

3.
2.
0.

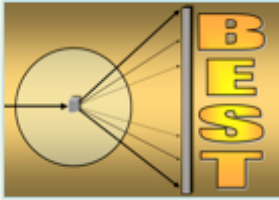
Gleb Bourenkov

EMBL-Hamburg

Alexander Popov

ESRF

<http://www.embl-hamburg.de/BEST/>



Software for Macromolecular Crystallography

BEST Copyright 2006

by European Molecular Biology Laboratory

Authors: [Alexander Popov](#) and [Gleb Bourenkov](#)

BEST is a program for optimal planning of X-ray data collection from protein crystals

[Download](#)

[Manual](#)

Relevant literature:

- A.N. Popov and G.P. Bourenkov "Choice of data-collection parameters based on statistic modeling" *Acta Crystallogr.* (2003). D59, 1145-1153 [Abstract](#)
- G.P. Bourenkov and A.N. Popov "A quantitative approach to data-collection strategies" *Acta Crystallogr.* (2006). D62, 58-64 [Abstract](#)

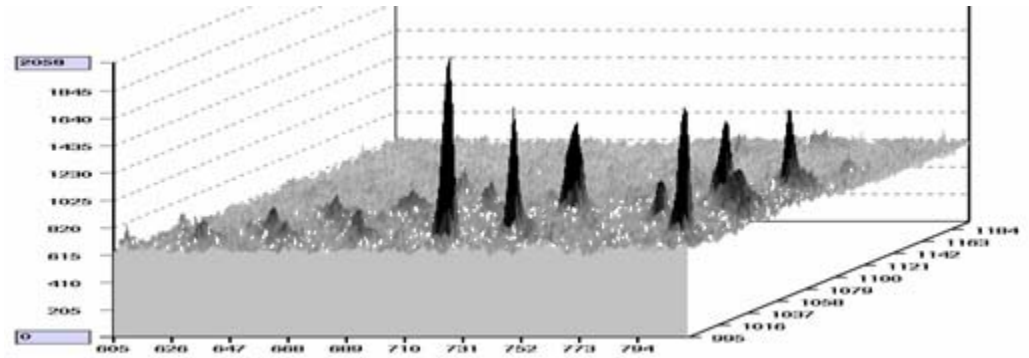
The BEST page is hosted at the [EMBL Hamburg Outstation](#)

Pages on this server were created by Alexander Popov and Gleb Bourenkov . GB and AP accept full responsibility for the content of these pages, for which EMBL is not responsible by any means.

Last Edited: 23.10.2006

Counting statistics is the major factor affecting the overall data statistics. It is mostly governed by the background (rather than the peak) counting statistics.

$$\sigma_{I_p}^2 = \left(I_p + I_b \cdot \frac{m \cdot (m+n)}{n} \right) \cdot G$$



The probability density function for diffraction intensity magnitudes follow the Wilson distribution

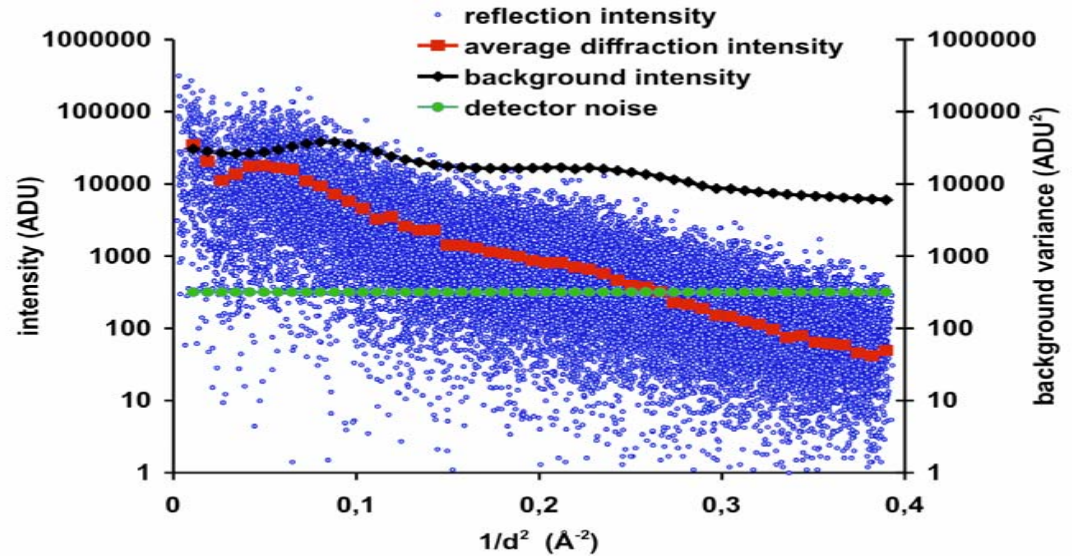
According to Wilson statistic

$$p_a(J) = \frac{1}{\hat{J}} \cdot \exp\left(-\frac{J}{\hat{J}}\right)$$

Acentric reflections

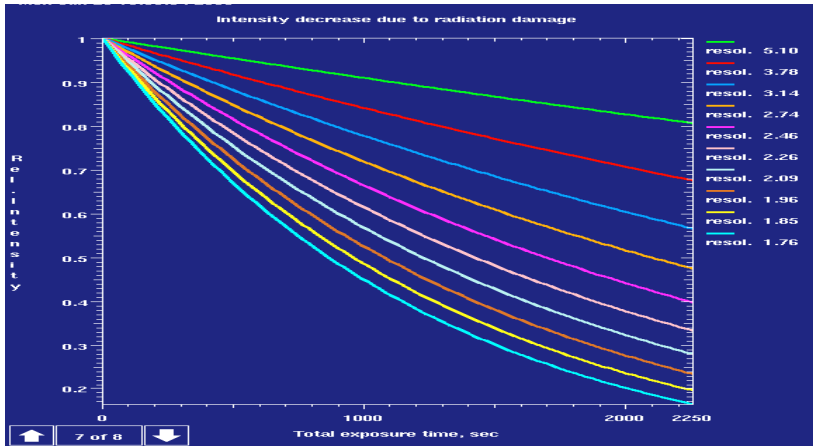
$$p_c(J) = \left(\frac{1}{2\pi \cdot \hat{J} \cdot J} \right)^{1/2} \cdot \exp\left(-\frac{J}{\hat{J}}\right)$$

