CCP4 School Chicago 2014

Data Processing with XDS

Tim Grüne
Georg-August-Universität
Institut für Strukturbiochemie

http://shelx.uni-ac.gwdg.de
tg@shelx.uni-ac.gwdg.de
X-ray Diffraction in a Nutshell
Model Building & Refinement — Ideal Crystal Data

- Purpose of crystal structure determination: Molecular Model

- Refinement of model against $h, k, l, F_{\text{ideal}}(hkl), \sigma_F$ or $h, k, l, I_{\text{ideal}}(hkl), \sigma_I$

- Ideal: independent of machine, of wavelength, of crystal shape and size
Experimental Errors

Misfocused Beam

Poorly diffracting crystals

Sources of Errors

- Crystal imperfections
- Radiation damage
- Overloads
- Detector Background Noise
- Missettings (Note down wavelength and distance !!!)

P.S.: Nukri’s beam is NOT an experimental error: this is the unfocused beam
Intensities and Amplitudes — Experiment vs. Theory

Intensities $I_{\text{exp}}(hkl)$ are experimental quantities measured from the detector.*

$$I_{\text{exp}}(hkl) = \frac{e^4}{me^2c^4} \frac{V_{\text{crystal}}}{V_{\text{u.c.}}} \lambda^3 I_0 L P T E I_{\text{ideal}}(hkl)$$  \hspace{1cm} (1)

Data Processing:

**Data integration:** Determination of intensities $I_{\text{exp}}(hkl)$ from frames

**Scaling / Merging:** Determination of amplitudes $I_{\text{ideal}}(hkl)$ from $I_{\text{exp}}(hkl)$ and experimental settings

Understanding your last Experiment

The Ewald Sphere Construction
Ewald Sphere Construction

Reciprocal Lattice:
\[ \vec{a}^* = \frac{\vec{b} \times \vec{c}}{(\vec{a} \times \vec{b}) \cdot \vec{c}} \]
\[ \vec{b}^* = \frac{\vec{c} \times \vec{a}}{(\vec{a} \times \vec{b}) \cdot \vec{c}} \]
\[ \vec{c}^* = \frac{\vec{a} \times \vec{b}}{(\vec{a} \times \vec{b}) \cdot \vec{c}} \]

Reciprocal Lattice:
\[ h\vec{a}^* + k\vec{b}^* + l\vec{c}^* \]
(hollow circles)

The crystal rotates about the origin of the reciprocal lattice.
Ewald Sphere Construction

Draw a sphere with radius $1/\lambda$ that touches the lattice origin. The sphere centre lies aligned with the X-ray source.

This sphere is the **Ewald Sphere**.
Ewald Sphere Construction

- The scattering vector $\vec{S}$ points from the origin to the lattice point.

- Exactly those lattice points on the surface of the Ewald sphere fulfil the Laue conditions.

\[
\begin{align*}
\vec{a} \cdot \vec{S} &= h \\
\vec{b} \cdot \vec{S} &= k \\
\vec{c} \cdot \vec{S} &= l \\
(\vec{S} &= \vec{S}_{\text{in}} - \vec{S}_{\text{out}})
\end{align*}
\]
Ewald Sphere Construction

Reflections on the Ewald Sphere can be recorded.

Some of these reflections hit the detector.
Ewald Sphere Construction

Crystal rotates = (0, 2)
Lattice rotates =
New reflections
(Rot. axis perpendicular to slide)
Ewald Sphere Construction

Large Cell = Small Lattice = More Reflections

$11 \times 13 \times 18 \text{Å}$

$78 \times 78 \times 78 \text{Å}$
Example Application: Inverse Beam for Anomalous Phasing

Wedges collected in the order
1. $0^\circ - 5^\circ$
2. $180^\circ - 185^\circ$
3. $5^\circ - 10^\circ$
4. $185^\circ - 190^\circ$
5. $10^\circ - 15^\circ$
6. $190^\circ - 195^\circ$
7. . . .

