XDS/AutoProcess Tutorial

This tutorial will proceed in two parts. In Part I, we will learn how to run XDS from scratch, and how to diagnose problems with the dataset. In part II, we will use AutoProcess to process another dataset automatically. Participants should be able to go through the tutorial independently although the instructor will be available to guide the tutorial and provide assistance to those needing additional help. This tutorial requires some familiarity with Linux/Unix commands. If you are new to Linux, you can learn more at *https://ryanstutorials.net/linuxtutorial/*

Datasets

Thaumatin:

- 1800 frames collected on a Pilatus 6M. What is special about these data?
- tiny spots (1-2 pixel)
- the header is not accurate (we will try to diagnose the problem)
- located in the folder /users/school/thaum10k

Insulin S-SAD:

- 360 frames collected on a Rayonix CCD.
- located in the folder /users/school/insulin_s_sad/normal/

TIPS:

- ***bold-italics-monospaced*** means type a command into the terminal
- Use the remote Data Processing terminal for running XDS and XSCALE (***xds_par, xscale_par, pointless, etc***) on the HPC cluster, and a normal local terminal for editing and viewing files (***gedit, xds-viewer, etc***).

Part I: Getting Started with XDS (Thaumatin):

1. Create a working directory to run XDS in and change into the directory

   cd ~
   mkdir xds-tutorial
   cd xds-tutorial

2. Create an XDS.INP

   generate_XDS.INP "/users/school/thaum10k/thaum15_????.cbf"

OR

auto.inputs /users/school/thaum10k/thaum15_0001.cbf

3. Open the XDS.INP file using a text editor:

   gedit XDS.INP &

4. We will start by changing the spot-finding parameters, we will be asking XDS to use three sets of 5 degrees of data around 0°, 45° and 90° oscillation angles. We do this by editing the file. Change the **SPOT_RANGE=** parameter to (*add extra lines if only one is present in the file, if you used* ***auto.inputs****, no changes need to be made*)

```
SPOT_RANGE= 1 19
SPOT_RANGE= 225 243
SPOT_RANGE= 450 468
```

5. Run XDS using the command

   xds_par

6. For the thaumatin dataset, this will fail at the IDXREF step. Inspect the files IDXREF.LP to diagnose the problem. Use information from the lecture:
   - *Are cluster indices integers?*
   - *how many subtrees were found?*
   - *is the selected index origin optimal?*
   - *is there a clear solution?*
7. Try to find the correct values for ORGX= and ORGY= from IDXREF by inspecting the list of possible index origins. Copy the correct values to XDS.INP and rerun XDS using the

   xds_par

8. This time, IDXREF should succeed and the processing should proceed to the end. Inspect the files IDXREF.LP, INTEGRATE.LP, CORRECT.LP using "***gedit***" and view the image FRAME.cbf using adxv or xds-viewer.

   - *Are the 3D profiles in INTEGRATE.LP centered?*
   - *what do the final statistics look like?*
   - *is there any anomalous signal?*
   - *What is the correct space group?*

9. Run POINTLESS to check the space group using the command

   pointless –xdsin XDS_ASCII.HKL

10. Repeat the CORRECT step of XDS using the space group found by pointless. Edit XDS.INP and use the space group number and unit cell constants corresponding to the correct lattice type from the table at the beginning of CORRECT.LP. Don't forget to add a REIDX= keyword and specify the re-indexing transformation as well. Then run XDS again.

    - *Does the data look good?*

11. Try to optimize the processing
    a. Save a copy of CORRECT.LP for comparison using the command

    cp CORRECT.LP CORRECT.LP.1

    b. Copy the refined parameters GXPARM.XDS to the file XPARM.XDS, using the command

    cp GXPARM.XDS XPARM.XDS

    c. From INTEGRATE.LP Copy two lines into XDS.INP

```
***** SUGGESTED VALUES FOR INPUT PARAMETERS *****
BEAM_DIVERGENCE=  0.54954  BEAM_DIVERGENCE_E.S.D.=  0.05495
REFLECTING_RANGE= 0.83218  REFLECTING_RANGE_E.S.D.= 0.11888
```

    d. Re-run XDS

    xds_par

    - *Did the statistics in CORRECT.LP improve over those in CORRECT.LP.1?*
    - *Can you identify anything wrong with the data?*

12. Scale the data using XSCALE. Create an XSCALE.INP file in the same directory containing the following lines

    gedit XSCALE.INP &

```
OUTPUT_FILE=XSCALE.HKL
INPUT_FILE=XDS_ASCII.HKL
```

13. Run **XSCALE** and then inspect XSCALE.LP to assess the quality of the dataset.

    xscale_par

14. Use pointless to convert XSCALE.HKL to MTZ format for use with PHENIX or CCP4

    pointless –copy –xdsin XSCALE.HKL hklout thaum.mtz

Part II: AutoProcess

15. Process the insulin_s_sad dataset using AutoProcess

    auto.process -a /users/school/insulin_s_sad/normal/normal_1_002.img

16. The output will be in the sub-directory 'normal_1-anom' Inspect the HTML report using the Firefox browser

    cd normal_1-anom
    firefox report.html

17. Optimize the processing and restrict the resolution to 2 Angstroms and check the updated report

    auto.integrate --o --r 2.0
    firefox report.html