Centric reflections



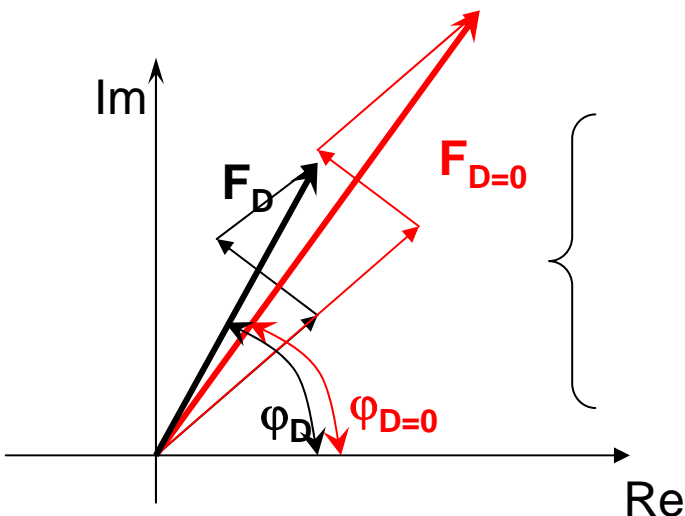
Radiation Damage Modelling in BEST

1. Intensity decay:



$$\frac{\langle I_D \rangle}{\langle I_{D=0} \rangle} \approx \left\langle e^{-2 \frac{dBi}{dD} Ds^2} \right\rangle \approx e^{-2\beta Ds^2}$$

2. Intensity variance (non-isomorphism):



$$\sigma_a \approx \left\langle e^{-\frac{dBi}{dD} Ds^2} \left\langle e^{-2 \frac{dBi}{dD} Ds^2} \right\rangle^{-1/2} \right\rangle \approx e^{-\alpha Ds^2}$$

$$R_{II} = \left\langle \left| \frac{I_{D=0}}{\langle I_{D=0} \rangle} - \frac{I_D}{\langle I_D \rangle} \right|^2 \right\rangle \approx (1 - \sigma_a^2)^{1/2}$$

Beamline Flux

RADDOSE

Absorbed dose

Gray per second

Shape factor

Susceptibility factor

Initial Images

MOSFLM XDS DENZO

Space group, Cell parameters, Orientation, Mosaicity

$I[(h,k,l), T_{\text{exposure}}], I_{\text{background}}$

The user choice

I/Sigma or max

Resolution

Anomalous

Redundancy

Geometry

Total time or dose

BEST 3.1.

