Tutorial

Crystallographic Data Processing with XDS

I) Introduction

XDS is a small suite of programs, very easy to use, dedicated to the processing of your crystallographic data. The suite of programs include:

XDS: data processing, from images to unmerged h,k,l,Intensities, sigma(Intensities)
XSCALE: scaling and merging Intensities from either one or several data sets.
XCSCONV: converts reflection data files as obtained from xds or xscale into various formats required by software packages for crystal structure determination like CCP4, CNS (X-PLOR), or SHELX.
2cbf: converts a detector image file to CBF format. (not often used)
merge2cbf: converts a series of detector image files to CBF format. (not often used)
cellparm: used to determine the mean of the cell parameters obtained from processing several data sets from the same crystal form. (not often used).

Only **XDS** and **XSCALE** will be used here.

XDS:

The items required to run XDS are:
- your diffraction images
- a file called XDS.INP that contains all the necessary information required by XDS about the experimental setup.

Most of the time, an XDS.INP file is generated automatically when your launch a data collection at a synchrotron site. However, the file XDS.INP needs some editing during the data processing, but only a few input parameters require to be looked at: the most important input parameters in the XDS.INP file are described in the XDS.INP.pdf file.

The whole data data processing includes 7 steps defined in the "JOB=" command

XYCORR: computes a table of spatial correction values for each pixel: allow to precisely localize each pixel of the detector. Fully automatic, to be done once.
files created:
X-CORRECTIONS.cbf
Y-CORRECTIONS.cbf
XYCORR.LP

INIT: determines an initial background for each detector pixel and finds the trusted region of the detector surface. Needs 5 to 10 images to run properly (look at "BACKGROUND_RANGE=" command) .To be done once.
files created:
BKGINIT.cbf
BLANK.cbf
GAIN.cbf
INIT.LP

COLSPOT: collects strong diffraction spots from a specified subset of the data images (see "SPOT_RANGE= " command).
files created:
FRAME.cbf
SPOT.XDS
COLSPOT.LP

IDXREF: interprets observed spots by a reciprocal lattice and refines all diffraction parameters (cell dimensions, orientation matrix, crystal-detector distance, etc ...).
files created:
XPARM.XDS
IDXREF.LP

DEFPIX: defines the trusted region of the detector, recognizes and removes shaded areas, and eliminates regions outside the resolution range defined by the user.
files created:
BKGPIX.cbf
ABS.cbf
DEFPIX.LP

XPLAN: helps planning data collection. Tells you what data to collect in order to get the most complete data set. Only useful when at the synchrotron beamline, before launching the data collection.
files created:
XPLAN.LP

INTEGRATE: collects 3-dimensional profiles of all reflections present in the data images and estimates their intensities (I)
files created:
INTEGRATE.HKL
INTEGRATE.LP

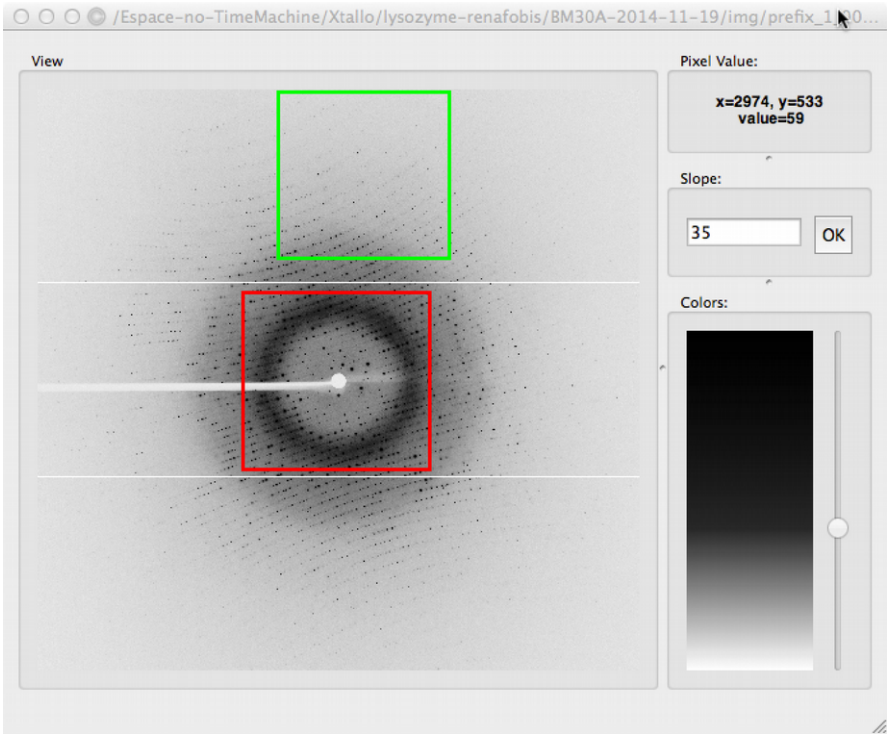
CORRECT: corrects intensities for decay, absorption and variations of detector surface sensitivity, reports statistics of the collected data set and refines the diffraction parameters using all observed reflections (much more that in the IDXREF step).
files created:
ABSORP.cbf
DECAY.cbf
DX-CORRECTIONS.cbf
DY-CORRECTIONS.cbf
GX-CORRECTIONS.cbf
GY-CORRECTIONS.cbf
MODPIX.cbf
GXPARM.XDS
XDS_ASCII.HKL
CORRECT.LP

