Using *XDS* - a practical perspective

Kay Diederichs

Universität Konstanz

Protein Crystallography /
Molecular Bioinformatics

Overview

- *XDS* is a data reduction program for X-ray data collected by the oscillation method on area detectors
- author: Wolfgang Kabsch (MPI Heidelberg)
- Basics: information flow within *XDS*
- Advanced: identifying systematic errors by „mining“ information from *XDS* output file XDS_ASCII.HKL (with examples)
- Current and future work: radiation damage; simulation
- XDSwiki; references

The *XDS* program suite

binary distribution (by W. Kabsch) for Linux & Mac:

- *XDS*: the main program (indexing, integrating, scaling)
- *XSCALE*: scale several *XDS* intensity data sets together, statistics
- *XDSCONV*: convert to *CCP4*, *CNS*, *SHELX*, ... format

source code available from sourceforge.net:

- *XDS-Viewer* : inspect control images written by *XDS*, or (single) data frames (alternatively, latest *adxv* may be used)

my own programs:

- *XDSSTAT*, *generate_adx* (both in XDSwiki)

How to use *XDS* ?

- Prepare a single input file XDS.INP with parameters describing data reduction
- XDS.INP often written by beamline software
- Parameters and their keywords have the form e.g. DETECTOR_DISTANCE= 120.
- There are about 30 relevant parameters, but only about 15 are required (and change between projects). All parameters have reasonable defaults where possible.

Example for MarCCD 225 @ SLS PX-III

```
JOB= XYCORR INIT COLSPOT IDXREF DEFPIX INTEGRATE CORRECT
ORGX=1546 ORGY=1552        !Detector origin (pixels); e.g. NX/2 NY/2
DETECTOR_DISTANCE=180      !(mm)
OSCILLATION_RANGE=0.50     !degrees (>0)
X-RAY_WAVELENGTH=0.980243 !Angstroem
NAME_TEMPLATE_OF_DATA_FRAMES=frms/wga2-27_1_???.img
DATA_RANGE=1 360           !Numbers of first and last data image collected
BACKGROUND_RANGE=1 10      !Numbers of first and last data image for background
SPACE_GROUP_NUMBER=  19    !0 for unknown crystals; cell constants are ignored.
UNIT_CELL_CONSTANTS= 44.4   86.4   104.5  90 90  90
REFINE(IDXREF)=BEAM AXIS ORIENTATION CELL DISTANCE
REFINE(INTEGRATE)=DISTANCE BEAM ORIENTATION CELL ! AXIS
ROTATION_AXIS= 1.0 0.0 0.0
INCIDENT_BEAM_DIRECTION=0.0 0.0 1.0
FRACTION_OF_POLARIZATION=0.99                    ! SLS X06SA
POLARIZATION_PLANE_NORMAL= 0.0 1.0 0.0
DETECTOR=CCDCHESS        MINIMUM_VALID_PIXEL_VALUE=1     OVERLOAD=65000
DIRECTION_OF_DETECTOR_X-AXIS= 1.0 0.0 0.0
DIRECTION_OF_DETECTOR_Y-AXIS= 0.0 1.0 0.0
VALUE_RANGE_FOR_TRUSTED_DETECTOR_PIXELS= 7000 30000 !Used by DEFPIX
                                  !for excluding shaded parts of the detector.
INCLUDE_RESOLUTION_RANGE=50.0 1.3 !Angstroem; used by DEFPIX,INTEGRATE,CORRECT
```

Bold keyword/parameter pairs are required. Complete documentation at http://xds.mpimf-heidelberg.mpg.de/html_doc/xds_parameters.html

Templates for many detectors at http://xds.mpimf-heidelberg.mpg.de/html_doc/INPUT_templates

Using *XDS* - principles I

- *simple, if basic idea is understood*
- There is one JOB= line in XDS.INP which does not specify a parameter, but instead a list of tasks:

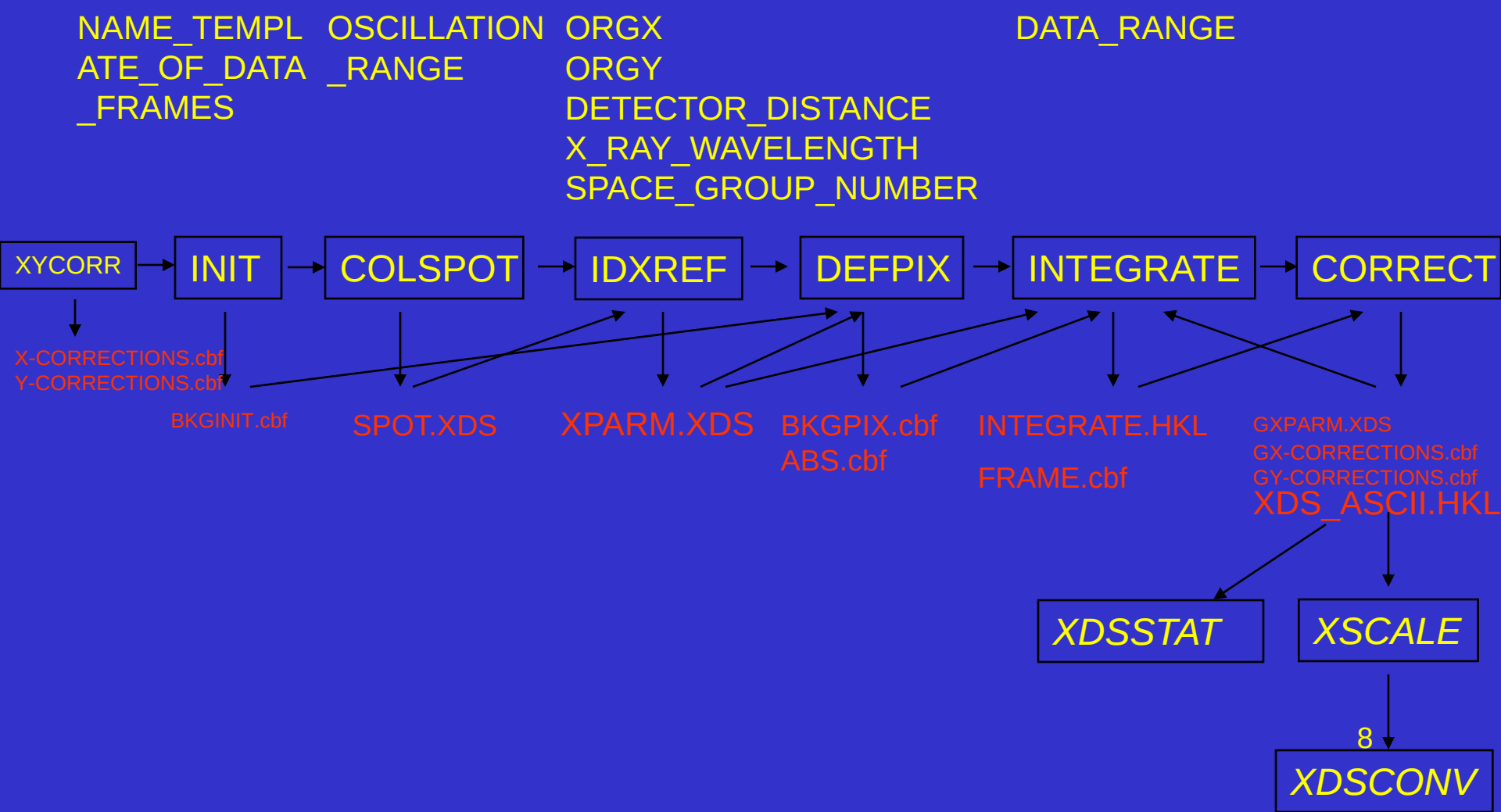
```
JOB= XYCORR INIT COLSPOT IDXREF DEFPIX INTEGRATE CORRECT
```

- data reduction is divided into tasks
- information storage/exchange/flow between tasks by data files
- each task needs the result from the previous tasks
- fine-tuning of a task does *not* require previous tasks to be repeated