Geometry

Optimal starting spindle angle and scan range

Maximum rotation angle without spot overlap

Multiplicity

Reconstruction of average intensity vs resolution

Statistics modeling based on Wilson distribution

Search for the optimal combination of data collection parameters

Statistics calculation

Low resolution path

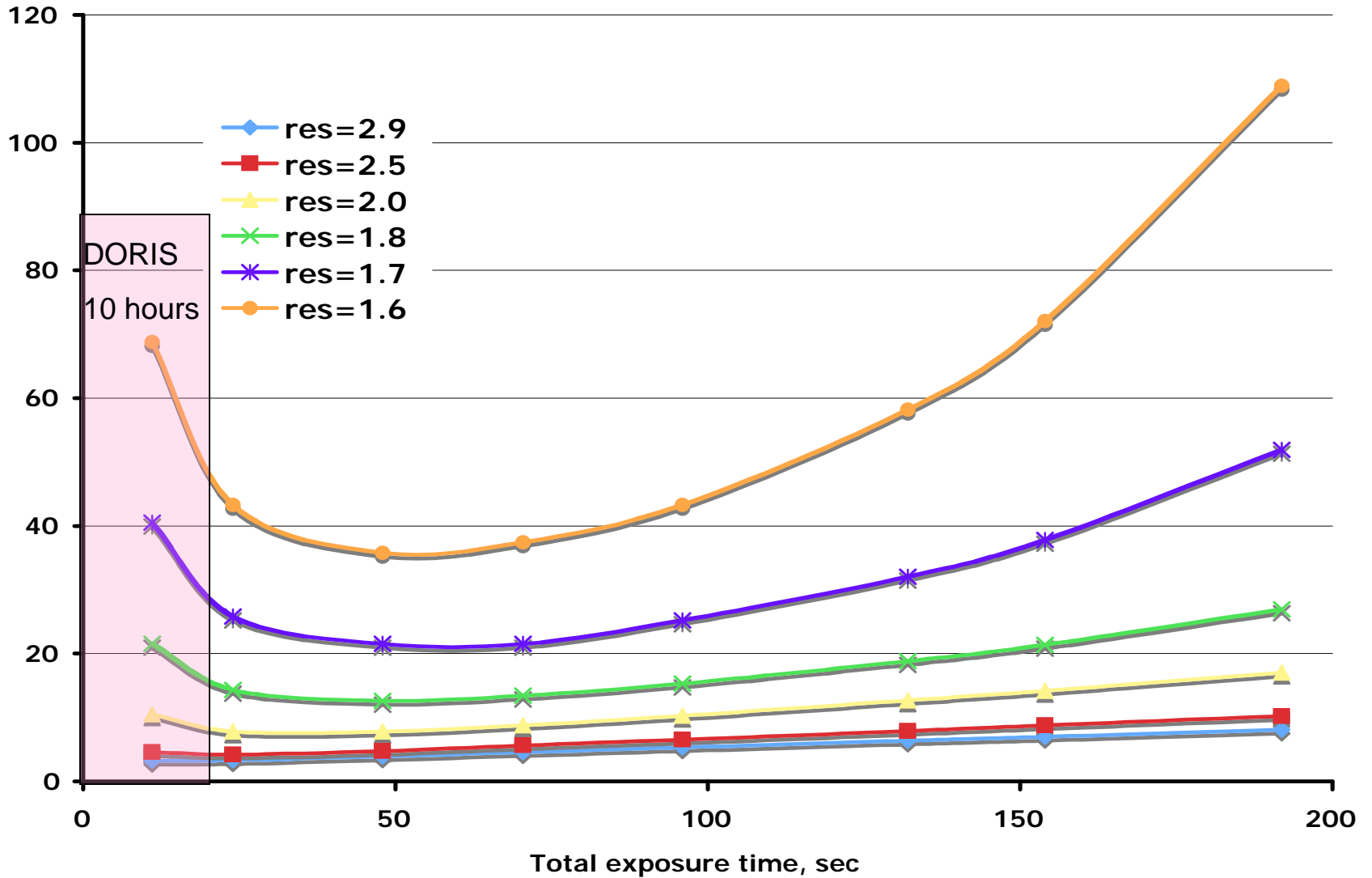
Plan(s) of data collection

Statistics

B-factor

Effect of radiation damage

R-factor vs. exposure time



Run BEST3.0 data collection strategy

Title

Run BEST to show graphs

Input from

mosflm dat in

mosflm par in

hkl #1 in

Change Symmetry

Reference Image parameters

Detector Exposure time (sec) Preset counts Read-outs

Radiation damage parameters

Enable radiation damage corrections

Dose rate * Shape factor Susceptibility

Use RADDOSE

Major optimization parameters

Target $\langle I \rangle / \langle \text{Sig} \rangle$ in the last shell = Anomalous data

Maximum resolution Angstrom

Minimize time, and limit it to sec

Rotation range parameters

Total rotation range

for completeness = and redundancy

Minimum rotation range/frame, deg:

Output plan parameters

Complexity level of data collection strategy

Save strategy to file

Rotation speed/exposure time limitations

Maximum scan speed, deg/sec:

Minimum exposure time/frame, sec:

mosfilm data in
hkl #1 in

Change Symmetry

Reference Image parameters

Detector Exposure time (sec) Preset counts Read-outs

Radiation damage parameters

Enable radiation damage corrections
Dose rate * Gray/second Shape factor Susceptibility

Use RADDOSE

Crystal cell & symmetry :
Unit Cell: Space Group

Crystal composition:
Monomers in assymmetric unit:
Amino acid residues per monomer:
DNA nucleotides per monomer:
RNA nucleotides per monomer:

Bound heavy atoms:
Atom: # per monomer

Solvent composition:
Atom: Concentration: mM

Crystal dimensions (across rotation axis?): x micron

Incident X-ray beam:

Title BEST 3.0

Input from HKL/DENZO

image in Full path.. /h/sasha/BEST3.0/pp_3_001.mccd

Browse

View

x #1 in Full path.. /h/sasha/BEST3.0/pp_3_001.x

Browse

View

Edit list

Add input x file

 Use sca file Change Symmetry ORIGINAL*Reference Image parameters*Detector ccd Exposure time (sec) 40.07 Dose mode dose (kHz-sec) 200.0 Read-outs 1*Radiation dose/damage parameters*

Gray per Second 0.005 x10e6 Gray/second Shape factor 1.0 Susceptibility factor 1.0

