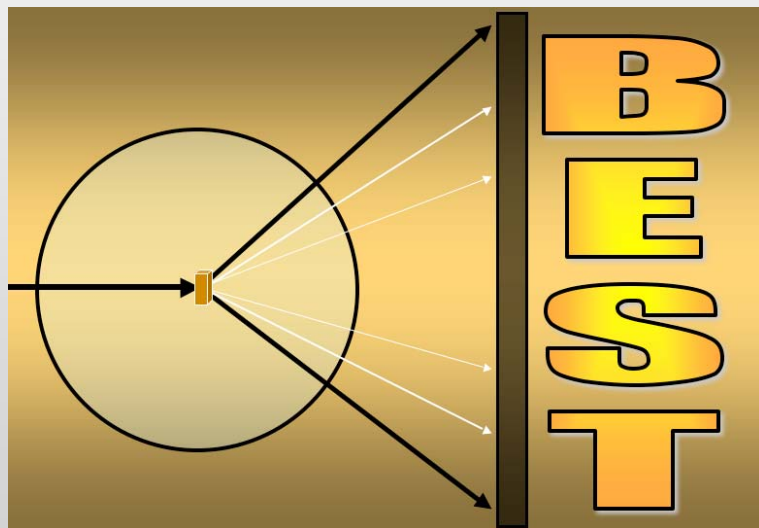


Calculating strategies for data collection

Alexander Popov



research papers

Acta Crystallographica Section D
Biological
Crystallography
ISSN 0907-4449

Choice of data-collection parameters based on statistic modelling

Alexander N. Popov^{a,c,*} and
Gleb P. Bourenkov^{b,c}

^aEuropean Molecular Biology Laboratory (EMBL)
Hamburg Outstation, c/o DESY, Notkestrasse 85,
22603 Hamburg, Germany, ^bMax-Planck-
Arbeitsgruppe für Strukturelle Molekularbio-
logie, Arbeitsgruppe Proteindynamik, Notke-
strasse 85, 22603 Hamburg, Germany, and ^cA.
V. Shubnikov Institute of Crystallography,
Russian Academy of Sciences, Leninski
Prospekt 59, 117333 Moscow, Russia

Correspondence e-mail:
sasha@embl-hamburg.de

A new method and the software program **BEST** for optimal planning of X-ray data collection from protein crystals using the rotation method are presented. From one or a few initial diffraction images, **BEST** estimates the statistical characteristics of the data set for different combinations of data-collection parameters and suggests the most optimal ones. The anisotropy in diffraction and the permitted width of oscillation without spatially overlapping reflections are taken into account. According to the option chosen, the optimal set of parameters provides a given average signal-to-noise ratio at a given resolution either in the shortest time or with the minimum total radiation dose. **BEST** has been successfully used at the protein crystallography beamlines at DORIS (DESY). The software proved to be extremely useful in using the available data-collection time in the most efficient way.

Received 19 February 2003
Accepted 10 April 2003

1. Introduction

The quality of X-ray diffraction data depends drastically on

Data quality

Completeness, Resolution, Statistics

Detector

*Size
N. of pixels
T readout
Noise
Sensitivity
Dynamic range*

Crystal

*Space group
Cell parameters
Orientation
B-factor
Mosaicity
Anisotropy
Radiation damage*

X-ray source

*Intensity
Divergence
Wavelength
Beam size*

Crystal-to-detector distance

*Resolution
 I_p/I_b
spot overlap*

Starting angle and total rotation range

*Completeness
Multiplicity
d.c. time*

Rotation width

*I_p/I_b
spot overlap,
total d.c. time*

Exposure time

*statistics
overloads
total d.c. time*

Table 1

Rotation range (°) required to collect a complete data set in different crystal classes

The direction of the spindle axis is given in parentheses; *ac* means any vector in the *ac* plane.

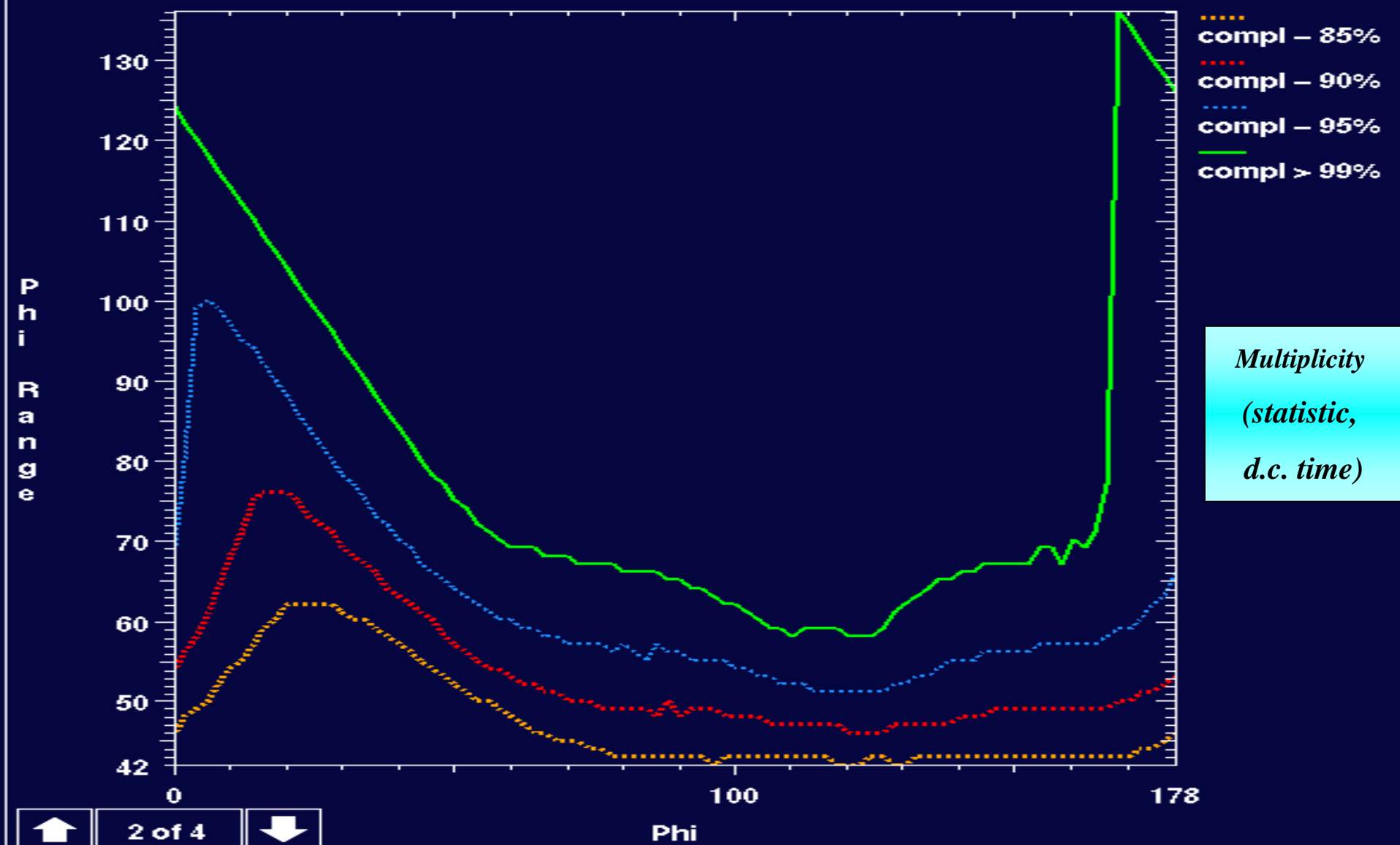
Point group	Native data	Anomalous data
1	180 (any)	$180 + 2\vartheta_{\max}$ (any)
2	180 (<i>b</i>); 90 (<i>ac</i>)	180 (<i>b</i>); $180 + 2\vartheta_{\max}$ (<i>ac</i>)
222	90 (<i>ab</i> or <i>ac</i> or <i>bc</i>)	90 (<i>ab</i> or <i>ac</i> or <i>bc</i>)
4	90 (<i>c</i> or <i>ab</i>)	90 (<i>c</i>); $90 + \vartheta_{\max}$ (<i>ab</i>)
422	45 (<i>c</i>); 90 (<i>ab</i>)	45 (<i>c</i>); 90 (<i>ab</i>)
3	60 (<i>c</i>); 90 (<i>ab</i>)	$60 + 2\vartheta_{\max}$ (<i>c</i>); $90 + \vartheta_{\max}$ (<i>ab</i>)
32	30 (<i>c</i>); 90 (<i>ab</i>)	$30 + \vartheta_{\max}$ (<i>c</i>); 90 (<i>ab</i>)
6	60 (<i>c</i>); 90 (<i>ab</i>)	60 (<i>c</i>); $90 + \vartheta_{\max}$ (<i>ab</i>)
622	30 (<i>c</i>); 90 (<i>ab</i>)	30 (<i>c</i>); 90 (<i>ab</i>)
23	~60	~70
432	~35	~45

Z. Dauter “Data-collection strategies”, 1999, Acta Cryst. D55, 1703-1717

Optimal starting angle and total rotation range

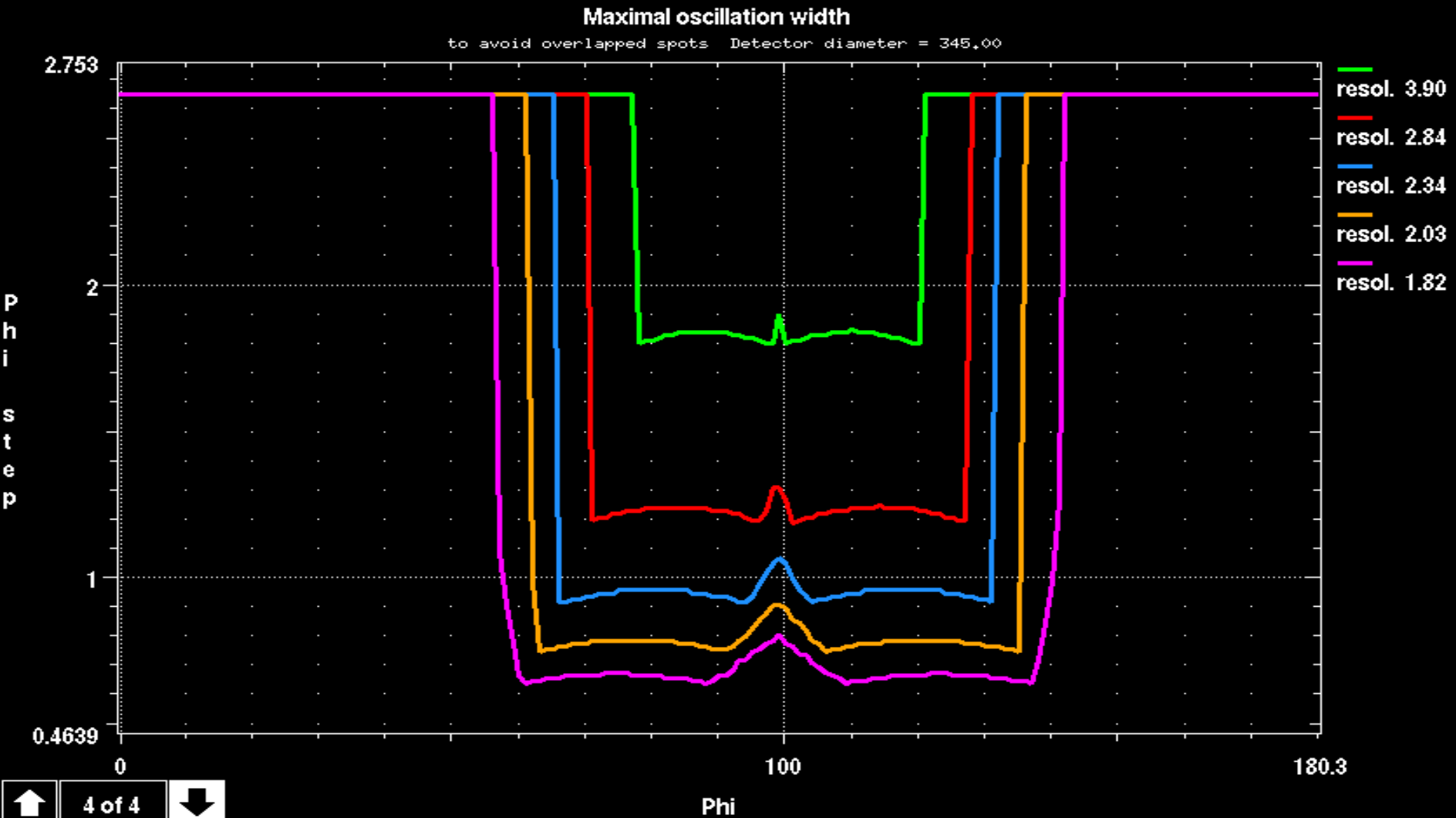
Tue Aug 28 20:30:52 2001

Minimal oscillation ranges for different completenesses



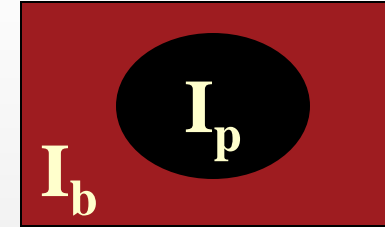
Maximum rotation angle without spot overlap

Wed Aug 14 17:52:41 2002

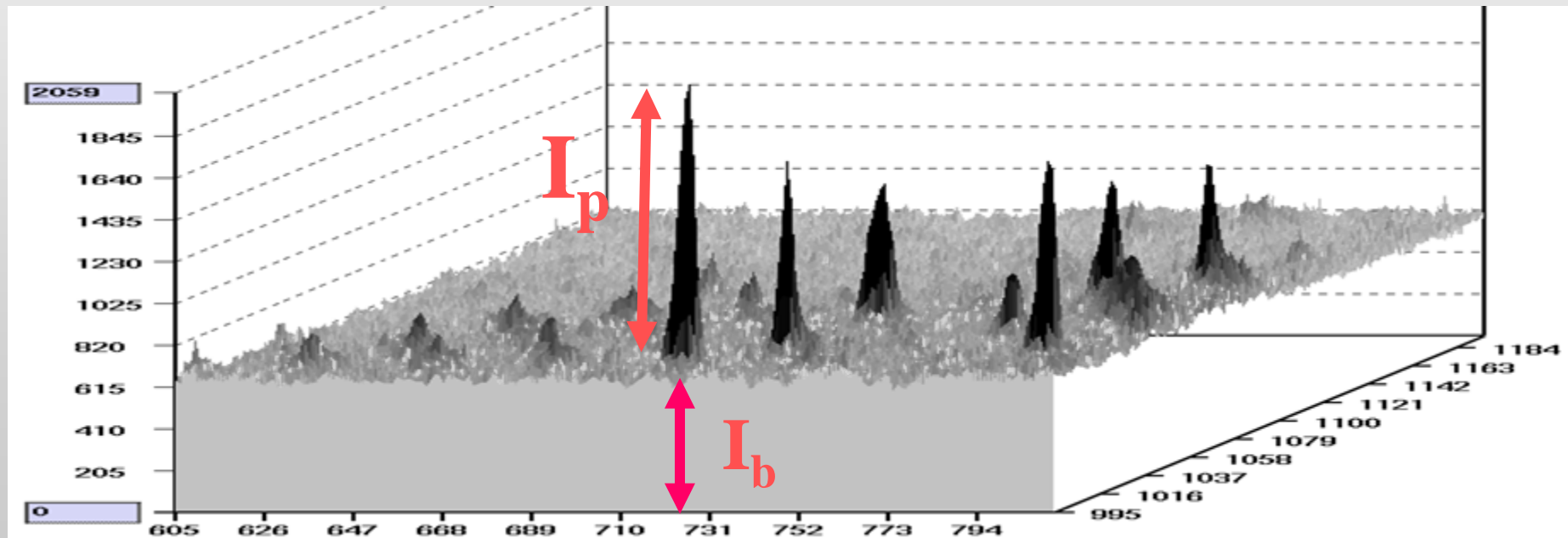


Main uncertainties of the observed intensities are determined by counting statistics

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



where m and n are number of pixels in the peak and background region of the measurement box respectively. G is the detector gain, which converts pixel counts to equivalent X-ray photons. K_{ins} is a proportionality constant for the instrument-error term