Using *XDS* - principles II

- XYCORR : write positional correction files (X-CORRECTIONS.cbf, Y-CORRECTIONS.cbf)
- INIT : find background pixels (defaults usually OK)
- COLSPOT: find reflection positions
- IDXREF : "index" reflections; user may supply/choose spacegroup
- (XPLAN : [not required] completeness as function of ...)
- DEFPIX : find beamstop shadow (defaults mostly OK)
- INTEGRATE : evaluates intensities on all frames, writes INTEGRATE.HKL and FRAME.cbf
- CORRECT : scales, rejects outliers, statistics, writes XDS_ASCII.HKL (and other files)

Information flow



XDS output file:

INTEGRATE.HKL

```
!OUTPUT_FILE=INTEGRATE.HKL    DATE= 3-Oct-2006
!Generated by INTEGRATE    (XDS VERSION  August 18, 2006)
!PROFILE_FITTING= TRUE
!SPACE_GROUP_NUMBER=   92
!UNIT_CELL_CONSTANTS=    57.69    57.69   150.03  90.000  90.000  90.000
!NAME_TEMPLATE_OF_DATA_FRAMES= ../series_2_????.img
!DETECTOR=ADSC
!NX=  3072  NY=  3072    QX=  0.102600  QY=  0.102600
!STARTING_FRAME=       1
!STARTING_ANGLE=    30.000
!OSCILLATION_RANGE=  0.500000
!ROTATION_AXIS=  0.999995  0.002515 -0.001722
!X-RAY_WAVELENGTH=  0.939010
!INCIDENT_BEAM_DIRECTION=  0.001723 -0.002233  1.064948
!DIRECTION_OF_DETECTOR_X-AXIS=  1.000000  0.000000  0.000000
!DIRECTION_OF_DETECTOR_Y-AXIS=  0.000000  1.000000  0.000000
!ORGX=   1541.53  ORGY=   1535.28
!DETECTOR_DISTANCE=   189.221
!UNIT_CELL_A-AXIS=   -11.482    53.781   -17.431
!UNIT_CELL_B-AXIS=   -17.974   -20.337   -50.906
!UNIT_CELL_C-AXIS=  -139.398   -12.226    54.103
!BEAM_DIVERGENCE_E.S.D.=     0.037
!REFLECTING_RANGE_E.S.D.=     0.113
!NUMBER_OF_ITEMS_IN_EACH_DATA_RECORD=20
!H,K,L,IOBS,SIGMA,XCAL,YCAL,ZCAL,RLP,PEAK,CORR,MAXC,
!             XOBS,YOBS,ZOBS,ALF0,BET0,ALF1,BET1,PSI
!Items are separated by a blank and can be read in free-format
!END_OF_HEADER
 -45 -9 -60 -3.755E+01 4.144E+01 3066.2 3053.3 273.5 0.75268 100 -10 46 0.0 0.0 0.0 -49.52 0.16 44.87 49.40 -29.89
 -45 -9 -59 8.133E+00 4.372E+01 3044.3 3056.1 274.5 0.75525 100 10 46 0.0 0.0 0.0 -49.52 0.16 45.34 49.22 -29.95
 -45 -8 -60 6.502E+01 4.327E+01 3046.6 3054.5 271.3 0.75438 100 14 47 3051.0 3057.7 272.0 -49.52 0.16 45.26 49.23 -30.66
............
```

XDS output file: XDS_ASCII.HKL

```
!FORMAT=XDS_ASCII    MERGE=FALSE    FRIEDEL'S_LAW=TRUE
!OUTPUT_FILE=XDS_ASCII.HKL        DATE= 3-Oct-2006
!Generated by CORRECT   (XDS VERSION  August 18, 2006)
!PROFILE_FITTING= TRUE
!SPACE_GROUP_NUMBER=   92
!UNIT_CELL_CONSTANTS=    57.71    57.71   150.08  90.000  90.000  90.000
!NAME_TEMPLATE_OF_DATA_FRAMES= ../series_2_????.img
!DATA_RANGE=       1     399
!X-RAY_WAVELENGTH=  0.939010
!INCIDENT_BEAM_DIRECTION=  0.001872 -0.002230  1.064947
!FRACTION_OF_POLARIZATION=   0.980
!POLARIZATION_PLANE_NORMAL=  0.000000  1.000000  0.000000
!ROTATION_AXIS=  0.999995  0.002477 -0.001917
!OSCILLATION_RANGE=  0.500000
!STARTING_ANGLE=    30.000
!STARTING_FRAME=       1
!DETECTOR=ADSC
!DIRECTION_OF_DETECTOR_X-AXIS=   1.00000   0.00000   0.00000
!DIRECTION_OF_DETECTOR_Y-AXIS=   0.00000   1.00000   0.00000
!DETECTOR_DISTANCE=   189.286
!ORGX=   1541.25  ORGY=   1535.30
!NX=  3072  NY=  3072    QX=  0.102600  QY=  0.102600
!NUMBER_OF_ITEMS_IN_EACH_DATA_RECORD=12
!ITEM_H=1
!ITEM_K=2
!ITEM_L=3
!ITEM_IOBS=4
!ITEM_SIGMA(IOBS)=5
!ITEM_XD=6
!ITEM_YD=7
!ITEM_ZD=8
!ITEM_RLP=9
!ITEM_PEAK=10
!ITEM_CORR=11
!ITEM_PSI=12
!END_OF_HEADER
     0     0     4  4.287E-01  2.814E-01  1501.6  1514.4     99.4 0.00920 100  27   75.39
     0     0    -4  2.243E-01  2.386E-01  1587.4  1548.6     91.6 0.00920 100  30  -79.02
     0     0     5  5.976E-03  3.443E-01  1490.9  1510.2    100.4 0.01150 100  22   74.94
```

XDS : feedback of information from later steps to previous steps (postrefinement)

- To optimize data quality rename GXPARM.XDS (written by CORRECT) to XPARM.XDS
- Then run the DEFPIX/INTEGRATE/CORRECT steps again – this improves statistics quite a bit if geometry not accurately known on 1st pass.
- Use the same procedure to correct detector distortions: rename GX-CORRECTIONS.cbf written by CORRECT to X-CORRECTIONS.pck (same for y)
- The latter usually works very well with strong datasets. Files resulting from strong datasets may be used for weak datasets.
- More in XDSwiki (article „Optimization“)

„Mining“ *XDS* data

- Example: two kinds of detector distortions
- **Geometric distortion**: a pixel is not at the position where it should be : at rectangular coordinates given by the pixel size and its x,y indices
- **Nonlinear response of pixel, or different amplification** (different response of pixels)
- Quantify both kinds of distortion by "mining" *XDS* data files
- Determination of distortions with the data themselves: requires good data & redundancy

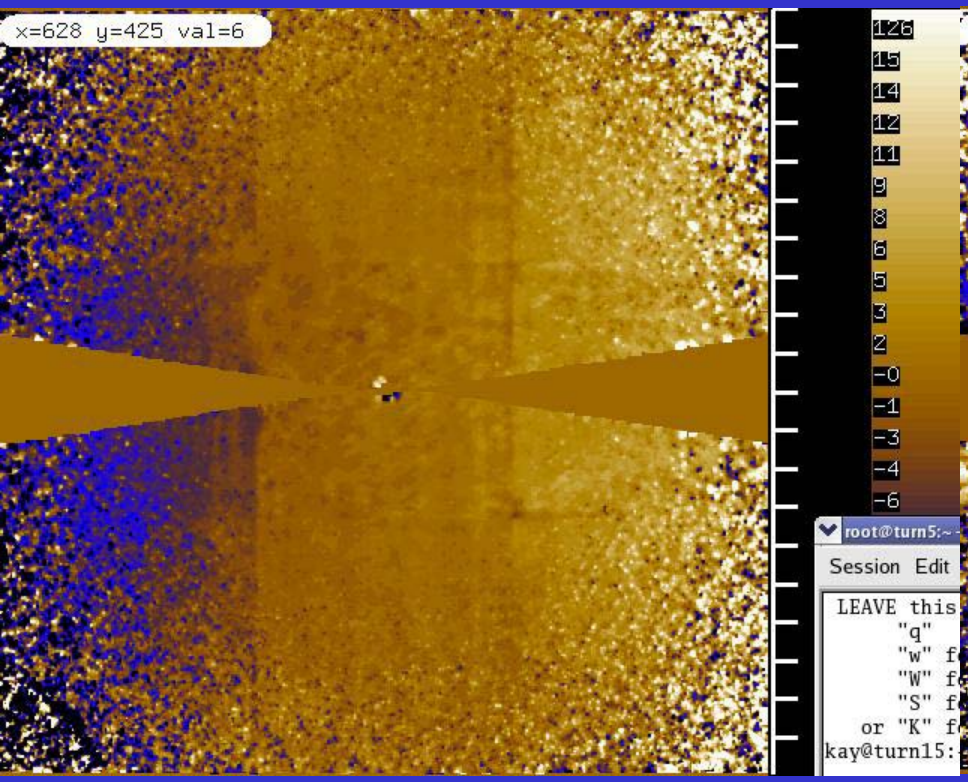
Visualizing Distortions and scaling problems