Select Task

Check Statistics

Show Graphs *Task parameters*

bst input PROJECT pp_3_001_plan2.bst

Browse

View

Reset

Attenuation: 1.000000 Resolution: 2.187622 Anomalous Data

#wedge Start Phi #frames Width Exposure

1 85.00 62 0.80 2.1226

2 134.60 27 0.55 4.8796

3 149.45 12 0.40 11.2471

Edit list

Add wedge

Hardware limitations

Run

Save or Restore

Close

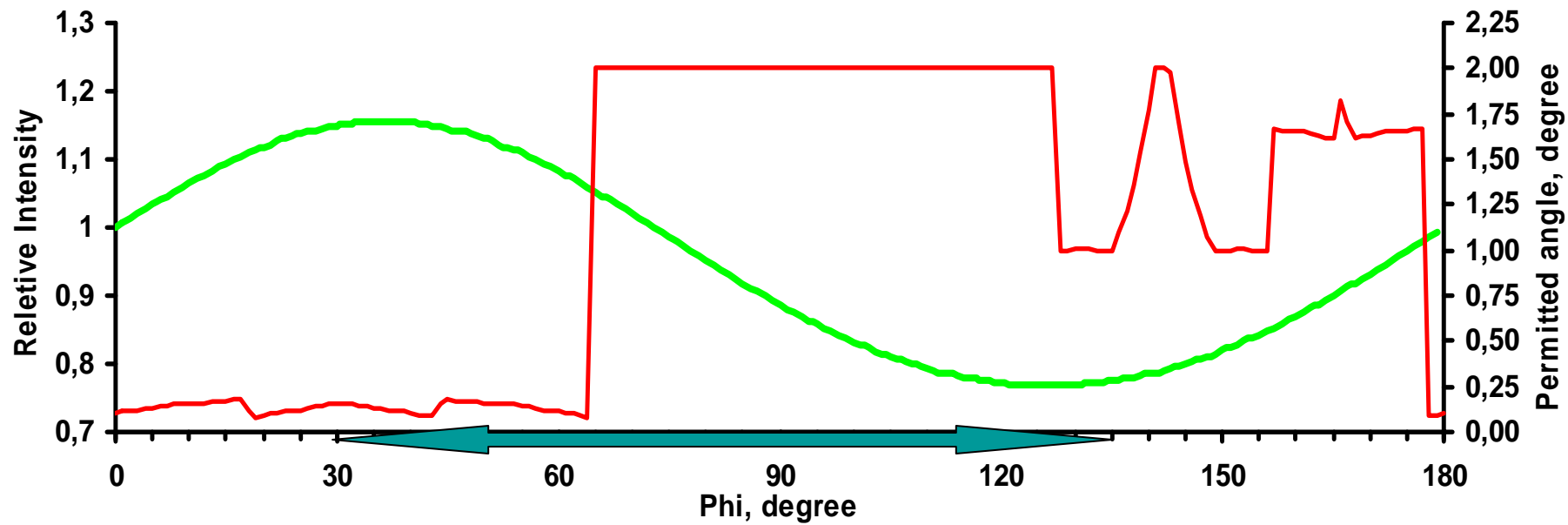
Optimal Plan of data collection

=====
Resolution limit is set by the radiation damage

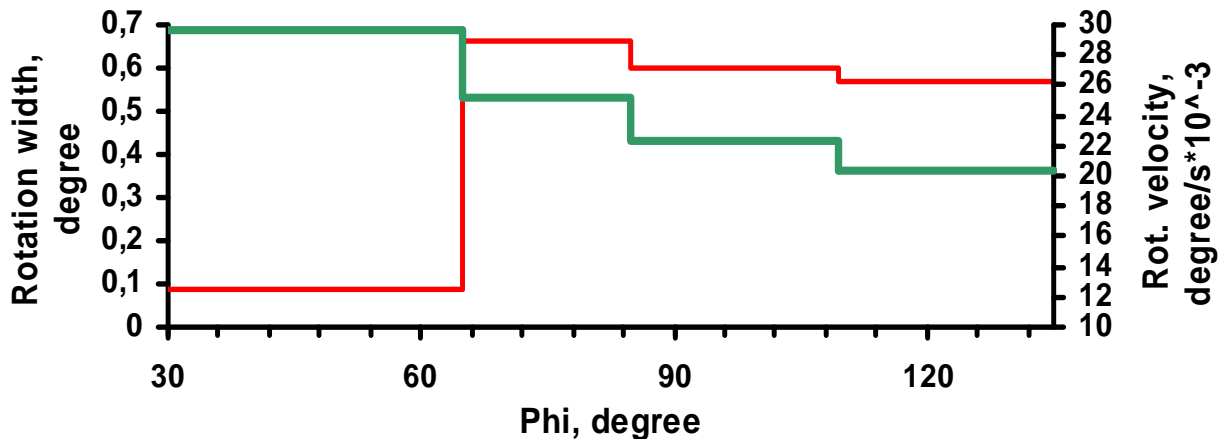
Attenuation = 1.0000

N	Phi_start	N.of.images	Rot.width	Dose	Distance	Overlap
1	180.00	183	0.30	33.16	162.1	No
2	234.90	76	0.20	67.89	162.1	No

Resolution limit : 1.87 Angstrom
Anomalous data : Yes
Phi_start - Phi_finish : 180.00 - 250.00
Total rotation range : 70.1 degree
Total N.of images : 259
Overall Completeness : 99.2%
Redundancy : 2.31
R-factor (outer shell) : 5.9% (23.2%)
I/Sigma (outer shell) : 20.3 (4.3)
Rel.decrease of intensity : 0.205 for outer resolution shell
Total Exposure time : 2249.7 sec (0.625 hour)
Total Data Collection time : 5098.7 sec (1.416 hour)



(a)



Plan of data collection

Data collection statistics according to the plan
=====

Resolution Lower Upper	Compl. %	Average Intensity	Sigma	I/Sigma stat	I/Sigma /Chi	Chi**2	R-fact %	Ranom %	Overload %
12.00 7.15	94.5	16009.6	611.2	26.2	25.5	1.06	3.8	3.7	0.02
7.15 5.57	97.6	8635.2	342.5	25.2	23.9	1.11	4.0	4.7	0.00
5.57 4.72	97.2	11871.6	471.7	25.2	22.9	1.20	4.2	5.3	0.00
4.72 4.17	97.3	15338.3	612.9	25.0	22.0	1.30	4.4	5.9	0.00
4.17 3.78	97.6	12968.2	522.9	24.8	21.0	1.40	4.7	6.5	0.00
3.78 3.48	97.4	10315.0	418.7	24.6	19.8	1.54	4.9	6.9	0.00
3.48 3.24	98.0	7715.0	318.4	24.2	18.7	1.68	5.3	7.4	0.00
3.24 3.04	98.1	5434.8	231.0	23.5	17.5	1.82	5.6	8.0	0.00
3.04 2.88	98.9	3988.7	174.2	22.9	16.5	1.93	6.0	8.3	0.00
2.88 2.74	98.4	2988.1	138.4	21.6	15.0	2.07	6.6	8.8	0.00
2.74 2.62	98.9	2391.1	117.2	20.4	14.2	2.08	6.9	9.2	0.00
2.62 2.51	99.5	1989.6	104.8	19.0	12.9	2.15	7.5	9.8	0.00
2.51 2.42	99.5	1690.5	95.8	17.6	12.1	2.14	8.1	10.3	0.00
2.42 2.33	100.0	1487.5	90.1	16.5	11.3	2.12	8.7	10.7	0.00
2.33 2.26	100.0	1308.6	86.9	15.1	10.4	2.09	9.3	11.1	0.00
2.26 2.19	100.0	1157.2	84.2	13.7	9.7	2.01	10.0	11.8	0.00
2.19 2.12	100.0	982.7	81.5	12.1	8.8	1.86	10.9	12.2	0.00
2.12 2.07	100.0	843.6	79.3	10.6	8.0	1.77	11.9	13.0	0.00
2.07 2.01	100.0	687.1	77.8	8.8	7.0	1.60	13.6	14.1	0.00
2.01 1.96	100.0	550.9	76.0	7.2	6.1	1.43	15.5	15.4	0.00
1.96 1.92	100.0	433.0	74.1	5.8	5.1	1.32	18.5	17.5	0.00
1.92 1.87	100.0	340.3	75.4	4.5	4.1	1.20	22.1	20.4	0.00
All data	99.2	3437.7	169.0	20.3	15.5	1.73	5.9	7.6	0.00

R-fact = $\text{SUM}(\text{ABS}(I - \langle I \rangle)) / \text{SUM}(I)$

Chi**2 = $\text{SUM}((I - \langle I \rangle)^2) / (\text{Error}^2 * N / (N-1))$

measure of the non-isomorphism due to radiation damage

Ranom = $\text{SUM}(\text{ABS}(\langle I+ \rangle - \langle I- \rangle)) / \text{SUM}(\langle I+ \rangle + \langle I- \rangle)$

measure of the statistical discrepancy between anomalous pairs

Additional information

=====

Blind region for given crystal orientation

Relative n.of unmeasurable unique reflections : 0.00%

Scaling

Relative scale : 3.01

Overall B-factor : 19.47 Angstrom²

B-factor eigenvalues : 18.36 18.36 22.15 Angstrom²

Resolution vs. measurement time/dose

=====

(Max.resolution is restricted by the radiation damage)

