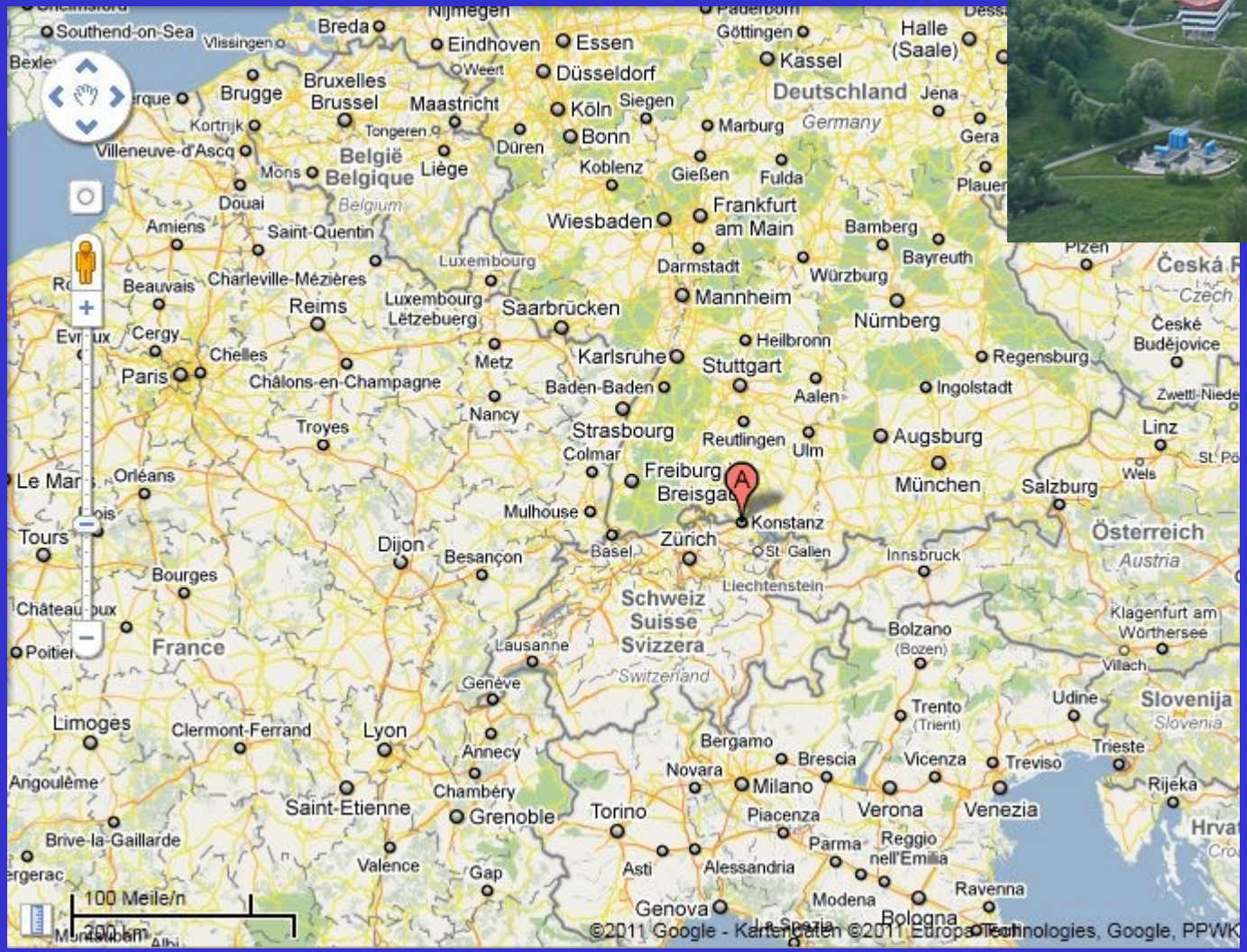
Data Processing Using *XDS*

Kay Diederichs

Protein Crystallography /
Molecular Bioinformatics
University of Konstanz, Germany

