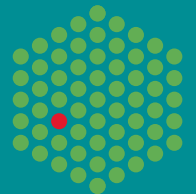




Automated data analysis on ESRF BM29

Martha Brennich (EMBL Grenoble)

EMBL

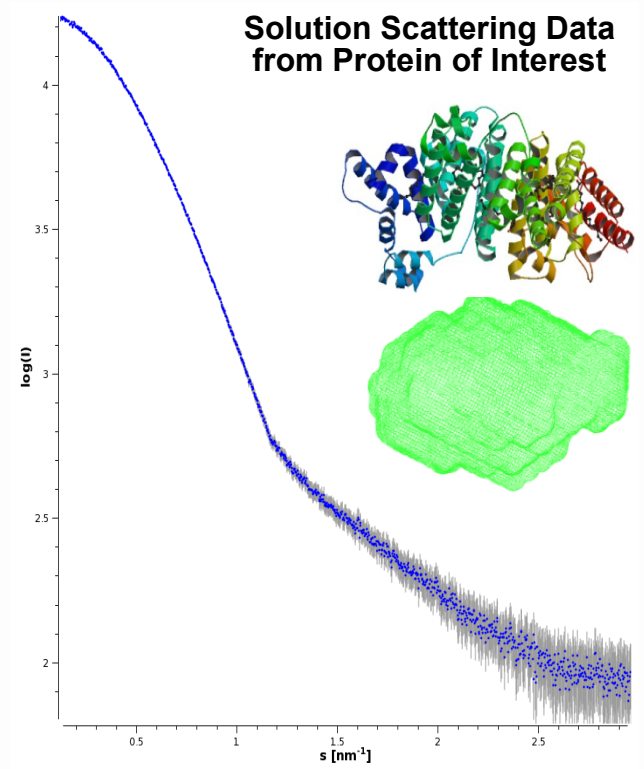


Idealized bio-SAS experiment



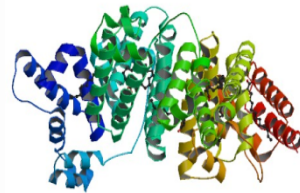
Black Box

Neutron source/beamline
homesource

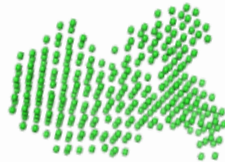
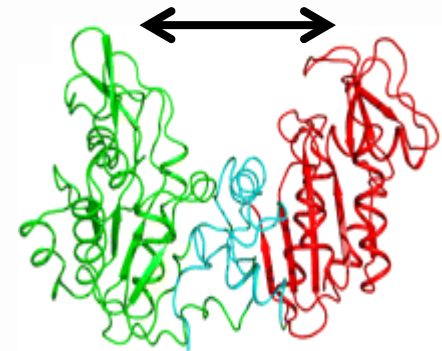


What can we learn from BioSAXS?

- Low-resolution structural information – shape, overall fold
- Mean molecular weight, oligomeric state
- Mixing ratios
- Model validation
- Domain placement
- Complex structures
- *Ab-initio* models
- ...



?

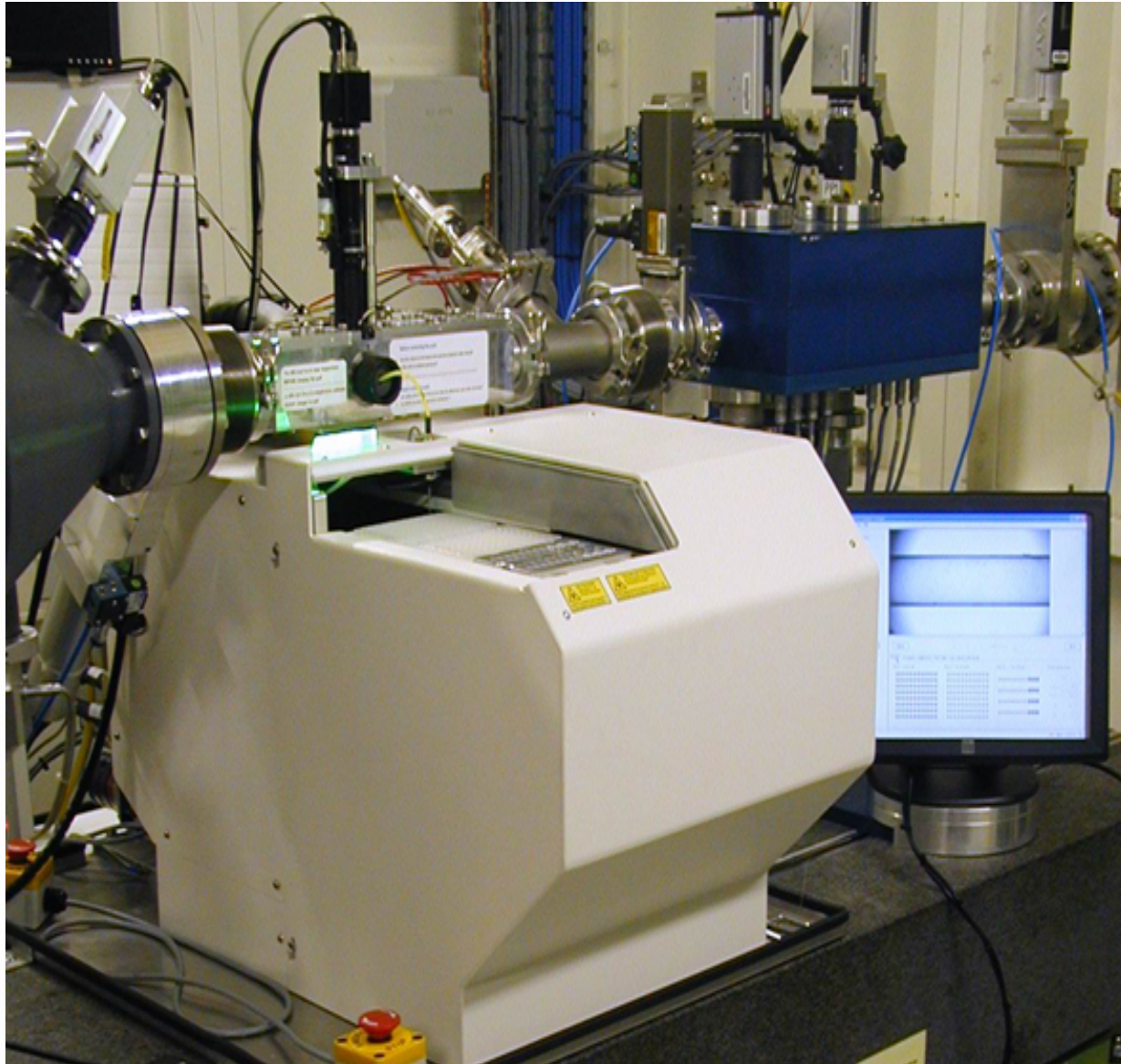




BM29

- Dedicated solution scattering beamline
- Optimized for macromolecules (4kDa -1MDa)
- Many “non-expert” users, short visits