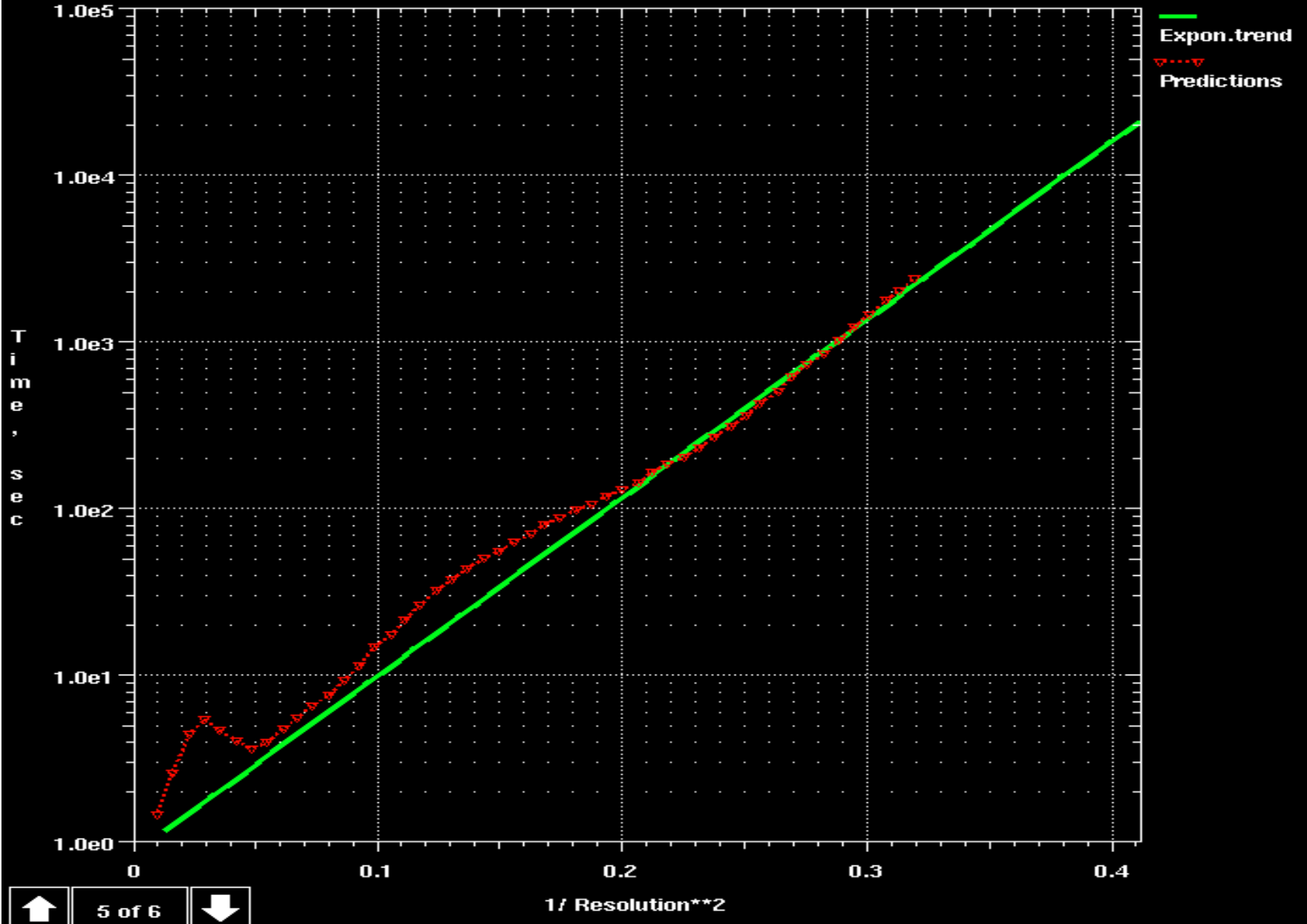
Parameters used for calculations:

Phi_start = 180.00 degree Phi_finish = 250.00 degree

Overall Completeness = 99.30% and Redundancy = 2.31

Resolution [Angst.]	Compl. %	Redund	DataCol_Time h: min	Exposure_Time h: min	I/Io	Rdamage	Attenuation
12.00 7.15	94.5	2.48	0: 6.11	0: 0.24	1.000	0.003	0.2E+00
7.15 5.57	97.6	2.38	0: 6.11	0: 0.24	1.000	0.006	0.3E+00
5.57 4.72	97.2	2.48	0: 6.11	0: 0.24	1.000	0.006	0.3E+00
4.72 4.17	97.3	2.34	0: 6.11	0: 0.24	0.999	0.007	0.3E+00
4.17 3.78	97.6	2.36	0: 6.65	0: 0.23	0.999	0.009	0.4E+00
3.78 3.48	97.4	2.34	0: 6.29	0: 0.24	0.999	0.012	0.5E+00
3.48 3.24	98.0	2.28	0: 6.29	0: 0.24	0.998	0.016	0.8E+00
3.24 3.04	98.1	2.20	0: 6.51	0: 0.28	0.996	0.021	0.1E+01
3.04 2.88	98.9	2.34	0: 7.01	0: 0.41	0.993	0.027	0.1E+01
2.88 2.74	98.4	2.31	0: 7.18	0: 0.58	0.989	0.034	0.1E+01
2.74 2.62	98.9	2.37	0: 7.54	0: 0.76	0.984	0.040	0.1E+01
2.62 2.51	99.5	2.31	0: 8.53	0: 1.01	0.977	0.048	0.1E+01
2.51 2.42	99.5	2.27	0: 9.73	0: 1.30	0.968	0.057	0.1E+01
2.42 2.33	100.0	2.35	0:11.08	0: 1.55	0.959	0.065	0.1E+01
2.33 2.26	100.0	2.29	0:12.76	0: 1.94	0.945	0.075	0.1E+01
2.26 2.19	100.0	2.30	0:14.34	0: 2.42	0.928	0.086	0.1E+01
2.19 2.12	100.0	2.31	0:18.39	0: 3.17	0.901	0.102	0.1E+01
2.12 2.07	100.0	2.26	0:20.30	0: 4.16	0.866	0.120	0.1E+01
2.07 2.01	100.0	2.28	0:23.18	0: 5.95	0.805	0.147	0.1E+01
2.01 1.96	100.0	2.32	0:28.73	0: 8.93	0.709	0.184	0.1E+01
1.96 1.92	100.0	2.28	0:44.22	0:15.62	0.533	0.247	0.1E+01
1.92 1.87	100.0	2.30	1:24.98	0:37.50	0.205	0.383	0.1E+01