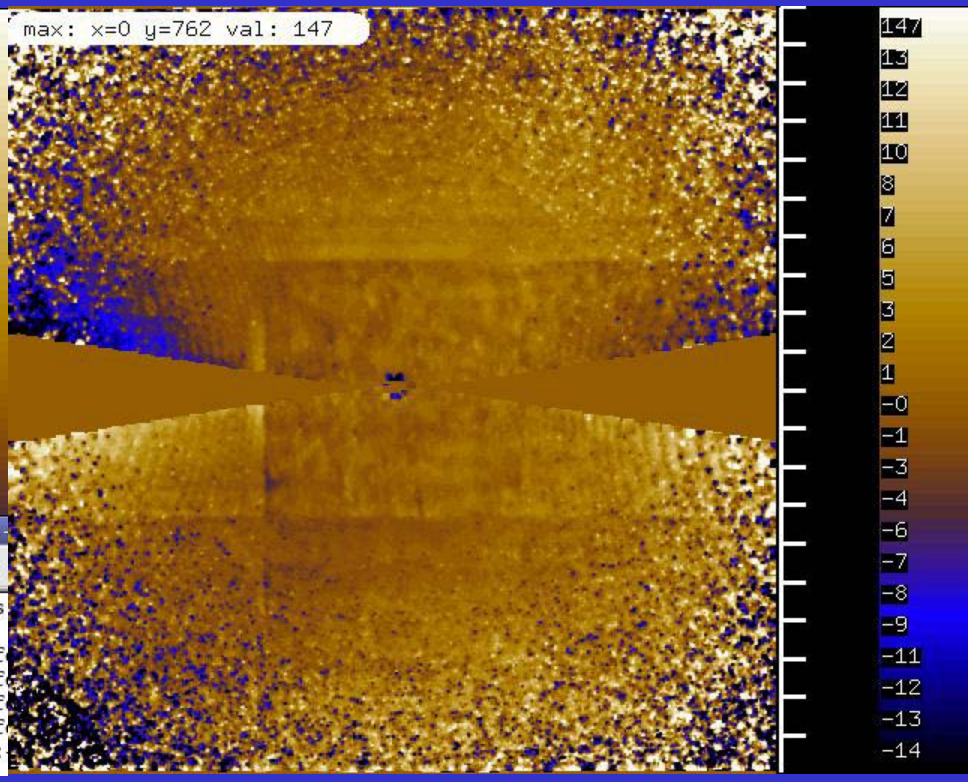
- *XDS* writes .cbf files for control purposes
- *XDS-Viewer* (or *adxv*) can display these files
- If not corrected: systematic errors, many rejections, reduced data quality, bad anomalous signal

X/Y- distortions

- GX-CORRECTIONS.cbf (from CORRECT task) has $10*(x_{obs}-x_{cal})$ as a function of position
- Similar for y: GY-CORRECTIONS.cbf