Synchrotrons often offer ‘inverse beam’ collection strategies to accurately determine Bijvoet pairs for phasing based on anomalous dispersion.
Bijvoet Pairs and Inverse Beam

- Bijvoet are collected close together in space and time.
- Reduces random errors but may pronounce systematic errors.
Data Processing with XDS
Data Processing

**Processing** your Data = Getting $I_{\text{ideal}}$ from your experiment

**Understanding** your Reduction Program(s) = Getting the best $I_{\text{ideal}}$ from your experiment
XDS & xdsGUI

XDS controlled with single input file XDS.INP
Command: xds_par
**XDS.INP**

XDS is controlled by one single input file: XDS.INP.

- Name cannot be changed
- **Each data set must be run in separate directory to avoid overwriting of files.**
- Contains about 100 Keywords\(^a\) of the form
  
    `KEYWORD=VALUE`

- Only about 10 Keywords must be modified for most data sets (e.g. image names, detector distance, number of images, etc.)
- Most important one:
  
    `JOB= XYCRR INIT COLSPOT IDXREF DEFPIX INTEGRATE CORRECT`

  Each name stands for one of the steps XDS carries out during data integration.

\(^a\)also called “cards” for historical reasons
Generating XDS.INP

1. `generate_XDS.INP` as part of xdsGUI: MarCCD, ADSC, Pilatus detectors (+ some Rigaku detectors)
2. `adxv` Reads virtually all frame formats → manual editing of XDS.INP
3. `sfrmtools` Convert Bruker sfrm to XDS coordinate systems (available upon request)
Templates for Manual XDS.INP

- Templates of input scripts for all supported detector formats
- Only very few adjustments necessary to get started
- Beamlines often generate appropriate input scripts
- It is worth learning how to set it up from scratch!
Steps of XDS

- **red:** Text files with parameters or data
- **cyan:** control images (use xds-viewer)
- **black:** data files for further processing

Diagram showing the steps of XDS with corresponding output files.
The Steps

XYINIT writes files for positional corrections of the detector plane. Most modern detectors provide already corrected images so that these to files are normally flat.

INIT determines initial detector background

COLSPOT Strong reflections for indexing

IDXREF indexing: unit cell dimensions and crystal orientation

DEFPIX set active detector area (exclude resolution cut-off, beam stop shadow, . . .)

INTEGRATE extract reflection intensities from frames $\rightarrow I_{\text{exp}(hkl)}$

CORRECT applies corrections (polarisation, Lorentz-correction, . . .), scales reflections, reports data statistics $\rightarrow I_{\text{ideal}(hkl)}$
Program Flow

- Each step must be passed at least once - the subsequent steps depend on files produced by the previous steps.
- Log-files for each step (XYINIT.LP, INIT.LP, ...).
- IDXREF = main hurdle - once unit cell and crystal orientation are determined, integration usually runs smoothly.
- CORRECT scaling and quality tables
- xdsGUI illustrates important graphs
IDXREF

- Indexing step: Find cell parameters and cell orientation.

- First refinement of experimental parameters (Detector distance, ...)

- Writes solution to XPARM.XDS

```
XPARM.XDS VERSION November 11, 2013

1 0.0000 0.1000 0.999978 0.006046 0.002667
0.826568 0.000543 0.001864 1.209820
199 78.0597 78.0597 78.0597 90.000 90.000 90.000
77.801140 -0.752043 6.303378
2.363492 75.369080 -20.179979
-5.891697 20.303986 75.142204
1 2463 2527 0.172000 0.172000
1224.162720 1249.473389 170.145401
1.000000 0.000000 0.000000
0.000000 1.000000 0.000000
0.000000 0.000000 1.000000
1 1 2463 1 2527
0.00 0.00 0.00 1.00000 0.00000 0.00000 0.00000 1.00000 0.00000
```

