X-ray crystallography practical

Oleron 2018

Last update: 21 Mai 2018

Table of Contents

Schedule of the X-ray crystallography practical

Monday June 4, 16:30 - 19:00 & 20:45 - 21:45

Presentation of the practical, group constitution (~10 min)

Data processing (~45 min)

Data :
15-05-13-lyso/lyso-Gd_SAD (anomalous Gd, 300 frames, 1.65 Å)
Soft : xds, xdsgui
Support : tutorial_xds

SAD phasing, phase improvement & automated model building (~45 min)

Données : lyso-Gd_SAD
Soft : ccp4
Support : tutorial_SAD, tutorial_SAD_bis

Completing automatically built model / refinement (~60 min)

Data : lyso-Gd_SAD, partial lysozyme model
Soft : coot, refmac
Support : tutorial_SAD, tutorial_SAD_bis

Locating Gd atoms, completing Lysozyme-Gd model (~60 min)

Data : lyso-Gd_SAD, partial lysozyme model
Soft : coot, refmac
Support : tutorial_SAD, tutorial_SAD_bis

Data collection on a Gd derivative on beamline FIP

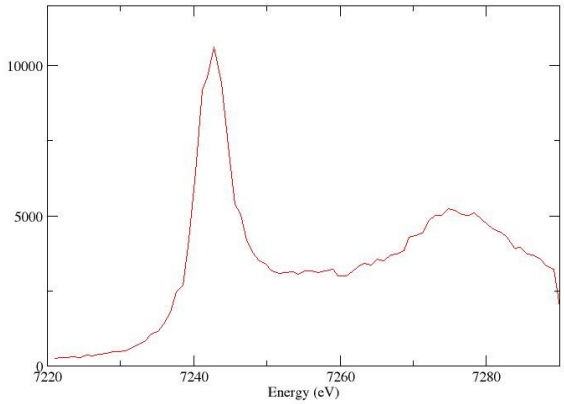
Reason for the choice of gadolinium:
at the LIII edge (λ = 1.711 Å) f'' = 28 e-, at λ = 1.54, f'' = 12e-

The fluorescence of Gd was measured with a Roentec MCA at the Gd LIII edge. Raw data are in

`Edge/lyso_1_Gd1 (columns 5 and 7)`

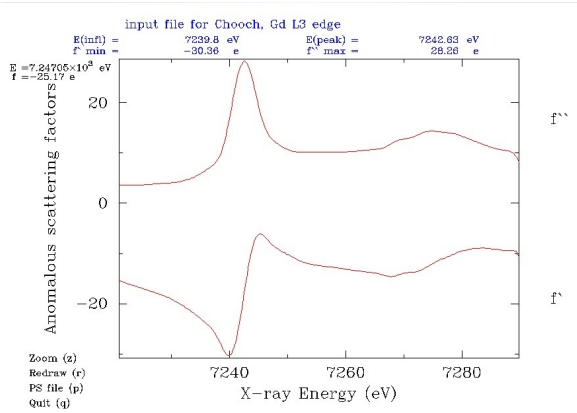
and the plot vs Energy in

`Edge/lyso_1_Gd1.jpg`



The spectrum was processed with Chooch. Final drawing of calculated f' and f'' is in

`Edge/final.jpg`



Based on that, beam energy was tuned to 7242.6 eV, and a single-wavelength dataset was collected (300 frames, 1 deg each). Frames (compressed with bzip2) are named

`img/e000_prefix_1_00xxx.img.bz2`

Data reduction with XDS package

Introduction

XDS is a suite of programs dedicated to the reduction of macromolecular crystallography data. The suite of programs includes:

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.
xdsconv: 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**, **xscale** and **xdsconv** will be used here.

xds requires
- diffraction images
- a parameter file called XDS.INP that contains all the necessary information regarding the experimental setup.

Most of the time, an XDS.INP file is generated automatically when you launch a data collection at a synchrotron. However, the file XDS.INP needs some editing during the data processing, but only a few input parameters require to be looked at. See the commented XDS.INP file for further details.

The whole data processing includes 7 steps define in the JOB= command line. Each step generates a log file named with the .LP suffix.

XYCORR: computes a table of spatial correction values for each pixel: allows to precisely localise 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 from COLSPOT 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 occurring in the data images and estimates their intensities
files created:

INTEGRATE.HKL
INTEGRATE.LP

CORRECT: corrects intensities for decay, absorption and variations of detector surface sensitivity, merge symmetric observations (but do not store them) and reports statistics of the collected data set and refines the diffraction parameters using all observed spots.
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

The different steps are presented in a series of directories, for sake of clarity. In practice, they can be performed in a single directory by successive modifications of the input files and running the XDS package programs at the command line. Another method is to use xdsgui (graphical interface for XDS) to run the different steps of XDS and have some graphics to checks data quality: this

interface will be used.