Mar225 @ Bessy BL1

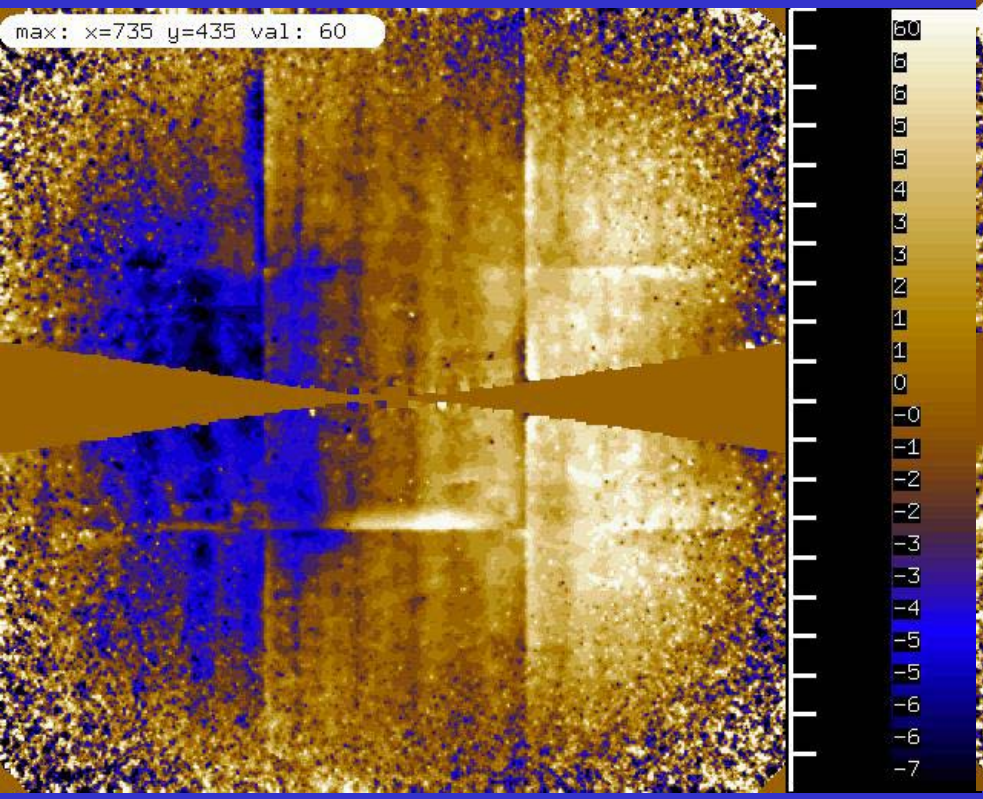


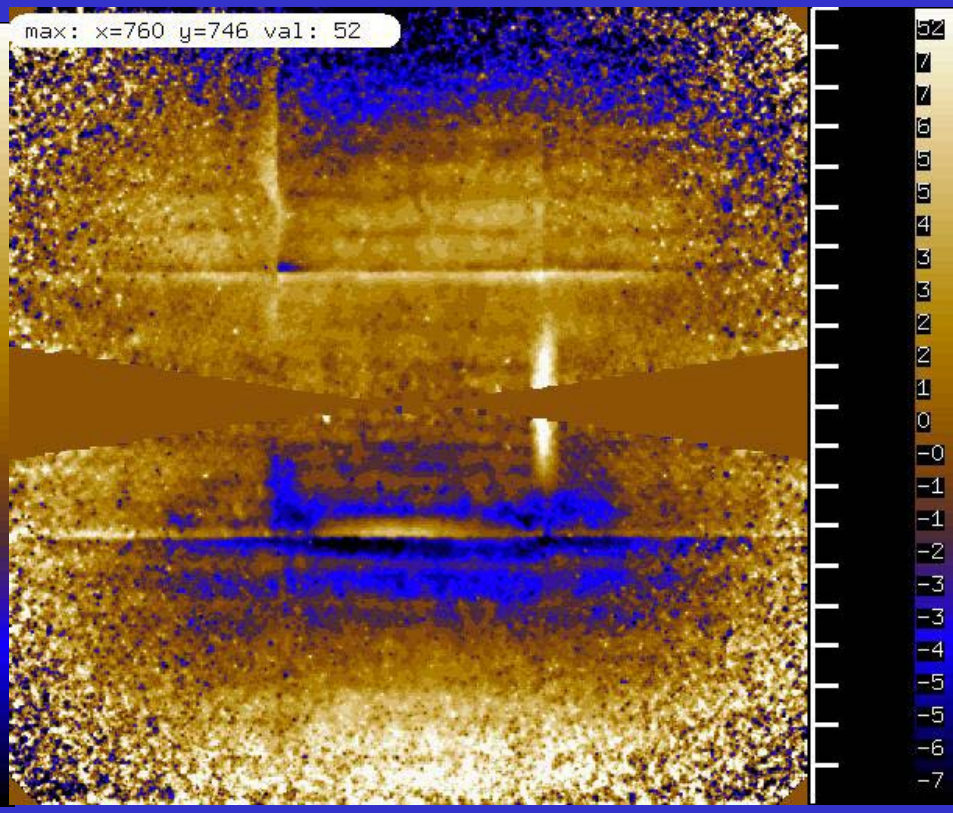


GX-CORRECTIONS.cbf

GY-CORRECTIONS.cbf

Mar225 @ SLS

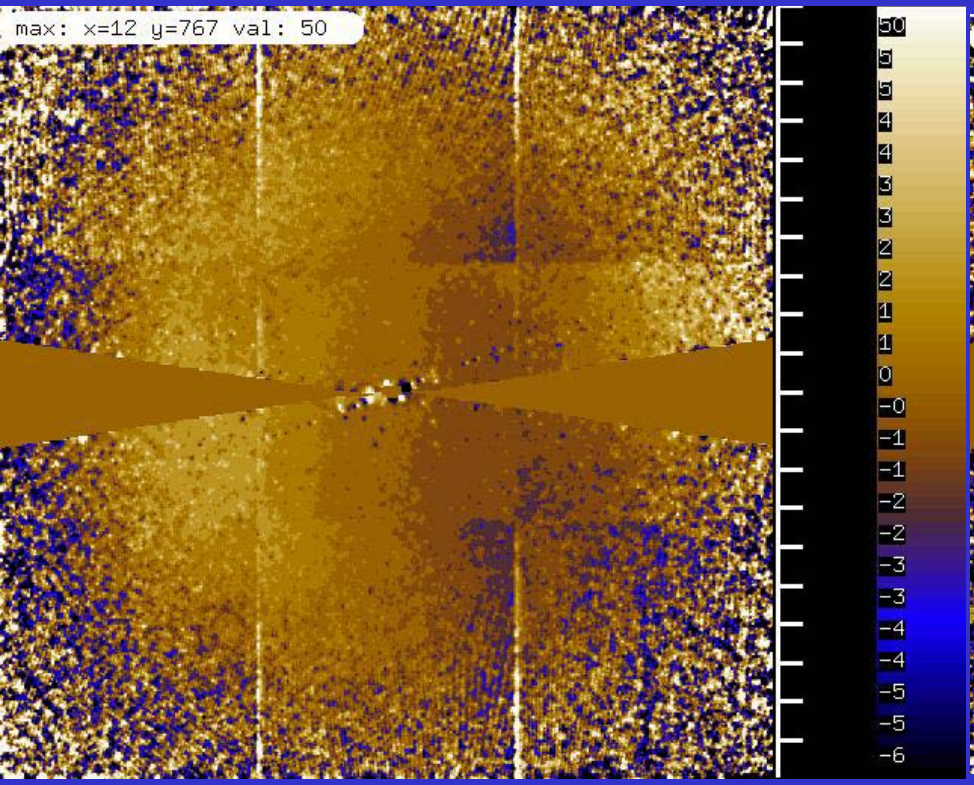


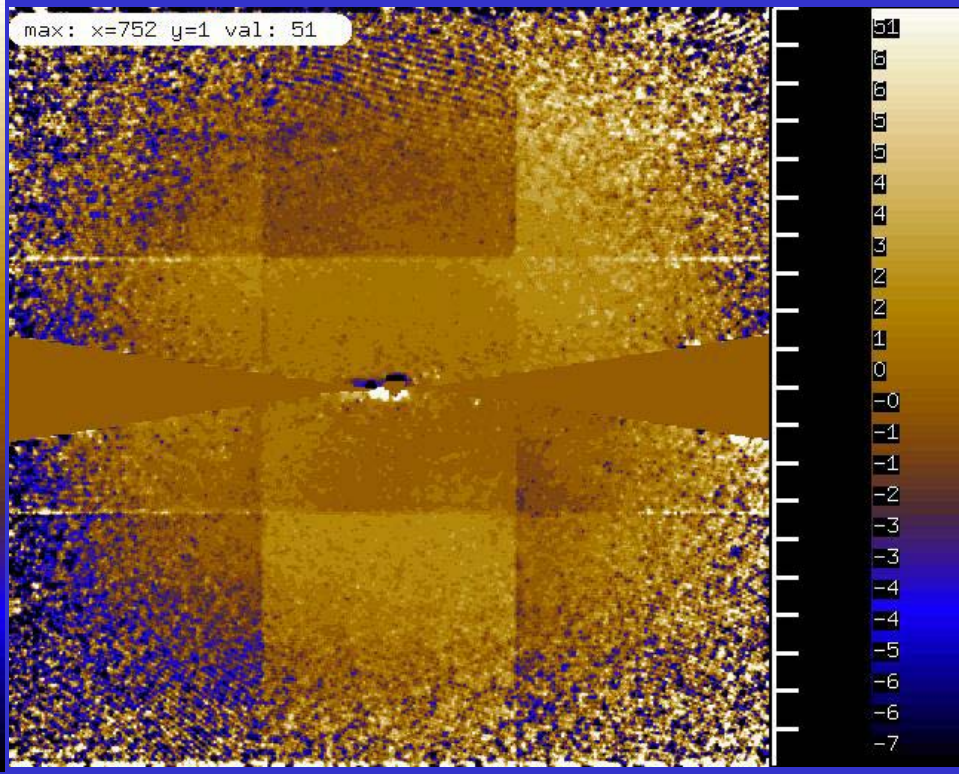