Automated sample Handling



Inline HPLC

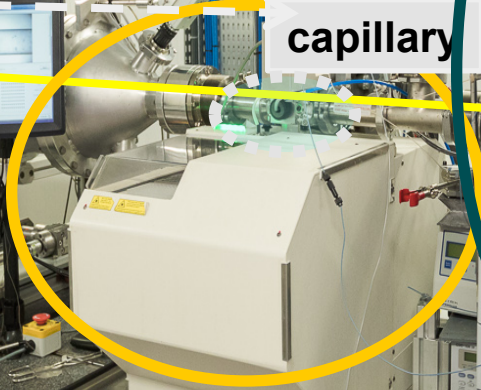
detector

3 m

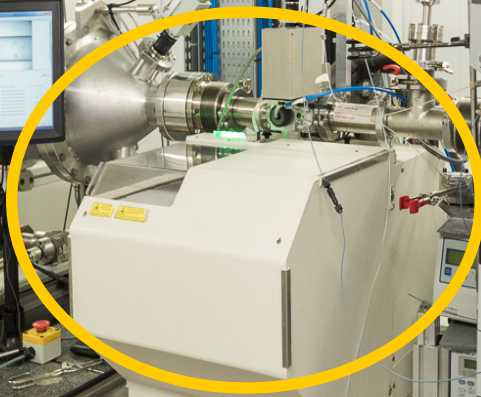
capillary

EMERGENCY STOP
ARRÊT D'URGENCE
x-rays

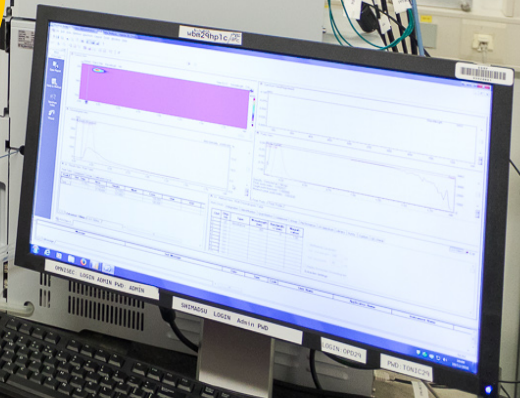
sample
changer



sample
changer



EMERGENCY
STOP
ARRÊT
D'URGENCE



Automated data acquisition

Parameters

File:

Sample type:

Storage temperature:

Extra flow time:

Optimization:

Initial Cleaning:

Buffer mode:

History

```
[2011,05/13 15:36:12] INFO: Finish collecting 'May'...
[2011,05/13 15:36:12] INFO: Waiting for end of flow...
[2011,05/13 15:36:27] INFO: Cleaning...
[2011,05/13 15:36:58] INFO: The data collection is done!
```

	Use	Type	Plate	Row	Well	Concentration	Comments	Code	Viscosity	Buffername	Transmission	Volume	SEU Temp	Flow	Recup.	Wait Time	D
1	<input checked="" type="checkbox"/>	Buffer	2	A	9	0.00 mg/ml	hepes ph7	BUFF1	Low	Buf1	100.0 %	30 u/l		<input checked="" type="checkbox"/>	<input type="checkbox"/>		
2	<input checked="" type="checkbox"/>	Sample	2	A	1	1.00 mg/ml	Construct A	A	Low	Buf1	100.0 %	30 u/l	5.00 C	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	0 sec	
3	<input checked="" type="checkbox"/>	Sample	2	A	2	2.50 mg/ml	Construct A	A	Low	Buf1	100.0 %	30 u/l	5.00 C	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	0 sec	
4	<input checked="" type="checkbox"/>	Sample	2	A	3	5.00 mg/ml	Construct A	A	Low	Buf1	100.0 %	30 u/l	5.00 C	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	0 sec	
5	<input checked="" type="checkbox"/>	Sample	2	A	4	1.50 mg/ml	Construct B	B	Low	Buf1	100.0 %	30 u/l	5.00 C	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	0 sec	
6	<input checked="" type="checkbox"/>	Sample	2	A	5	3.00 mg/ml	Construct B	B	Low	Buf1	100.0 %	30 u/l	5.00 C	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	0 sec	
7	<input checked="" type="checkbox"/>	Sample	2	A	6	6.00 mg/ml	Construct B	B	Low	Buf1	100.0 %	30 u/l	5.00 C	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	0 sec	
8	<input checked="" type="checkbox"/>	Sample	2	B	1	1.00 mg/ml	Construct AB	AB	Low	Buf1	100.0 %	30 u/l	5.00 C	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	0 sec	
9	<input checked="" type="checkbox"/>	Sample	2	B	1	2.00 mg/ml	Construct AB	AB	Low	Buf1	100.0 %	30 u/l	5.00 C	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	0 sec	
10	<input checked="" type="checkbox"/>	Sample	2	B	1	4.00 mg/ml	Construct AB	AB	Low	Buf1	100.0 %	30 u/l	5.00 C	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	0 sec	
11	<input checked="" type="checkbox"/>	Sample	2	B	1	8.00 mg/ml	Construct AB	AB	Low	Buf1	100.0 %	30 u/l	5.00 C	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	0 sec	
12	<input checked="" type="checkbox"/>	Buffer	2	C	9	0.00 mg/ml	pes ph7 ATP	BUFF2	Low	Buf2	100.0 %	30 u/l		<input checked="" type="checkbox"/>	<input type="checkbox"/>		
13	<input checked="" type="checkbox"/>	Sample	2	C	1	1.00 mg/ml	AB + ATP	ABATP	Low	Buf2	100.0 %	30 u/l	5.00 C	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	0 sec	
14	<input checked="" type="checkbox"/>	Sample	2	C	2	2.00 mg/ml	AB + ATP	ABATP	Low	Buf2	100.0 %	30 u/l	5.00 C	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	0 sec	
15	<input checked="" type="checkbox"/>	Sample	2	C	3	4.00 mg/ml	AB + ATP	ABATP	Low	Buf2	100.0 %	30 u/l	5.00 C	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	0 sec	
16	<input checked="" type="checkbox"/>	Sample	2	C	4	8.00 mg/ml	AB + ATP	ABATP	Low	Buf2	100.0 %	30 u/l	5.00 C	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	0 sec	

4

About 3 minutes per
buffer/sample/buffer set

Actual acquisition rate:
10 frames/minute

ISPyB: Prepare your acquisition from

BIOSAXS Experiment Designer

Define Measurements

Define only the macromolecule's measurement you want to make. This wizard will add buffers' measurement needed for subtraction automatically.

Single Measurement | **Concentration Series**

Macromolecules: PGK | Buffer: AMP

How many unknown concentrations do you have?: 3

Exposure. Temp.: 4 | Vol. To Load (µl): 50 | Transmission (%): 100

Wait Time: 0 | Viscosity: low | Flow:

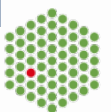
Measurements

Specimen				Parameters						Comments	
Macromo.	Conc. (mg/ml)	Buffer	Exp. Temp.	Vol. Load	Trans.	Wait T.	Flow	Viscosity			
PGK	1.000	AMP	4.00 c	50.00 µl	100 %		yes	low	<input type="button" value="REMOVE"/>		
PGK	2.000	AMP	4.00 c	50.00 µl	100 %		yes	low	<input type="button" value="REMOVE"/>		
PGK	3.000	AMP	4.00 c	50.00 µl	100 %		yes	low	<input type="button" value="REMOVE"/>		