Tim Grüne XDS 24/46
**SPOT.XDS**

- **COLSPOT**: Detector coordinates and Intensity of strong spots to be used for indexing:

<table>
<thead>
<tr>
<th>X(pixel)</th>
<th>Y(pixel)</th>
<th>#image</th>
<th>counts</th>
</tr>
</thead>
<tbody>
<tr>
<td>1056.11</td>
<td>1529.51</td>
<td>15.35</td>
<td>2544.</td>
</tr>
<tr>
<td>1895.52</td>
<td>1525.49</td>
<td>9.19</td>
<td>2481.</td>
</tr>
<tr>
<td>1913.43</td>
<td>1547.90</td>
<td>2.63</td>
<td>1999.</td>
</tr>
</tbody>
</table>

- **IDXREF**: Miller-Indices according to XPARM.XDS

<table>
<thead>
<tr>
<th>X(pixel)</th>
<th>Y(pixel)</th>
<th>#image</th>
<th>counts</th>
<th>H</th>
<th>K</th>
<th>L</th>
</tr>
</thead>
<tbody>
<tr>
<td>1056.11</td>
<td>1529.51</td>
<td>15.35</td>
<td>2544.</td>
<td>14</td>
<td>13</td>
<td>18</td>
</tr>
<tr>
<td>1895.52</td>
<td>1525.49</td>
<td>9.19</td>
<td>2481.</td>
<td>-7</td>
<td>-13</td>
<td>-14</td>
</tr>
<tr>
<td>1913.43</td>
<td>1547.90</td>
<td>2.63</td>
<td>1999.</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

- 0 0 0: not indexed with current cell
"!!! ERROR !!! SOLUTION IS INACCURATE"

Correct indexing is crucial for data integration. If XDS indexes 50 % of all spots in SPOT.XDS it stops with the above error message.

Most common reasons:

1. Wrong ORGX, ORGY
2. Wrong Parameter settings (Detector distance, wavelength)
3. Poor data quality
Carry on regardless

The XDS error message \texttt{SOLUTION IS INACCURATE} does not necessarily mean that something is seriously wrong. The step IDXREF still refines the experimental parameters and writes them into \texttt{XPARM.XDS}. This is all we need to continue.

\textbf{Check IDXREF.LP for}

- Number of indexed reflections

  \begin{verbatim}
  ***** INDEXING OF OBSERVED SPOTS IN SPACE GROUP # 1 *****
  1909 OUT OF 2506 SPOTS INDEXED.
  0 REJECTED REFLECTIONS (REASON: OVERLAP)
  597 REJECTED REFLECTIONS (REASON: TOO FAR FROM IDEAL POSITION)
  \end{verbatim}

- Does the cell make sense? (check with \textit{e.g.} adxv)
- refined detector distance and detector origin do not shift too much

Set \texttt{JOB = DEFPIX INTEGRATE CORRECT} and integrate your data.

If something \textit{did} seriously go wrong, the file \texttt{XPARM.XDS} would not have been written.
**DEFPIX: active detector mask**

DEFPIX sets the area of the detector which is taken into account during integration. It takes into account:

1. `INCLUDE_RESOLUTION_RANGE` default: 20 Å to detector edge.
2. `VALUE_RANGE_FOR_TRUSTED_DETECTOR_PIXELS` exclude shadowed regions, *e.g.* beamstop, cryostream nozzle
3. `UNTRUSTED_RECTANGLE` exclude gaps of CCD chips *e.g.* Pilatus detector (automatic)
4. `EXCLUDE_RESOLUTION_RANGE` exclude ice rings from data integration
VALUE RANGE FOR TRUSTED DETECTOR PIXELS

ABS.cbf  BKGPIX.cbf, all included

VALUE_RANGE_FOR_TRUSTED_DETECTOR_PIXELS= 1000 30000
VALUE_RANGE_FOR_TRUSTED_DETECTOR_PIXELS

ABS.cbf  BKGPIX.cbf, shadows removed

VALUE_RANGE_FOR_TRUSTED_DETECTOR_PIXELS= 6100 30000
INTEGRATE and CORRECT: Intensity determination and fine tuning