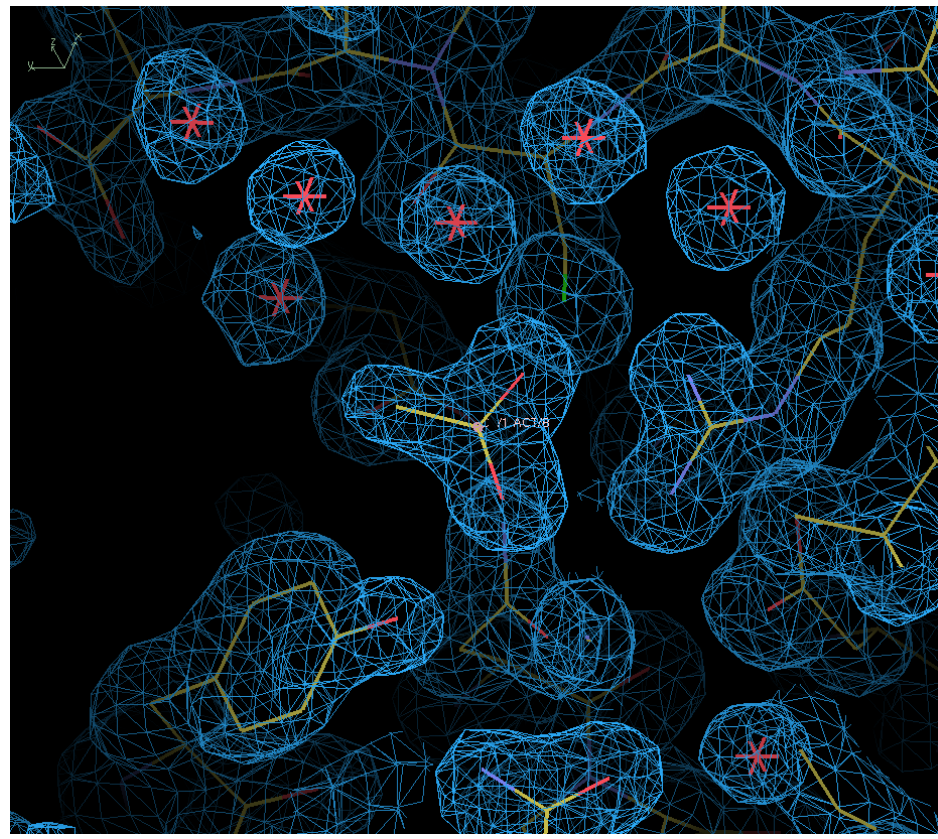
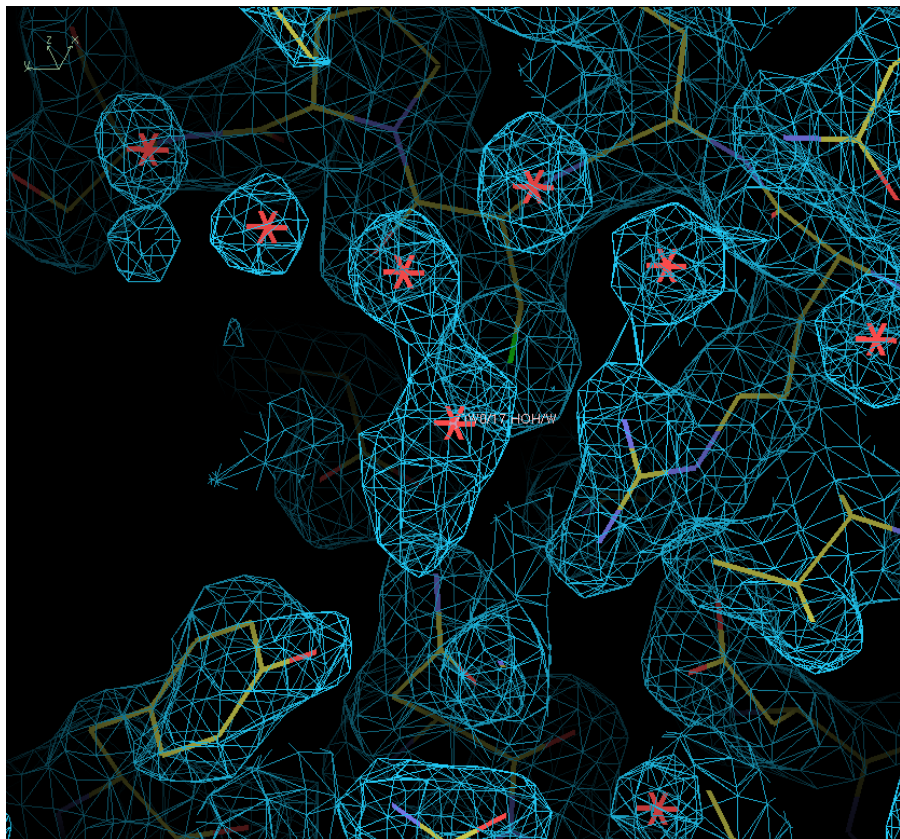
Total exposure time vs resolution
and exponential fit using last 20 bins

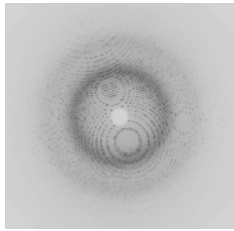


Mtb GltS

Resolution 2.1 Å

Resolution 1.65 Å





Initial Images

Beamline Flux



RADDDOSE

- Absorbed dose
- Gray per second
- Shape factor
- Susceptibility factor



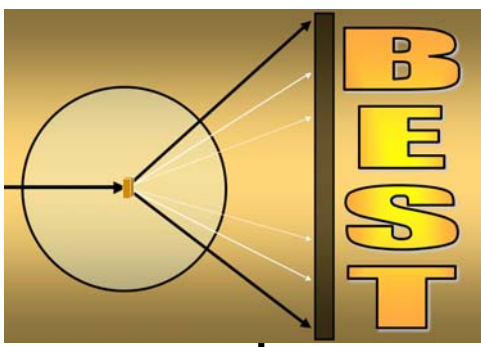
MOSFLM XDS DENZO

Space group, Cell parameters, Orientation, Mosaicity
 $I[(h,k,l), T_{\text{exposure}}], I_{\text{background}}$