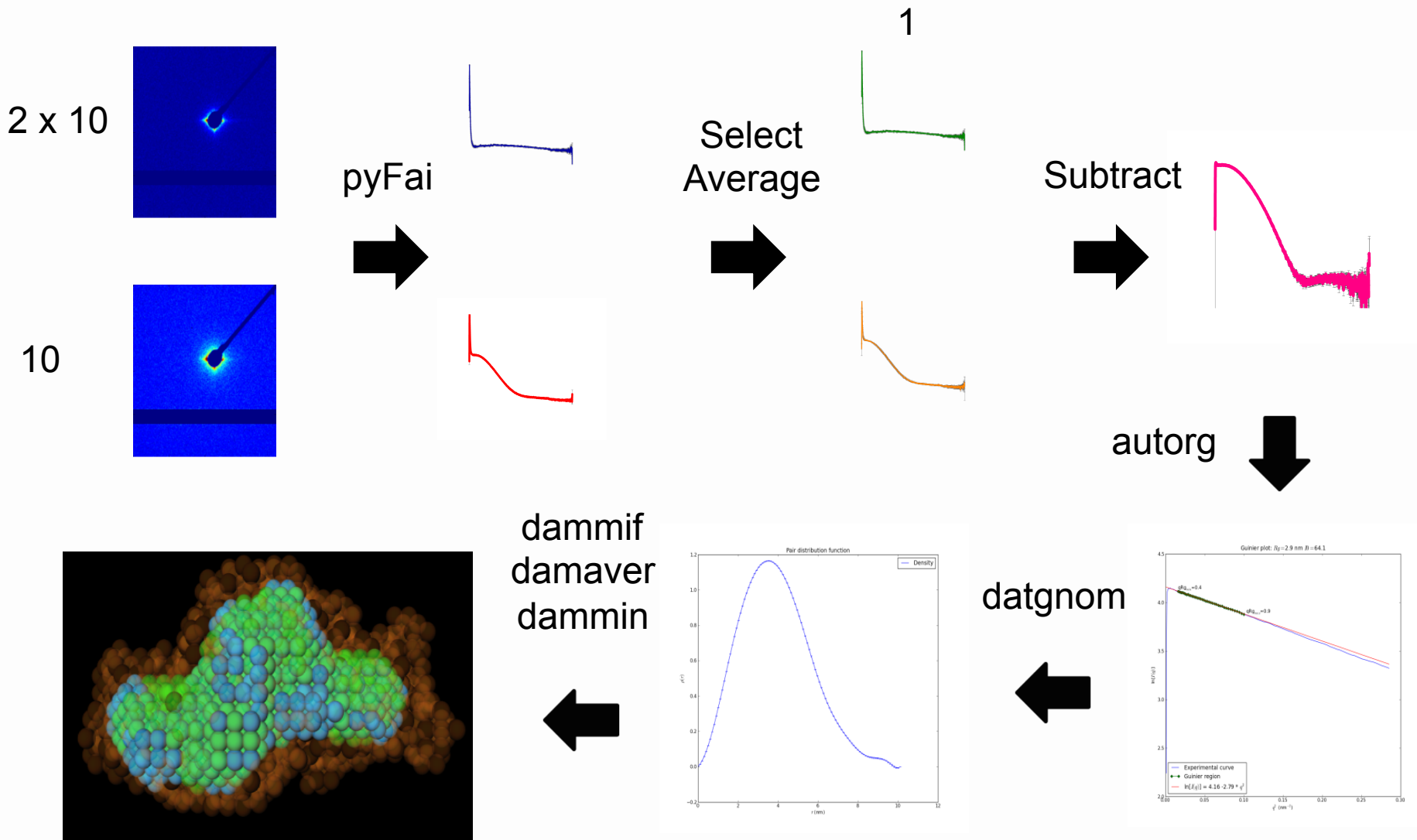
ISPyB: Information System for Protein CrystallographY Beamlines



