

Radiation Damage Analysis with

$$R_{\text{cp}}(d)$$

ACA 2012

Graeme Winter

Diamond Light Source

July 2012



Overview

- Strategy, background, caveats
- Review of statistics
- Example cases
- New statistic - $R_{cp}(d)$
- More example cases
- Discussion, acknowledgements

Caveats

- All data described processed with xia2 / XDS i.e. using “standard methods”
- All this will show is a new tool and some instances where it could be useful
- All calculations independent of data analysis program used
- Program to perform these calculations included with xia2: pychef
- Discussion assumes sample bathed in beam

Strategy

- Radiation damage gives rise to lots of changes
- Will only discuss changes to measured *intensities* i.e. not
 - Changes to unit cell
 - Sample *B*-factor
 - Vanishing diffraction
- Assume sufficient data for scaling, analyse after corrections applied
- Make no assumptions about scaling *program*
- Looking to answer the question: would this data set have been more useful if I stopped collecting data earlier - balancing:
 - Gains from additional measurements i.e. multiplicity
 - Losses due to systematic changes reducing signal
- Assume sufficient multiplicity that above question is meaningful c.f. zero-dose

$$R_{\text{merge}} = \frac{\sum_{\underline{h}} \sum_j |I_{\underline{h}j} - \bar{I}_{\underline{h}}|}{\sum_{\underline{h}} \sum_j I_{\underline{h}j}} \quad (1)$$

Measures: how well reflections on frame j agree with the average values of those reflections.



$$R_d = \frac{\sum_{\underline{h}} \sum_{|b_j - b_i| = d} |I_{\underline{h}j} - I_{\underline{h}i}|}{\sum_{\underline{h}} \sum_{|b_j - b_i| = d} \frac{1}{2} |I_{\underline{h}j} + I_{\underline{h}i}|} \quad (2)$$

Measures: how well reflections separated by d frames agree.



Application to Example Sets

- Straightforward thaumatin example
- Radiation-damaged SAD
- Radiation-damaged MAD

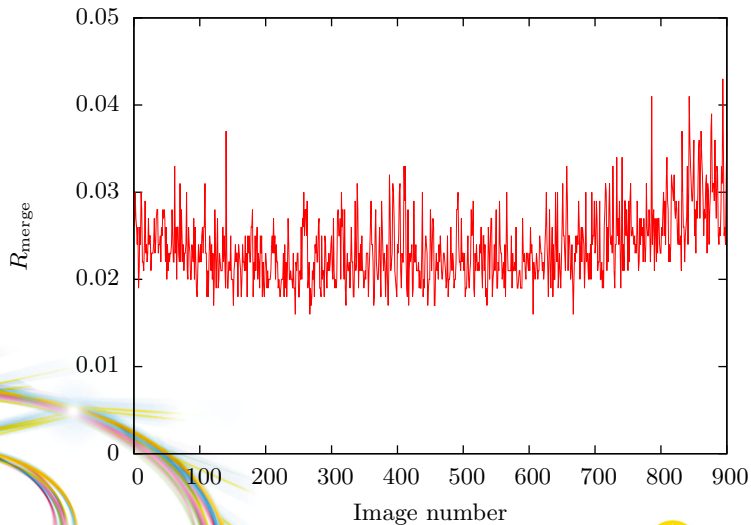


Example 1: thaumatin

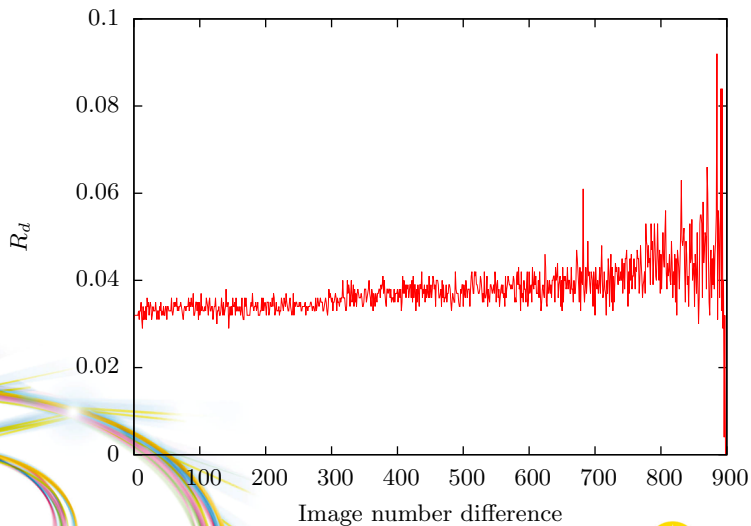
- A dull, native, good-quality sample
- Recorded during beamline setup at Diamond I03
- 900 frames $\times 0.1^\circ$
- Purpose: illustrate properties of existing residuals