Before lauching xdsgui:

going the proper directory:
check that you are in the home directory by typing:
```
pwd
```
(/home/tp should be the result)
then type:
```
cd Data/RX/lyso-Gd_15May2013
```
list the content of this directory by typing:
```
ls
```

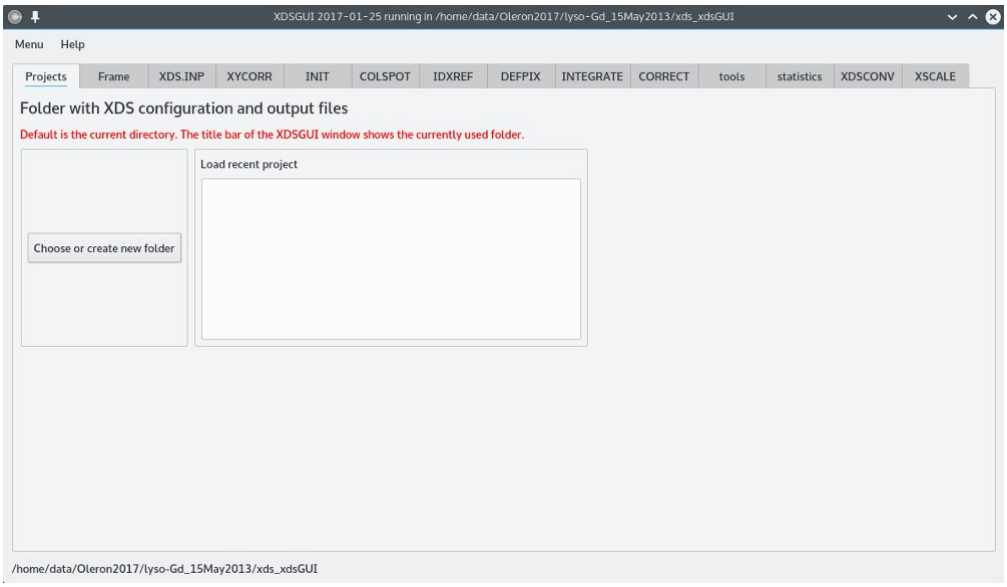
img directory contains diffraction images
xds_step0_default_XDSINP directory contains the XDS.INP input file necessary for data processing with XDS and an annotated input file for explanations (XDS.INP_sav)

now, dupplicate xds_step0_default_XDSINP and name it xds_2018:
```
cp -r  xds_step0_default_XDSINP  xds_2018
```
then go in the directory xds_2018:
```
cd xds_2018
```