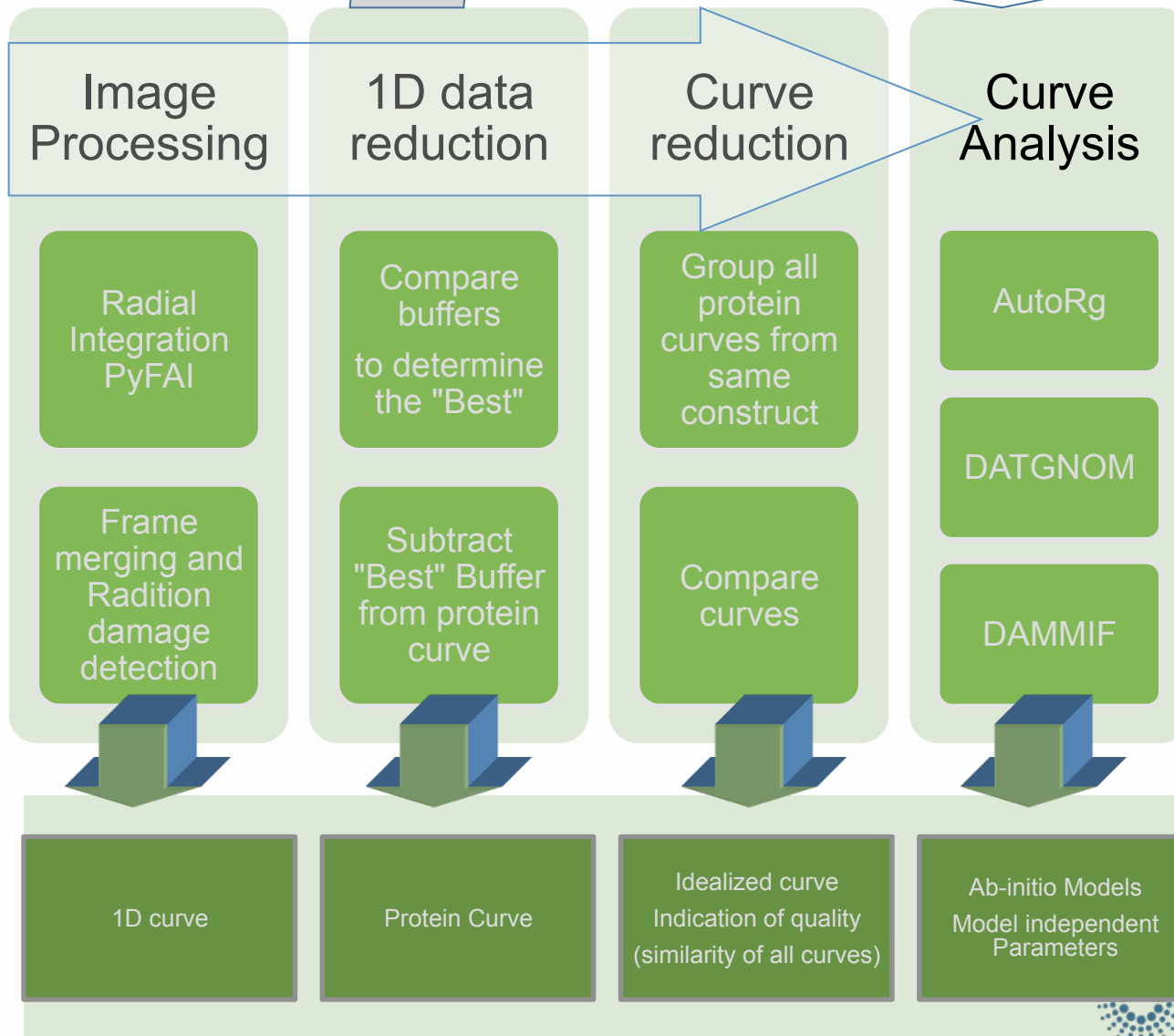
EMBL



Data Processing - EDNA



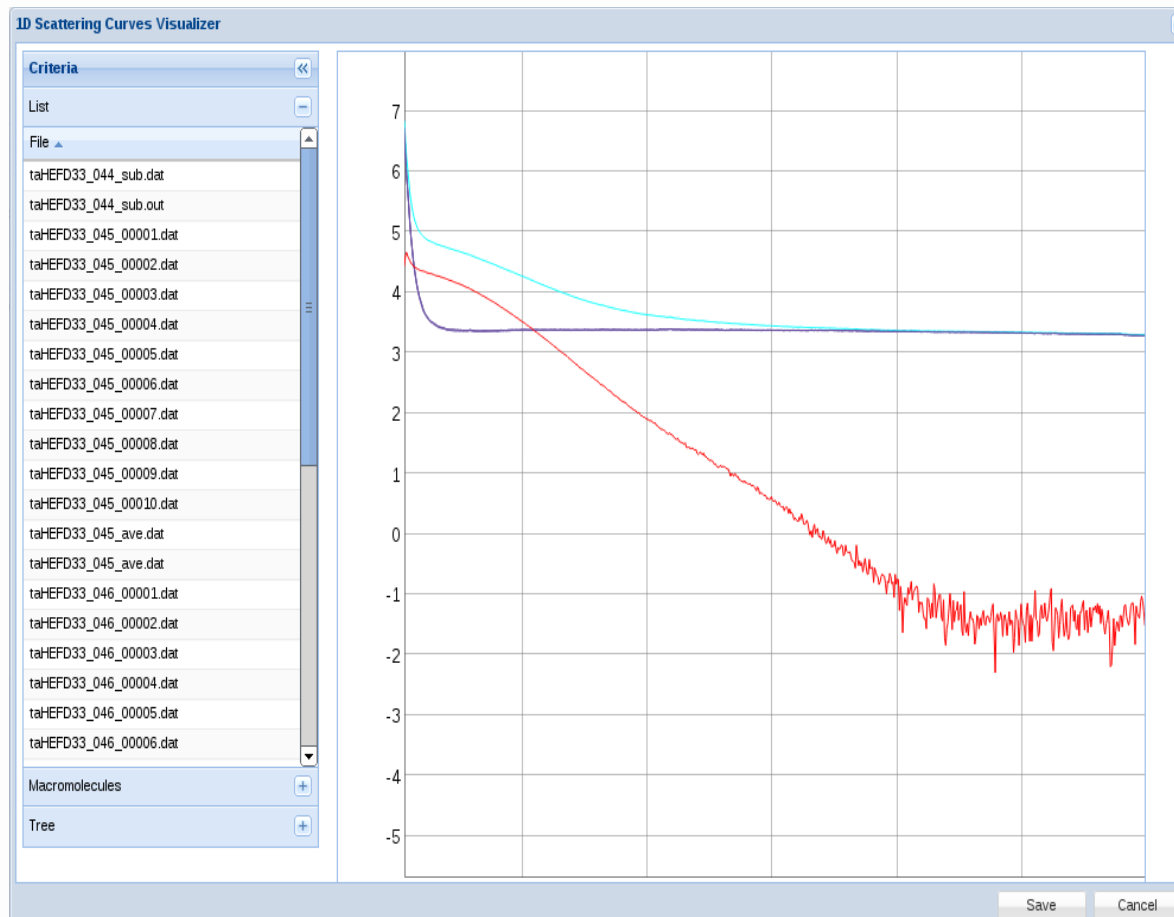
Data Processing - EDNA



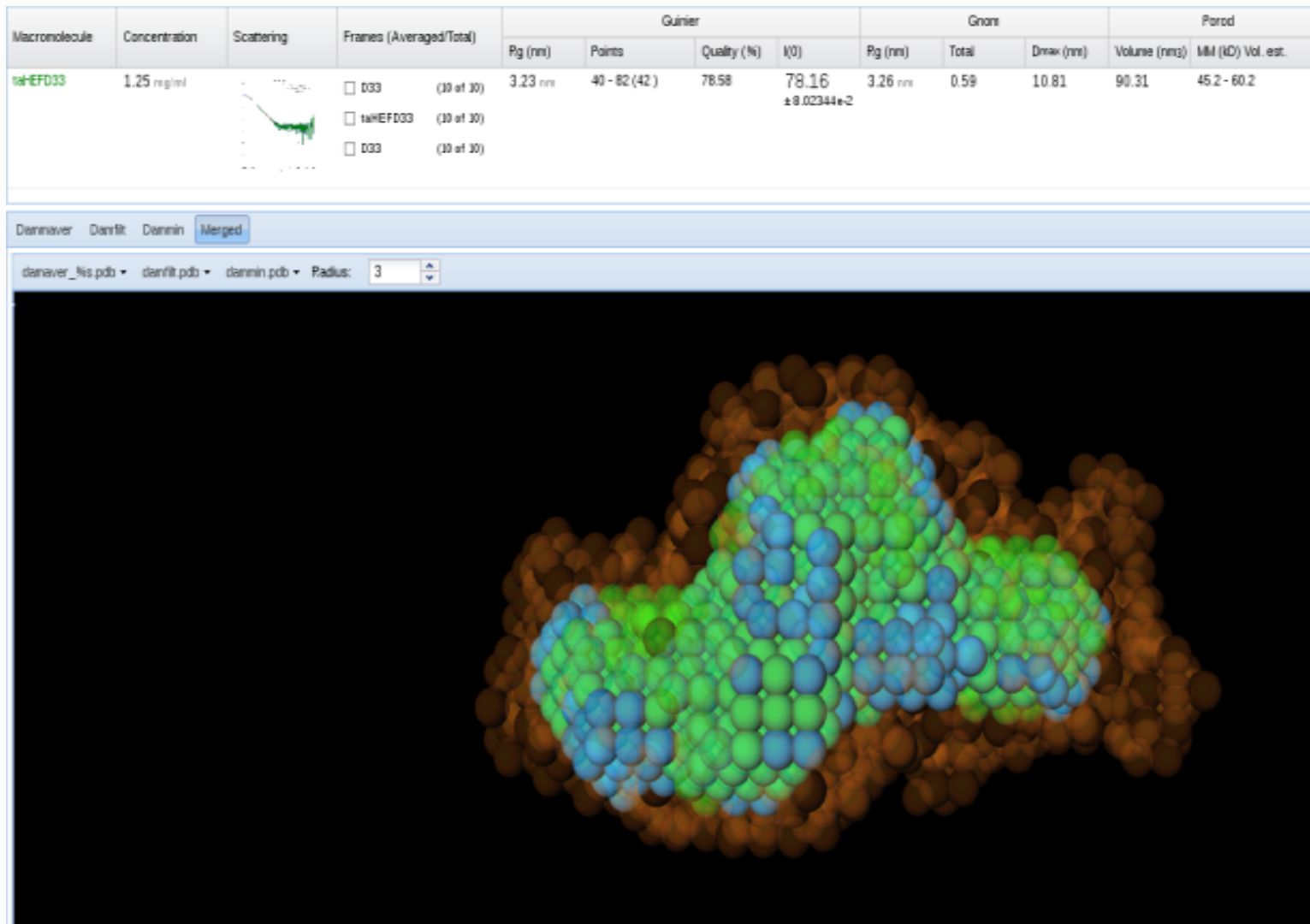
ISPYB: Data Analysis Overview

Overview Measurements Analysis 1D Viewer												
Macromolecule	Concentration	Scattering	Frames (Averaged/Total)	Guinier				Gnom			Porod	
				Rg (nm)	Points	Quality (%)	I(0)	Rg (nm)	Total	Dmax (nm)	Volume (nm ³)	MM (kD) Vol. est.
taHEFD33	14.00 mg/ml		<ul style="list-style-type: none"> D33 (10 of 10) taHEFD33 (10 of 10) D33 (1 of 10) 	4.75 nm	19 - 37 (18)	83.95	90.78 ± 6.88492...	4.94 nm	0.51	24.09	154.27	77.1 - 102.8
taHEFD33	7.00 mg/ml		<ul style="list-style-type: none"> D33 (1 of 10) taHEFD33 (5 of 10) D33 (5 of 10) 	3.97 nm	12 - 42 (30)	92.14	71.21 ± 4.1859e-2	3.91 nm	0.44	13.90	112.54	56.3 - 75.0