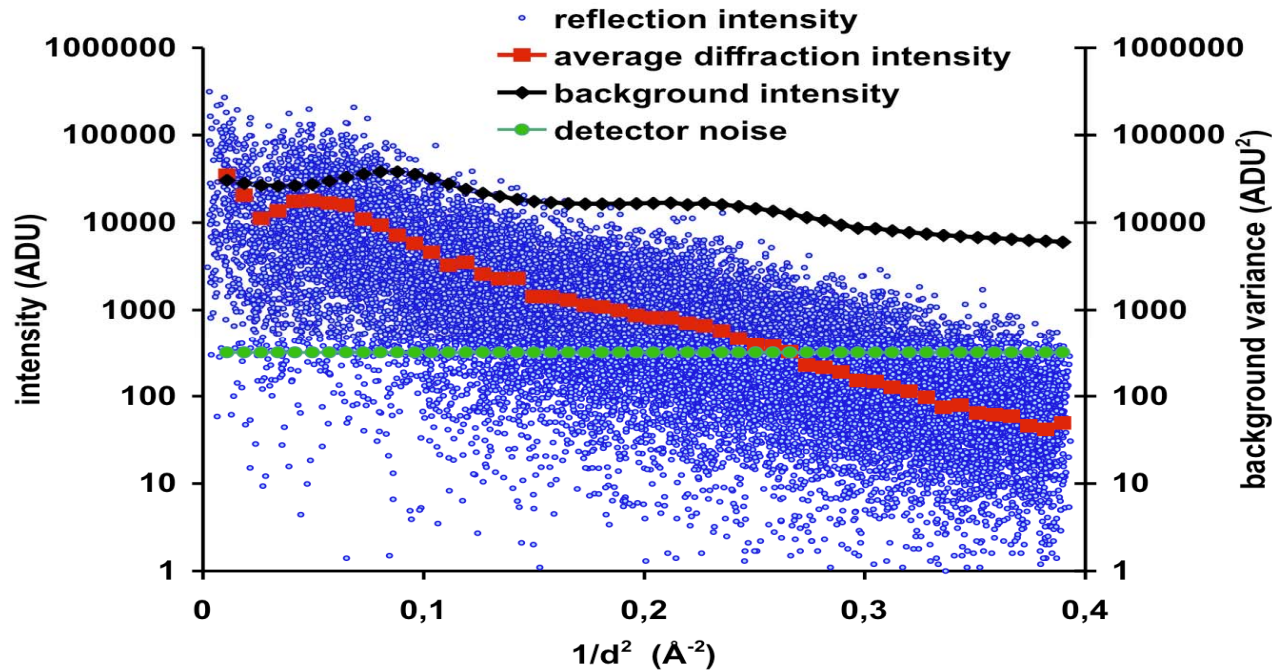
According to Wilson statistic

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

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



ALCOHOL DEHYDROGENASE

ID23-1, $\lambda=0.973$ Å, Flux= $2.5 \cdot 10^{12}$

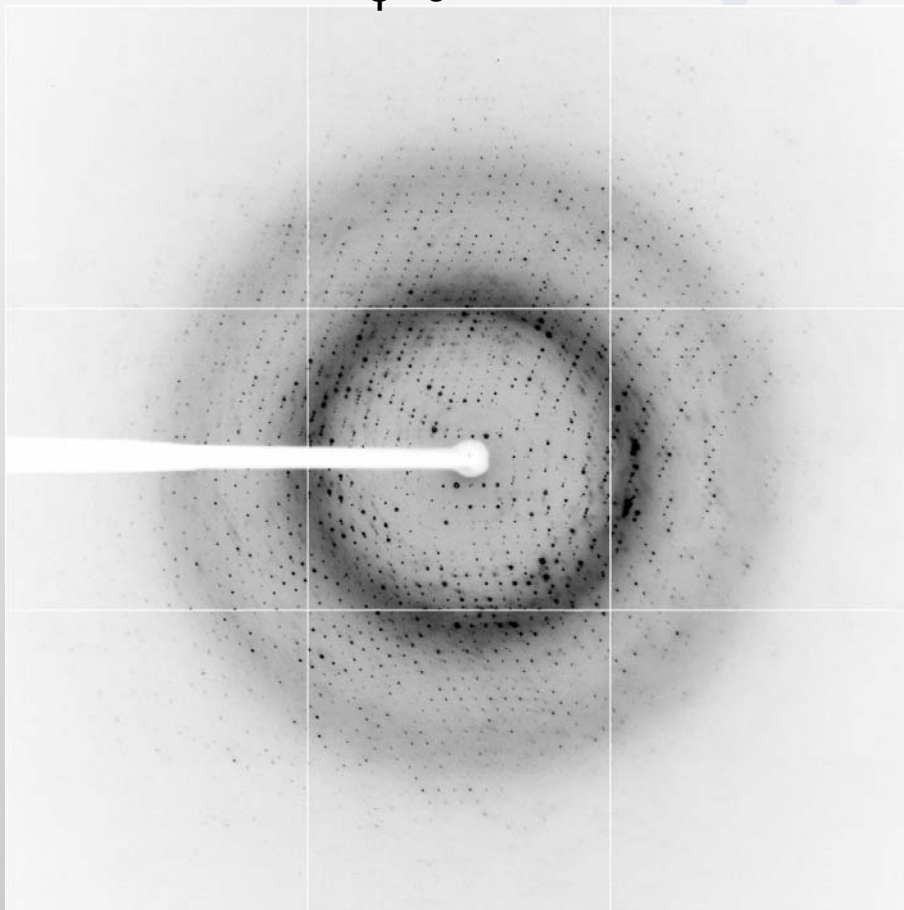
resolution = 1.45 Å, $\Delta\phi=1^\circ$, t exposure= 0.2 s,