INTEGRATE  determine spot intensities based on parameters (saved in XPARM.XDS)
CORRECT  experimental corrections (e.g. Lorentz- and Polarisation-correction), refine parameters and determine space group (saved to GXPARM.XDS)
CORRECT writes FRAME.cbf:
  • predicted spot positions encircled
  • check correctness of predictions
Recycling

- Parameters (in GXPARM.XDS) depend on measured intensities
- Intensities (including corrections) depend on Parameters
  ⇒ rename GXPARM.XDS to XPARM.XDS and rerun XDS (JOB = DEFPIX INTEGRATE CORRECT) to improve results.

This way one should also set the correct high- and low-resolution cut-offs
Resolution Cut-Off

The default resolution range in XDS is 20 Å to the detector edge

`INCLUDE_RESOLUTION_RANGE=20.0 0.0`

- Medium to low resolution data: increase 20 Å to 30 Å or even 50 Å (check BKGPIX.cbf)
- After second round of integration: determine high-resolution cut-off.

Why after second round?

- Correct space group rather than $P1$
  - more symmetry related reflections
  - more reliable data statistics, especially $I/\sigma_I$
### High Resolution Cut-Off

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

<table>
<thead>
<tr>
<th>RESOLUTION LIMIT</th>
<th>NUMBER OF REFLECTIONS</th>
<th>COMPLETENESS</th>
<th>R-FACTOR COMPARED I/SIGMA</th>
<th>R-meas</th>
<th>CC(1/2)</th>
<th>Anom SigAno</th>
<th>SigAno Nano</th>
</tr>
</thead>
<tbody>
<tr>
<td>13.63</td>
<td>9181</td>
<td>1459</td>
<td>1511</td>
<td>96.6%</td>
<td>2.5%</td>
<td>2.9%</td>
<td>9180</td>
</tr>
<tr>
<td>9.64</td>
<td>17093</td>
<td>2592</td>
<td>2672</td>
<td>97.0%</td>
<td>2.7%</td>
<td>3.1%</td>
<td>17091</td>
</tr>
<tr>
<td>7.87</td>
<td>23196</td>
<td>3381</td>
<td>3433</td>
<td>98.5%</td>
<td>3.2%</td>
<td>3.4%</td>
<td>23195</td>
</tr>
<tr>
<td>6.81</td>
<td>27489</td>
<td>3965</td>
<td>4049</td>
<td>97.9%</td>
<td>4.5%</td>
<td>4.4%</td>
<td>27483</td>
</tr>
<tr>
<td>6.09</td>
<td>31672</td>
<td>4498</td>
<td>4565</td>
<td>98.5%</td>
<td>5.8%</td>
<td>5.7%</td>
<td>31671</td>
</tr>
<tr>
<td>5.56</td>
<td>35456</td>
<td>4969</td>
<td>5040</td>
<td>98.6%</td>
<td>6.8%</td>
<td>6.6%</td>
<td>35451</td>
</tr>
<tr>
<td>5.15</td>
<td>37852</td>
<td>5387</td>
<td>5465</td>
<td>98.6%</td>
<td>7.2%</td>
<td>7.1%</td>
<td>37847</td>
</tr>
<tr>
<td>4.82</td>
<td>40555</td>
<td>5829</td>
<td>5894</td>
<td>98.9%</td>
<td>7.2%</td>
<td>7.1%</td>
<td>40546</td>
</tr>
<tr>
<td>4.54</td>
<td>41126</td>
<td>6125</td>
<td>6213</td>
<td>98.6%</td>
<td>7.8%</td>
<td>7.7%</td>
<td>41105</td>
</tr>
<tr>
<td>4.31</td>
<td>45588</td>
<td>6568</td>
<td>6647</td>
<td>98.8%</td>
<td>9.5%</td>
<td>9.5%</td>
<td>45578</td>
</tr>
<tr>
<td>4.11</td>
<td>47633</td>
<td>6842</td>
<td>6900</td>
<td>99.2%</td>
<td>12.4%</td>
<td>12.4%</td>
<td>47622</td>
</tr>
<tr>
<td>3.93</td>
<td>50103</td>
<td>7192</td>
<td>7284</td>
<td>98.7%</td>
<td>16.8%</td>
<td>16.8%</td>
<td>50091</td>
</tr>
<tr>
<td>3.78</td>
<td>51899</td>
<td>7414</td>
<td>7519</td>
<td>98.6%</td>
<td>23.0%</td>
<td>22.9%</td>
<td>51889</td>
</tr>
<tr>
<td>3.64</td>
<td>54983</td>
<td>7766</td>
<td>7842</td>
<td>99.0%</td>
<td>31.6%</td>
<td>31.8%</td>
<td>54975</td>
</tr>
<tr>
<td>3.52</td>
<td>56772</td>
<td>8033</td>
<td>8114</td>
<td>99.0%</td>
<td>43.3%</td>
<td>43.5%</td>
<td>56759</td>
</tr>
<tr>
<td>3.41</td>
<td>58909</td>
<td>8313</td>
<td>8412</td>
<td>98.8%</td>
<td>59.1%</td>
<td>59.8%</td>
<td>58897</td>
</tr>
<tr>
<td>--&gt; 3.31</td>
<td>60300</td>
<td>8607</td>
<td>8702</td>
<td>98.9%</td>
<td>83.0%</td>
<td>84.5%</td>
<td>60285</td>
</tr>
<tr>
<td>3.21</td>
<td>62086</td>
<td>8804</td>
<td>8904</td>
<td>98.9%</td>
<td>116.7%</td>
<td>119.8%</td>
<td>62073</td>
</tr>
<tr>
<td>3.13</td>
<td>64144</td>
<td>9057</td>
<td>9159</td>
<td>98.9%</td>
<td>156.4%</td>
<td>160.9%</td>
<td>64134</td>
</tr>
<tr>
<td>3.05</td>
<td>61751</td>
<td>9009</td>
<td>9395</td>
<td>95.9%</td>
<td>206.2%</td>
<td>212.5%</td>
<td>61650</td>
</tr>
<tr>
<td>total</td>
<td>877788</td>
<td>125810</td>
<td>127720</td>
<td>98.5%</td>
<td>10.5%</td>
<td>10.7%</td>
<td>877522</td>
</tr>
</tbody>
</table>