The user choice

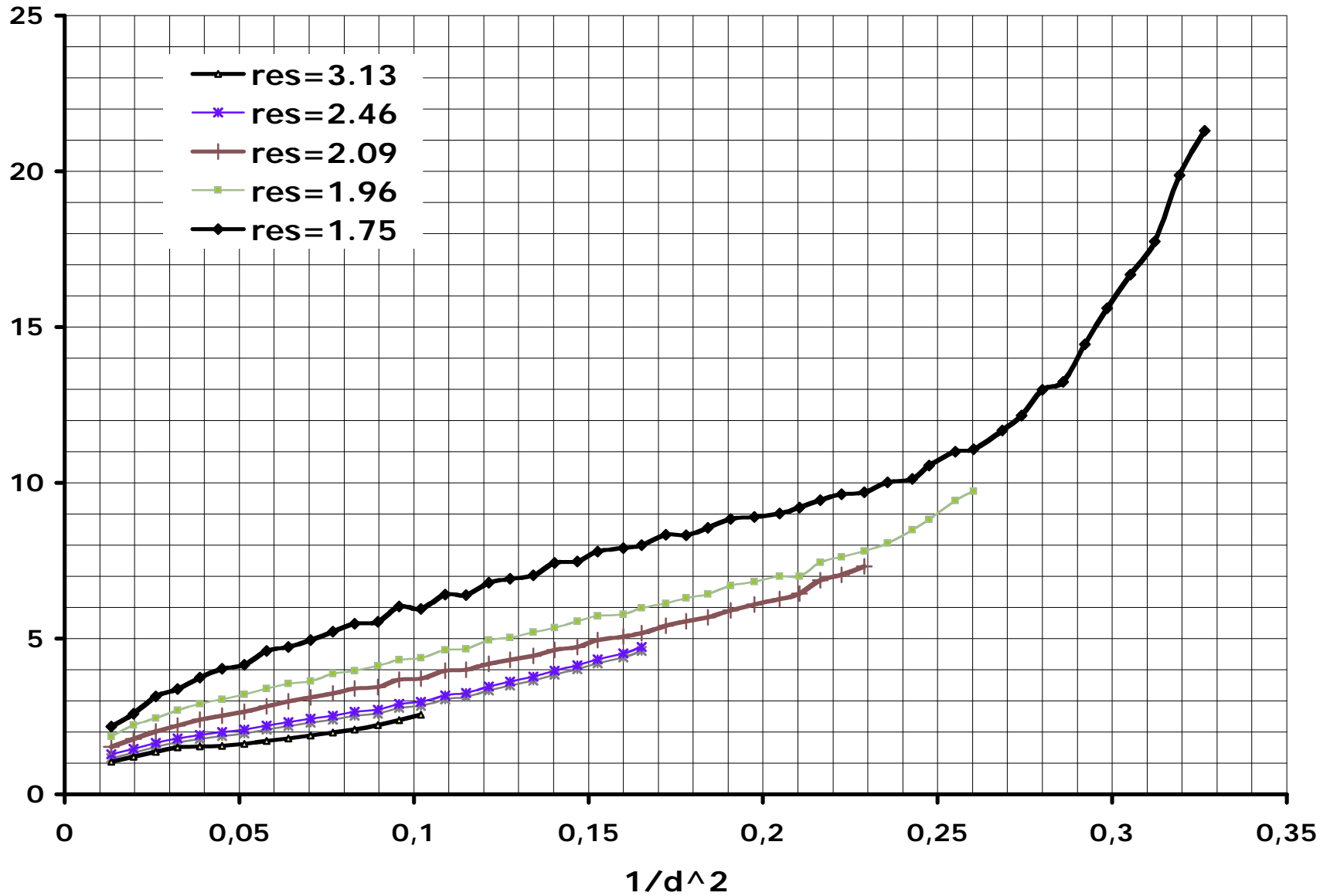
- I/Sigma or max Resolution
- Anomalous Redundancy
- Geometry
- Total time or dose