GX-CORRECTIONS.cbf

GY-CORRECTIONS.cbf

ADSC @ ID14-4

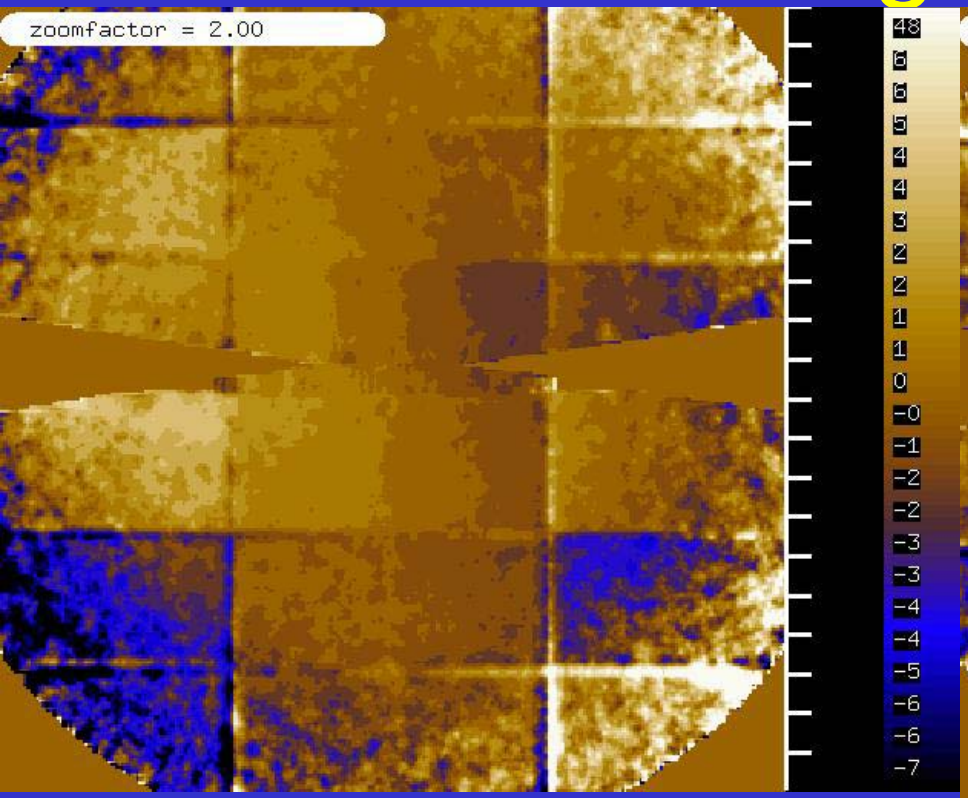


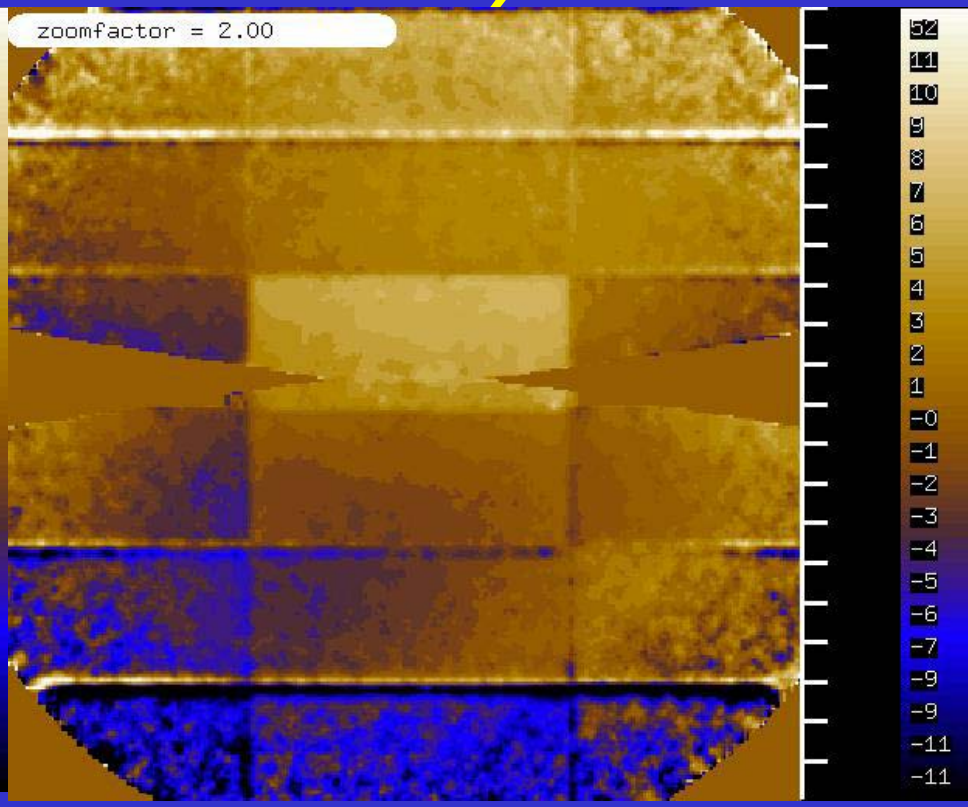


GX-CORRECTIONS.cbf

GY-CORRECTIONS.cbf

Pilatus (pixel detector @ Swiss Light Source)