Space Group : C 1 2 1

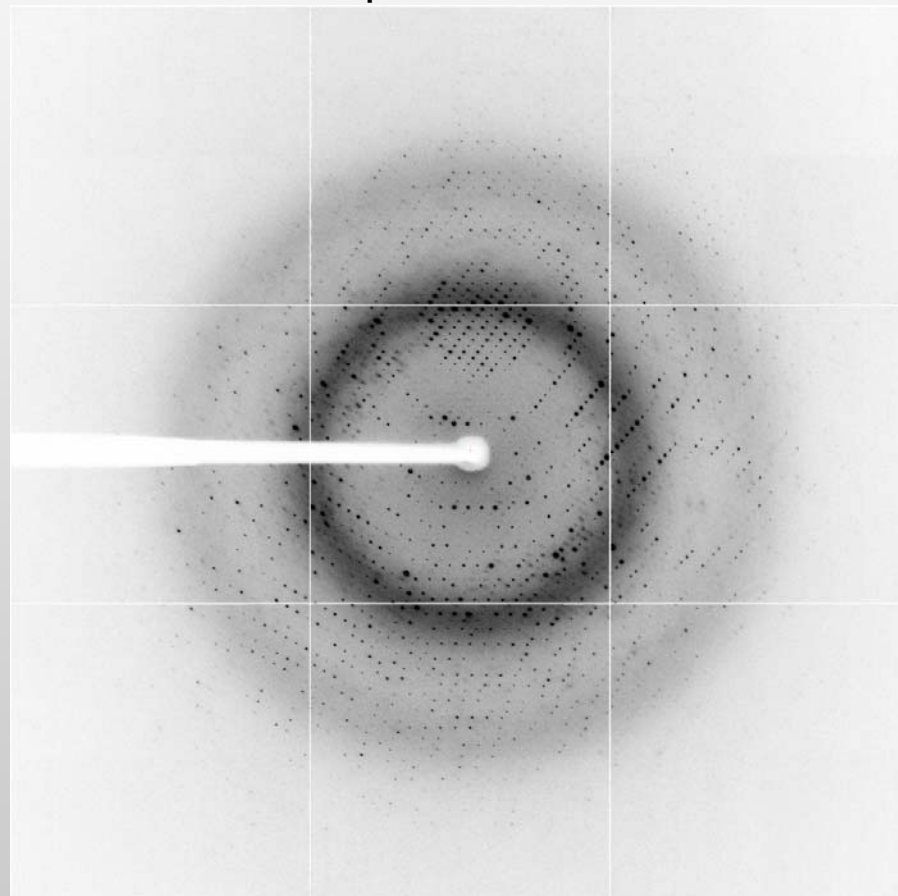
Cell : 148.7 53.6 76.5 90.0 103.6 90.0

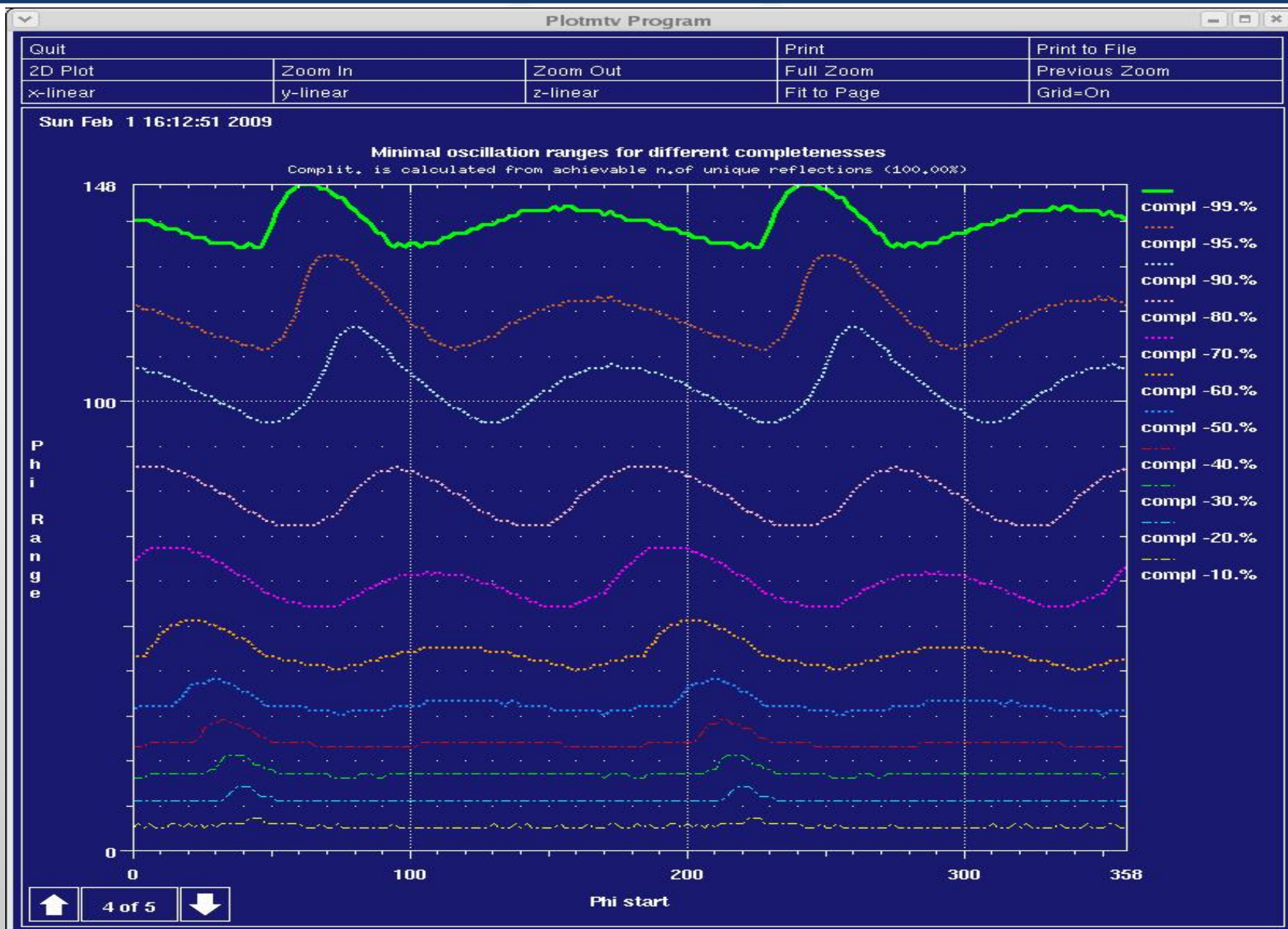
Mosaicity : 0.40 degree

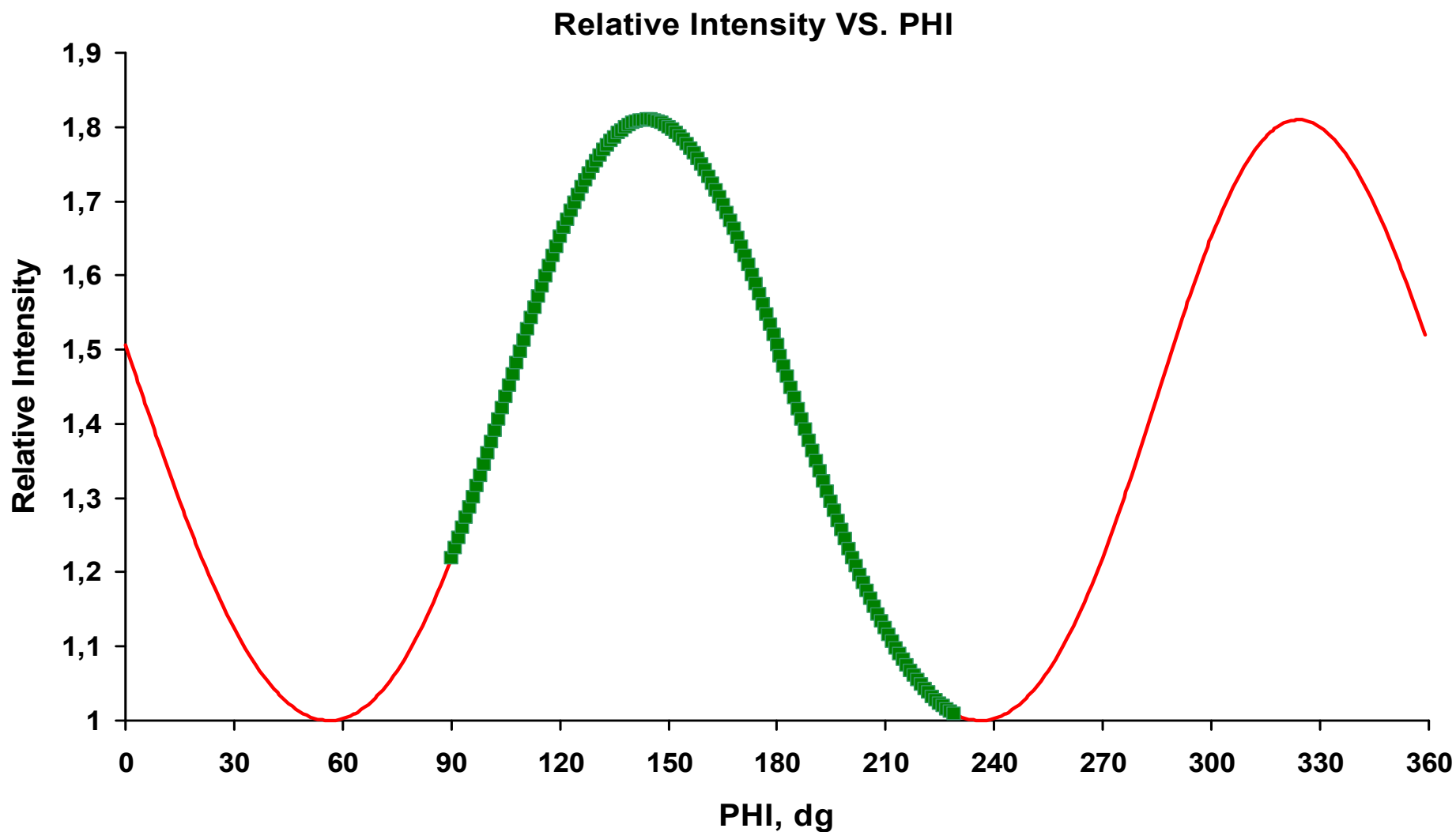
$\phi=0^\circ$

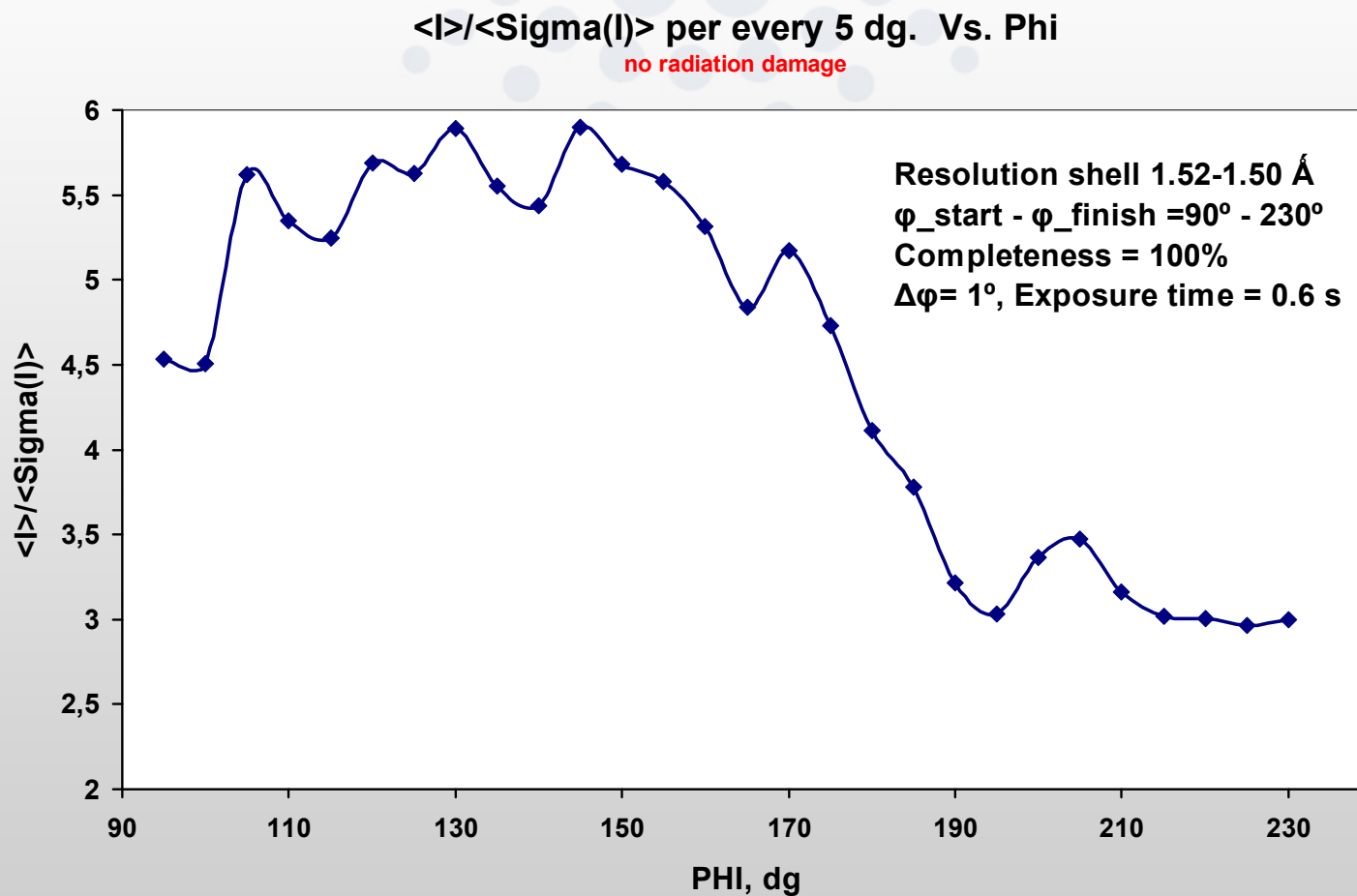


$\phi=90^\circ$









$\langle I \rangle / \langle \Sigma(I) \rangle$ per every 5 dg. Vs. Phi

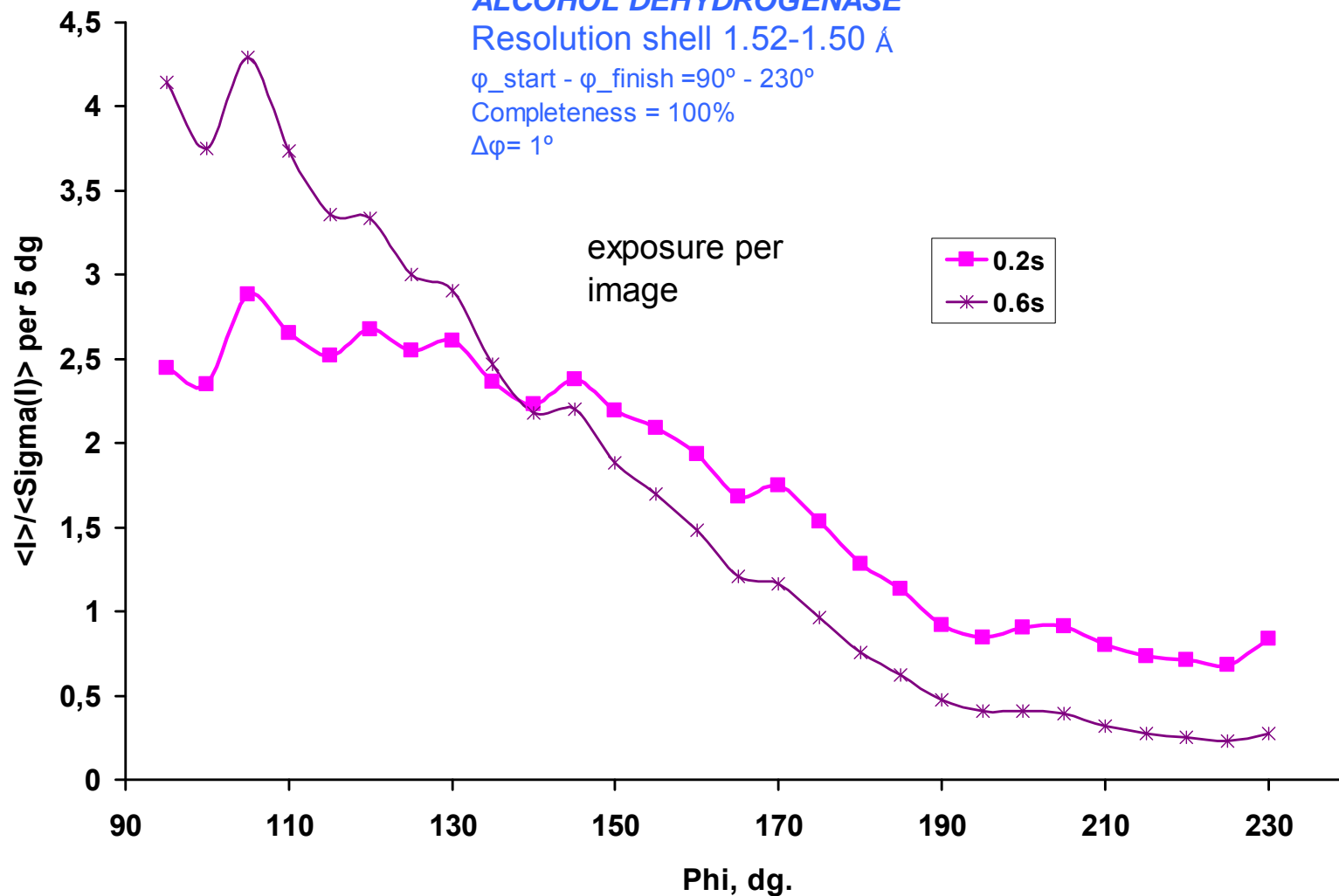
ALCOHOL DEHYDROGENASE

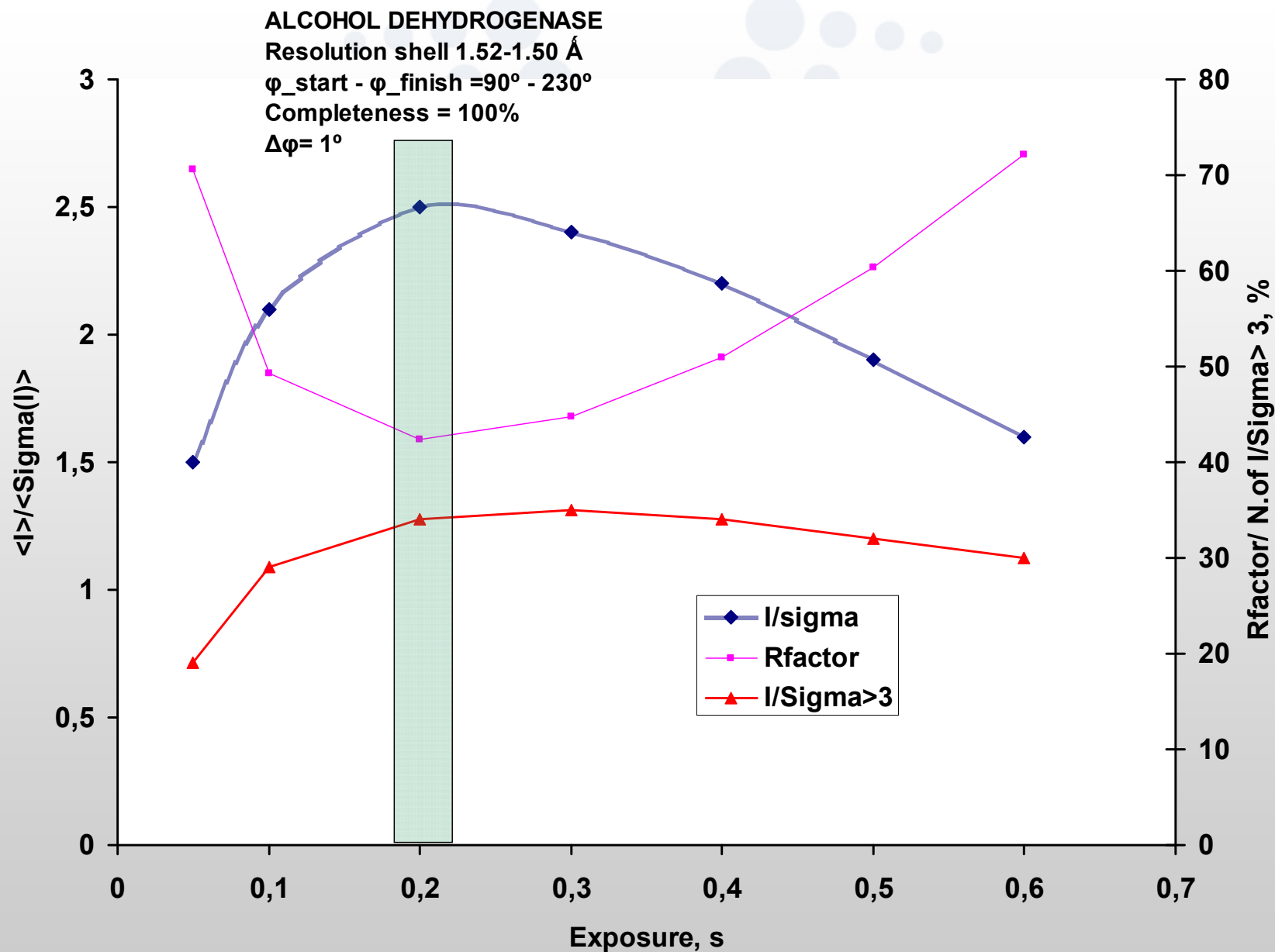
Resolution shell 1.52-1.50 Å

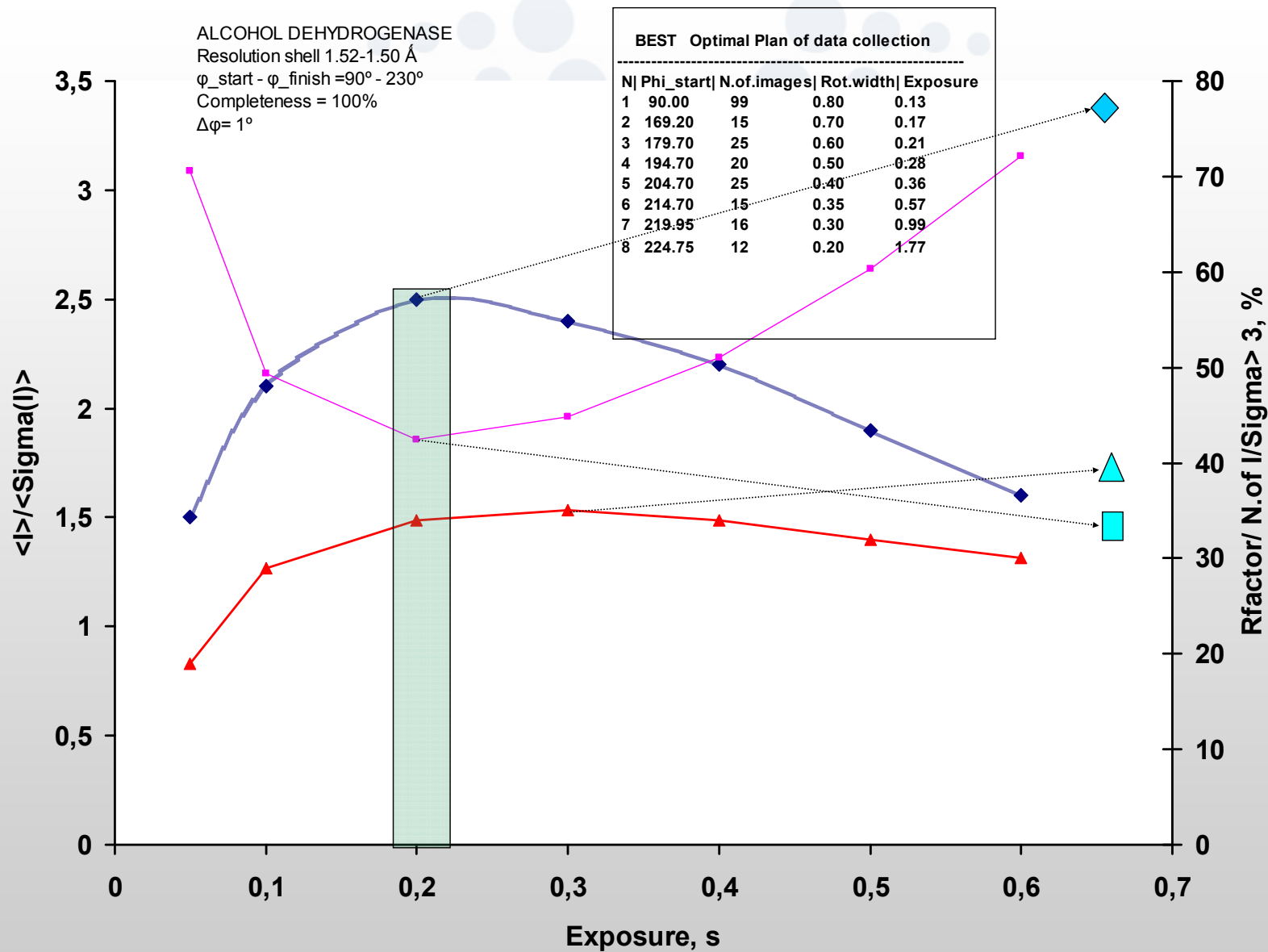
$\phi_{\text{start}} - \phi_{\text{finish}} = 90^\circ - 230^\circ$

Completeness = 100%

$\Delta\phi = 1^\circ$







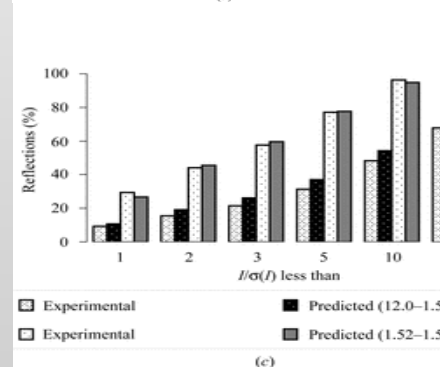
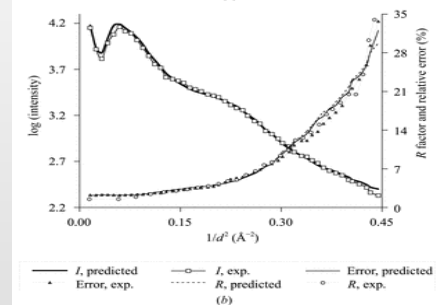
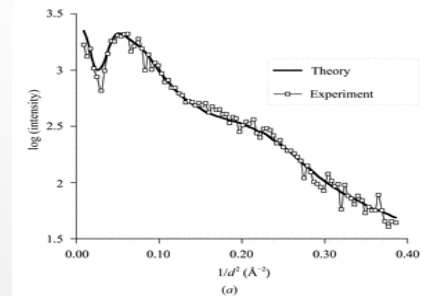
- semi-empirical model for diffraction intensity vs reciprocal space coordinate

$$\hat{J}(\mathbf{h}) = \frac{1}{S} \cdot \hat{J}_u(h) \cdot \text{Exp}(-\mathbf{h} \cdot \mathbf{B} \cdot \mathbf{h}^T)$$

- semi-empirical model of variance vs integrated intensity (Bragg peak, background - ADU/ADU)

$$\sigma^2(I) = \alpha_0 + \alpha_1 I + \alpha_2 I^2$$

- factorization on the basis of few initial diffraction images (background and Bragg scattering) and few instrumental parameters (calibration)
- relationship *via* diffraction geometry (rotation method) – scan speed, oscillation range, mosaicity, spot shape, Lorentz, polarization
- integration over the scanned reciprocal space + Wilson distribution: **data statistics** ($\langle I/\text{SigI} \rangle$ /distributions, R_{merge} , R_{friedel} etc.)
- **optimal data collection parameters** :
total rotation range, oscillation ranges/frame, exposure time providing *required* data statistics in a *minimal* total measurement time or exposure dose



Popov & Bourenkov, Acta Cryst. D 2003

Average standard uncertainty of the observed squared structure factors for a sampled volume v is approximated by:

$$\hat{\sigma}_J(h, \varphi) = \frac{1}{2N} \sum_{i=1}^N \int_0^\infty \sqrt{k_{0i} + k_{1i} J + k_2 J^2} \left(p(J | \hat{J}(\mathbf{h}_{i1})) + p(J | \hat{J}(\mathbf{h}_{i2})) \right) dJ$$

$$k_0(h, \zeta, \varphi) = \frac{Gk_b \Delta\Phi (\Delta S^2 PA \rho_s \omega + \rho_{dark} \omega + \rho_{readout} \omega^2 / \Delta\varphi)}{(\text{LPA})^2}$$

Scattering background component

$$k_1(h, \zeta, \varphi) = \frac{G\omega}{\text{LPA}}$$

the contribution of peak counting statistics

$$k_2 = k_{ins}$$

the contribution of the instrumental error

After merging symmetry equivalent/redundant observations over a large rotation range subdivided into m equally small intervals centred at spindle positions, and assuming that redundant measurements are uniformly distributed over the rotation range, the expectation values of standard uncertainties can be estimated as:

$$\hat{\sigma}_J(h) \cong M^{-1/2} m^{-1} \sum_{j=1}^m \hat{\sigma}_J(h, \varphi_j)$$

M is the mean redundancy of the measurements in the total rotation interval

merging equivalents

$$1/\sigma^2_{\langle E^2 \rangle} = 1/\sigma^2_1 + 1/\sigma^2_2 + \dots + 1/\sigma^2_n$$

$$\langle E^2 \rangle_{hkl} = \sigma^2_{\langle E^2 \rangle} (E^2_1/\sigma^2_1 + E^2_2/\sigma^2_2 + \dots + E^2_n/\sigma^2_n)$$

Variance $E^2_{k hkl}$ about $\langle E^2 \rangle_{hkl}$ $\rightarrow R_{\text{merge}}$

$$\begin{pmatrix} 1/\sigma^2_1 \\ 1/\sigma^2_2 \\ \dots \\ 1/\sigma^2_{k-1} \\ \dots \\ 1/\sigma^2_n \end{pmatrix}^T \left[\begin{pmatrix} \sigma_B^{RD}(D_i) \sigma_B^{RD}(D_j) \\ \text{Corr}^{RD}(|D_i - D_j|) \end{pmatrix} + \frac{1}{2}\pi \begin{pmatrix} \sigma^2_1 \\ \sigma^2_2 \\ \dots \\ \sigma^2_k \\ \dots \\ \sigma^2_n \end{pmatrix} \right] \begin{pmatrix} 1/\sigma^2_1 \\ 1/\sigma^2_2 \\ \dots \\ 1/\sigma^2_{k-1} \\ \dots \\ 1/\sigma^2_n \end{pmatrix}$$

$$V_{\kappa} = V_{\text{damage}} + V_{\text{stat}} = \sum_i^n \sum_j^n \left(\frac{\hat{\sigma}^2}{\sigma_i^2} - \delta_{ik} \right) \left(\frac{\hat{\sigma}^2}{\sigma_j^2} - \delta_{jk} \right) C_{ij} \sigma_{bi} \sigma_{bj} + (\sigma_k^2 - \hat{\sigma}^2)$$

$$R = \frac{\sum \sum \left(V_{\text{damage}} + \frac{2}{\pi} V_{\text{stat}} \right)^{1/2}}{\sum \sum \hat{I}_i}$$

A variety of different tasks, crystal characteristics and specific instrument conditions make it impossible to define rigid protocols for data collection that would be applicable in all cases. The appropriate decision has to be a result of a compromise between several competing requirements.

Structure refinement

resolution and completeness are more important than accuracy.

Molecular replacement

is usually limited to low-resolution data, for which experimental random errors are not significant. Very sensitive to systematically missing the strongest intensities owing to detector saturation

Experimental phasing

is based on measuring differences of same-index (or symmetry-related) reflections. The magnitude of the phasing power is resolution dependent. The phasing-power magnitude can be improved equally well by an increase in phasing signal or by reducing errors associated with the signal.

Ligand binding study, Drug Discovery

Crystal ranking

Beamline Flux
Crystal contents

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

RADDPOSE

Absorbed dose

BEST 3.2.