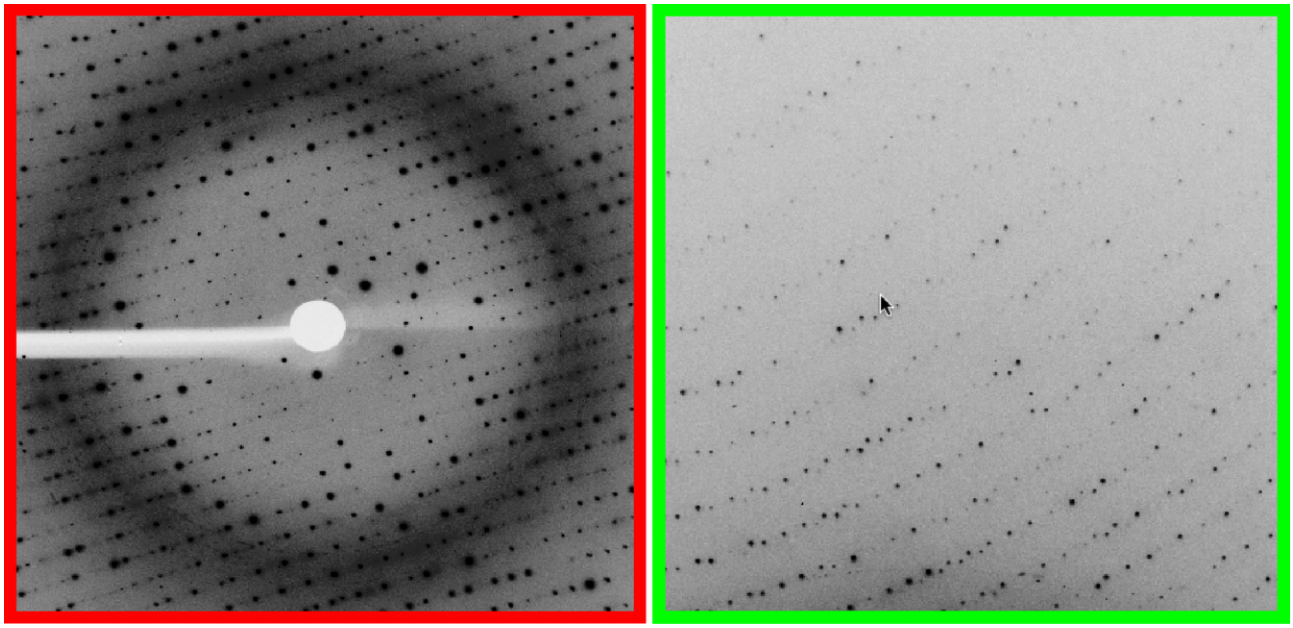
II) Practical

Data processing of native lysozyme data

0) Have a look at one diffraction image with XDS-VIEWER

Open a terminal window
type xds-viewer
go in the "file" menu, then "Load Image" and select one image of the data





crystal-detector distance 141.02 mm
wave length 0.979263 Å
size of the detector 6140x6140 pixels
pixel size 0.051 mm
maximum resolution: edge: 1.20 Å, corner 1.01 Å

Question to be answered, among others:

What is the dark ring in the middle of the diffraction image
Is there ice diffraction spots or rings?

Is there diffraction up to the edge of the detector?
Does the crystal seems to be unique?

1) Indexation

Open a terminal window
go to the directory where the XDS.INP file is (folder xds01_1).
Edit XDS.INP with your favorite text editor
check the input parameters (see XDS.INP.pdf commented file)
choose the JOB command JOB= XYCORR INIT COLSPOT IDXREF and comment all the others.
save the XDS.INP file
run xds (type xds or xds_par in your terminal window)

Open COLSPOT.LP and figure out how many spots where selected.
Open IDXREF.LP and try to figure out the results of the indexation step. See the IDXREF-ALL-P1.LP.pdf for a commented output.

Question to be answered, among others:
Are most of the spots explained by a reciprocal unit cell?
Is the fit between actual spots and predicted spots satisfactory?
Is the origin of the reciprocal space (beam position) correct.
Is your crystal unique or multiple?
What are the possible Bravais lattices for your crystal and possible related space groups?

2) Initial automated data processing

Edit XDS.INP with your favorite text editor
choose the JOB command JOB=ALL and comment all the others.
save the XDS.INP file.
run xds (type xds or xds_par in your terminal window)

Check whether the program stop after the IDXREF or continue until the end of the CORRECT step. If it does perform all the steps (in principle it should) open the CORRECT.LP file and check it. See the CORRECT-ALL-P1.LP.pdf for a commented output.