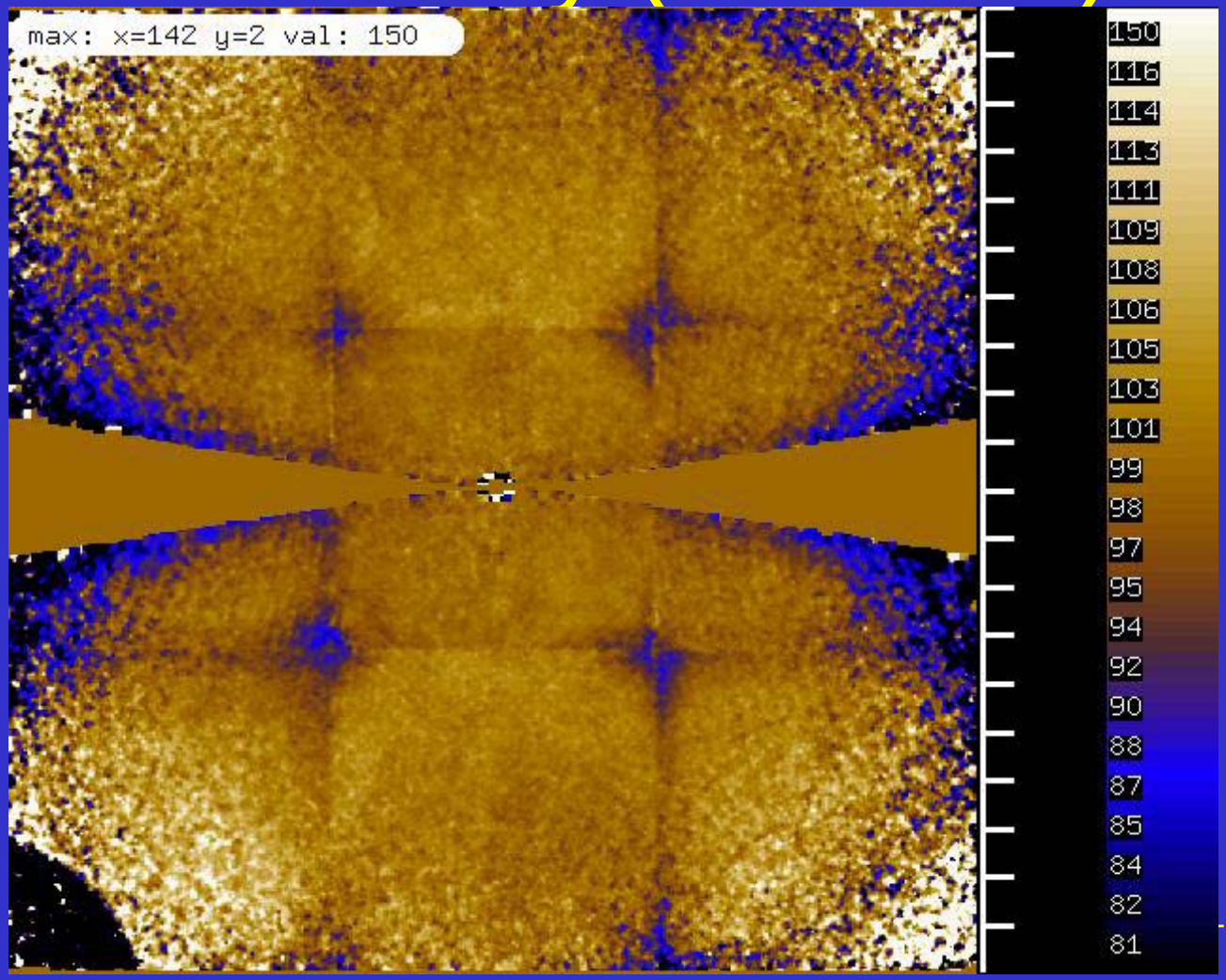




GX-CORRECTIONS.cbf

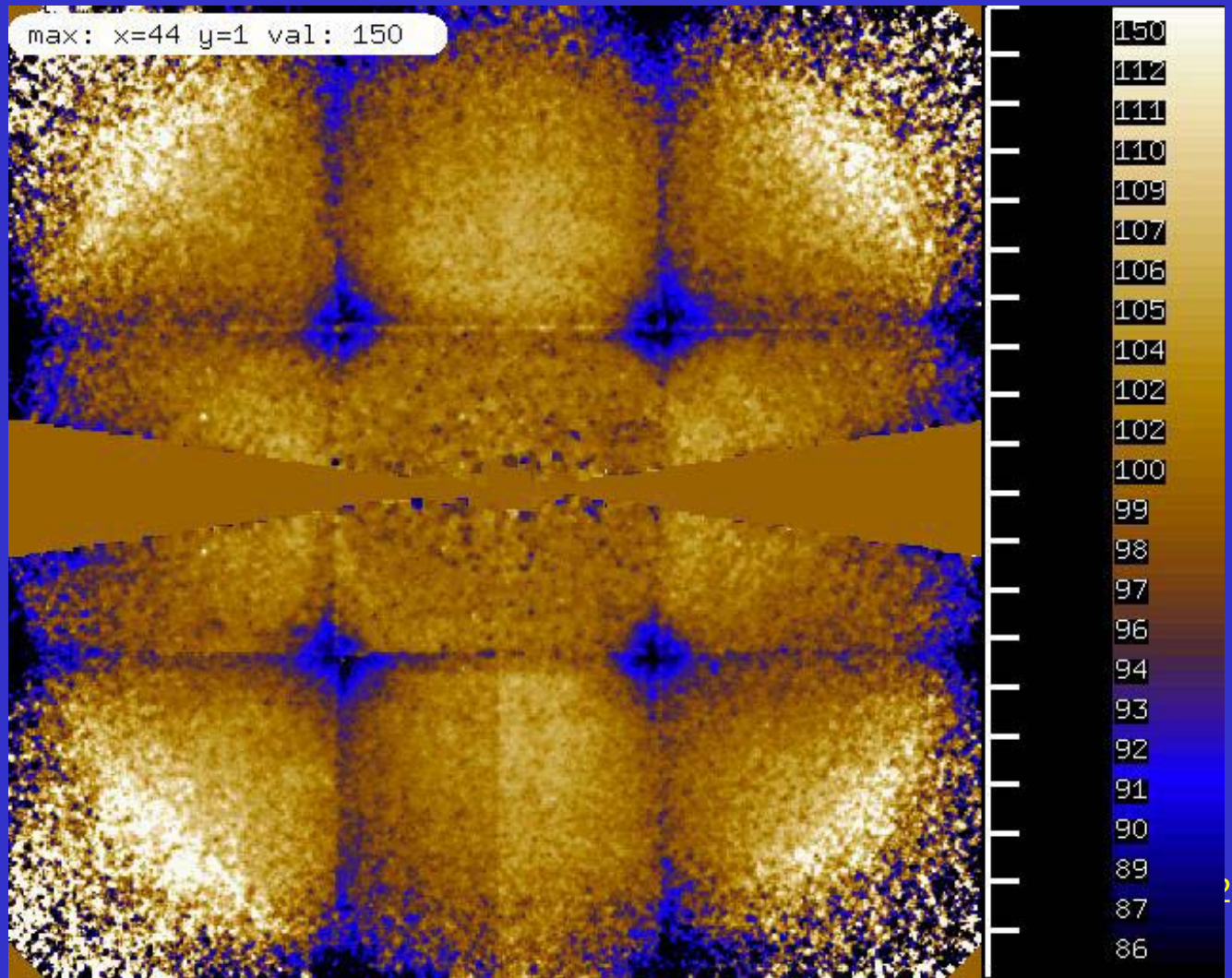
GY-CORRECTIONS.cbf

Scale factors on detector surface: Bessy (Mar 225)