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

Crystallographic phenomenology

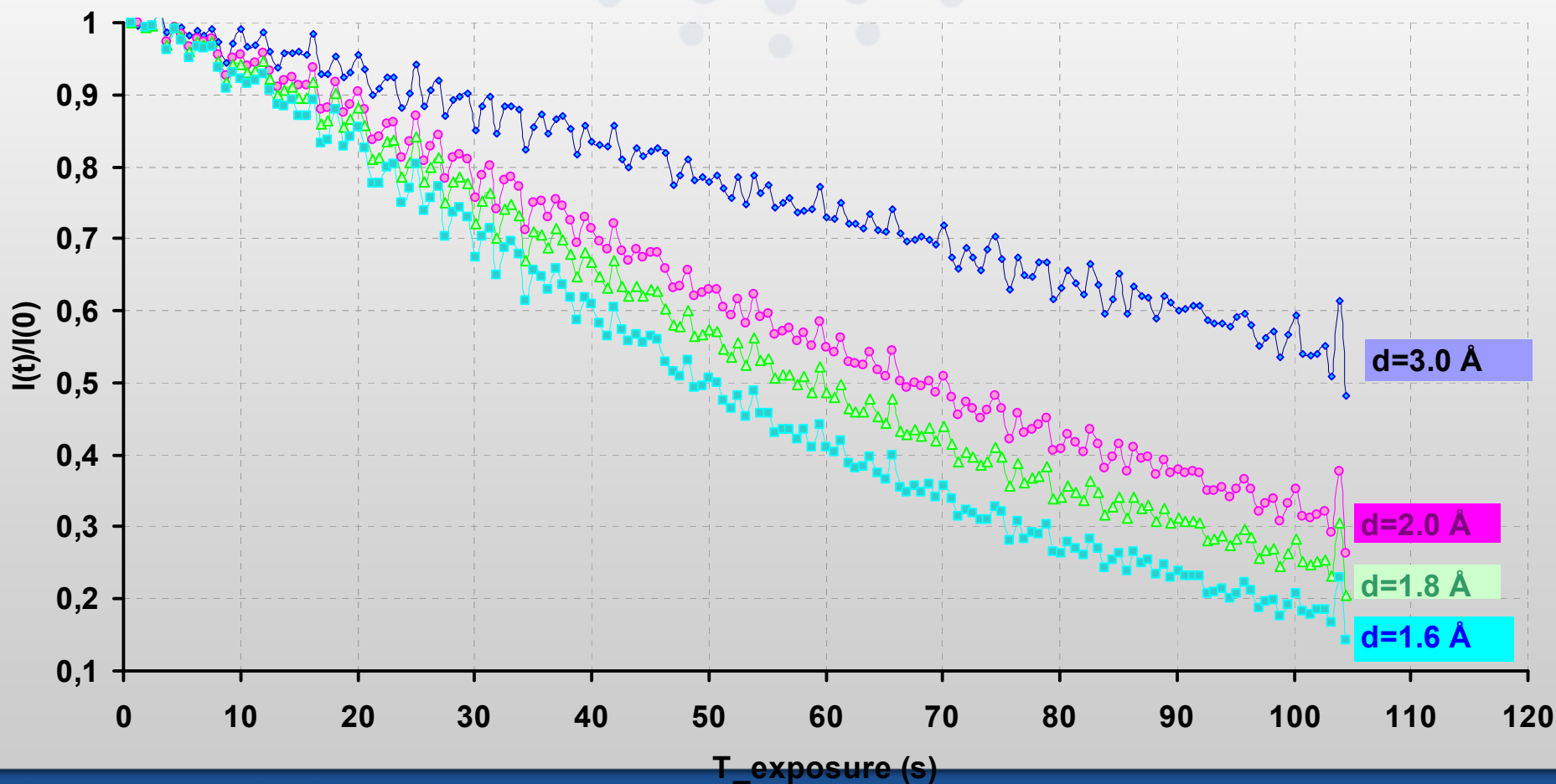
After high dose ...

- cell dimensions/orientation change *
- mosaicity increase *
- R-factors increase *
- phasing signals washed away *
- diffraction disappears *
- what can be reduced will be reduced
- holes punched
- warmed-up samples explode
-

(*irreproducible)

ID29 ESRF

Oxoanion polyreductase



Radiation Damage (Bulk MX, Cryo-T)

- Diffraction Intensity is a function of dose

$$I(hkl) = \text{scale}(\text{Dose}, |hkl|) \cdot I(hkl) + \Delta(\text{Dose})$$

- overall Debye-Waller factor (B) grows by **1 Å²** per **1 MGy**
- Luzzatti isomorphism factor (Log σ_A) decays by **0.1 Å²** per **1 MGy**

Kmetko et al. (2006)

Quantifying X-ray radiation damage in protein crystals at cryogenic temperatures

Acta Crystallographica Section D
Biological
Crystallography
DOI: 10.1107/S0907464806004441

Ian Kmetko,*† Naji S. Hussein,
Matthew Nades, Yevgeniy
Kalinin and Robert E. Thorne

The dependence of radiation damage to protein crystals at cryogenic temperatures upon the X-ray absorption cross-section of the crystal has been examined. Lysozyme crystals containing various concentrations of heavy atoms were irradiated.

Received 4 January 2006
Accepted 23 June 2006

Table 2

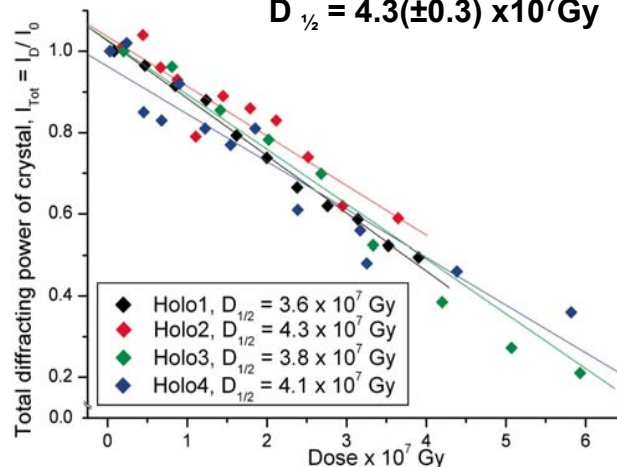
Coefficients of sensitivity, $s_{AD} = (\Delta\langle u^2 \rangle / \Delta D) = (\Delta B_{\text{rel}} / 8\pi^2 \Delta D)$, for various proteins as determined from dose curves.

Protein	Lysozyme	Catalase	Thaumatococcus	Apoferitin
PDB code†	1l28	8cat	1ly0	1ier
Space group	$P4_32_12$	$P3_221$	$P4_12_12$	$F432$
MW (kDa)	14.4	230	22.2	476
Solvent content‡ (%)	39	53	56	61
Photon energy (keV)	8.82	10.0	9.26	9.66
Mass-energy absorption coefficient μ_{en}/ρ (cm ² g ⁻¹)	8.3	4.3	6.0	7.0
Coefficient of sensitivity s_{AD} (Å ² MGy ⁻¹)	0.012	0.012	0.018	0.017

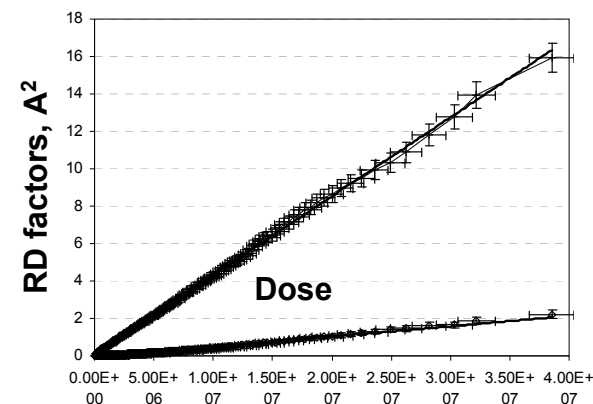
$$\beta = 8\pi^2 s_{AD} = 0.95 \quad 0.95 \quad 1.4 \quad 1.3 \text{ Å}^2$$

Owen et al. (2006)

$$D_{1/2} = 4.3(\pm 0.3) \times 10^7 \text{ Gy}$$

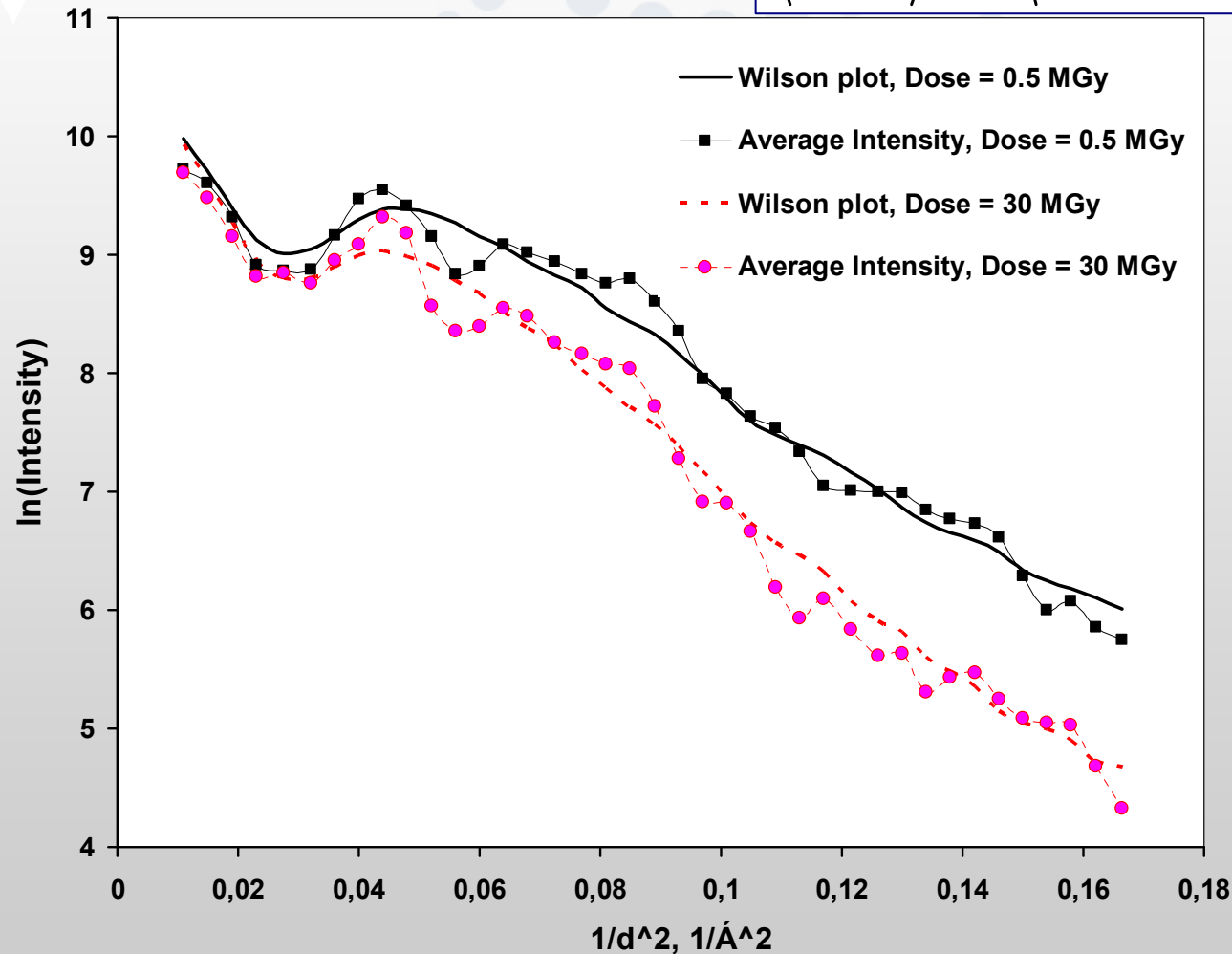


Popov et al. (2006)

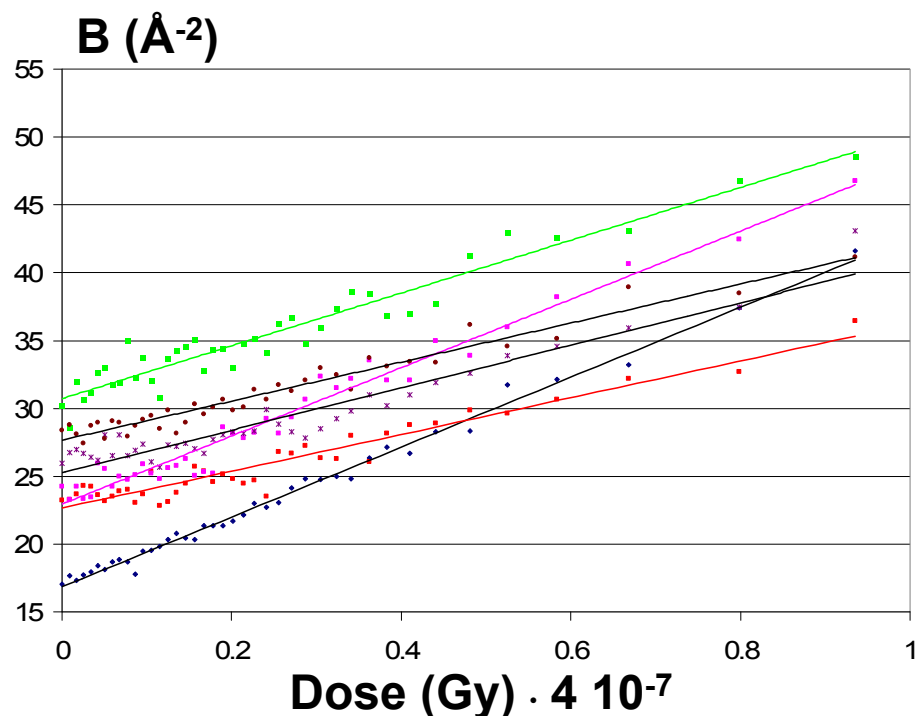


1. Intensity decay:

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



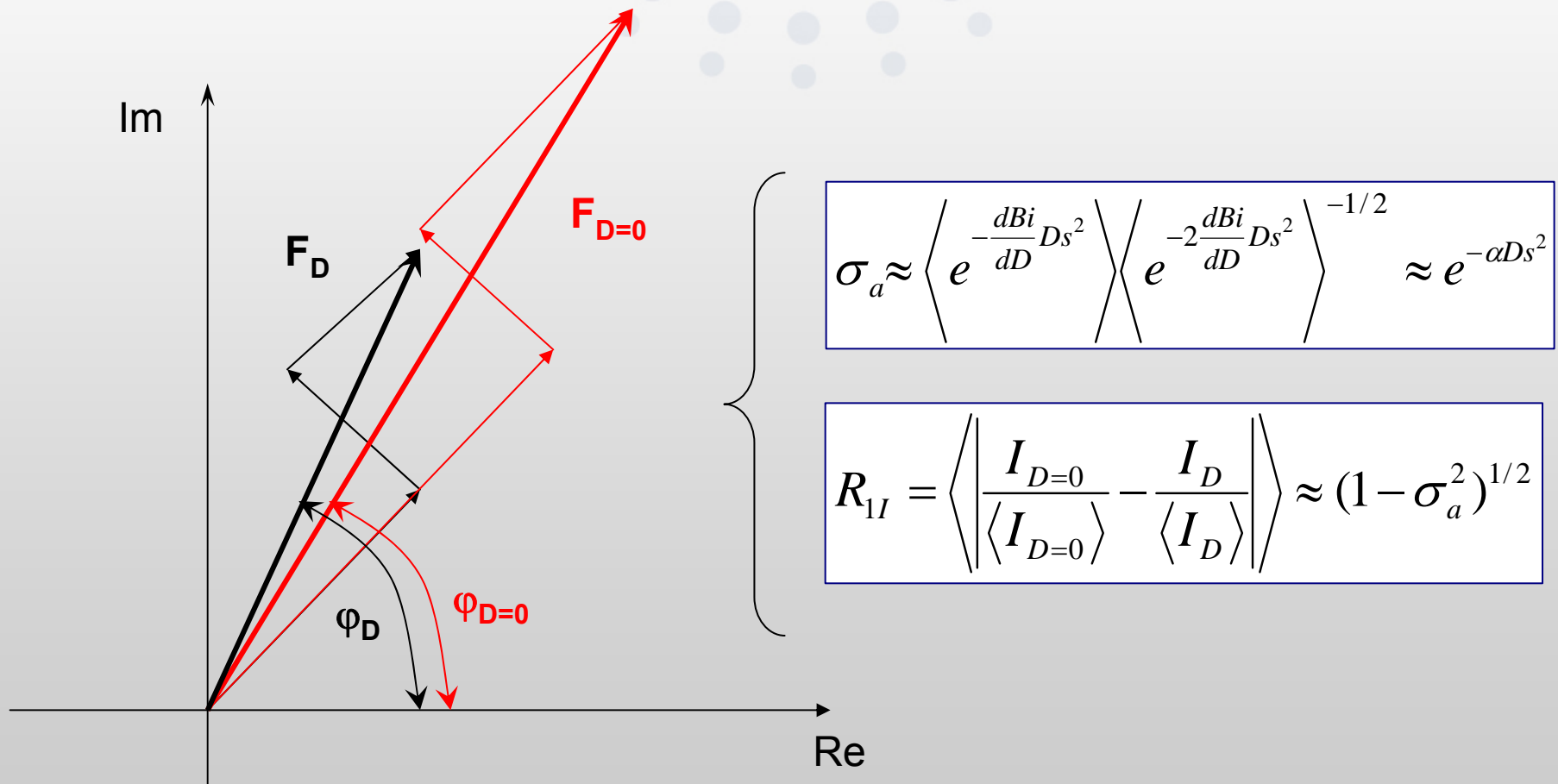
Non-specific Radiation Damage model: Atomic Debye-Waller factor variation with Dose