Thaumatoin: R_{merge}



Thaumatoin: R_d

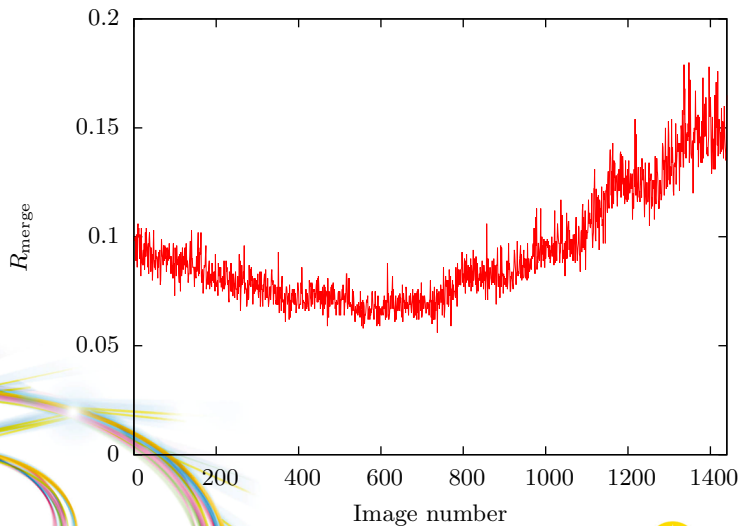


Example 2: radiation damaged SAD data

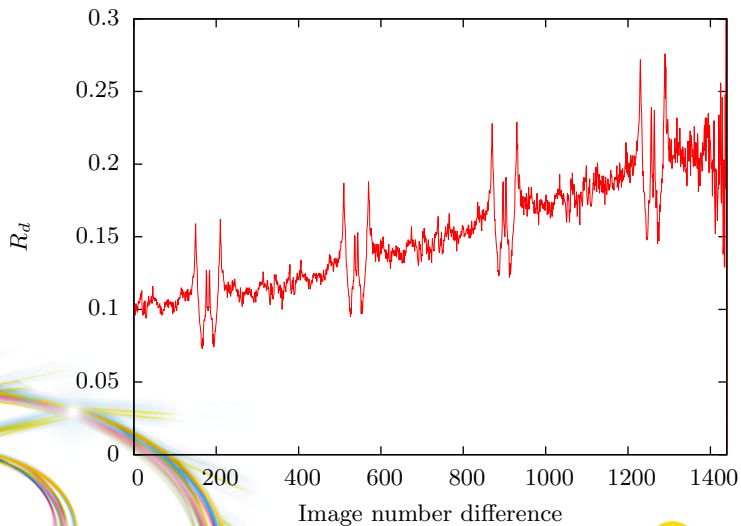
- High multiplicity SAD data set recorded from *Helix pomatia* Agglutinin at ESRF beamline ID14EH2 by Ed Mitchell as part of ongoing research
- 1440° of data, of which 720° were used in the structure solution due to radiation damage
- Symmetry is $H32$, so the data have ~ 80 -fold multiplicity