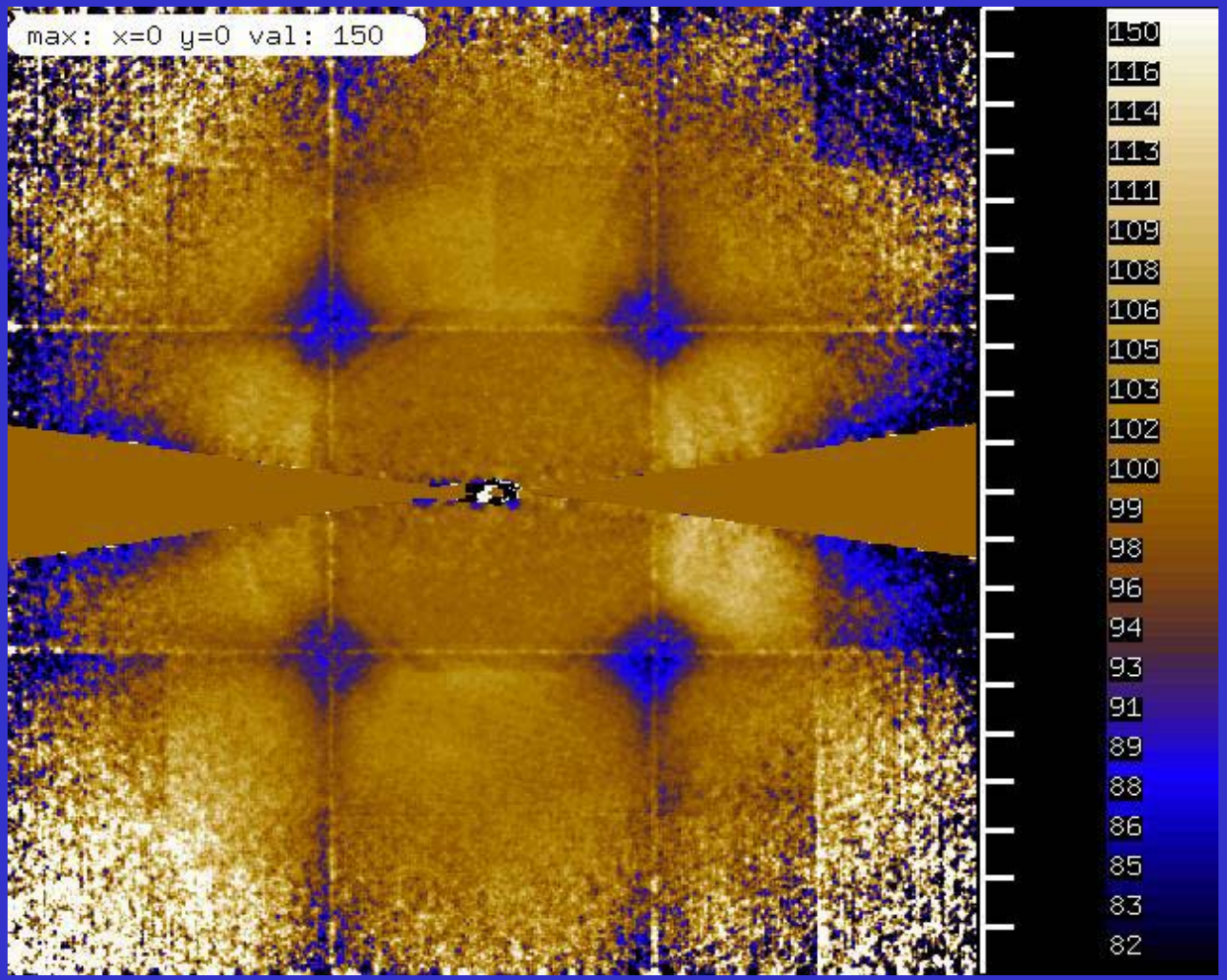
MODPIX.cbf

Scale factors on detector surface: SLS



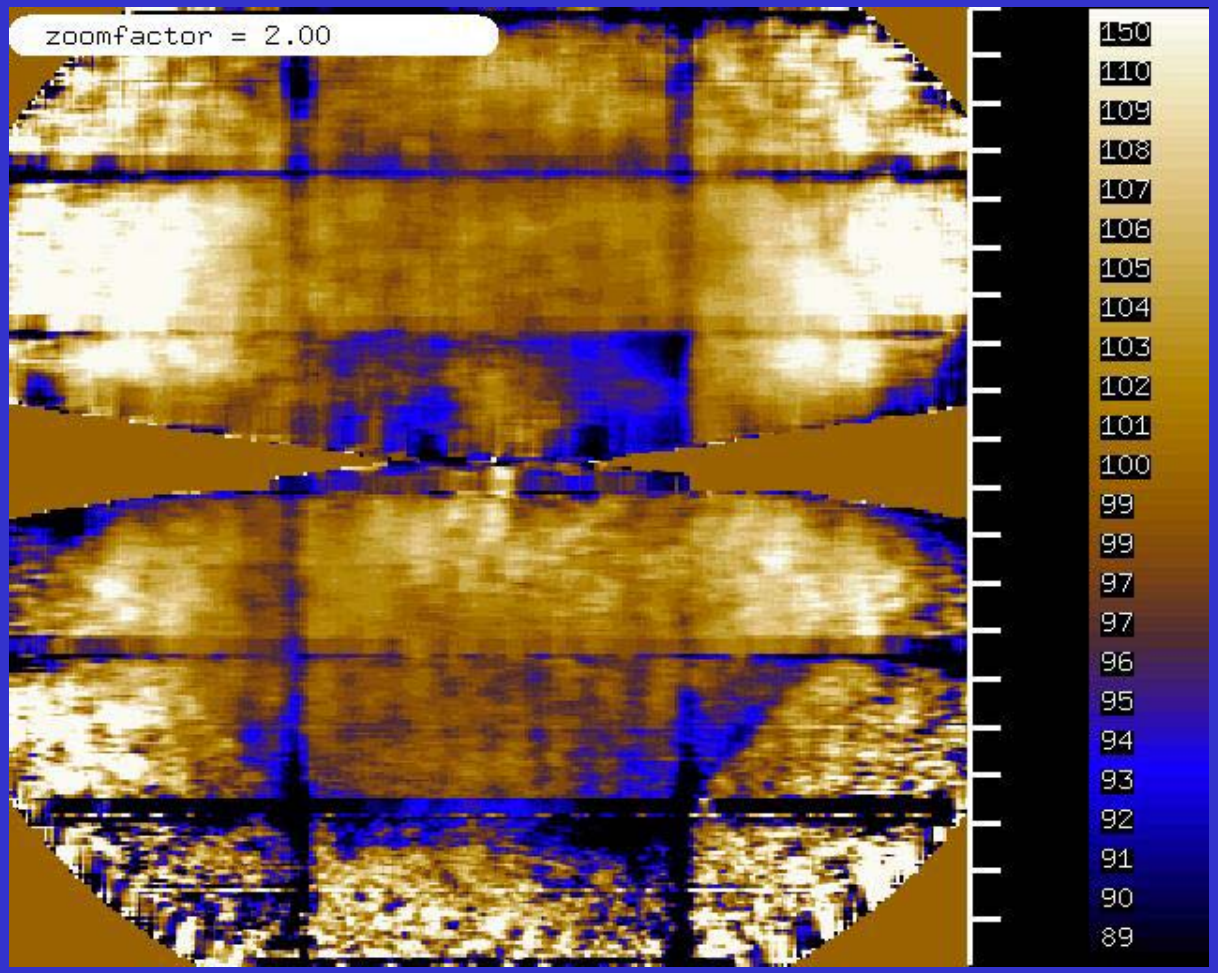
MODPIX.cbf

Scale factors on detector surface: ADSC @ ID14-4



MODPIX.cbf

Scale factors on detector surface: Pilatus



MODPIX.cbf

Further information from *XDSSTAT*

- writes XDSSTAT.LP (visualize with CCP4 *loggraph*)
- scales.pck shows scale factor in percent as a function of position (after correction in *XDS*)
- misfits.pck shows outliers mapped on detector
- rf.pck shows R-factor mapped on detector
- anom.pck shows anomalous difference mapped on detector
- These files may be displayed with *adxv*, *XDS-Viewer*, or *VIEW* (distributed with old versions of XDS)

XDSSTAT.LP

Frame	#refs	#misfits	Iobs	sigma	Iobs/sigma	Peak	Corr	Rmeas	#rmeas	#unique
1	11434	96	137.	21.0	6.53	97.97	42.97	0.1419	11429	5
2	8727	107	125.	19.9	6.27	99.86	41.05	0.1434	8725	2
3	8826	58	131.	20.6	6.36	99.86	41.05	0.1353	8824	2
4	8636	116	127.	20.1	6.31	99.89	40.57	0.1361	8633	3
5	8776	59	131.	20.8	6.30	99.06	40.06	0.1287	8773	3
6	8713	78	132.	21.1	6.24	99.61	38.41	0.1426	8710	3

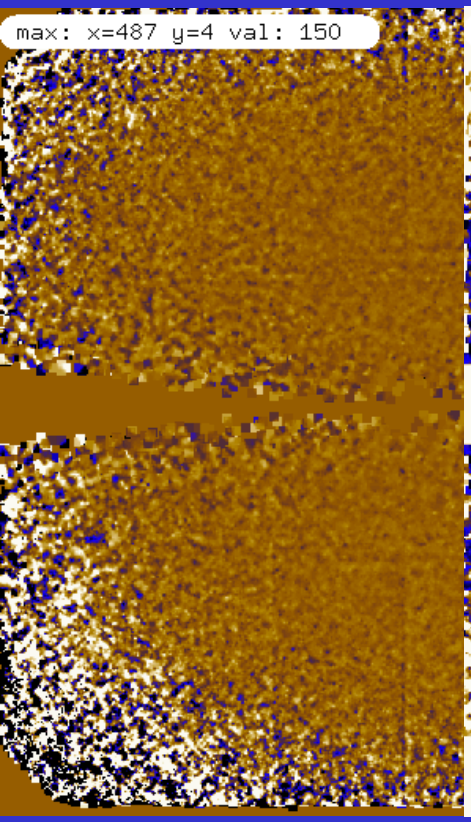
…

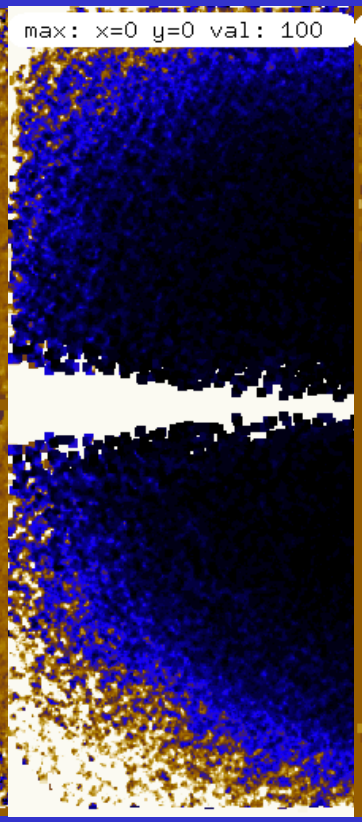
R_d factor as a function of frame number difference

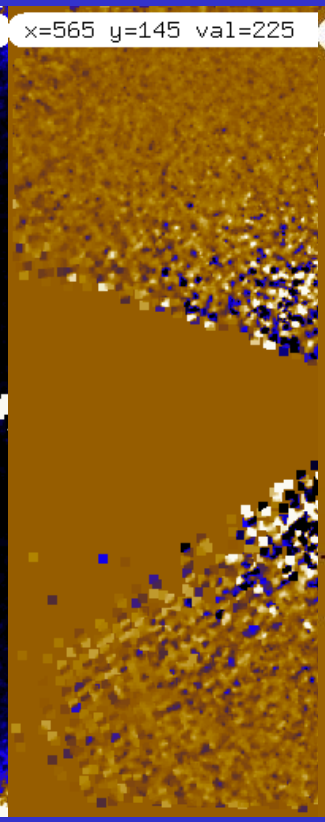
framediff	n-all	Rd-all	n-notfriedel	Rd-notfriedel	n-friedel	Rd-friedel	
0	26160	0.1720	10856	0.1698	15304	0.1736	DIFFERENCE
1	51943	0.1738	21047	0.1695	30896	0.1768	DIFFERENCE
2	50238	0.1626	20888	0.1648	29350	0.1612	DIFFERENCE
3	47429	0.1645	20297	0.1639	27132	0.1649	DIFFERENCE
4	46395	0.1679	20095	0.1695	26300	0.1666	DIFFERENCE
5	44861	0.1649	19505	0.1665	25356	0.1637	DIFFERENCE
6	43656	0.1633	19279	0.1658	24377	0.1615	DIFFERENCE