Konstanz


Konstanz
Freiburg im Breisgau
Zürich
Basel
Stuttgart
Karlsruhe
Mannheim
Frankfurt am Main
Wiesbaden
Köln
Düsseldorf
Essen
Eindhoven
Nijmegen
Breda
Brugge
Bruxelles
Brussel
Maastricht
Luxembourg
Saarbrücken
Metz
Nancy
Strasbourg
Colmar
Mulhouse
Reutlingen
Ulm
Augsburg
München
Nürnberg
Regensburg
Ingolstadt
Aalen
Heilbronn
Würzburg
Darmstadt
Bamberg
Bayreuth
Kassel
Göttingen
Halle (Saale)
Jena
Gera
Plauen
Marburg
Siegen
Koblenz
Gießen
Fulda
Bonn
Düren
Liège
Mons
Douai
Amiens
Saint-Quentin
Beauvais
Charleville-Mézières
Reims
Châlons-en-Champagne
Troyes
Paris
Chelles
Cergy
Orléans
Tours
Bourges
Dijon
Besançon
Lausanne
Genève
Schweiz
Suisse
Svizzera
Switzerland
France
Innsbruck
Österreich
Austria
Liechtenstein
St. Gallen
Bolzano
(Bozen)
Trento
(Trient)
Milano
Bergamo
Brescia
Verona
Vicenza
Venezia
Treviso
Udine
Trieste
Slovenija
Slovenia
Rijeka
Klagenfurt am Wörthersee
Villach
Linz
Wels
Salzburg
Zwettl-Niede
České Budějovice
Česká
Czech
Pizen
Torino
Novara
Piacenza
Parma
Reggio nell'Emilia
Modena
Bologna
Ravenna
Genova
La Spezia
Asti
Alessandria
Grenoble
Saint-Etienne
Chambéry
Annecy
Lyon
Clermont-Ferrand
Limoges
Angoulême
Brive-la-Gaillarde
Valence
Gap
100 Meile/n
200 km
©2011 Google - Kartendaten ©2011 EuroGeographics, Google, PPWK

Outline

- „*XDS* is a data processing program for X-ray data collected by the oscillation method on area detectors“. Use: stand-alone or in „pipeline“ mode. No GUI, but visual feedback available
- How to get information about *XDS*
- Problems and their diagnostics
- Highlights of 3CSL, 2QVO, 1Y13, 2VB1, 1G1C-simulated

throughout this talk: *program*, file

The *XDS* program suite

Original author: Wolfgang Kabsch (Max-Planck-Institute Heidelberg); I joined ~ 4 yrs ago

Distribution for Linux & Mac: new version every ca. 6 months (latest: Apr-2011)

- *XDS*: the main program (indexing, integrating, scaling)
- *XSCALE*: scale several *XDS* intensity data sets together; zero-dose extrapolation; statistics
- *XDSCONV*: convert to other programs' formats

  (the following are independant of the main distribution:)
- *XDS-Viewer* - inspect diagnostic images written by *XDS*, or (single) data frames; alternatively, latest *adxv* may be used (sourcecode available from sourceforge.net)
- *XDSSTAT* - additional statistics (download and use: see XDSwiki; not part of main distribution)

Interfaces

- GUIs: *XDSi* (P. Kursula), *XDSi* (M. Krug)
- CCP4: *pointless*, *xdsconv* (type CCP4, or CCP4_I, or CCP4_F)
- *CNS / PHENIX / SHELX*: *xdsconv*
- pipelines: *xia2* (CCP4), *autoPROC* (Globalphasing), *autoxds* (SSRL), *autoprocess* (CMCF) ...