**Logfile XSCALE.LP**
Low Resolution Cut-Off

- Default resolution range (20 Å - edge)
  - includes noise at edge
  - loose low resolution reflections

- Adjusted resolution range (40 Å - 2.85 Å)
  - $I/\sigma_I \approx 2$ in outer shell
  - low resolution reflection important for “shape” of molecule (MR, refinement)
XDS and Friedel’s Law

- XDS.INP and XSCALE.INP allow the keyword FRIEDEL’S LAW=FALSE

- Frequent belief: ‘TRUE’ = merging of Bijvoet pairs = loss of anomalous signal

- Correct:
  - XDS.ASCII.HKL always unmerged
  
  
  - FRIEDEL’S LAW=TRUE only affects scaling and statistics

- FRIEDEL’S LAW=TRUE: twice as many reflections
  
  
  - more reliable statistics

  
  
  - better anomalous signal
Multiple Datasets: REFERENCE_DATA_SET

In many space groups, indexing is not unique. \( I2_13 \): 24 possibilities.

```
# 1 2 3 4 5 6 7 8 9 10 11 12
1 1 0 0 0 0 1 0 0 0 0 1 0
2 0 0 1 0 1 0 0 0 0 1 0 0
...
24 0 -1 0 0 0 0 -1 6 1 0 0 0
```

REFERENCE_DATA_SET= ../A1/XDS_ASCII.HKL in XDS.INP takes care of everything.