taHEFD33	3.50 mg/ml		<ul style="list-style-type: none"> D33 (5 of 10) taHEFD33 (10 of 10) D33 (10 of 10) 	3.37 nm	50 - 77 (27)	72.77	59.53 ± 6.55654...	3.44 nm	0.53	11.81	95.25	47.6 - 63.5
taHEFD33	1.25 mg/ml		<ul style="list-style-type: none"> D33 (10 of 10) taHEFD33 (10 of 10) D33 (10 of 10) 	3.23 nm	40 - 82 (42)	78.58	78.16 ± 8.02344e-2	3.26 nm	0.59	10.81	90.31	45.2 - 60.2
taHEFD33	0.61 mg/ml		<ul style="list-style-type: none"> D33 (10 of 10) taHEFD33 (10 of 10) D33 (10 of 10) 	3.16 nm	27 - 78 (51)	86.16	78.86 ± 9.98563...	3.20 nm	0.75	11.06	84.35	42.2 - 56.2

ISPYB: 1d Visualisation



ISPYB: Model Visualisation



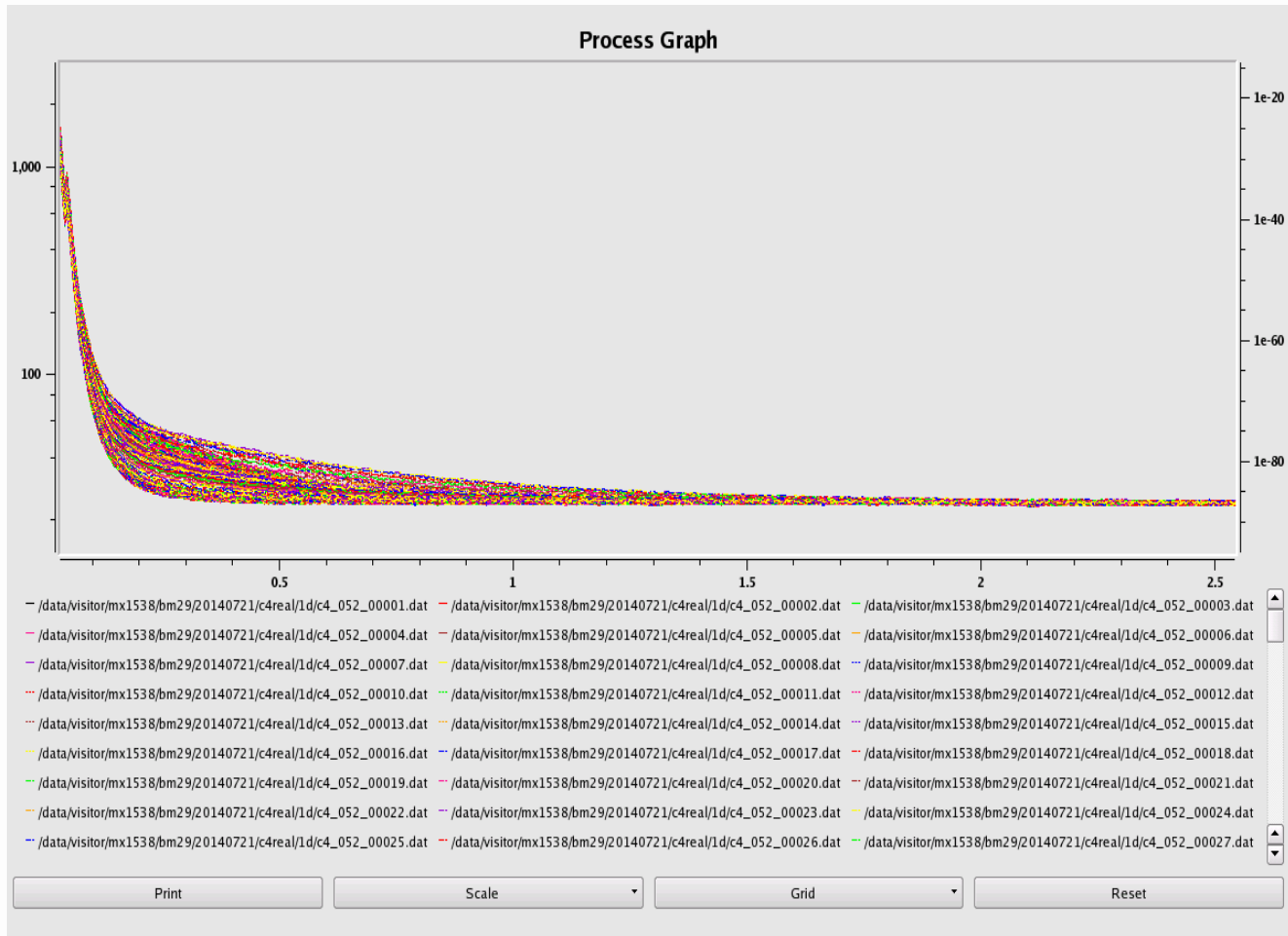
Inline HPLC

sample
changer

EMERGENCY
STOP
ARRÊT
D'URGENCE
x-rays

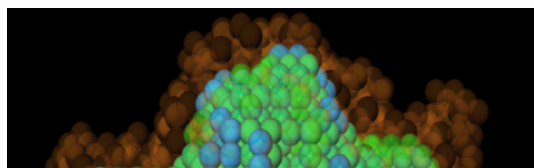
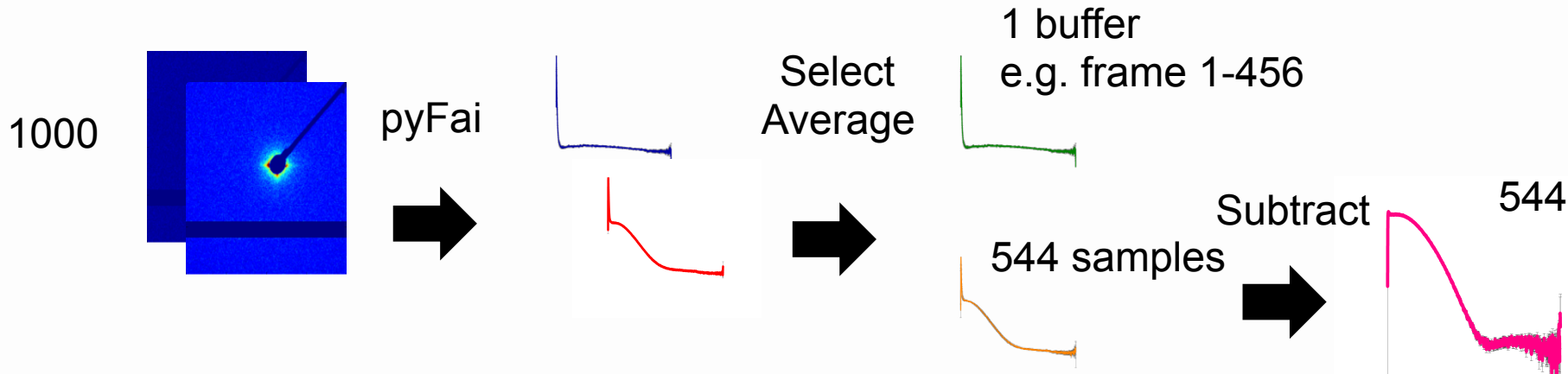