XDS algorithms

Features:

- 3D - profiles of reflections are transformed into their own coordinate systems which makes them highly similar
- Pixel-labelling method
- Smooth scaling
- Zero-dose extrapolation (*XSCALE*)
- Two levels of parallelization

Sources of information

- *XDS* main website: http://xds.mpimf-heidelberg.mpg.de - complete, accurate, up-to-date documentation; download
- XDSwiki: http://strucbio.biologie.uni-konstanz.de/xdswiki/index.php/Main_Page
- CCP4 bulletin board
- “XDS webinar” (at Rigaku website)
- Email to kay.diederichs@uni-konstanz.de

*XDS*wiki

- started Feb 2008; ~ 150 pages at

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

- e.g. „Optimization“; explanations of task output
- „Tips and Tricks“, „FAQ“
- „Quality Control“ with datasets and results, and links to the five projects of this workshop
- anybody can contribute!
  (same holds for CCP4wiki: ~ 300 pages at
  http://strucbio.biologie.uni-konstanz.de/ccp4wiki/index.php/Main_Page)

Some typical questions ...

- “How to scale & merge different datasets from similar or same xtal(s), using XDS?”
- “What about twinning?”
- “Is it possible to integrate small molecule data as well?”
- …
- Questions and Answers in FAQ article of XDSwiki

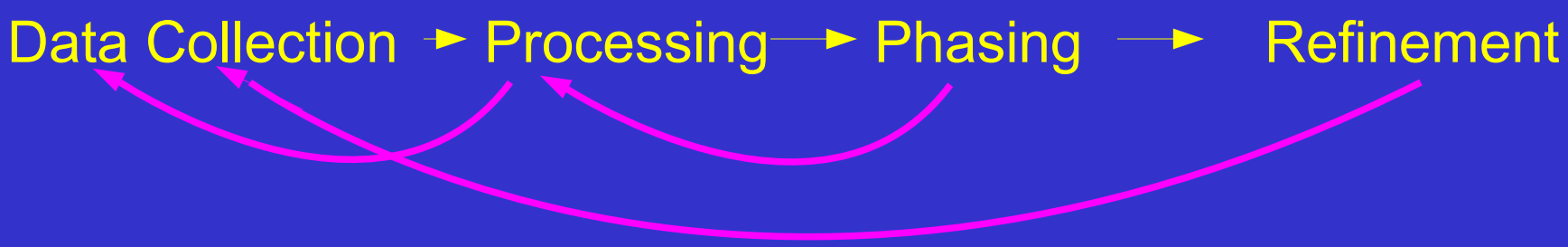
Easy vs. difficult

Data Collection → Processing → Phasing and Refinement

Easy:

1) *generate_XDS.INP "mydata_???.img"* (see XDSwiki)
2) *xds* (or rather *xds_par*)
3) *xdsconv* (or *pointless xdsin XDS_ASCII.HKL hklout mydata.mtz)*

Difficult



Information for making decisions?

What can go wrong?

Beamline: beam center wrong (90%) or unusual convention, rotation backwards; shutter jitter, vibrations ...

Experiment: crystal with split reflections; ice rings; radiation damage

Interpretation of data: twinning overlooked, or wrong spacegroup

Phasing and refinement: anomalous signal too weak, resolution too low, disorder, anisotropy

Which diagnostics to look at?

