	11.8	74	45	82	15:29
36	zipk180_036.dat	1.0	zip kinase mutant	3.56±4%	54.25	50 - 111 (62)	12.5	29	133	68	15:44
41	zipk265_041.dat	1.0	zip kinase mutant	3.51±3%	74.51	50 - 113 (64)	11.2	39	45	65	16:40
43	drp1k1_043.dat	2.0	drp kinase full	5.46±0%	277.50	42 - 63 (22)	17.9	147	306	65	16:55
45	drp1k2_045.dat	2.0	drp kinase full	3.54±0%	912.90	60 - 98 (39)	12.4	482	180	80	17:10
47	drp1atp_047.dat	2.0	drp 1 kinase	3.34±0%	177.00	34 - 111 (78)	10.9	93	32	49	17:24 Low quality.

Developed at EMBL Hamburg

Image courtesy of A. Kikhney

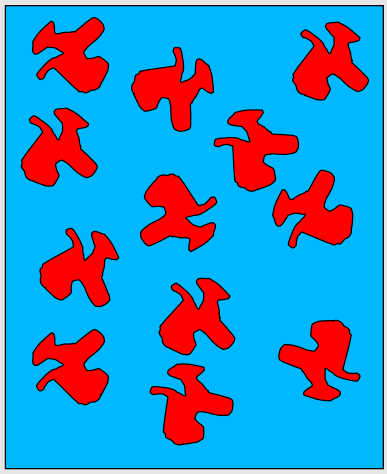
Part 3

Preparation for Bio-SAXS Data collection

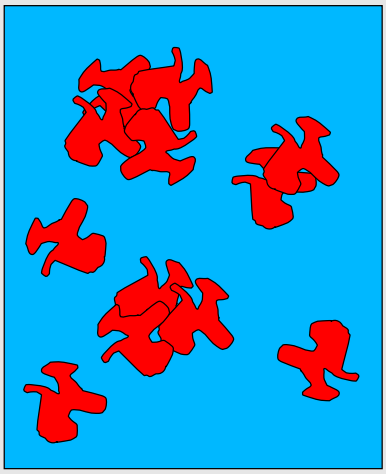
Sample Preparation

In solution SAXS we observe the **Average**

Monodisperse



Mixture



Average

Sample Preparation

If your samples are pure the scattering represents the individual subunit

With this assumption we have one single shape that represents the scattering data and we can calculate models of the low resolution shape of unknown proteins

Contamination will affect the average make the assumption of an individual shape invalid and although you will get models they will be increasingly different from the true shape the more contamination you have.

Sample Preparation

A good experiment requires:

MONODISPERSE!

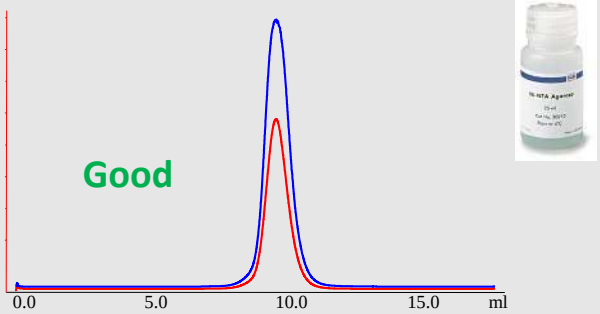
samples in solution

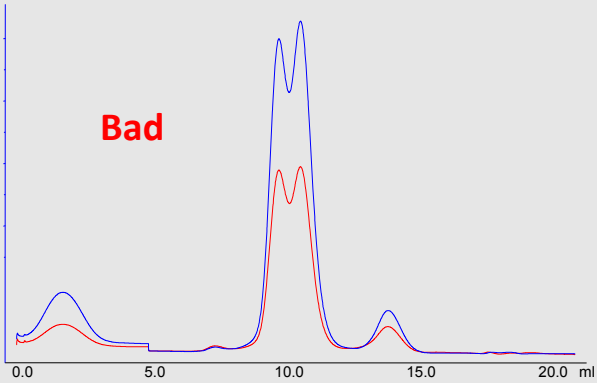
Pure protein (>90%)!
In only 1 oligomeric state!
NO agregation!
Free from antiparticle effects!

Check before you go to the beamline

MALS/DLS
Analytical ultra centrifugation

Sample Preparation





ESRF EMBL

Sample Preparation

Know your protein!

In solution samples can degrade, aggregate, react!

Know the best buffer conditions for stability!

Know what storage temperature is required!
Take adequate precautions when shipping samples
Extra ice! It may take longer than you expect!

If you cant purify and store stock solution for shipping you can request access to preparation facilities to do a final purification immediately prior to measurements

ESRF EMBL

Summary

- Bio-SAXS gives:
  - Size! Rg, Dmax, Volume and MM estimates.
  - Basic shape (Extended or Globular)
  - Information on why crystallisation isn't happening (find optimum conditions)
  - Validation of crystal structure in solution under physiological conditions
  - *ab-initio* low resolution shape
  - Quaternary structure using docking/rigid body refinement
  - Dynamic investigations under physiological conditions
- ESRF Bip-SAXS Beamline is easy to use:
  - High level of automation for
    - Data collection (collaboration between ESRF and EMBL GR and HH outstations)
    - Data processing (ESRF BLISS)
    - Data Analysis (EMBL-Hamburg data processing pipeline)
- Sample preparation and experimental design are important:
  - Monodispersity
  - Interparticle effects
- Information from complimentary techniques is very helpful:
  - MX, NMR, Electron Microscopy

ESRF EMBL

Acknowledgments

ESRF

Sine Larson
Serge Perez
Sean McSweeney
Gordon Leonard
Matthew Bowler
Petra Pernot
Sasha Popov
Daniele de Sanctis
Christoph Müller-Dieckmann
Joanna Timmins
Didier Nurizzo
David Flot
Elspeth Gordon
Debby Davison
Pascal Theveneau
Menhard Menyhert Kocsis
John Surr
Thierry Giraud
Ricardo Nogueira Fernandes
Darren Spruce

EMBL

Grenoble

Stephen Cusack
Florent Cipriani
Andrew McCarthy
Sandor Brockhauser
Franck Felisaz
Lukas Födinger
Alexandre Gobbo
Julien Huet
Bernard Lavault
Cyril Villard

Hamburg

Dmitri Svergun
Manfred Roessle
Stefan Fiedler
Uve Ristau
Clement Blanchet
Daniel Franke
Michal Jan Gajda
Christian Gorba

Data Examples

Julia Santiago
Marcello Clerici
Florine Dupeux
Regina Antoni
Sang-Youl Park
Sean R. Cutler
Pedro Luis Rodriguez
José Antonio Márquez
Daniel Panne

Teemu Ikonen
Alexey Kikhney
Haydyn Mertens
Maxim Petoukhov
Weifeng Shang
Aliaksandr Shkumatau
Alexey Zozulya

And many more people who gave advise and help with testing hardware and software at ID14-3

ESRF EMBL