In-situ HPLC – data acquisition



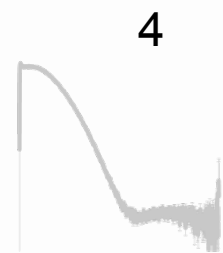
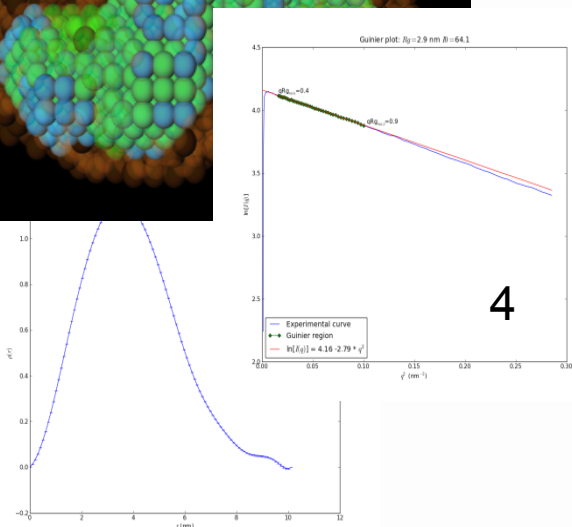
1000 or more single measurements in a dataset