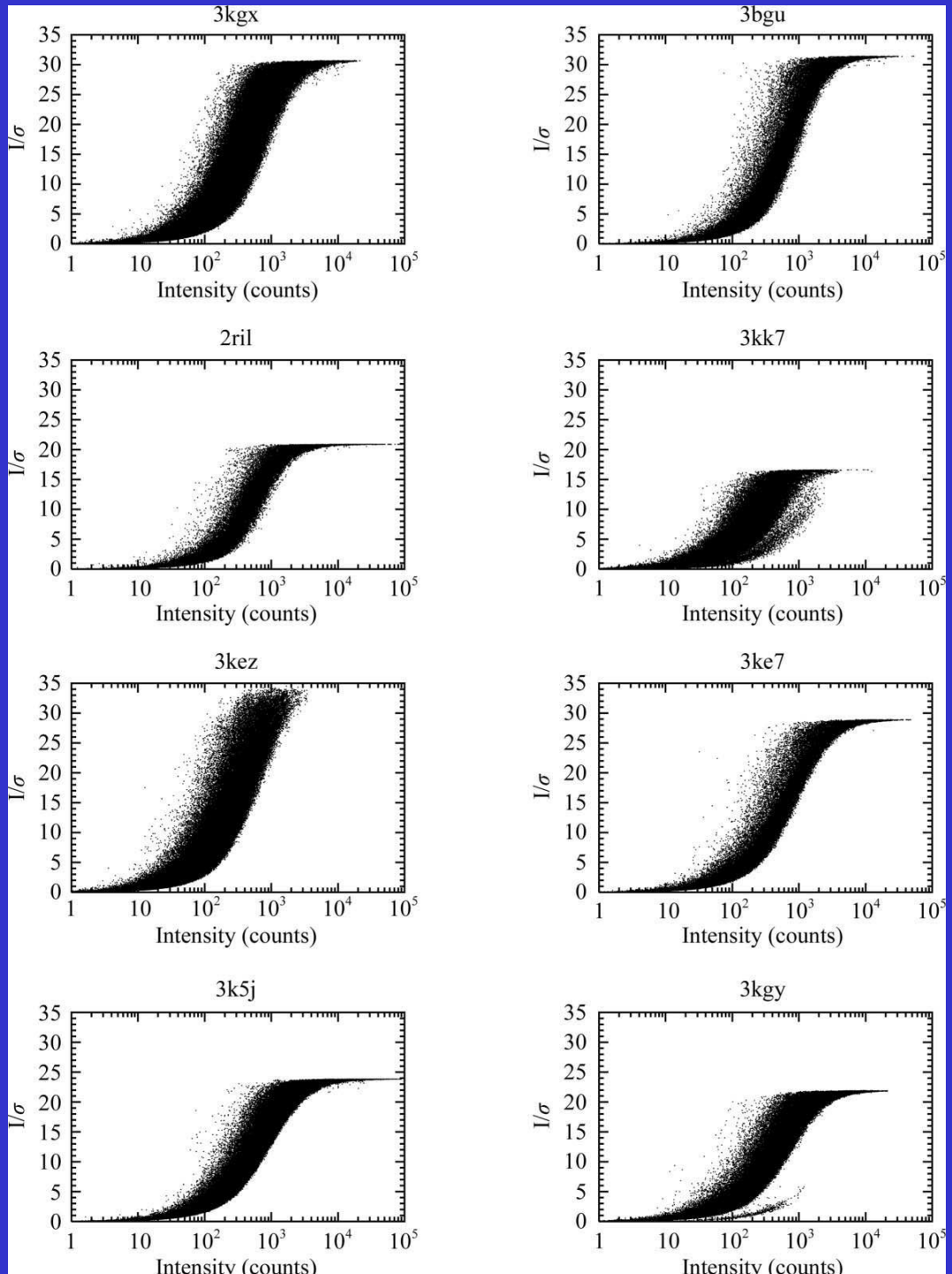
IDXREF.LP - how many lattices? r.m.s.d. between observed and calculated positions?

INTEGRATE.LP - by frame: scales, mosaicity, cell …
FRAME.cbf

CORRECT.LP – shutter stats; spacegroup-related stats; R-factors and I/sigma; twinning; ISa = $(I/\sigma)_{asymptotic}$
MODPIX.cbf, DECAY.cbf, ABSORP.cbf

XDSSTAT.LP (use *loggraph*) – by frame: R-factors, I/sigma, # rejections, ...
scales.pck, rf.pck, R_d