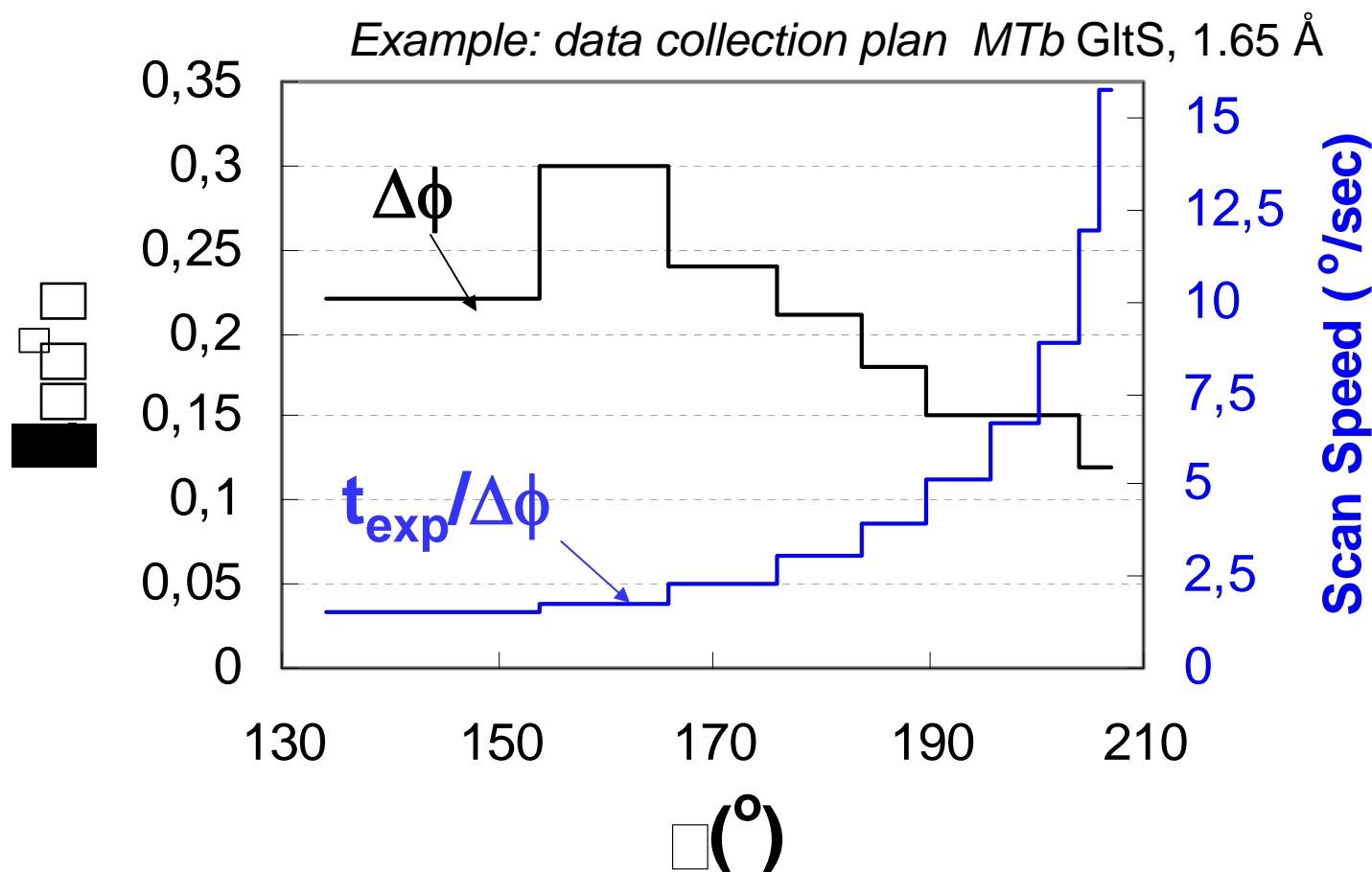
$R_{\text{friedel}} = \text{Minimal}$
 Best orientation?

Plan(s) of data collection
Statistics
B-factor

SAD – optimal Rfriedel



PROBLEM 1 - too complicate ...



SOLUTION - EDNA

James M. Holton

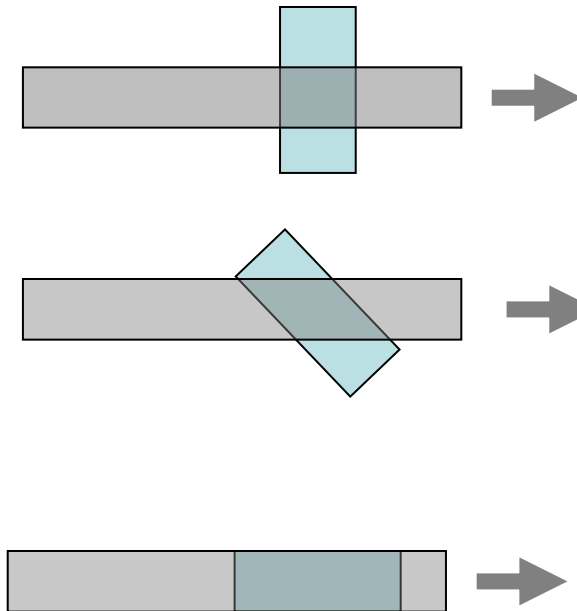
Why are we not solving more structures?

Most X-ray data sets collected at synchrotron sources do not produce usable results. An analysis of data collected in 2003 at the ALS beamline 8.3.1 shows that 2346 datasets were collected and 48 structures were deposited in the PDB. Although it is understandable that not every dataset leads to a published structure, it is hard to explain how ~98% of them do not.

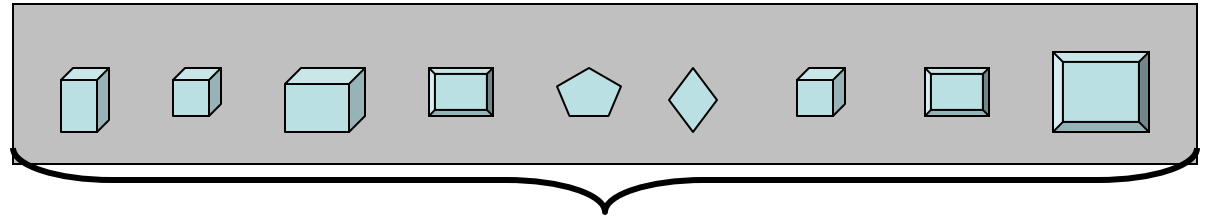
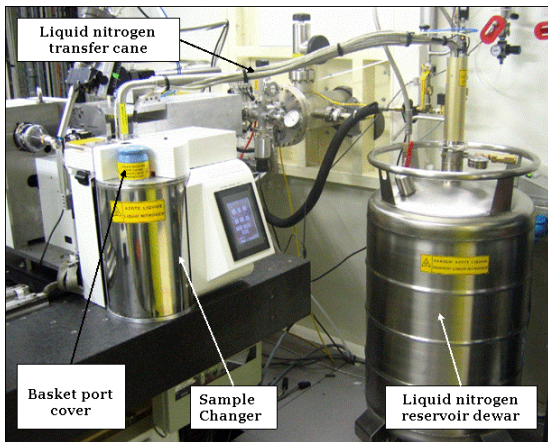
PROBLEM 2

sample > beam

Irradiation crystal volume VS Phi ????



Data collection using several crystals



Test images

Auto-indexing

Experimental aim

BEST

Crystal characterization and ranking
Determination of maximal achievable resolution
Optimal crystal orientation - Kappa goniometer

Crystal 3 ⇒ Completeness 23%

Crystal 5 ⇒ Completeness 58%

Crystal 1 ⇒ Completeness 91%

Crystal 8 ⇒ Completeness 99.7%

D.C. plan