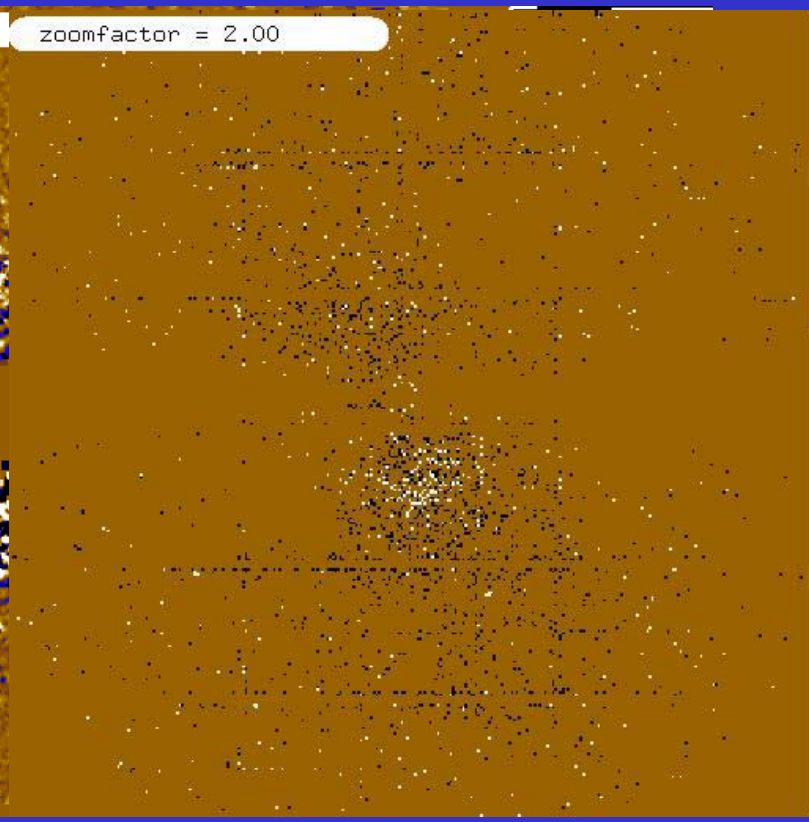
...

Examples of *XDSSTAT*-written files









Scales.pck rf.pck anom.pck misfits.pck

XDS References

- Kabsch, W. (1988). *Evaluation of single-crystal X-ray diffraction data from a position-sensitive detector. J. Appl. Cryst.* **21**, 916-924.
- Kabsch, W. (1993). *Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. J. Appl. Cryst.* **26**, 795-800.
- Kabsch, W. (2001) Chapter **11.3.** *Integration, scaling, space-group assignment and post refinement*

  Kabsch, W. (2001) Chapter **25.2.9.** *XDS*

  both in *International Tables for Crystallography*, Volume F. Crystallography of Biological Macromolecules, Rossmann, M.G. and Arnold, E. (2001). Editors. Dordrecht: Kluwer Academic Publishers.
- http://www.mpimf-heidelberg.mpg.de/~kabsch/xds

XDSwiki

- started Feb 2008; ~ 100 pages at

http://strucbio.biologie.uni-konstanz.de/xdswiki/index.php/Main_Page

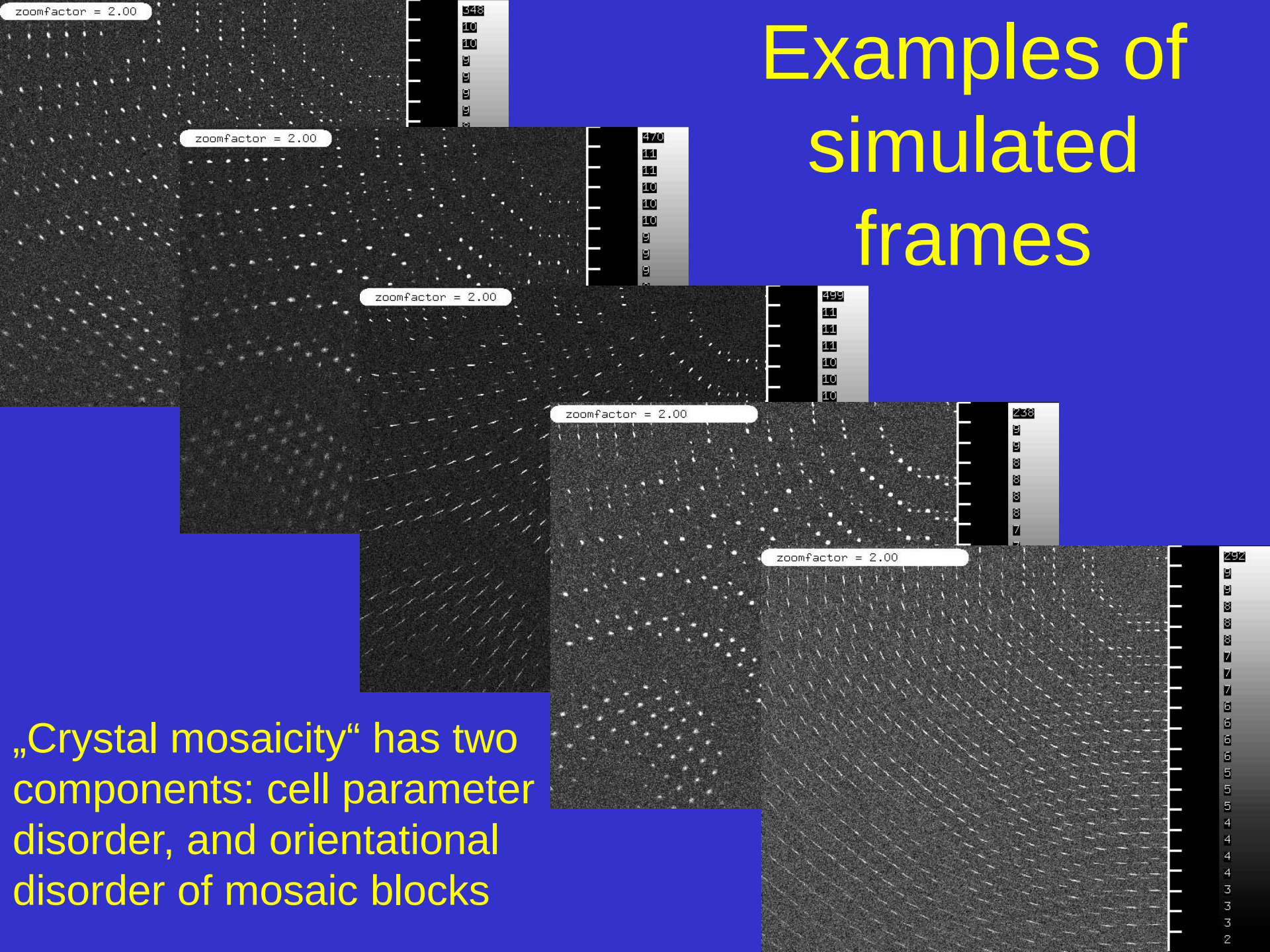
- e.g. „Optimization“; explanations of task output
- „Tips and Tricks“
- „Quality Control“ with datasets and results
- anybody can contribute (same holds for CCP4wiki: ~ 220 pages at

  http://strucbio.biologie.uni-konstanz.de/ccp4wiki/index.php/Main_Page)

current work

- Radiation damage and its computational correction: Diederichs, K., Junk, M. (2009) „Post-processing intensity measurements at favourable dose values“ *J. Appl. Cryst.* **42**, 48-57
- „Simulation of X-ray frames from macromolecular crystals using a ray-tracing approach“ Diederichs K. (2009) Acta Cryst. D65, 535-42

Examples of simulated frames



„Crystal mosaicity“ has two components: cell parameter disorder, and orientational disorder of mosaic blocks

Potential benefits from simulation of raw data

- Test (debug) the whole data reduction / structure solution pipeline with *known* data
- Limits of data quality, and influence of data quality on refinement results
- Evaluate alternative data collection strategies (e.g. fine-slicing) before the actual data collection
- Understand physical principles behind mosaicity
- Simulate certain kinds of systematic errors
- Teaching ...