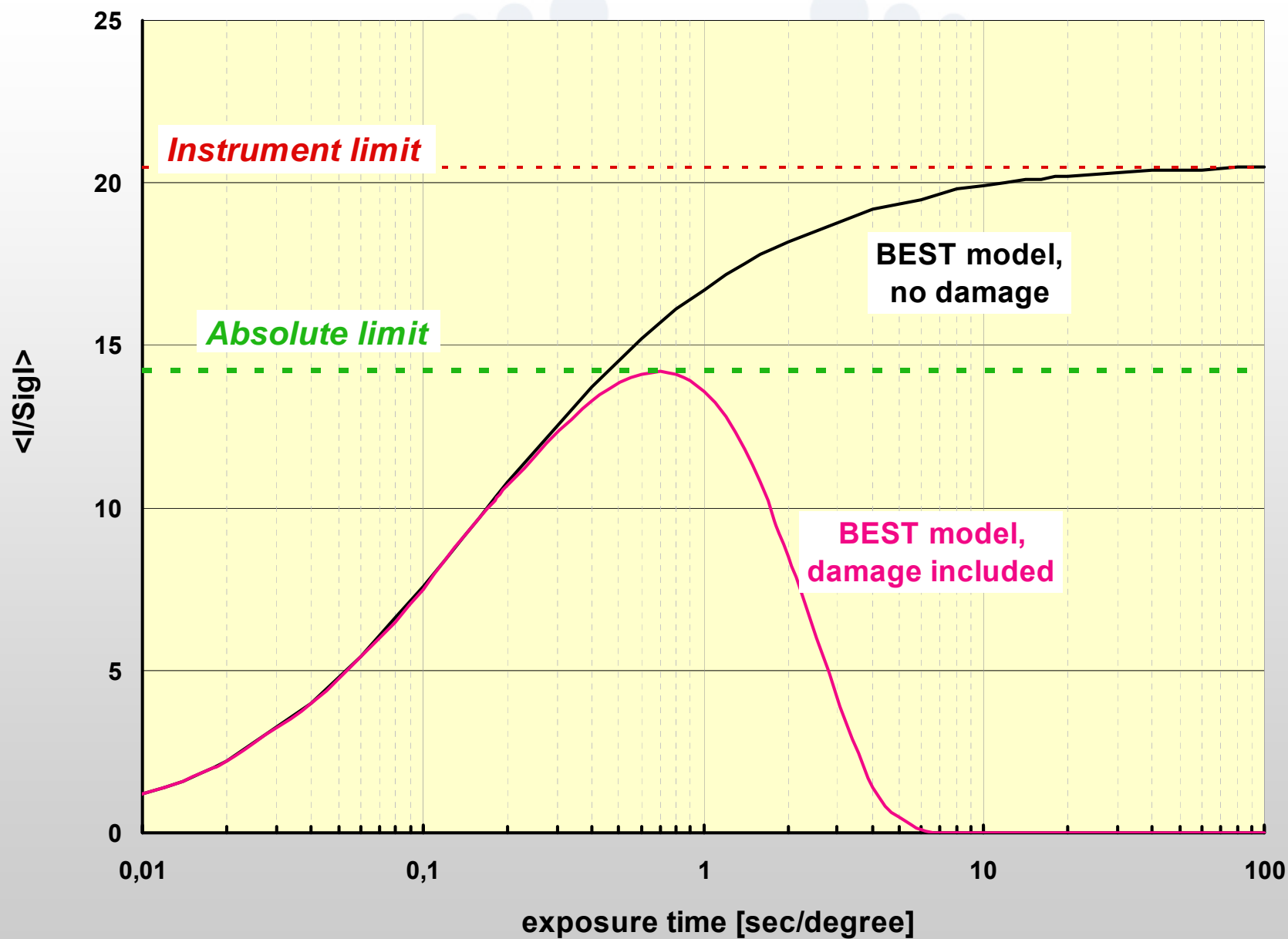


- Derived from the series of refined structures as a function of Dose @ constant I/SigI/resolution
- Not visible in electron density maps (other than blurring)
- Generates extreme non-isomorphism ($>70\%$ in R_E^2)

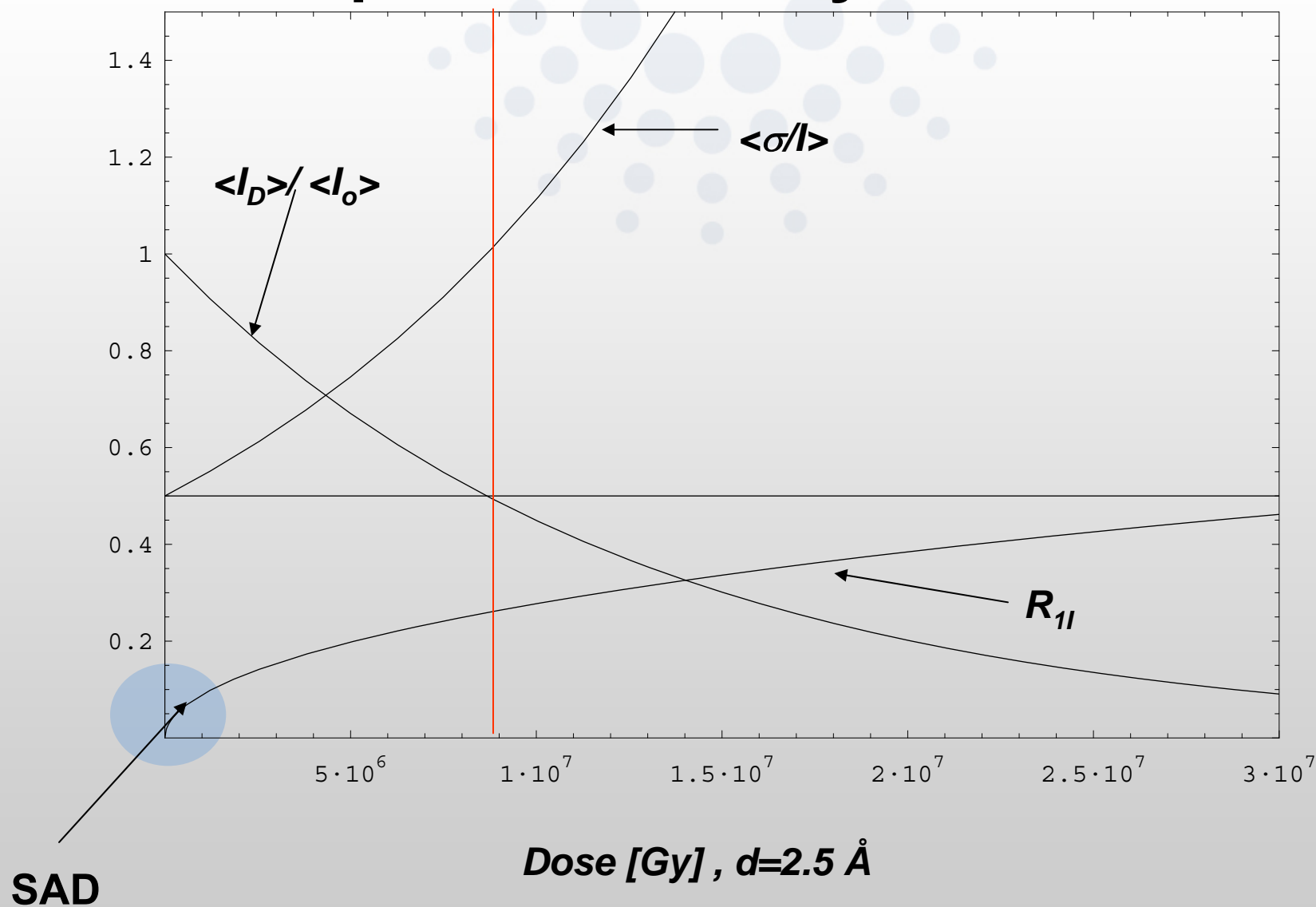
2. Intensity variance (non-isomorphism)

Non-specific Radiation Damage model: Atomic Debye-Waller factor variation with Dose





Expected Intensity Variation



Interface: CCP4I BEST

Run BEST3.0 data collection strategy

Choose function

Title: BEST 3.2.0.z

Run BEST to: **optimize data collection** ☐ show graphs

Input from: **optimize data collection**

mosflm dat in: **find optimal crystal orientation**

mosflm par in: **check minimum Rfiedel versus resolution**

hkl #1 in: **Full** **optimize SAD data collection**

find optimal crystal orientation for SAD

estimate data statistics

☐ Change Symmetry

Reference image parameters

Detector ESRF_ID23-1 Exposure time (sec) 0.2 ☐ Preset counts Read-outs 1

Radiation damage parameters

☒ Enable radiation damage corrections

☐ Use RADDOS

Dose rate 4.1 * 10⁵ Gray/second Shape factor 1.0 Susceptibility 1

Exposure time to reach 2*10⁷ Gray = 48.78 sec = 0 hrs : 0 min : 48.78 sec

Major optimization parameters

Target <I>/<SigI> in the last shell = 3.0 ☐ Anomalous data

Maximum resolution 3.5 Angstrom

Minimize exposure time, and limit it to sec

Rotation range parameters

Total rotation range auto

for completeness = 0.99 and auto redundancy

Minimum rotation range/frame, deg: 0.05

Output plan parameters

Complexity level of data collection strategy few lines (recommended for DC GUI's users)

☐ Save strategy to file

Rotation speed/exposure time limitations

Maximum scan speed, deg/sec: unlimited

Minimum exposure time/frame, sec: unlimited

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)

☐ Change Symmetry

Reference Image parameters

Detector mccd225 Exposure time (sec) 32.4 ☐ Preset counts Read-outs 1

Radiation damage parameters

☒ Enable radiation damage corrections

Dose rate 0.5 * 10⁴ Gray/second Shape factor 1 Susceptibility 1.5

Use RADDOSE

Crystal cell & symmetry :

Unit Cell: 50.02 57.80 74.41 90.000 90.000 90.000 Space Group P21212

Crystal composition:

Monomers in assymetric unit:

Amino acid residues per monomer:

DNA nucleotides per monomer:

RNA nucleotides per monomer:

Bound heavy atoms:

Atom: S # per monomer

Edit list

Add heavy atom

Solvent composition:

Atom: Concentration: mM

Edit list

Add atom:

Crystal dimensions (across rotation axis?): x micron

Incident X-ray beam:

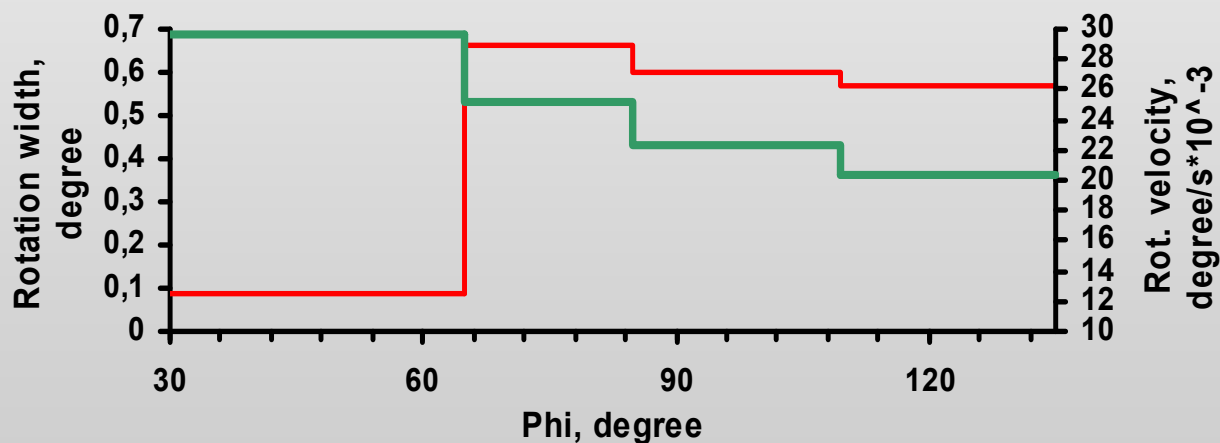
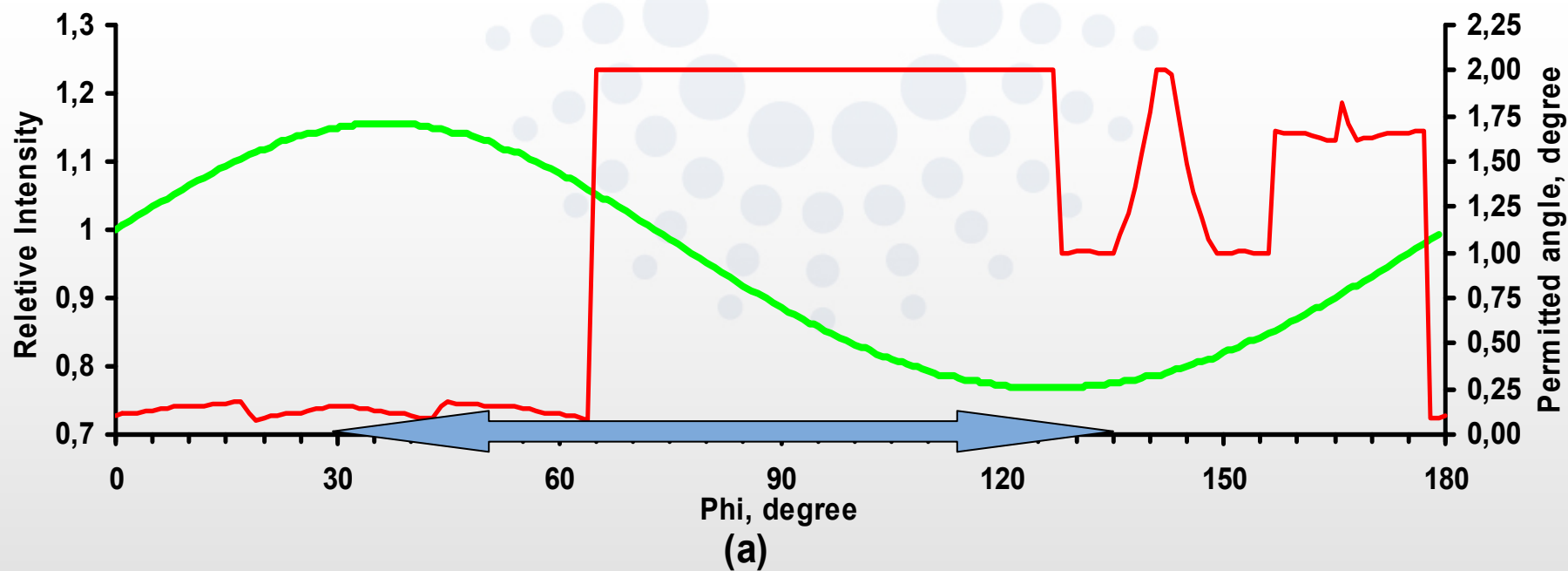
Run

Save or Restore

Close

Data Collection with Variable Exposure Time and Oscillation Width

- BEST optimizes the data collection parameters for *each crystal orientation* (i.e. spindle position reached after exposure to a certain dose) *individually*;
For convenience of data collection/processing the data collection "plan" is smoothed out to produce a *small number sub-wedges* with varying exposure/oscillation width
- Even without taking the Radiation damage into account, this is useful (e.g. severely anisotropic diffraction or long cell edge)
- For high-dose data collection, BEST suggest to increase the *exposure time* gradually during the data collection, in order to *compensate the loss of the diffraction signal* due to the radiation damage (according to the model-based expectations) and keep signal-to-noise at a required level.



Plan of data collection



Total dose $2.1 \cdot 10^7$ Gy

BEST, strategy+predictions

N	Phi_start	N.of.images	Rot.width	Exposure	Distance	Overlap
1	50.00	117	0.30	0.90	306.9	No
2	85.10	45	0.55	2.26	306.9	No
3	109.85	33	0.45	3.05	306.9	No
4	124.70	29	0.35	3.84	306.9	No
5	134.85	18	0.30	5.07	306.9	No

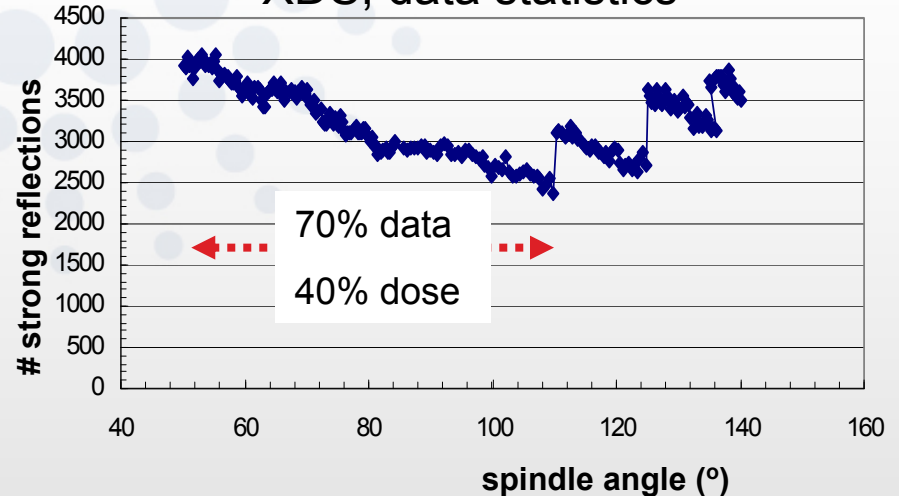
Resolution limit : 2.09 Angstrom
 Anomalous data : No
 Phi_start - Phi_finish : 50.00 - 140.25
 Total rotation range : 90.25 degree
 Total N.of images : 242
 Overall Completeness : 99.8%
 Redundancy : 3.58
 R-factor (outer shell) : 6.8% (49.6%)
 I/Sigma (outer shell) : 26.2 (2.5)
 Rel.decrease of intensity : 0.101 for outer resolution shell
 Total Exposure time : 510.3 sec (0.142 hour)
 Total Data Collection time : 1115.3 sec (0.310 hour)

Data collection statistics according to the plan

Resolution	Lower	Upper	Compl.	Average	I/Sigma	I/Sigma	Chi**2	R-factor	Overload
			%	Intensity	Sigma	/Sigma	/Chi	%	%
12.00	7.21	98.7	98937.9	2752.3	35.9	33.8	1.13	3.4	0.29
7.21	5.63	99.1	48367.1	1323.2	36.6	31.7	1.33	3.9	0.00
5.63	4.78	98.3	59056.4	1590.0	37.1	29.6	1.57	4.3	0.00
4.78	4.22	100.0	71915.2	1916.9	37.5	26.9	1.95	4.8	0.00
4.22	3.82	99.3	57916.7	1542.5	37.5	25.1	2.24	5.3	0.00
3.82	3.52	99.4	42862.0	1189.2	36.0	22.7	2.51	5.9	0.00
3.52	3.28	99.5	30208.6	873.8	34.6	20.8	2.77	6.6	0.00
3.28	3.08	99.5	19816.4	634.9	31.2	18.8	2.77	7.3	0.00
3.08	2.92	99.5	13351.8	488.5	27.3	16.6	2.70	8.1	0.00
2.92	2.78	100.0	9301.2	396.5	23.5	14.9	2.48	9.0	0.00
2.78	2.65	100.0	6743.6	356.3	18.9	12.8	2.18	10.2	0.00
2.65	2.54	100.0	5144.5	347.4	14.8	10.8	1.87	11.8	0.00
2.54	2.45	100.0	4062.3	345.6	11.8	9.2	1.64	13.7	0.00
2.45	2.36	100.0	3218.2	368.6	8.7	7.3	1.44	17.1	0.00
2.36	2.29	100.0	2679.3	398.4	6.7	6.0	1.27	20.6	0.00
2.29	2.22	100.0	2175.6	464.8	4.7	4.3	1.16	28.0	0.00
2.22	2.15	100.0	1760.4	486.0	3.6	3.4	1.10	35.1	0.00
2.15	2.09	100.0	1386.3	512.6	2.7	2.6	1.06	46.4	0.00
All data		99.8	18040.4	689.7	26.2	19.4	1.81	6.8	0.02

R-factor = $\frac{\sum (|I - \langle I \rangle|)}{\sum (I)}$
 Chi**2 = $\frac{\sum (I - \langle I \rangle)^2}{(\text{Error}^2 * N / (N-1))}$
 measure of the non-isomorphism due to radiation damage