PROCESSING FOR HPLC

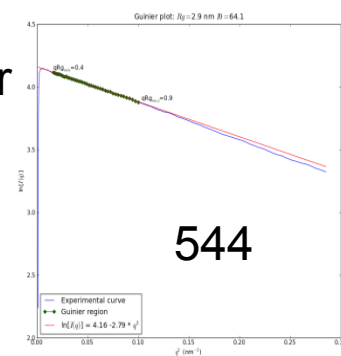


autorg
datgnom
dammif
damaver
dammin

autorg



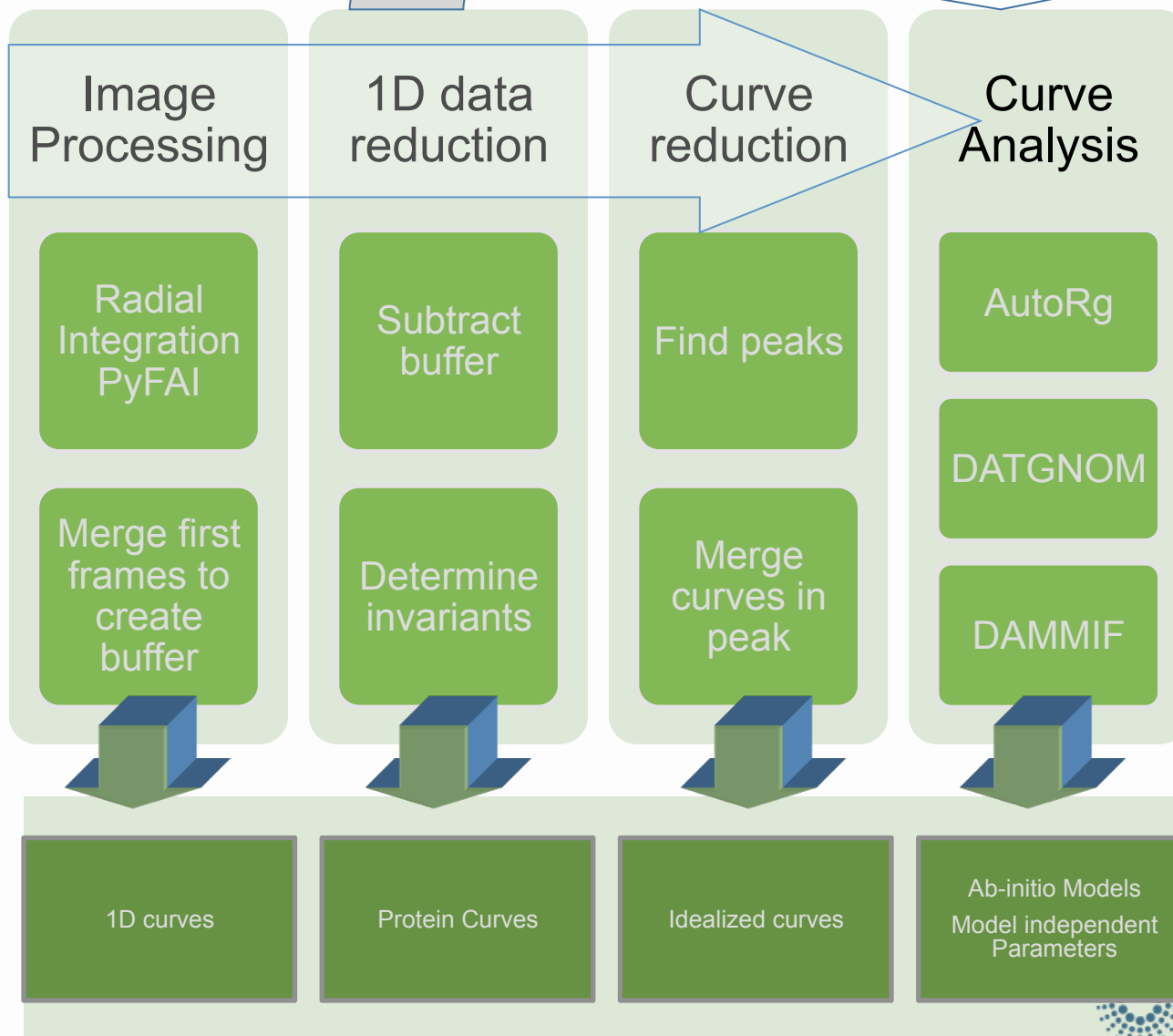
peak finder



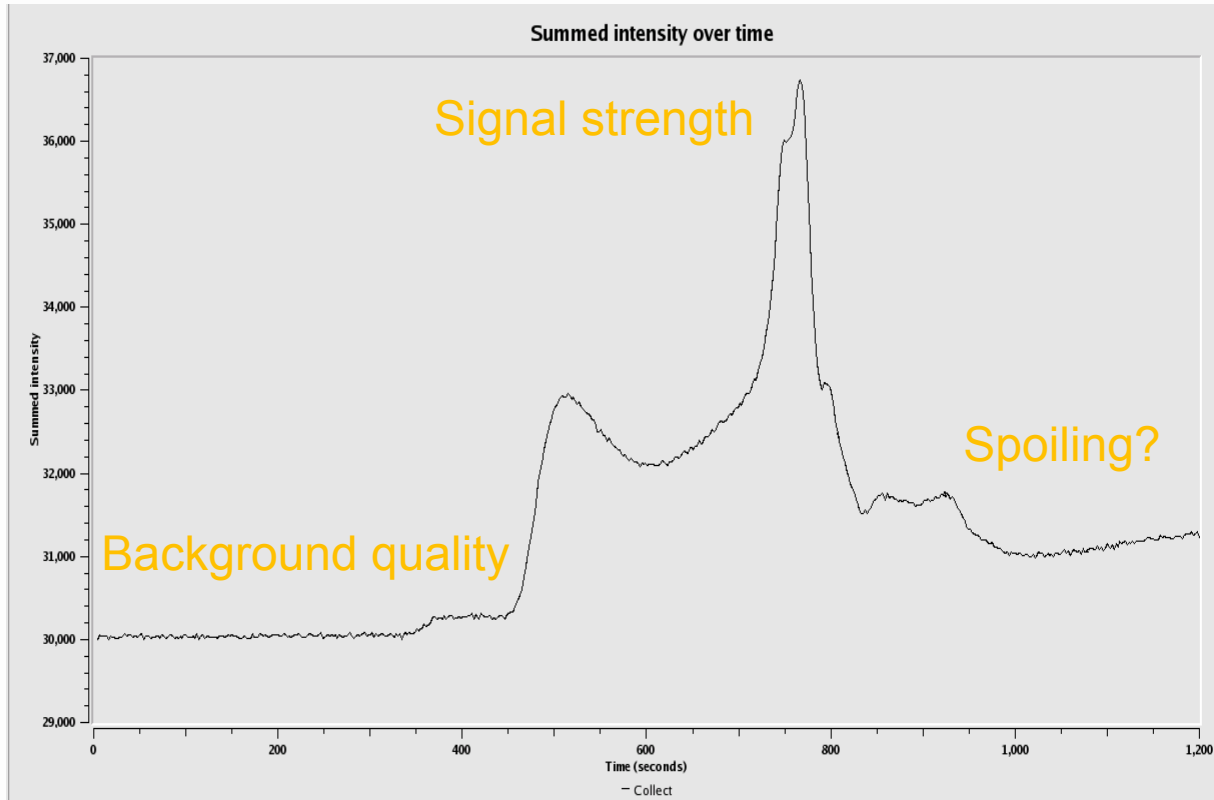
EMBL



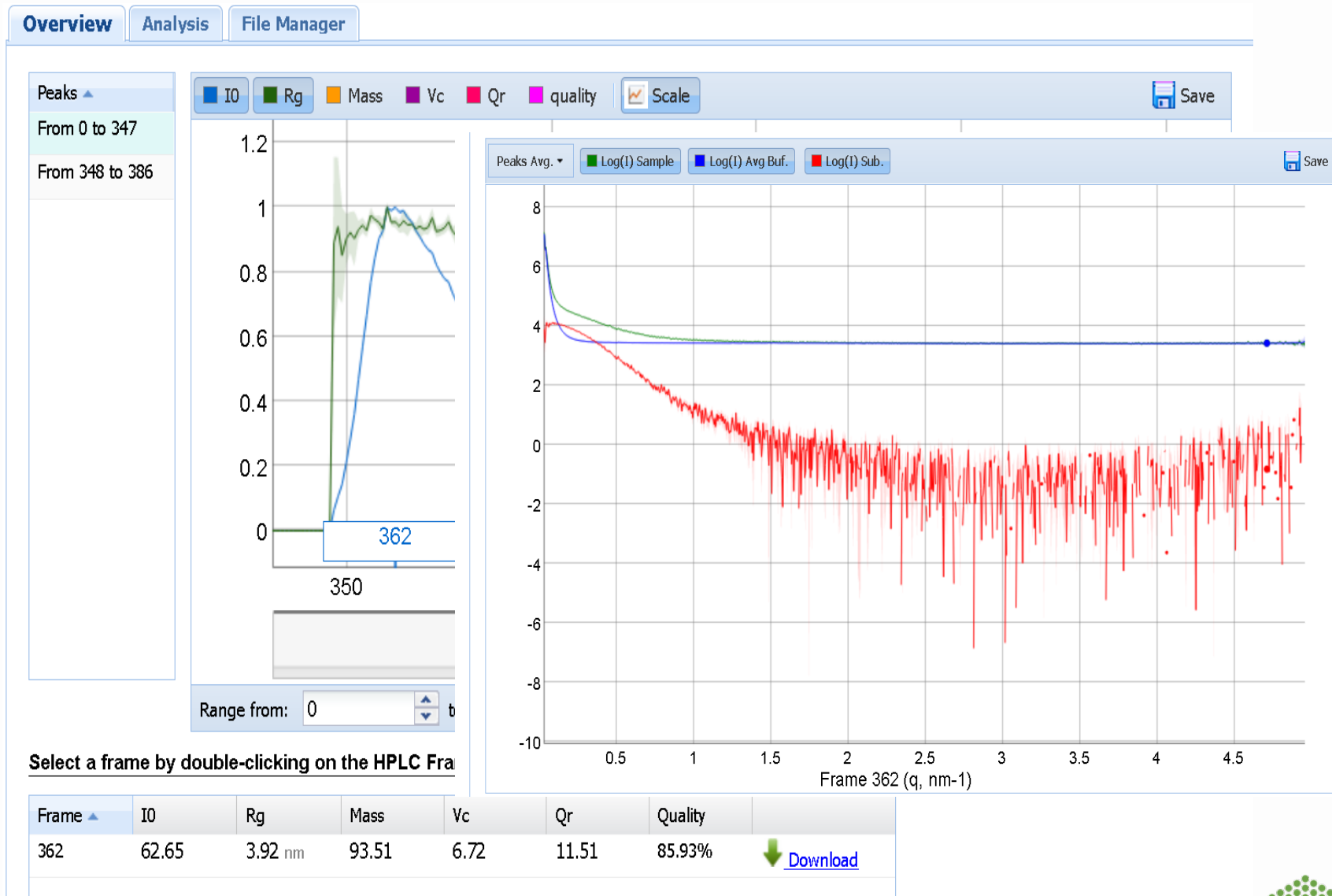
AutoMATED PROCESSING FOR HPLC



HPLC: Real Time feedback



ISPYB: HPLC overview



EDNA

- Data processing framework
- Collaboration between ESRF and Diamond
- Mostly used in macromolecular crystallography
- Python 2.7 based
- At BM29 as a TANGO device

- No direct user interaction: At BM29, the users only need to explicitly provide sample concentrations

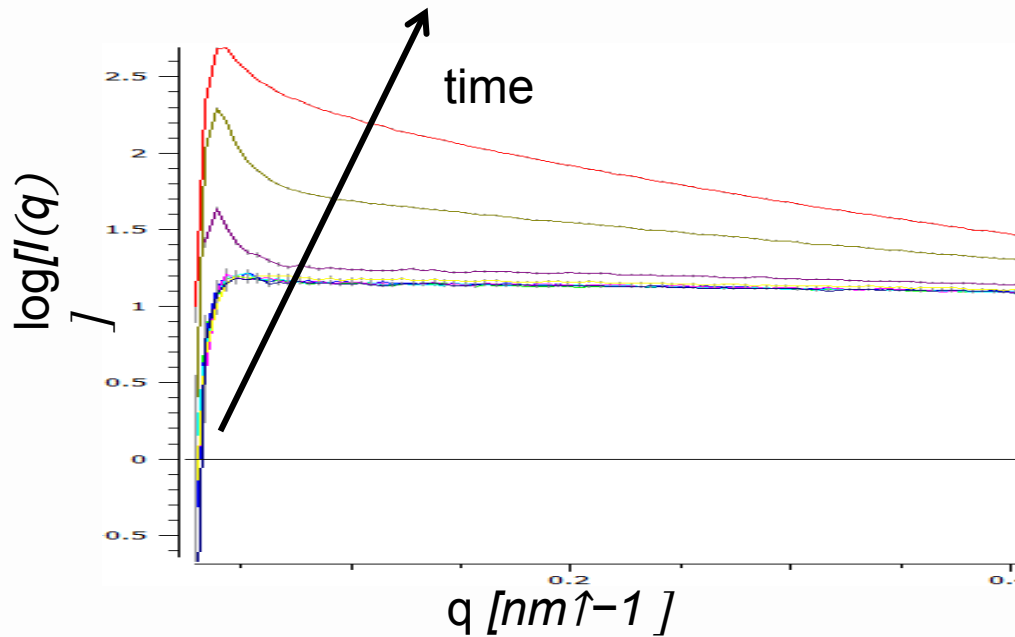
BM29 Data Analysis Hardware

3 local machines for online processing,
in principle each can do everything

Primary Processing	Bead modelling	HPLC processing
XEON 2 core, 3 GHz	2 x XEON 4 core, 2.26 GHz	XEON 6 core, 3.40 GHz
nVidia Quadro 4000, 2 GB memory	nVidia GeForce GTX 750 Ti, 2 GB memory	nVidia Quadro M2000, 4 GB memory
Before 2009	2011	2016

Why do we select frames?