HPA: R_{merge}



HPA: R_d

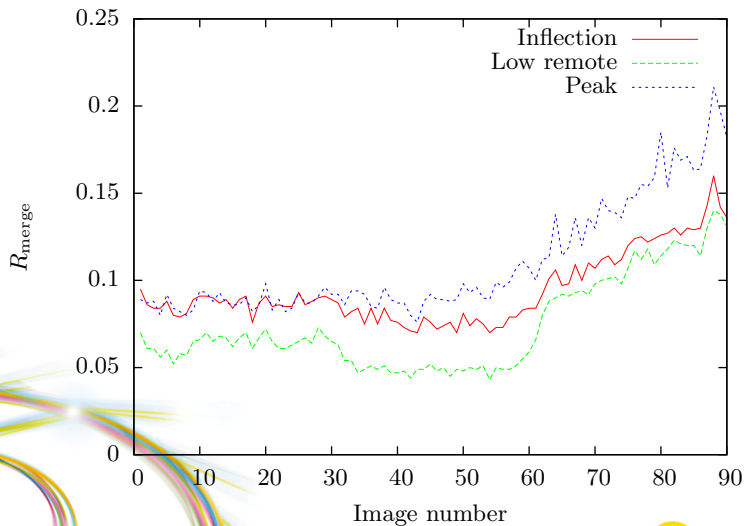


Example 3: radiation damaged MAD data

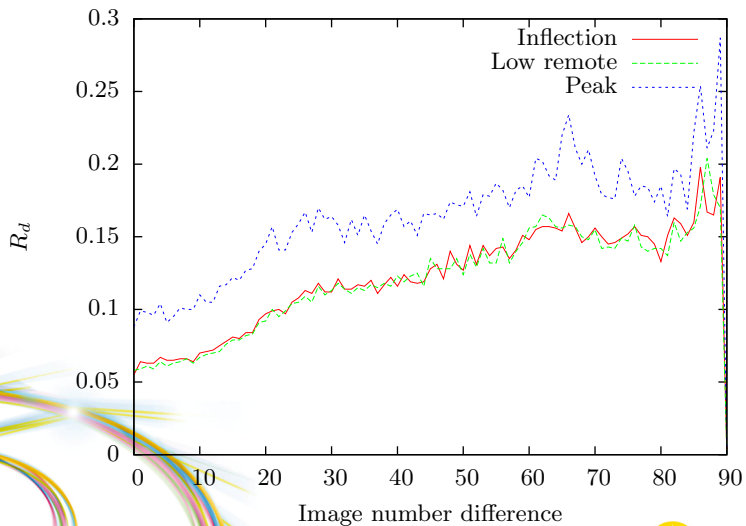
- JCSG target TB0541B
- 8 Se / 200 AA, 3 λ MAD with inverse beam on inflection and low energy remote
- Classic example of nasty radiation damage



2ISB: R_{merge}



2ISB: R_d



Review

- Stats behave as expected - presence or absence of radiation damage is clear
- No information provided so far about subset options i.e. when I clearly do have radiation damage, what to do?
- MAD data sets not considered very gracefully



Novel Statistic

- Question: could things have been better if I had stopped collecting data earlier? I.e. think about wall-clock *time*
- R_{merge} contains time information, R_d shows difference information, wouldn't it be great to have both?
- Including how *complete* the data are would also be important



$R_{cp}(d)$, Cumulative-Pairwise Residual

$$R_{cp}(d) = \frac{\sum_{\lambda} \sum_{\underline{h}} \sum_{i \neq j, i: d_i \leq d, j: d_j \leq d} |I_{\underline{h}i} - I_{\underline{h}j}|}{\sum_{\lambda} \sum_{\underline{h}} \sum_{i \neq j, i: d_i \leq d, j: d_j \leq d} \frac{1}{2} |I_{\underline{h}i} + I_{\underline{h}j}|} \quad (3)$$

Measures: including data to a point d , how internally consistent are they?