XDS, data statistics



SUBSET OF INTENSITY DATA WITH SIGNAL/NOISE ≥ -3.0 AS FUNCTION OF RESOLUTION									
RESOLUTION	NUMBER OF REFLECTIONS			COMPLETENESS	R-FACTOR	R-FACTOR	I/SIGMA		
LIMIT	OBSERVED	UNIQUE	POSSIBLE	OF DATA	observed	expected	COMPARED		
12.00	766	265	370	71.6%	3.2%	4.1%	766	26.59	
5.63	8749	2656	2718	97.7%	3.3%	4.4%	8749	25.93	
4.22	13255	3980	4033	98.7%	4.0%	4.5%	13255	25.92	
3.52	16942	4918	4954	99.3%	4.2%	4.8%	16942	23.74	
3.08	20222	5766	5784	99.7%	5.1%	5.5%	20222	19.22	
2.78	22280	6252	6258	99.9%	7.8%	8.0%	22280	13.39	
2.54	26218	7342	7347	99.9%	11.5%	11.8%	26218	9.51	
2.36	27112	7564	7577	99.8%	18.3%	19.1%	27112	6.40	
2.22	27554	7717	7725	99.9%	26.0%	25.6%	27554	4.88	
2.09	19409	7758	9112	85.1%	34.3%	34.0%	19409	3.09	
total	182507	54218	55878	97.0%	6.2%	6.8%	182507	12.36	

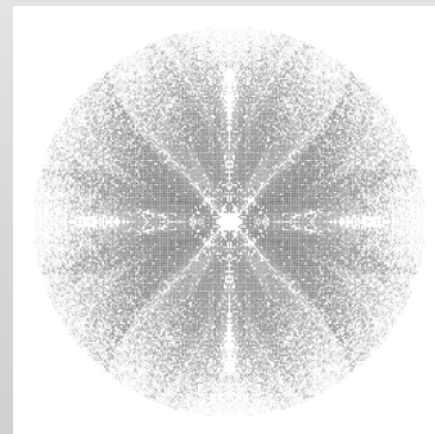
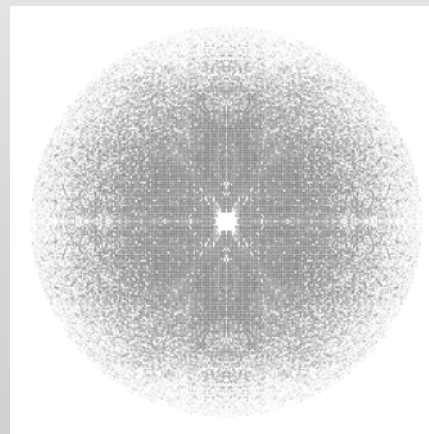
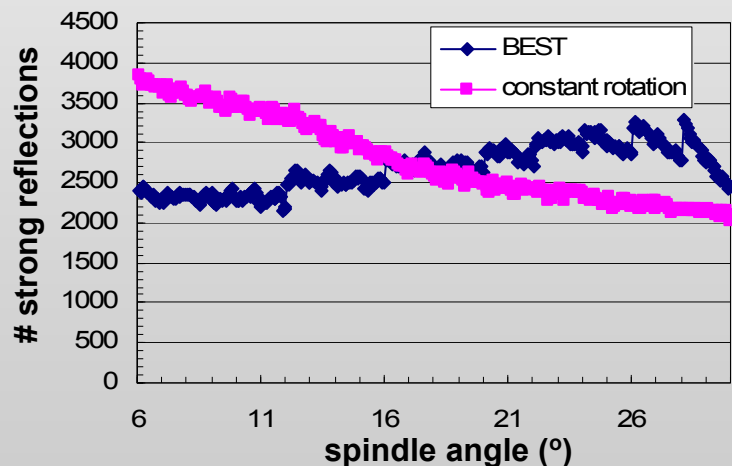
Chi², R_{merge} calculations involve Non-Isomorphism model

BEST data collection, SCALEPACK

Shell limit	Lower Angstrom	Upper Angstrom	Average I	Average error	stat. error	Norm. Chi**2	Linear R-fac	Square R-fac
99.00	4.36	4.36	23482.1	917.9	564.6	2.166	0.044	0.050
4.36	3.46	3.46	25783.9	973.8	547.2	2.279	0.045	0.051
3.46	3.02	3.02	14175.3	563.7	344.3	2.351	0.054	0.059
3.02	2.75	2.75	8797.3	389.8	265.1	2.206	0.061	0.066
2.75	2.55	2.55	6357.4	314.1	232.4	2.108	0.069	0.073
2.55	2.40	2.40	4780.3	266.1	210.9	1.973	0.077	0.081
2.40	2.28	2.28	3923.5	243.3	200.7	1.862	0.085	0.087
2.28	2.18	2.18	3339.1	232.8	199.3	1.831	0.097	0.098
2.18	2.10	2.10	3001.2	232.6	204.4	1.681	0.104	0.106
2.10	2.02	2.02	2479.7	224.5	203.5	1.546	0.120	0.123
2.02	1.96	1.96	2057.6	219.5	203.8	1.471	0.137	0.143
1.96	1.90	1.90	1582.7	210.1	199.5	1.370	0.167	0.173
1.90	1.85	1.85	1258.9	203.2	195.7	1.305	0.202	0.206
1.85	1.81	1.81	1023.9	199.1	193.7	1.280	0.243	0.253
1.81	1.77	1.77	832.5	195.6	191.7	1.210	0.292	0.295
1.77	1.73	1.73	696.2	195.4	192.6	1.210	0.351	0.356
1.73	1.70	1.70	616.3	196.7	194.4	1.170	0.396	0.405
1.70	1.66	1.66	562.9	199.6	197.7	1.176	0.456	0.484
All reflections			4207.4	298.3	242.0	1.225	0.049	0.054

Constant rotation, SCALEPACK

Shell limit	Lower Angstrom	Upper Angstrom	Average I	Average error	stat. error	Norm. Chi**2	Linear R-fac	Square R-fac
99.00	4.27	4.27	55916.7	2230.3	1419.7	3.226	0.057	0.067
4.27	3.39	3.39	58167.5	2318.5	1432.9	3.514	0.061	0.070
3.39	2.96	2.96	30537.1	1319.4	883.2	3.581	0.071	0.077
2.96	2.69	2.69	19545.7	949.6	689.0	3.316	0.081	0.085
2.69	2.49	2.49	13434.0	748.4	579.1	3.026	0.093	0.095
2.49	2.35	2.35	10232.7	662.1	542.7	2.799	0.108	0.109
2.35	2.23	2.23	8489.9	635.8	543.1	2.598	0.127	0.126
2.23	2.13	2.13	7485.7	637.1	559.1	2.471	0.147	0.152
2.13	2.05	2.05	6420.3	644.3	580.8	2.199	0.170	0.175
2.05	1.98	1.98	5287.1	639.3	591.8	1.984	0.204	0.221
1.98	1.92	1.92	4107.3	637.0	603.6	1.754	0.254	0.280
1.92	1.86	1.86	3194.9	635.4	612.1	1.621	0.325	0.364
1.86	1.81	1.81	2514.2	636.4	619.9	1.492	0.412	0.451
1.81	1.77	1.77	2053.0	649.6	637.4	1.401	0.519	0.563
1.77	1.73	1.73	1680.9	662.7	653.7	1.345	0.667	0.714
1.73	1.69	1.69	1482.4	689.0	681.7	1.295	0.779	0.855
1.69	1.66	1.66	1341.0	720.7	714.6	1.251	0.905	0.973
All reflections			9373.7	861.1	740.2	1.935	0.118	0.090



Pseudo-precession images, $I/\sigma I$ distribution in $\{hk0\}$ plane

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

Resolution		Compl. %	Average		I/Sigma stat	I/Sigma /Chi	Chi**2	R-fact %	Ranom %	Overload %
Lower	Upper		Intensity	Sigma						
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

SAD optimization

Minimum of $R_{Friedel} = \langle | \langle E^2+ \rangle - \langle E^2- \rangle | \rangle$ is a target
noise only, no anomalous scattering itself:
decay, non-isomorphism
exact pair-vice dose differences for Bijvoet mates

merging equivalents

$$1/\sigma^2_{\langle E^2 \rangle} = 1/\sigma^2_1 + 1/\sigma^2_2 + \dots + 1/\sigma^2_n$$

$$\langle E^2 \rangle_{hkl} = \sigma^2_{\langle E^2 \rangle} (E^2_1/\sigma^2_1 + E^2_2/\sigma^2_2 + \dots + E^2_n/\sigma^2_n)$$

Variance $\langle E^2 \rangle_{hkl}$ about $\langle E^2 \rangle_{hkl}$ $\rightarrow R_{\text{Fridel}}$

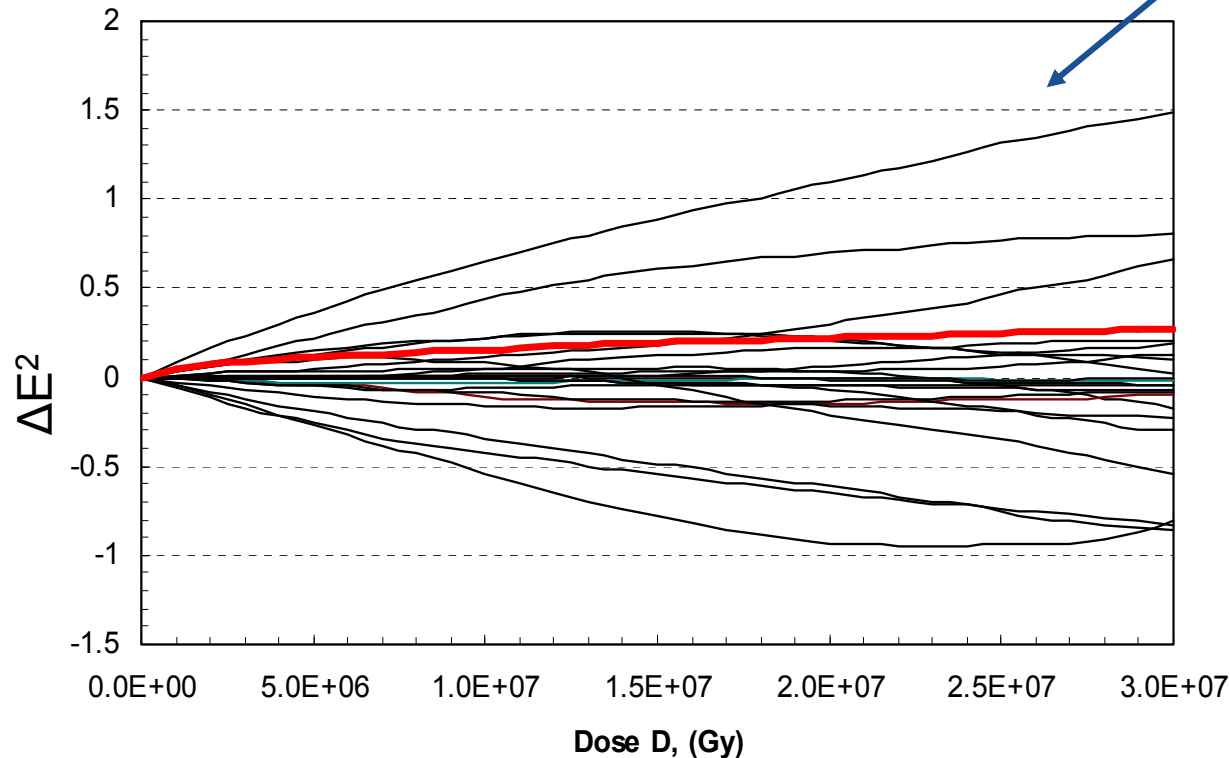
$$\begin{pmatrix} \pm 1/\sigma^2_{1\pm} \\ \pm 1/\sigma^2_{2\pm} \\ \dots \\ \pm 1/\sigma^2_{k\pm} \\ \dots \\ \pm 1/\sigma^2_{n\pm} \end{pmatrix}^T \left[\begin{array}{c} \sigma_B^{\text{RD}}(D_i) \sigma_B^{\text{RD}}(D_j) \\ \text{Corr}^{\text{RD}}(|D_i - D_j|) \end{array} + \begin{array}{c} \sigma^2_1 \\ \sigma^2_2 \\ \dots \\ \sigma^2_k \\ \dots \\ \sigma^2_n \end{array} \right] \begin{pmatrix} \pm 1/\sigma^2_{1\pm} \\ \pm 1/\sigma^2_{2\pm} \\ \dots \\ \pm 1/\sigma^2_{k\pm} \\ \dots \\ \pm 1/\sigma^2_{n\pm} \end{pmatrix}$$

$$(\Delta_{\text{anom}})_k^2 = \sum_i^n \sum_j^n \frac{\hat{\sigma}_i^2}{\sigma_i^2} \frac{\hat{\sigma}_j^2}{\sigma_j^2} C_{ij} \sigma_{bi} \sigma_{bj} \begin{bmatrix} -1 & + & - \\ +1 & + & + \\ & - & - \end{bmatrix} + \frac{2}{\pi} (\hat{\sigma}_+^2 - \hat{\sigma}_-^2)$$

$$R_{\text{anom}} = \frac{\frac{1}{N} \sum \Delta_{\text{anom},i}}{2\hat{I}}$$

$\Delta E^2 = E^2_{\text{Dose}} - E^2_{\text{Dose}=0}$ as a function of dose for a bunch of reflections in a narrow resolution shell, calculated from series of refined structures.

Stationary stochastic process (exponential distribution at any Dose)



Variance:

$$\sigma_B^{2\text{RD}} \approx 1 - \text{Exp}[-2 \cdot 10^{-7} D \text{ s}^2]$$

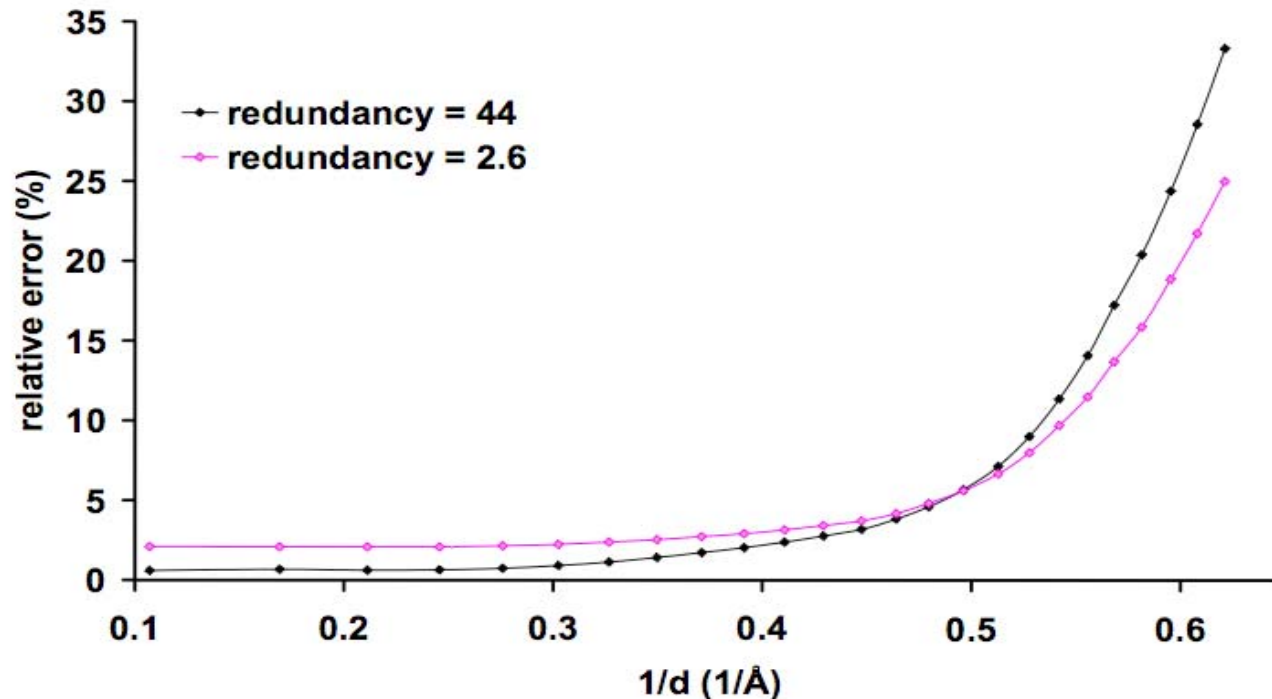
Correlation (D_1, D_2):

$\text{Corr}^{\text{RD}} \approx$

$$\text{Exp}[-10^{-8} |D_1 - D_2| \text{ s}^2]$$

Total rotation range

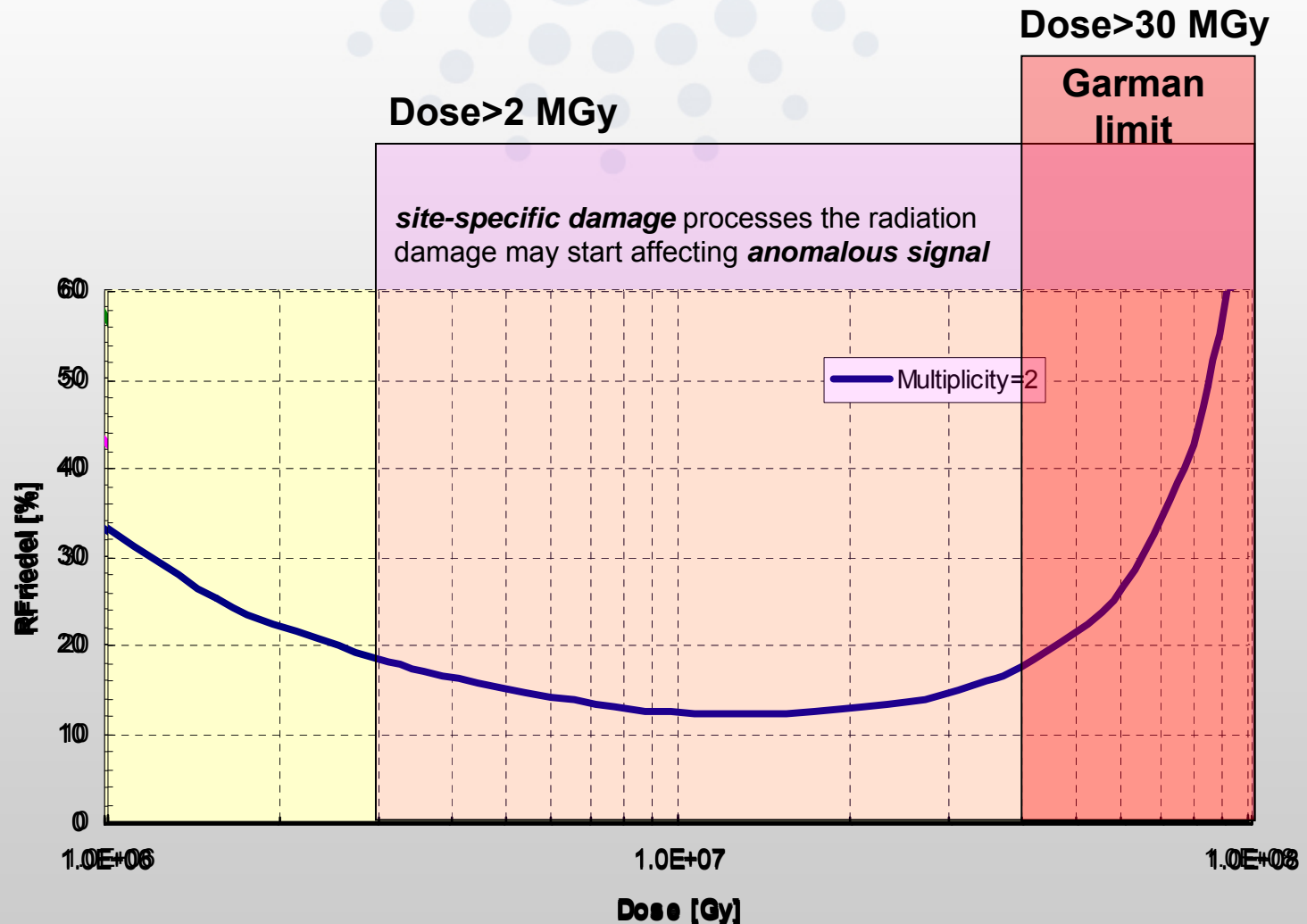
Predicted average relative error *versus* resolution for data collections with the same total radiation dose and different data redundancies