ISa = (I/sigma)$_{asymptotic}$



K. Diederichs

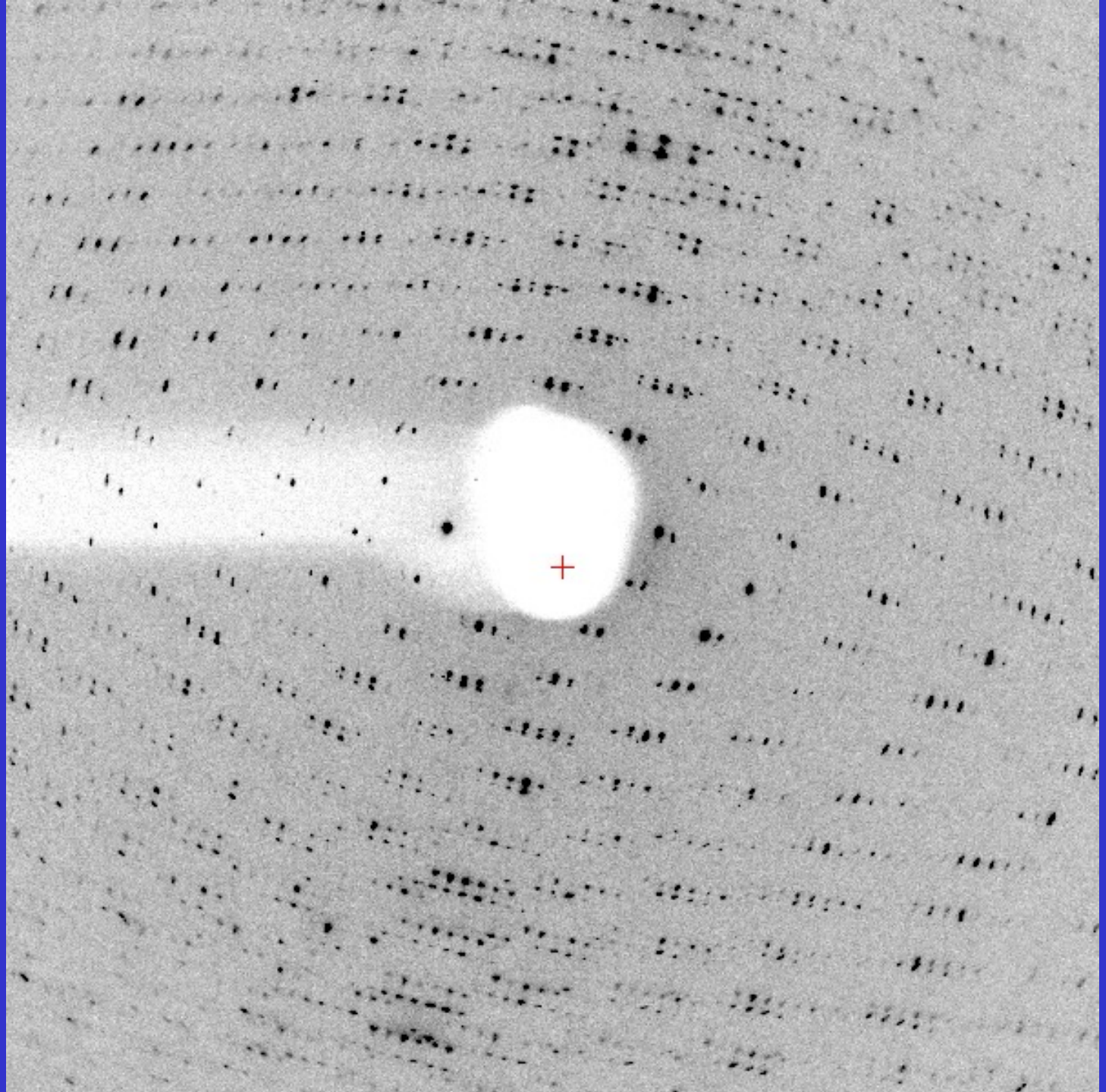
"Quantifying instrument errors in macromolecular X-ray data sets"