Question to be answered, among others:
In which space group has the integration been performed?
Check the number of reflections accepted.
Check the automatic space group assignment:
Quality index of possible Bravais lattice?
Choice of the space group (indicated by a *): Does it seem a reasonable choice ? (discuss the Rmeas value)
Is there other possible space groups belonging to the same Laue class (same symmetry in the reciprocal space)?
Check the refined unit cell dimensions: do they differ much from the one used for the integration?
Does the refinement seems OK?
What is the resolution limit of this data set?
Do you have systematic extinctions that could indicate the presence of screw rotation axes?
Any final guess about possible space group(s)?
Are the data complete?

3) Final data processing

Edit XDS.INP with your favorite text editor
specify the proper space group (#96 P4(3)2(1)2) and unit cell dimensions and uncomment these lines
specify the resolution limit chosen.
save the XDS.INP file.
run xds (type xds or xds_par in your terminal window)
Have a look at the new CORRECT.LP file.

Edit XSCALE.INP with your favorite text editor
check it
cun xscale (type xscale in your terminal window)
check the XSCALE.LP

Question to be answered, among others:
Assess the quality of the data after final processing in the P4(3)2(1)2 space group

4) Create an MTZ file with the output of xscale for using it in CCP4

a) With CCP4

edit the output file of xscale, lyso-1-1-A.hkl, that contains the unique reflections h k l I σ(I) in a fixed format (3i6,2e11.4) and remove the header (at the top) and the last line that contains "! END_OF_DATA" at the end. Save as a new file in the ccp4/unique directory.

edit the uniquemtz.com file (in the ccp4/unique directory)
update the values for cell, space group, resmax, etc... and save.
run the script by typing:
./uniquemtz.com

This script run several program of the CCP4 suite (scripting is an alternative to ccp4i):
f2mtz: reads your hkl file from xscale (ascii format, after some editing, see above) and creates an mtz file (binary format of ccp4)

unique: generates an mtz file with all the unique reflections (h, k, l indices only) for a given unit cell, space group and resolution range.

freerflag: adds a column to an exisiting mtz file which contain a flag used for free R factor calcultation (refinement step, essentially). It evenly divides your data set in either 10 (freerfrac 0.1) or 20 (freerfrac 0.05) subsets. By default, the subset 0 will be used for free R factor calculation and not used during the refinement. The other subsets will used for the refinement and for the work R factor calculation.

cad: merges information from several mtz files

truncate: calculates structure factors (F) and their standard deviation (σ(F)) from intensities (I, σ(I)). Provides also some statistics (Wilson plot, ...)

mtzdump: enables to printout the content of an mtz file (a binary file cannot be edited in a text editor)

Look at the uniquemtz.out output file.

check that the final mtz file contains what you need for the next steps (phasing, refinement, ...):

```
 OVERALL FILE STATISTICS for resolution range   0.000 -   0.826
  ======================


 Col Sort    Min    Max    Num      %     Mean     Mean   Resolution   Type Column
 num order              Missing complete          abs.   Low    High       label

   1 ASC      0      71      0  100.00     37.8     37.8  55.85   1.10   H  H
   2 NONE     0      50      0  100.00     15.6     15.6  55.85   1.10   H  K
   3 NONE     0      33      0  100.00     12.4     12.4  55.85   1.10   H  L
   4 NONE  -22.2 18038.9   718   98.50   307.58   307.71  33.45   1.10   J  I
   5 NONE    0.1   582.5   718   98.50     5.22     5.22  33.45   1.10   Q  SIGI
   6 NONE    0.0  1343.1   718   98.50   118.68   118.68  33.45   1.10   F  F
   7 NONE    0.3    40.9   718   98.50     3.01     3.01  33.45   1.10   Q  SIGF
   8 NONE    0.0     9.0     0  100.00     4.52     4.52  55.85   1.10   I  FreeR_flag


 No. of reflections used in FILE STATISTICS    47908
```

b) With xdsconv

Have a look XDSCONV.INP
run xdsconv (type xdsconv in the terminal window).
xdsconv generate a reflection file (ascii format, name specified in the "OUTPUT_FILE=" command of XDSCONV.INP) and an input file for f2mtz: F2MTZ.INP
then look at the XDSCONV.LP file (or terminal) to know what to do next in order to create the mtz file. e.g.:
type on your terminal:

```
f2mtz HKLOUT temp.mtz<F2MTZ.INP
cad HKLIN1 temp.mtz HKLOUT lyso-1-1-A-F.mtz<<EOF
LABIN FILE 1 ALL
DWAVELENGTH FILE 1 1    0.97926
END
EOF
```

Note: freeRflag (mandatory for several ccp4 programs) may be added at the xdsconv step (un-comment !GENERATE_FRACTION_OF_TEST_REFLECTIONS=0.1) or with the freerflag program (ccp4):
freerflag HKLIN lyso-1-1-A-F.mtz HKLOUT lyso-1-1-A-Ffree.mtz<< EOF
freerfrac 0.05
end
EOF