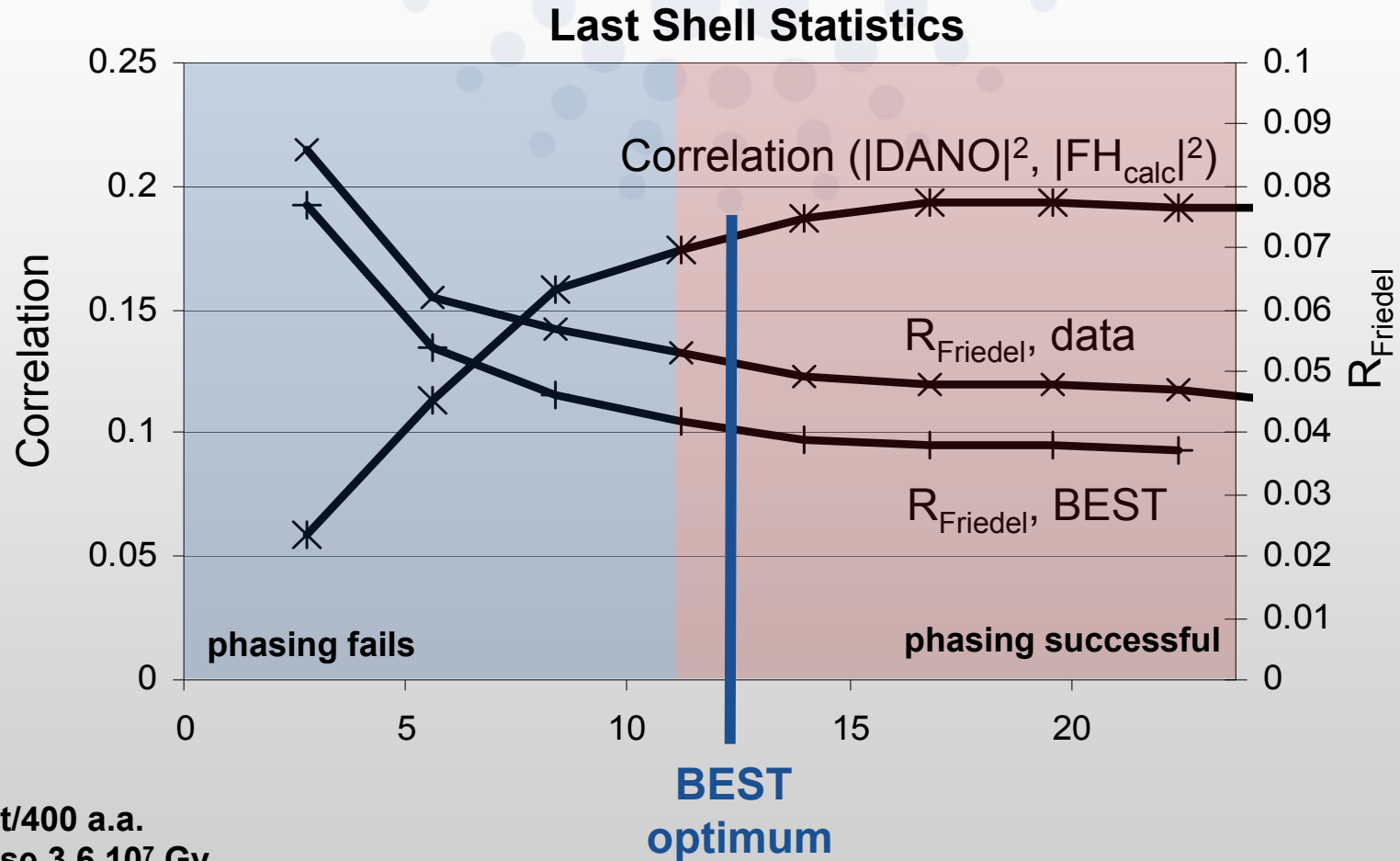


The choice of the starting and final angles of rotation range must assure the desired completeness of the data set. The smallest rotation range gives the data set with minimum redundancy. Two perfect experiments with the same total exposure dose used for different rotation ranges would provide the same data statistics. However, in the presence of the systematic error increasing the exposure dose per frame may not increase for strong signals over a certain limit. In our model, the contribution of the systematic error to the total intensity error equals 3% of intensity, i.e. for a single observation. Independent redundant measurements can improve the data statistics above this limit. On the other hand, increasing the number of frames by collecting redundant data results in stronger effects of the readout noise on weak signals.

SAD optimization

Minimum of $R_{Friedel} = \frac{\langle |E^2| \rangle - \langle E^2 \rangle^2}{\langle |E^2| \rangle}$ is a target



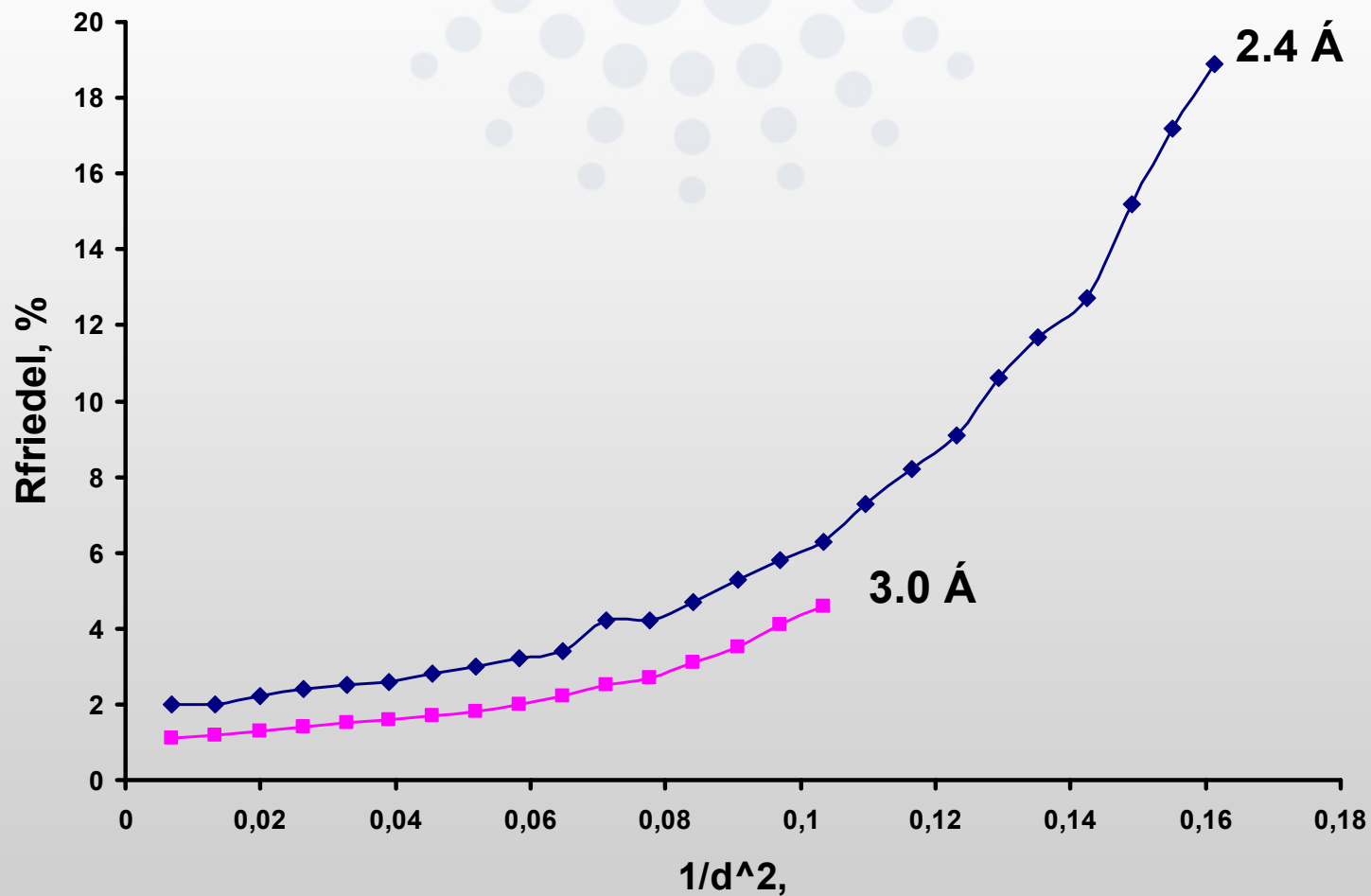


3.3 Å
11 SeMet/400 a.a.
Total Dose $3.6 \cdot 10^7$ Gy
Constant rotation

Minimum of RFriedel = $\langle | \langle E^2 + w \rangle - \langle E^2 - w \rangle | \rangle$ is a target
noise only, no anomalous scattering itself:
decay, non-isomorphism
exact pair-vice dose differences for Bijvoet mates

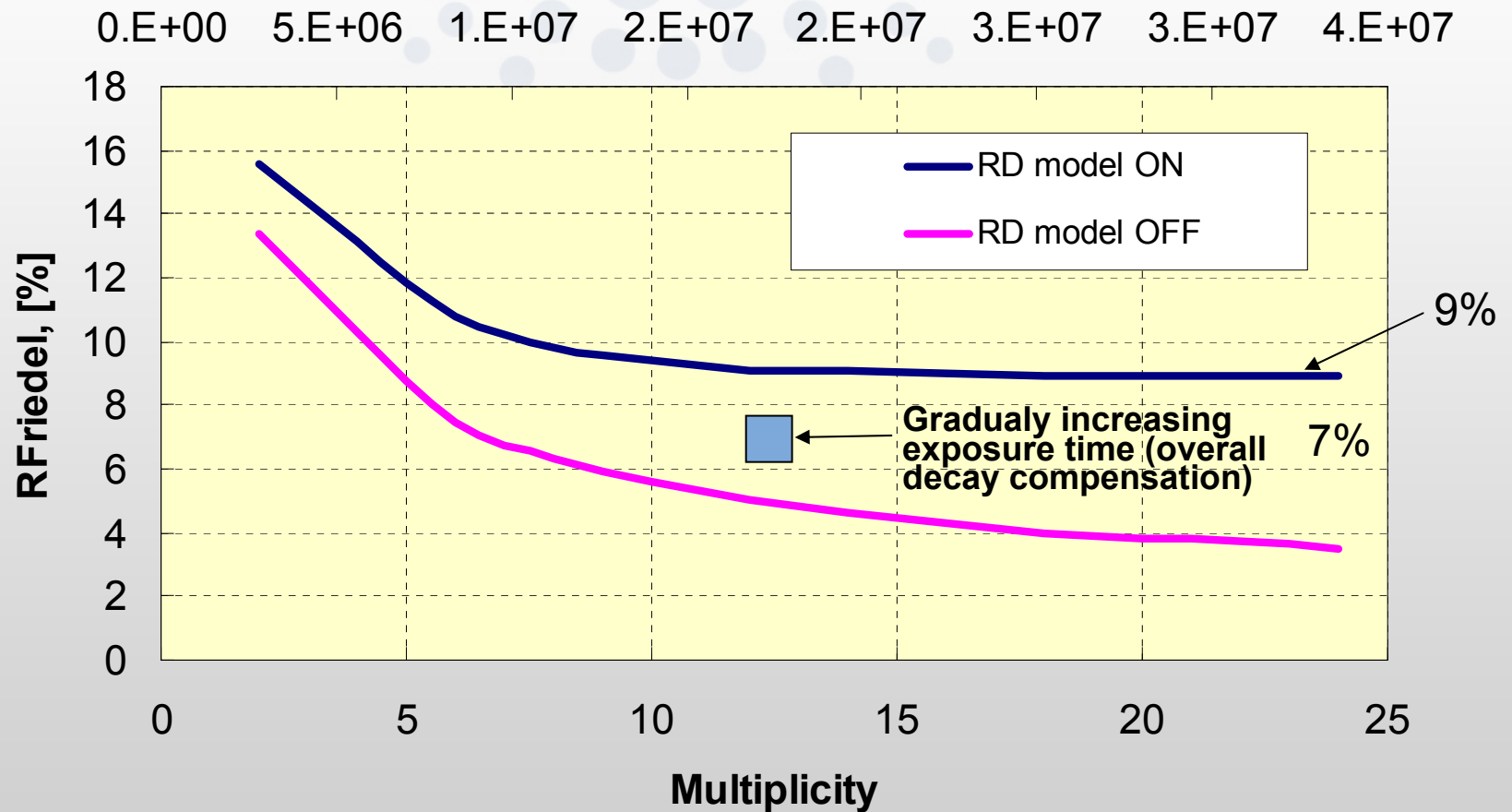
Resolution	RFriedel(%)	I/Sigma	Redundancy
10.12	0.8	74.1	23.7
6.90	0.8	43.6	23.7
5.34	1.1	48.4	23.0
4.51	1.2	47.5	23.5
3.98	1.6	34.5	20.6
3.60	2.5	22.4	13.9
3.31	4.0	14.0	11.9
3.08	6.6	8.3	7.0
2.89	10.5	5.2	6.1
2.73	15.6	3.7	2.5
2.60	23.0	2.4	3.8