Otherwise: data sets do not merge.
Scaling Multiple Datasets

- XDS_ASCII.HKL already scaled and ready to use

- Multiple data sets (e.g. inverse beam): Use XSCALE

Command: xscale_par; Control: XSCALE.INP

OUTPUT_FILE=insulin.HKL
INPUT_FILE=../A1/XDS_ASCII.HKL
INPUT_FILE=../A2/XDS_ASCII.HKL
Format Conversion

The final integrated data are written to the file XDS_ASCII.HKL for each run.

1. for phasing with shelx c/d/e

2. for refinement (mtz-file)
Phasing with shelx c/d/e: HKL → hkl

1. Phasing with shelx c/d/e requires no conversion

2. shelxc and xprep both read XDS and XSCALE output (HKL--) files

3. format detected automatically
HKL → mtz

How to (suggested by Clemens Vonrhein, Globalphasing):

1. Conversion to mtz, hklref: indexing correction if necessary (e.g. space group $C2$)

   ```bash
   pointless -copy xdsin xscale.HKL hklref freeRef.mtz hklout pointless.mtz
   ```

2. Merging of equivalent indices:

   ```bash
   aimless hklin pointless.mtz hklout aimless.mtz << eoaimless
   ONLYMERGE
   eoaimless
   ```

3. Intensities to amplitudes ($I \rightarrow |F|$):

   ```bash
   truncate hklin aimless.mtz hklout truncate.mtz << eotruncate | tee truncate.log
eotruncate
   ```
HKL $\rightarrow$ mtz (continued)

4. Copying $R_{\text{free}}$ flags if necessary

```bash
cad hklin1 truncate.mtz hklin2 freeRef.mtz hklout cad.mtz << eocad | tee cad.log
LABIN FILE 1 ALL
LABIN FILE 2 E1=FreeR_flag
eocad
```

5. Correction for reported resolution, e.g. by mtzdmp

```bash
sftools << eosftools | tee sftools.log
read cad.mtz
SELECT ONLY COLUMN IMEAN PRESENT
write sftools.mtz
eosftools
```

6. Extension of $R_{\text{free}}$ reflections to full resolution

```bash
freerflag hklin sftools.mtz hklout freerflag.mtz << eofreerflag | tee freerflag.log
COMPLETE FREE=FreeR_flag
eofreerflag
```
Documentation

XDS Program Package

for processing single-crystal monochromatic diffraction data recorded by the rotation method.

Version: December 6, 2010 Release_Notes
Author: Wolfgang Kabach

- Downloading and Installation
- Getting started
  - xde
    - Input parameters
    - Files
    - Coordinate systems
  - xscale
    - Input parameters
    - Files
    - xdsconv
      - Input Parameters
      - Output format
    - 2cbf
    - Cellparm
    - References

- Very well organised html-documentation
- Every STEP documented
- Every keyword documented
- Coordinate systems are explained

© 2005-2010, MPI for Medical Research, Heidelberg Imprint.
Wolfgang.Kabach@mpimf-heidelberg.mpg.de
page last updated: December 6, 2010
The XDS Resources

availability:

XDS http://xds.mpimf-heidelberg.mpg.de/ main program suite

Wiki and auxiliary programs http://strucbio.biologie.uni-konstanz.de/xdswiki/

GUIs

- xdsGUI http://strucbio.biologie.uni-konstanz.de/xdswiki/index.php/XDSgui
- XDSi http://cc.oulu.fi/~pkursula/xdsi.html
- Xdsapp http://www.helmholtz-berlin.de/forschung/funkma/soft-matter/forschung/bessy-mx/xdapp
References


xdsGUI  W. Brehm, K. Diederichs, M. Hoffer (©2013)

ADXV  Andrew Arvai, http://www.scripps.edu/~arvai/adxv.html


Acknowledgment

Lisandro Otero, Fundación Instituto Leloir, Argentina