Acta Cryst. (2010). D66, 733–740

3CSL: membrane protein complex between heme receptor HasR and its hemophore HasA, plus heme (my lab); SeMet MAD @ 3.5 Å; 2*(865+206) residues / ASU

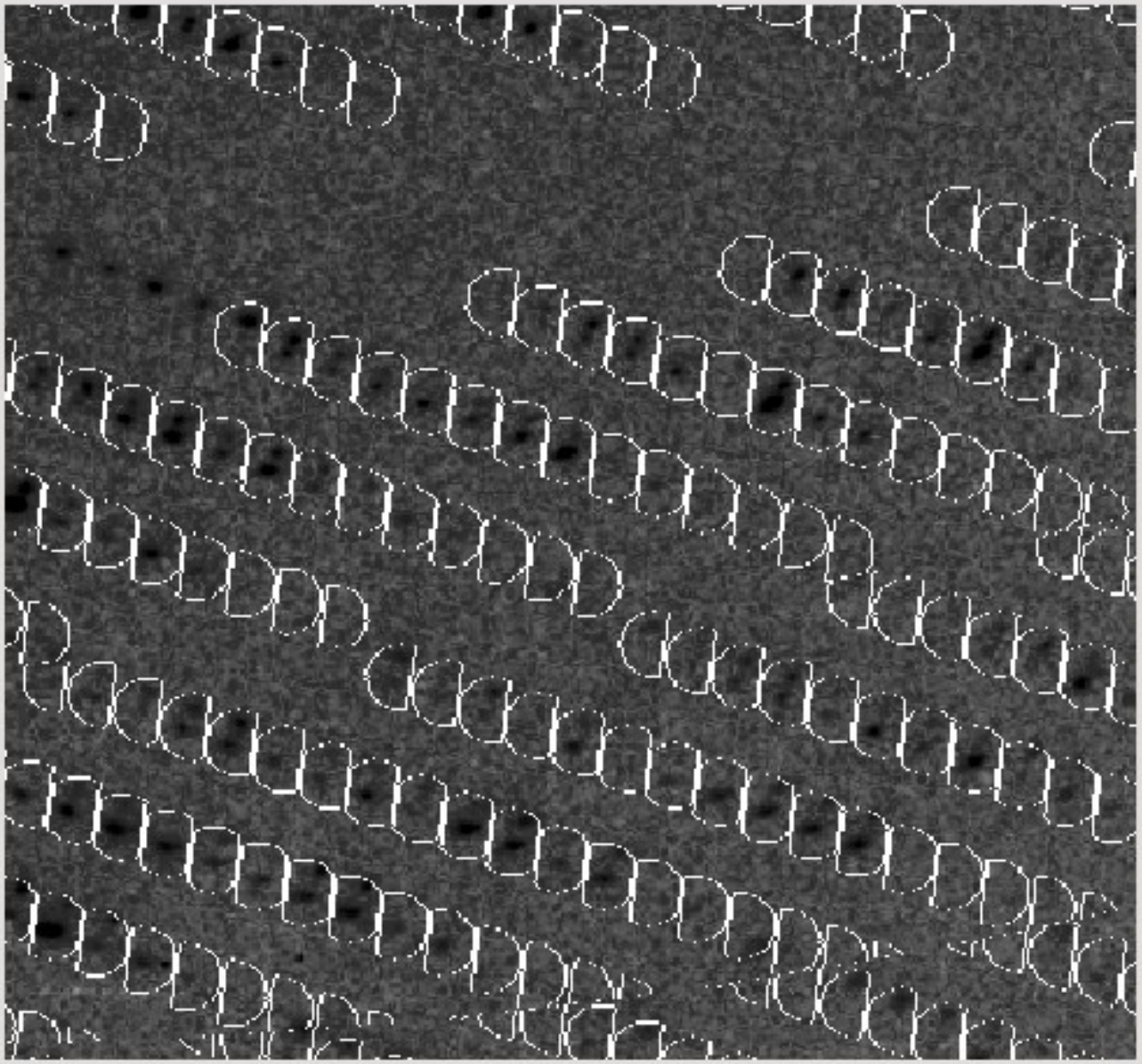
- cracked main xtal, plus smaller xtal
- frames 1-180: IDREF assigns 2898 of the 3000 strongest reflections (59% of all refs) to the major lattice (19% to minor). INTEGRATE is robust enough to give useful data, based on the major lattice.
- 3x3 CCD detector: weaker intensities (80%) at corners of segments; needs special ("MODULATION") correction in scaling, otherwise poor anom signal.