Rfriedel vs. resolution

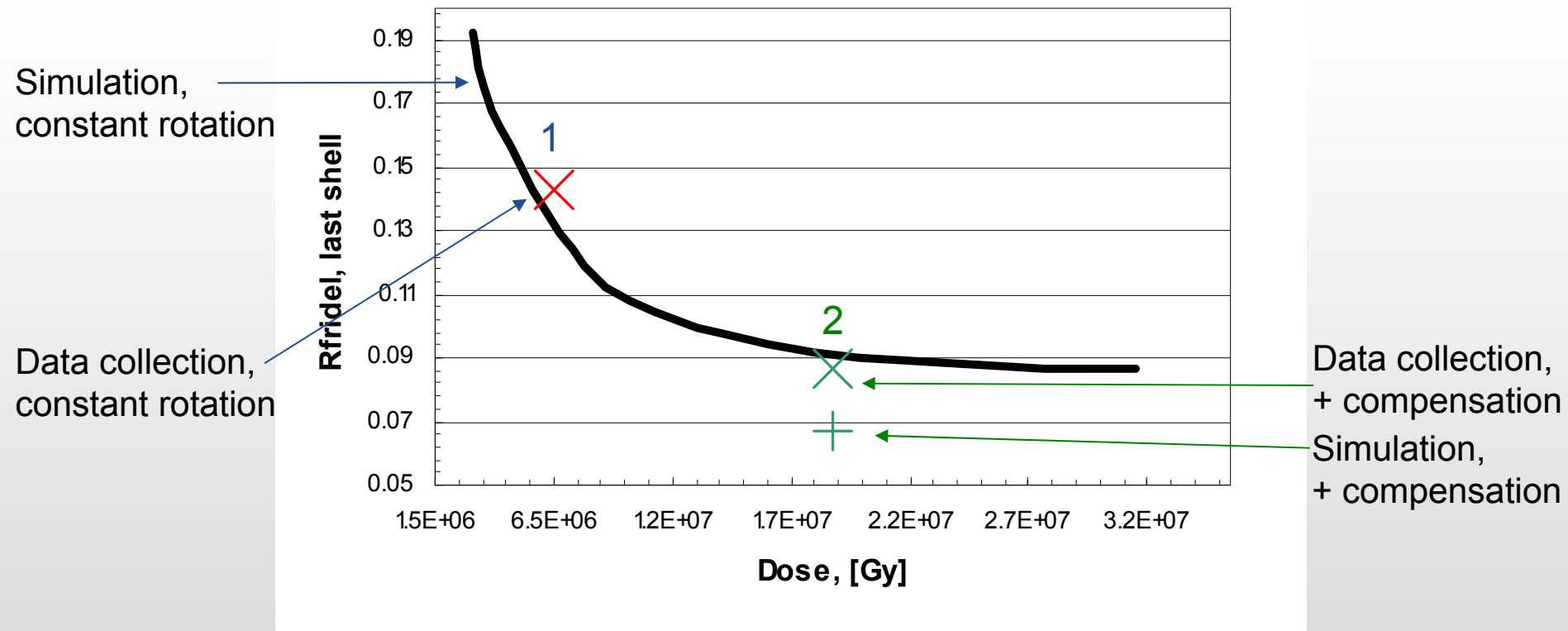


SAD optimization

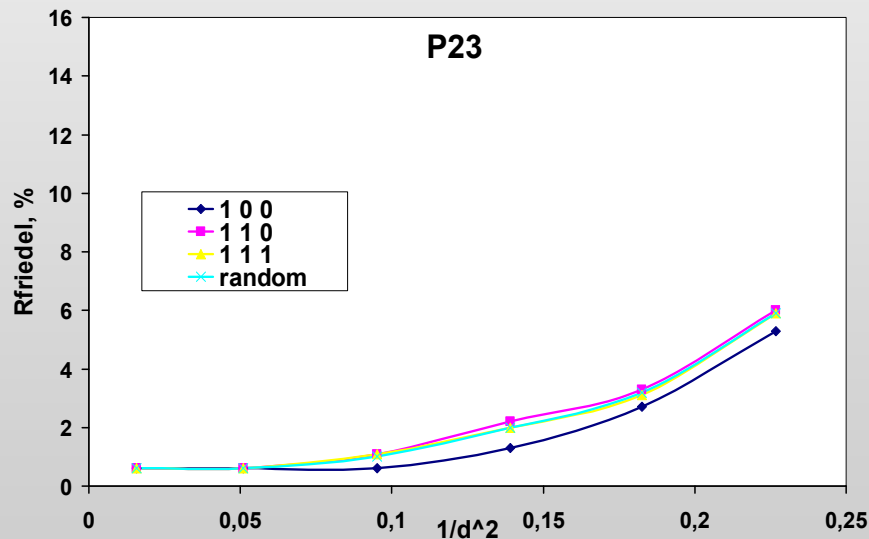
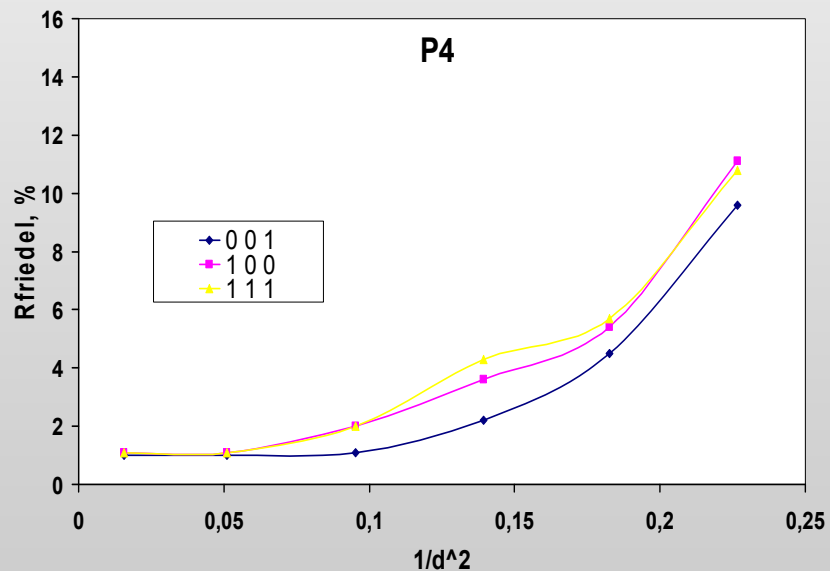
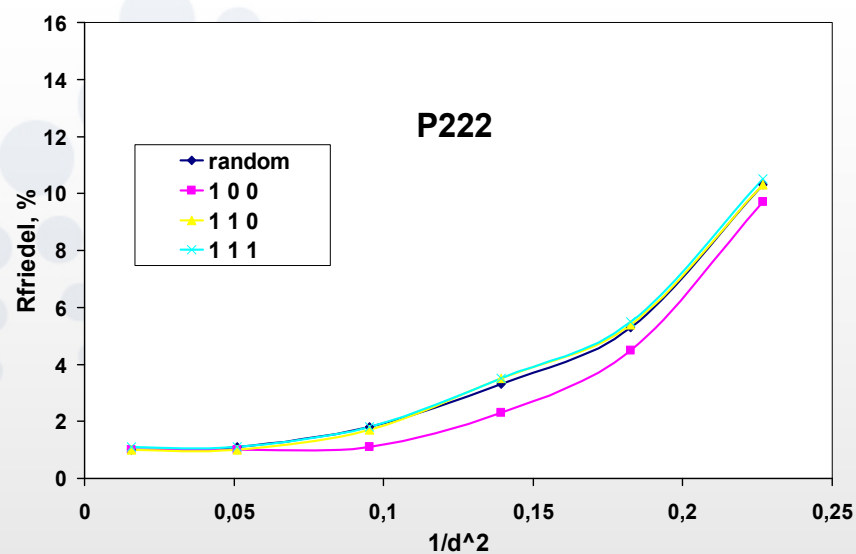
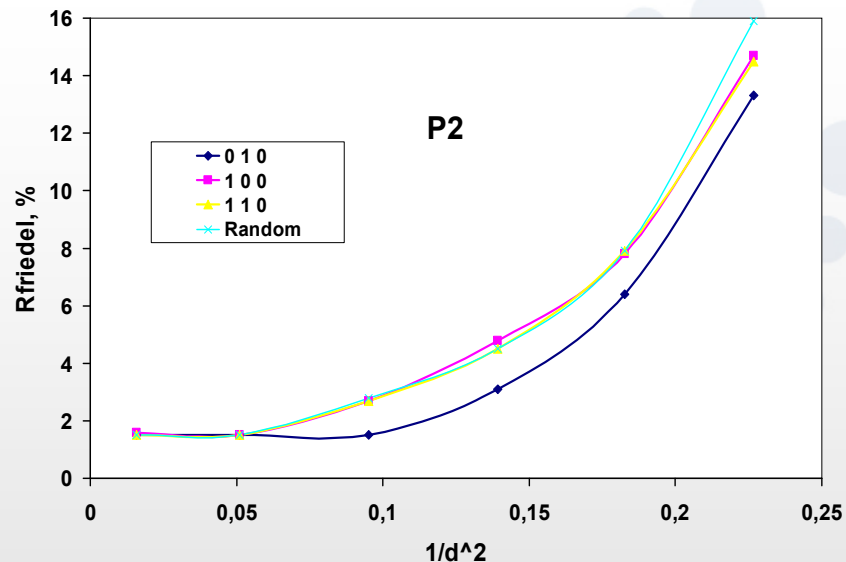
Minimum of $R_{\text{Friedel}} = \langle |\langle E^2+ \rangle - \langle E^2- \rangle| \rangle$ is a target
Dose, [Gy]



Testing: Dose dependency/Decay compensation



Data set	1	2
Correlation $ DANO , FH _{calc}$		
last shell, 3.3Å	0.09	0.17
overall, 20-3.3Å	0.26	0.38



```
>edna-prototype --flux 1.0e12 --beamSize 0.04 --images ref-my*img
```

```
-----
Indexing: MOSFLM : Version 7.0.1 for Image plate and CCD data 20th August 2007
```

```
..
```

```
Integration : MOSFLM :-----
```

```
..
```

```
Strategy : Summary of Strategy:
```

```
Strategy : -----
```

```
Strategy : Raddose : version 4.2 // 2007/07/05
```

```
..
```

```
Strategy : Raddose : Dose rate : 1.7e+05 [Grays/s]
```

```
Strategy :
```

```
Strategy : -----
```

```
Strategy : Best : Version 3.1.0.d // 16.07.2007
```

```
Strategy : Best :
```

```
Strategy : Best : Plan for data collection:
```

Strategy : Best :	N	Phi_start	N.of.images	Rot.width	Exposure	Distance	Overlap
Strategy : Best :	1	96.00	80	0.50	0.25	342.7	No
Strategy : Best :	2	130.00	40	0.50	0.45	342.7	No
Strategy : Best :	3	156.00	26	0.40	0.72	342.7	No

```
Strategy : Best :
```

```
Strategy : Best : Attenuation : 1.0
```

```
Strategy : Best : Total rotation range : 70.4 [degrees]
```

```
Strategy : Best : Total number of images : 146
```

```
Strategy : Best : Total exposure time : 56.6 [s]
```

```
Strategy : Best : Total data collection time : 421.6 [s]
```

```
Strategy : Best : Resolution : 2.2 [A]
```

```
Strategy : Best : Resolution reasoning : Resolution limit is set by the radiation damage
```

```
Strategy : Best : Ranking resolution : 2.2 [A]
```

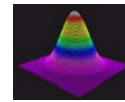
Strategy : Best : Predicted statistics:

Strategy : Best :	Resolution	Compl.	Average		I/Sigma	I/Sigma	Chi**2	R-fact	Overload	
Strategy : Best :	Lower Upper	%	Intensity	Sigma		/Chi		%	%	
Strategy : Best :	-----									
Strategy : Best :	12.00	7.82	95.0	36464.7	735.5	49.6	46.8	1.12	2.0	0.00
Strategy : Best :	7.82	6.23	97.0	18329.1	382.5	47.9	41.5	1.33	2.4	0.00
Strategy : Best :	6.23	5.33	98.0	14530.4	315.8	46.0	36.7	1.58	2.7	0.00
..										
Strategy : Best :	2.45	2.38	99.0	617.8	179.1	3.4	3.1	1.20	24.1	0.00
Strategy : Best :	2.38	2.32	99.0	519.1	183.9	2.8	2.6	1.15	28.6	0.00
Strategy : Best :	2.32	2.26	99.0	440.1	189.1	2.3	2.2	1.11	33.9	0.00
Strategy : Best :	2.26	2.21	99.0	365.2	194.7	1.9	1.8	1.09	40.7	0.00
Strategy : Best :	12.00	2.21	99.0	5684.9	229.8	24.7	17.7	1.96	5.4	0.00

XDS :

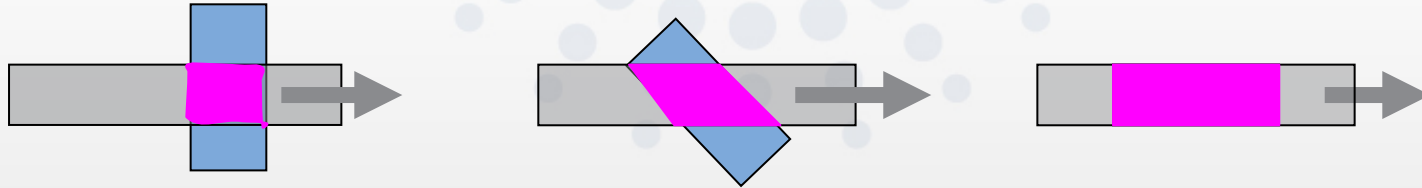
SUBSET OF INTENSITY DATA WITH SIGNAL/NOISE >= -3.0 AS FUNCTION OF RESOLUTION

RESOLUTION	NUMBER	OF REFLECTIONS		COMPLETENESS	R-FACTOR	R-FACTOR	COMPARED	I/SIGMA	R-meas
LIMIT	OBSERVED	UNIQUE	POSSIBLE	OF DATA	observed	expected			
10.00	502	240	263	91.3%	2.0%	3.0%	425	31.25	2.6%
9.00	207	90	93	96.8%	2.2%	3.0%	176	32.01	2.7%
8.00	323	149	150	99.3%	2.1%	3.2%	278	30.60	2.6%
7.00	516	239	241	99.2%	2.0%	3.3%	439	28.53	2.6%
6.00	937	419	428	97.9%	2.6%	3.4%	802	27.63	3.2%
..									
2.50	10249	4617	4639	99.5%	14.9%	15.0%	8665	4.39	18.5%
2.30	10146	4598	4622	99.5%	25.9%	26.3%	8580	2.56	32.4%
2.20	6085	2848	2913	97.8%	31.6%	31.9%	4984	1.97	39.4%
total	51356	23400	23682	98.8%	4.9%	5.4%	43259	10.39	6.0%

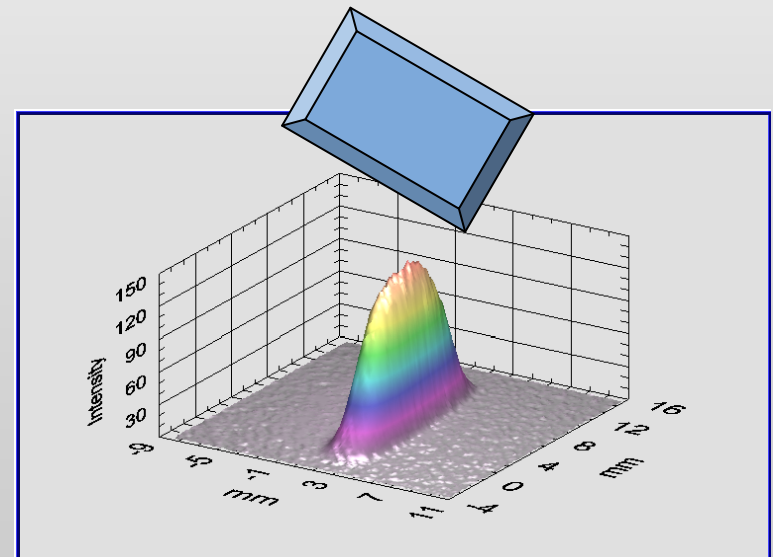
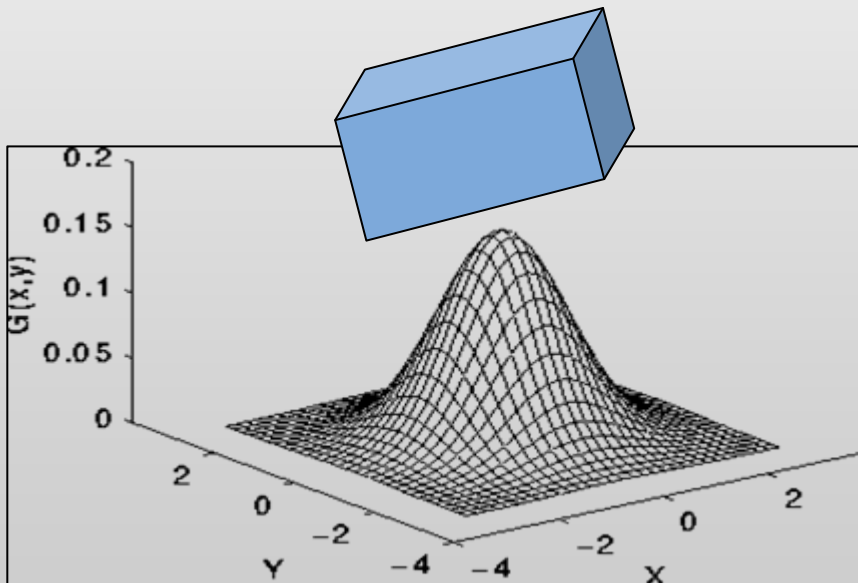


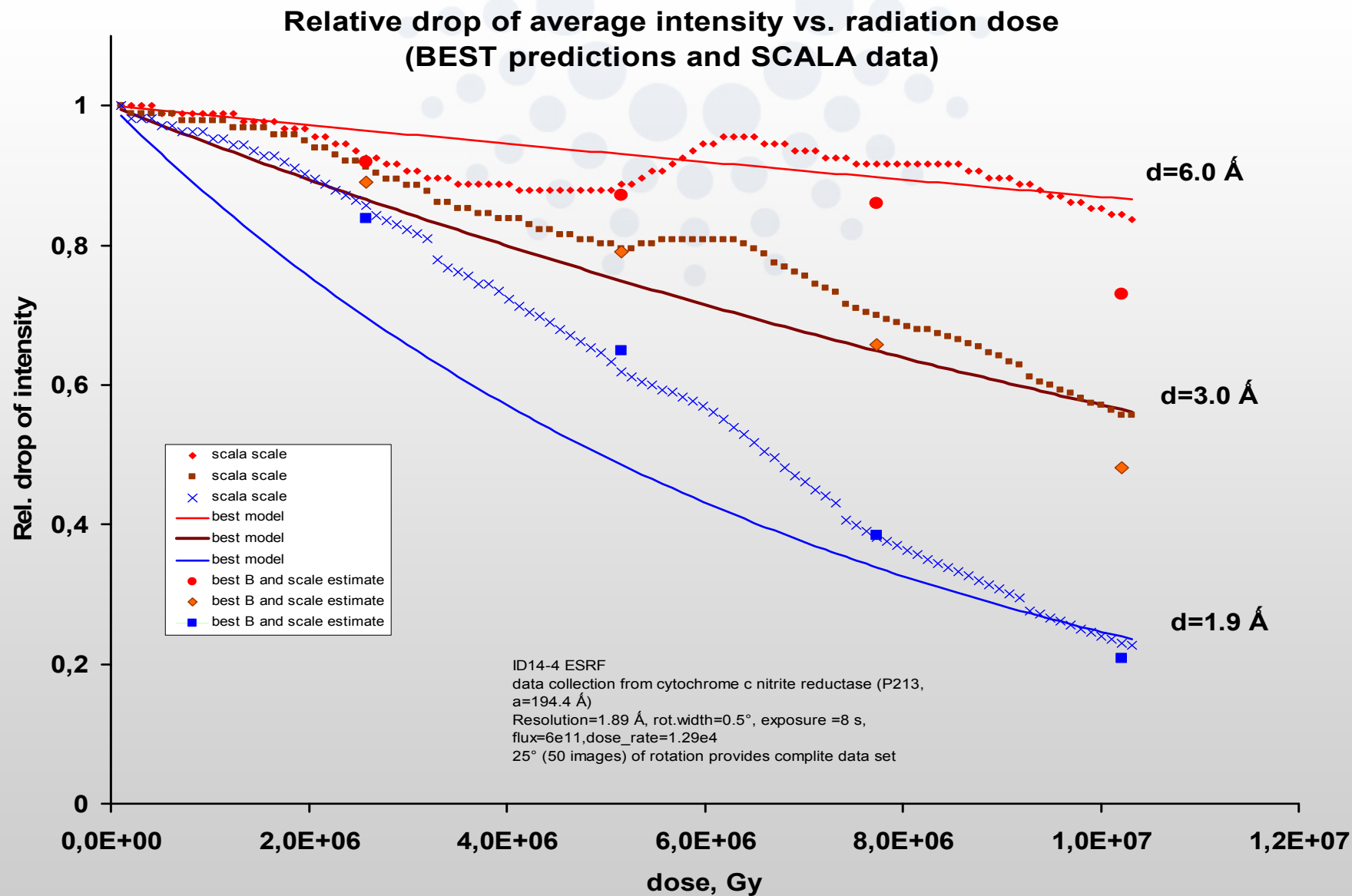
1. sample > beam

Irradiation crystal volume VS Phi ????



2. Beam profile and crystal shape?

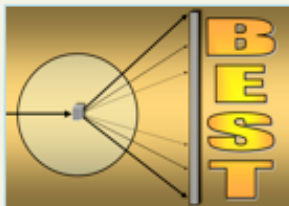




Gleb Bourenkov

- ESRF Beamlines : Sean McSweeney, Raimond Ravelli
- RADDOSE: Elspeth Garman
- Kappa workgroup
Johan Unge (EMBL-HH – MaxLab, Lund)
Sandor Brokhauser (EMBL-Grenoble)
Peter Keller, Gerard Bricogne (Global Phasing)
- EDNA Collaboration (www.edna-site.org)
- BEST users, all those who gives samples to us
- BIOXHIT before 05.08

Software available for download at:
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