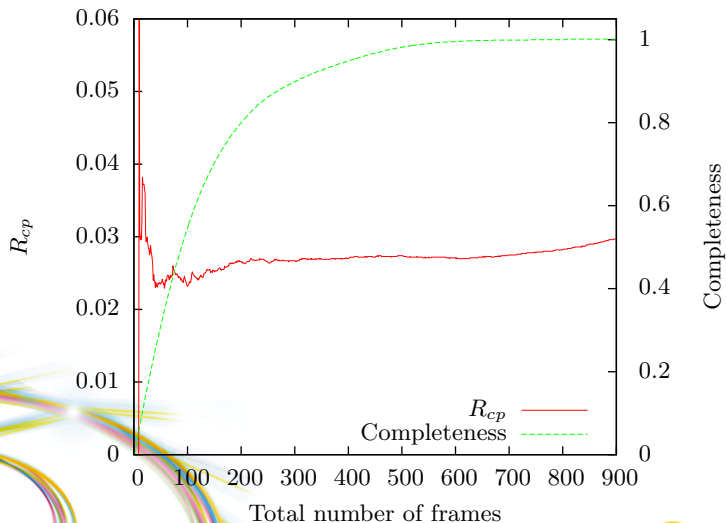


In English?

- For each wavelength, for each hkl , accumulate generally how different the intensities are up to some dose d
- Then plot this as a function of d
- And while you're there record how complete wavelengths are up to point d too...



Application to Thaumatin

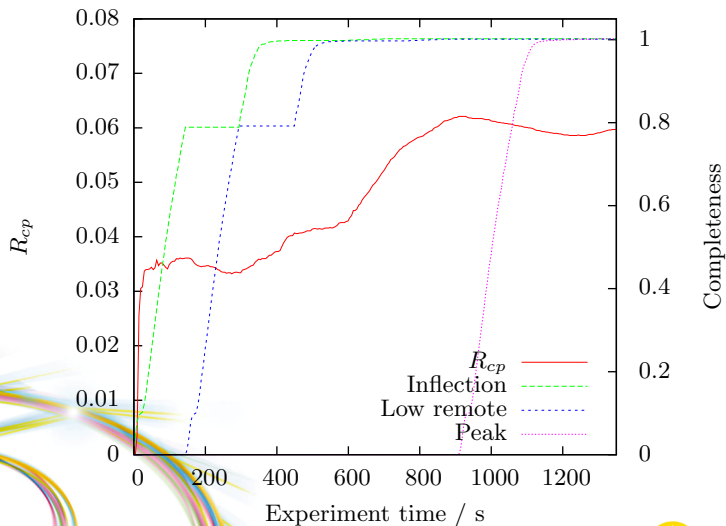


Comments and observations

- At the start always 'noisy' as there are few pairs
- Most interesting is what happens after the data are fairly complete - in this case, nothing
- Even more interesting is to use *dose* as the baseline rather than batch, though equivalent for SAD



Application to MAD data

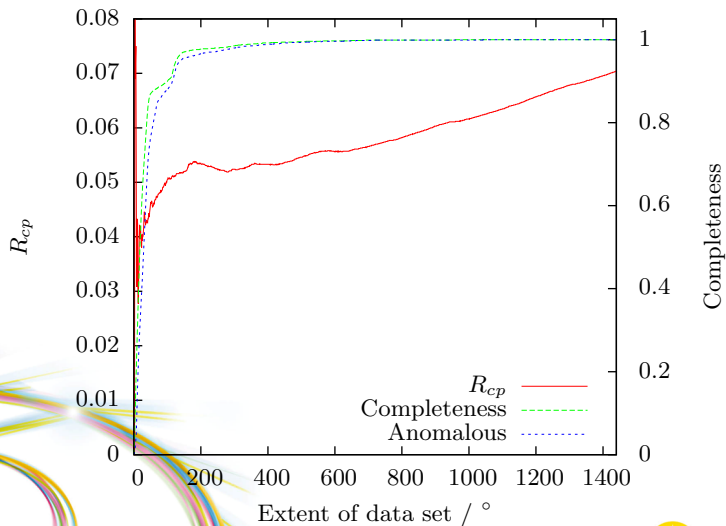


Interpretation

- Full three wavelength data has high residual (i.e. after 1350s)
- After 600s we have two complete wavelengths, much lower residual