- Reject radiation damaged data



- Identify peaks in HPLC mode

How do we select frames?

- Oversampled data, error bars of each data points non-ideal (correlated, ...)
- Correlation Map (CORMAP) test, originally proposed by Daniel Franke at EMBL Hamburg
- Core idea: If two frames come from “the same” sample, the difference between should be random!
- Hence the distribution of + and – differences corresponds to a series of coin tosses

CORMAP II

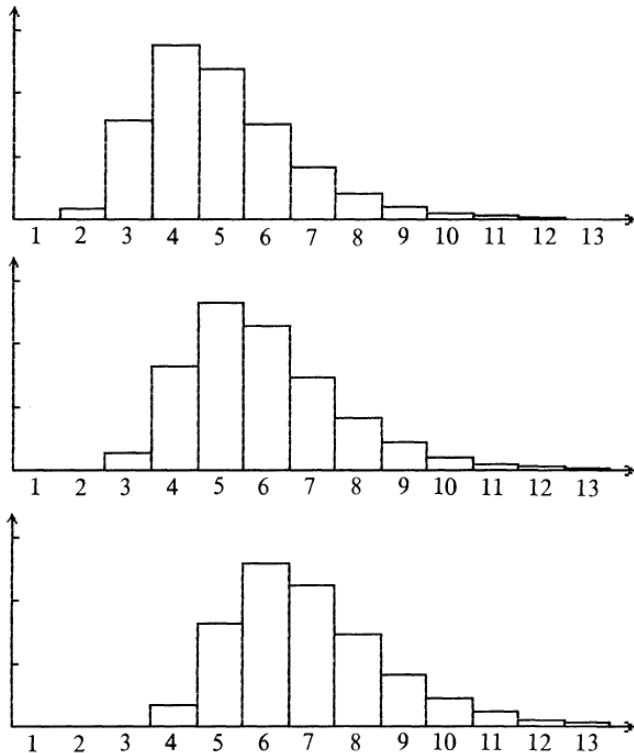


Figure 3

Distributions of R_n for (a) $n = 50$, (b) $n = 100$, (c) $n = 200$

- Distribution is recursive for the number of coin tosses
- The longest run is actually pretty short!
- e.g. at BM29 with 1043 q-bins in the range between 7 and 14 points
- Available in freesas

Mark F. Schilling

The College Mathematics Journal

Vol. 21, No. 3 (May, 1990), pp. 196-207

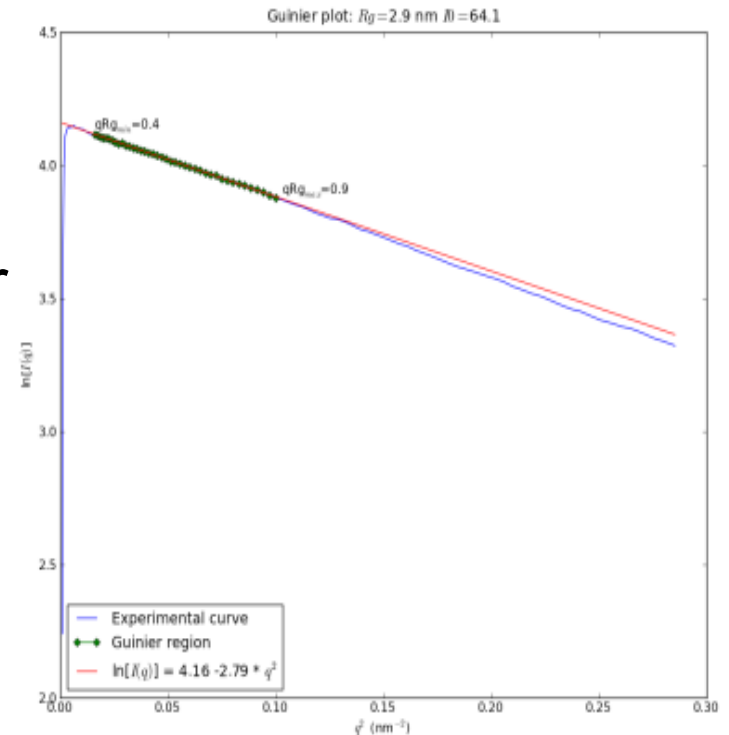


EMBL



AutoRg

- Forward scattering and radius of gyration are useful for identifying concentration effects on the scattering signal
- But the appropriate data range for the Guinier approximation is sample dependent and *a priori* unknown
- Score fits in different regions
- Originally used ATSAS version
- Moved to freeSAS implementation for HPLC



Beam center - the BM29 way

The screenshot displays the BM29 control interface, divided into two main sections: 'Operation' and 'Display'.

Operation Panel:

- Buttons:** Collect, Reprocess, Browse, Beamline.
- Read only:**
- Directory:** /data/visitor/mx1522/bm29/20130615
- Prefix:** test
- Run #:** 003
- Frame #:** 00001
- Time per frame:** 1.0 s
- Concentration:** 0.00 mg/ml
- Comments:** bsa buffer
- Code:** bsa
- Show Beamline Parameters:** (button)
- Radiation damage (10⁻⁴):** 0.00 (range 0.00 to 50.00)
- Collect using SC:** XML: BSA404.xml
- Collect using HPLC:**
- Notify when done:** Check beam: Energy adjust pilatus:
- Buttons:** Test, Collect, Abort (red bar), Collection status: running (yellow bar), sh: opened (green bar), Close (button)
- Transmission (current, new):** 0.00003 % (0.00000) (Filters)
- Energy [keV]:** 12.5000 keV (New Energy)
- Wavelength [Å]:** 0.9919 Angstrom (New Wavelength)

Display Panel:

- Buttons:** Collect using SC, EH Scan, OH Scan, 1D, 2D raw, Robot.
- Toolbar:** 800%, icons for zoom, pan, and other display functions.
- Display Area:** A large blue rectangle representing the detector image, with a small multi-colored spot (red, green, blue, yellow) in the center.
- Status Bar:** X: 796.00 Y: 161.50 Z: 0

Transmission $3 \cdot 10^{-7}$

X,Y

ACKNOWLEDGMENTS



Petra Pernot

Mark Tully

Benoit Maillot

Jérôme Kieffer

Staffan Ohlsson

Matias Guijarro

Antonia Betava

Alejandro De Maria Antolinos



Adam Round

Andrew McCarthy

***Thank you for
your attention!***



freesas



pipeline