3CSL
frame
closeup



File Help

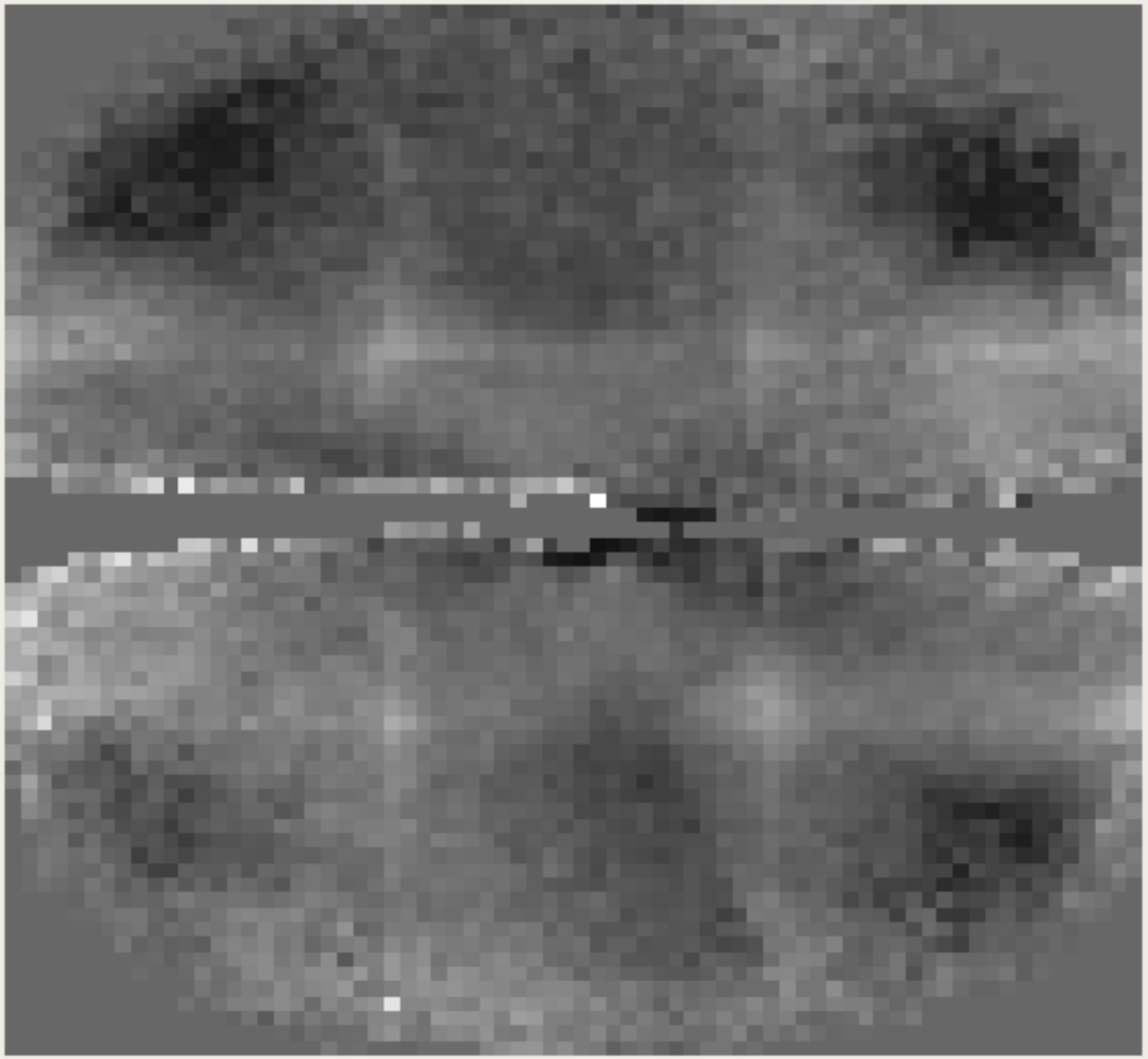
3CSL: FRAME.cbf



3CSL: MODPIX.cbf

File Help

View



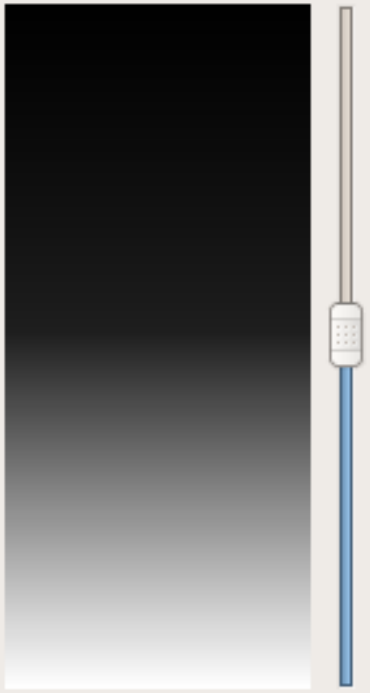
Pixel Value:

x=24, y=23
value=823

Slope:

100 OK

Colors:



1G1C: 2*99 residues / ASU; simulated data (James Holton); 100 datasets of 15 frames each @ 1.8 Å

- batch processing; can easily be scripted
- cell has a=b: don't let the cell fool you w.r.t. the spacegroup decision, it's the symmetry that counts!
- best data with the 100 first frames only
- anom signal incomplete: tweak data processing to include additional weak data & take care of radiation damage

1Y13: 3*181 residues / ASU, SGPP (Jürgen Bosch), SeMet-SAD @ 2.2Å

- problems: information “SeMet data collected at two different wavelengths (and two different beamlines)” inconsistent with results of data processing; radiation damage; ice rings
- diagnostics:
  a) frame-by-frame scalefactors (INTEGRATE): be alerted by jumps
  b) correlation coefficients (XSCALE): be critical towards given information, find better model
  c) R_d plot (XDSSTAT): radiation damage; this can be nicely corrected by XSCALE (0-dose extrapolation)

2QVO: 95 residues/ASU; SECSG (James T. Swindell), S-SAD @ 2.1 Å

- structure was already solved by me 2 years ago; it was indeed more difficult at that time and required more effort
- with current XDS version and default parameters, a straightforward solution is obtained from each dataset; even half of a dataset is enough
- *a posteriori* analyis (documented in XDSwiki): decreasing the number of misfits from the 2.5% of observations obtained with default WFAC1 to the intended 1 % of observations, the best CC All/Weak becomes 44.9 / 22.8 (from 37.3 / 21.4), and the number of successful trials goes from 60% to 91% . **One should note that all internal quality indicators get worse when increasing WFAC1 - but the external ones got significant better!**

2VB1: lysozyme in P1, 0.65Å native data

- straightforward in principle
- don't forget to use corners of detector for low-resol runs (TRUSTED_REGION)
- shadows on the detector need to be removed; there are several keywords for this: UNTRUSTED_RECTANGLE, UNTRUSTED_ELLIPSE, MINIMUM_VALID_PIXEL_VALUE. This requires inspection of frames.

Thank you!

These slides are at
http://strucbio.biologie.uni-konstanz.de/~dikay/XDS_Diederichs_ACA2011.pdf