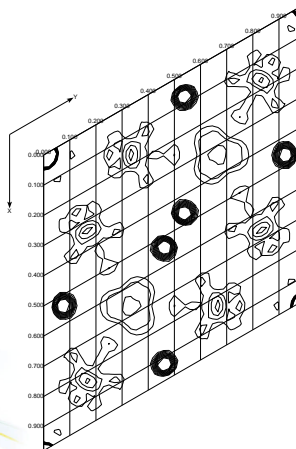
Application to HPA



Interpretation

- Time equivalent to number of frames
- Complete data (anomalous, native) pretty early on
- R_{cp} appears to increase systematically ~ 400 images - just over one full rotation
- One Zn site on special position - anomalous difference Patterson easy to interpret
- Processed data set 32 times - $0^\circ - 45^\circ$, $0^\circ - 90^\circ$, ..., $0^\circ - 1440^\circ$ to fixed resolution limit (1.65\AA), used 30\AA to 3\AA to compute Patterson

HPA Anomalous Difference Patterson

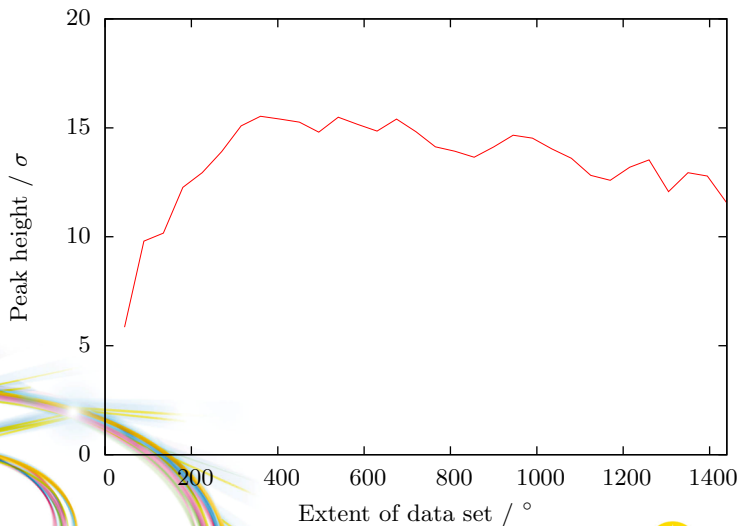


Scale = 2.500000 mme/Å

Map from fit

Section 0 (0.000) on z axis

Peak Height / σ



Interpretation

- After $\sim 360^\circ$ radiation damage becomes measurable
- Complete subset around here gives “most internally consistent” representation of the sample
- Peak in anomalous difference Patterson using this much data
- N.B. phasing works with subset, full data set, just using SHELX C/D/E - George have you fiddled the methods? This didn't used to work!
- Zero-dose gave peak height of 16.07σ - slightly better than the 360° set but with $4\times$ the measurements used

Summary to date

- Looked at some examples
- Showed new statistic
- Illustrated statistic may be useful, perhaps more useful than R_{merge} vs. batch and complementary to R_d
- No cases shown where the insoluble becomes soluble - let's fix that

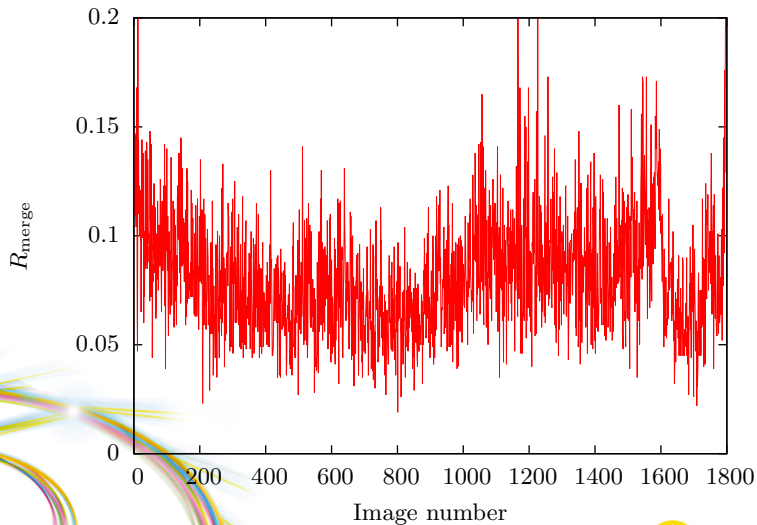
Unpublished Example - Au soak SAD

- Example from Christian Siebold at STRUBI
- Measured at Diamond Light Source I03 using Pilatus 6M
- Native data available, issue with this data set is substructure determination for phasing

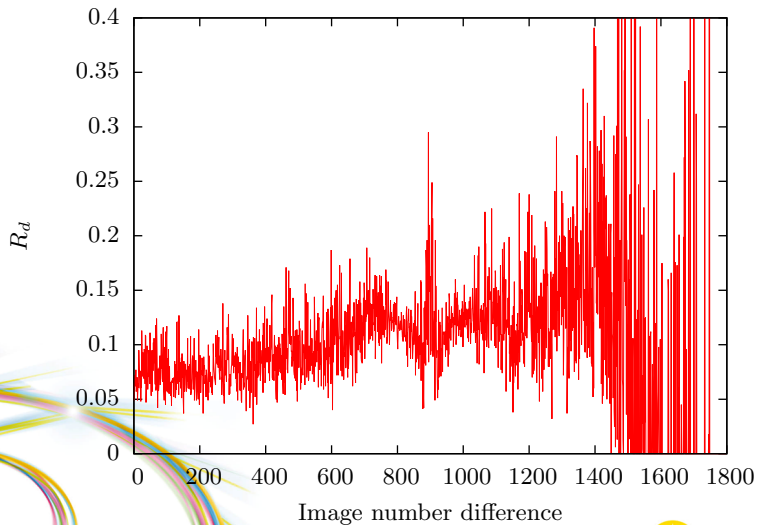


High resolution limit	2.01	8.98	2.01
Low resolution limit	29.85	29.84	2.06
Completeness	71.2	97.3	13.5
Multiplicity	5.3	6.0	2.1
I/sigma	14.7	21.6	2.3
Rmerge	0.063	0.043	0.405
Rmeas(I)	0.086	0.067	0.707
Rmeas(I+/-)	0.076	0.052	0.568
Rpim(I)	0.034	0.027	0.470
Rpim(I+/-)	0.041	0.028	0.397
Wilson B factor	33.264	.	.
Partial bias	0.000	0.000	0.000
Anomalous completeness	65.1	97.6	10.6
Anomalous multiplicity	2.9	3.3	1.1
Anomalous correlation	0.388	0.613	0.690
Anomalous slope	1.589	0.000	0.000
Total observations	40714.	775.	228.
Total unique	7685.	130.	111.

Au soak: R_{merge}



Au soak: R_d

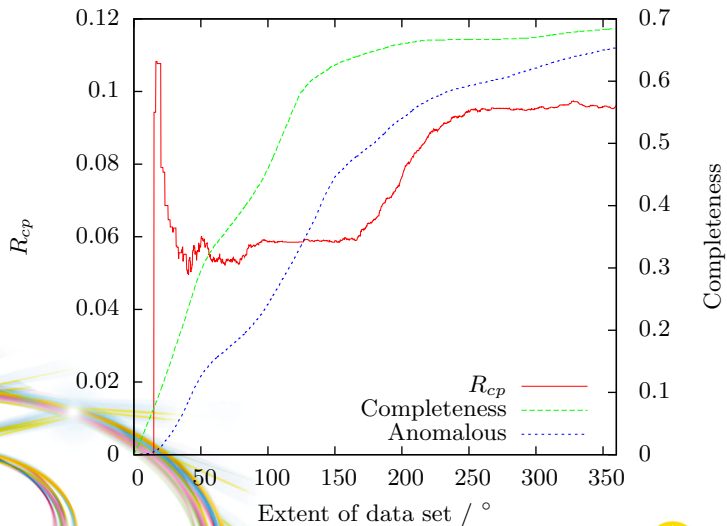


Interpretation

- Detector too far back - low completeness at high resolution
- Merging stats reasonable, clearly we have some radiation damage
- What does R_{cp} say?



Au soak: R_{cp}



Clear radiation damage

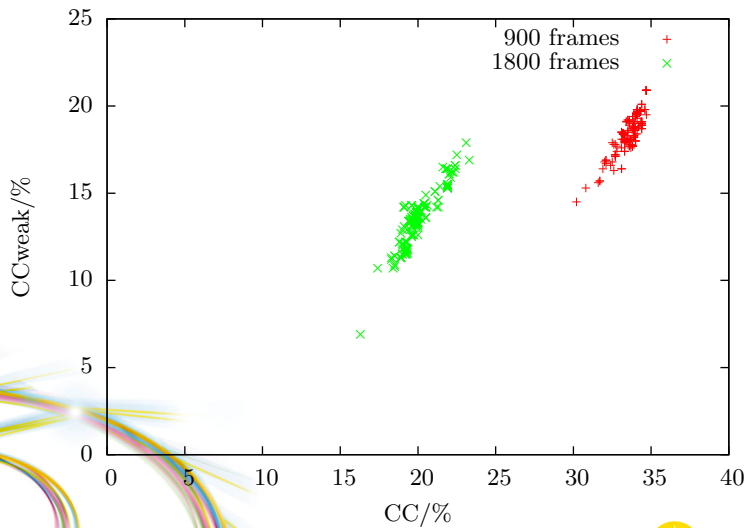
- Clear radiation damage according to this measure
- Data essentially as complete as they can be half way through
- Assert: using 900 frames of the data will work better than all 1800
- Test: use SHELXD (via “Fast EP” at Diamond)



High resolution limit	2.01	8.98	2.01
Low resolution limit	29.83	29.83	2.06
Completeness	70.9	95.1	12.5
Multiplicity	2.9	3.1	1.2
I/sigma	16.0	31.5	1.4
Rmerge	0.036	0.022	0.348
Rmeas(I)	0.063	0.051	0.703
Rmeas(I+/-)	0.049	0.030	0.493
Rpim(I)	0.035	0.027	0.497
Rpim(I+/-)	0.034	0.021	0.348
Wilson B factor	32.688	.	.
Partial bias	0.000	0.000	0.000
Anomalous completeness	55.7	86.7	2.0
Anomalous multiplicity	1.6	1.8	1.0
Anomalous correlation	0.609	0.835	0.000
Anomalous slope	1.312	0.000	0.000
Total observations	21812.	385.	123.
Total unique	7646.	126.	101.

High resolution limit	2.01	8.98	2.01
Low resolution limit	29.85	29.84	2.06
Completeness	71.2	97.3	13.5
Multiplicity	5.3	6.0	2.1
I/sigma	14.7	21.6	2.3
Rmerge	0.063	0.043	0.405
Rmeas(I)	0.086	0.067	0.707
Rmeas(I+/-)	0.076	0.052	0.568
Rpim(I)	0.034	0.027	0.470
Rpim(I+/-)	0.041	0.028	0.397
Wilson B factor	33.264	.	.
Partial bias	0.000	0.000	0.000
Anomalous completeness	65.1	97.6	10.6
Anomalous multiplicity	2.9	3.3	1.1
Anomalous correlation	0.388	0.613	0.690
Anomalous slope	1.589	0.000	0.000
Total observations	40714.	775.	228.
Total unique	7685.	130.	111.

SHELXD CC, CCweak



Discussion

- Subset of data currently being used with native for structure solution
- “Good” solutions from SHELXD were verified to be correct
- This tool possibly most useful with shutterless data collection
- Anisotropic diffraction: typically related to symmetry, symmetry used in calculation so not sure what effects are likely to be...
- For reference: used zero-dose extrapolation in XSCALE, get 27.23 / 13.57 for CC / CCweak

Conclusions

- Makes no changes to the intensity values - only asks the question “what happens if I stopped collecting earlier?”
- Makes no model of the radiation damage
- Works with your existing software in a graceful way



Acknowledgements

- Miroslav Papiz, Steve Prince for conversation which got this started, inspiration for the name *chef*
- Kay Diederichs for R_d , as well as 0-dose with Sean McSweeney and Raimond Ravelli
- Phil Evans for useful discussions on this subject
- Ed Mitchell, Diamond Staff, JCSG and Christian Siebold for the examples shown