Data reduction with XDS in graphic/automated mode

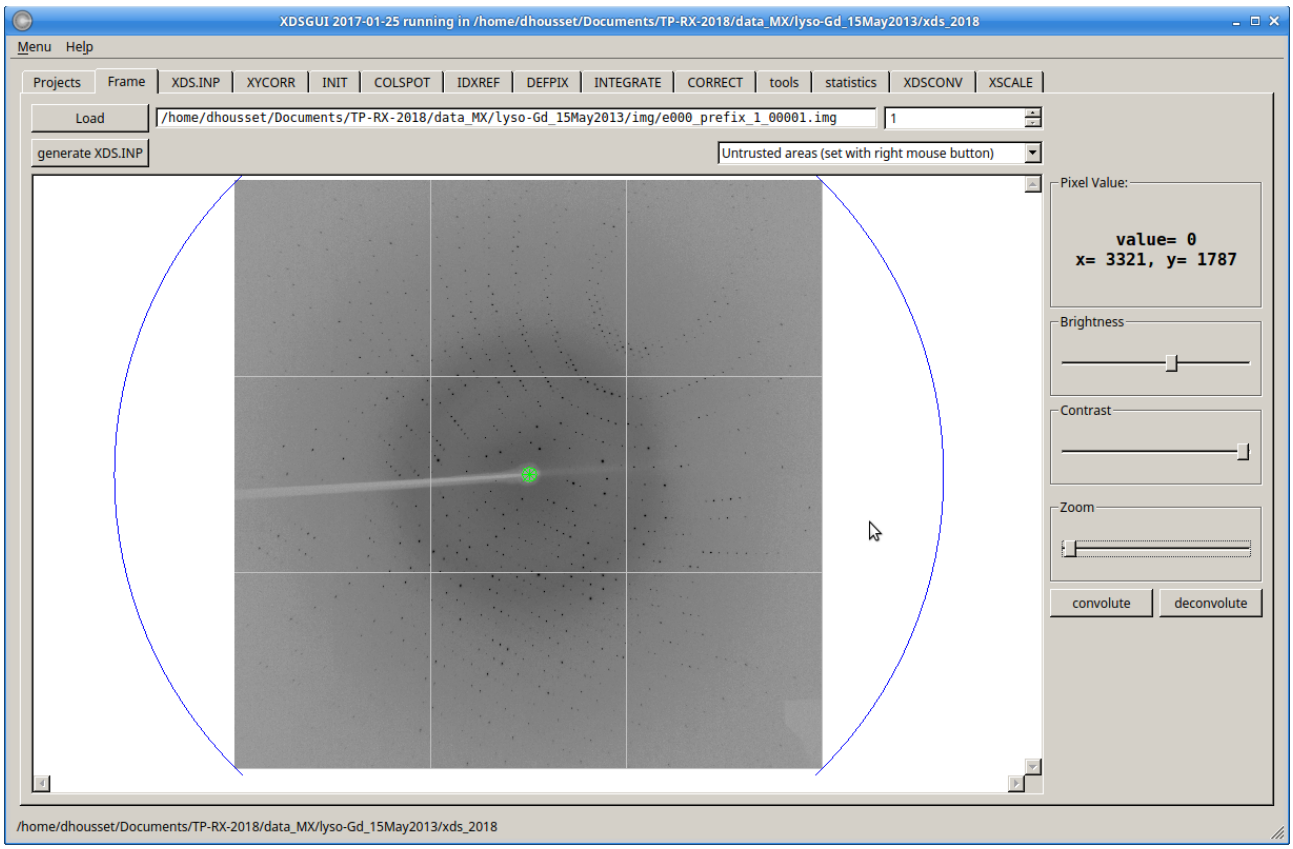
Graphic mode, using xdsgui

In xds_2018
Run xdsgui by typing:
`xdsgui &`



a) First click on “Choose or create a new folder”, and select xds_2018

b) Then look at one image by clicking on “Frame” tab, and then load (select e000_prefix_1_00001.img in the img folder)



Questions:

What is the darker ring in the middle of the diffraction image
Are there ice diffraction spots or rings?
Is there diffraction up to the edge of the detector?
Does the crystal seem to be unique?

c) Edit XDS.INP file by clicking on the “XDS.INP” tab.

From there, all steps described above can be performed, starting with the edition of the XDS.INP parameter file.

As we expect anomalous signal, the Friedel mates will differ. So uncomment the line

```
FRIEDEL'S_LAW=FALSE
```

You may add a spot_range, in order to have 2 ranges, 90° apart: this should improve unit cell accuracy in IDXREF step

```
SPOT_RANGE=91 110
```

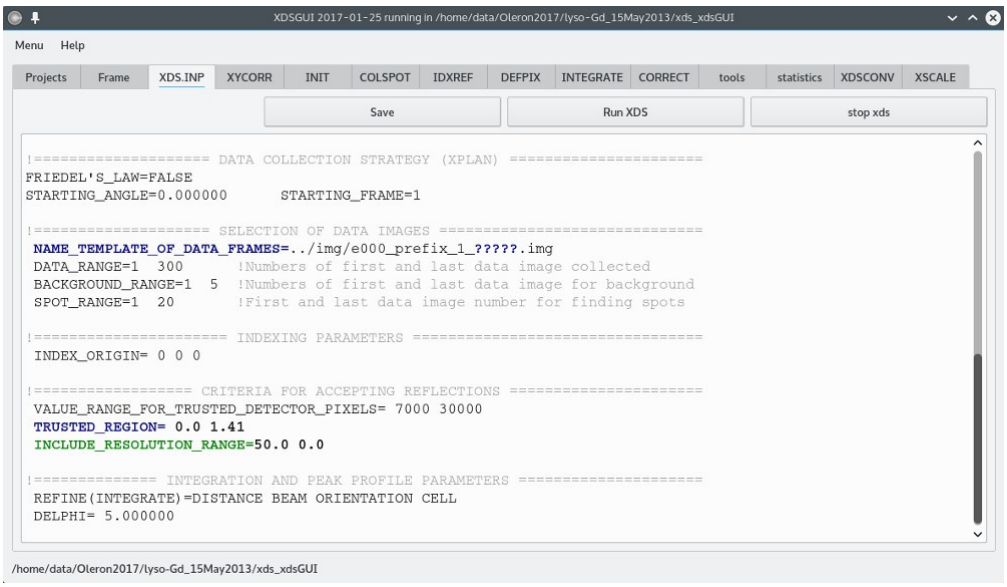
You may choose which parameters are refined during IDXREF, INTEGRATE and CORRECT steps by adding (if not already present) the following commands:

```
REFINE(IDXREF)=BEAM ORIENTATION CELL AXIS POSITION
REFINE(INTEGRATE)=DISTANCE BEAM ORIENTATION CELL
REFINE(CORRECT)=POSITION BEAM AXIS ORIENTATION CELL
```

If not specified, default values are used (may change with versions of XDS). For these data it is important to refine crystal-detector distance at the IDXREF step, as the value provided in XDS.INP is not very accurate. Thus, it may be wise to add:

REFINE(IDXREF)=BEAM ORIENTATION CELL AXIS POSITION

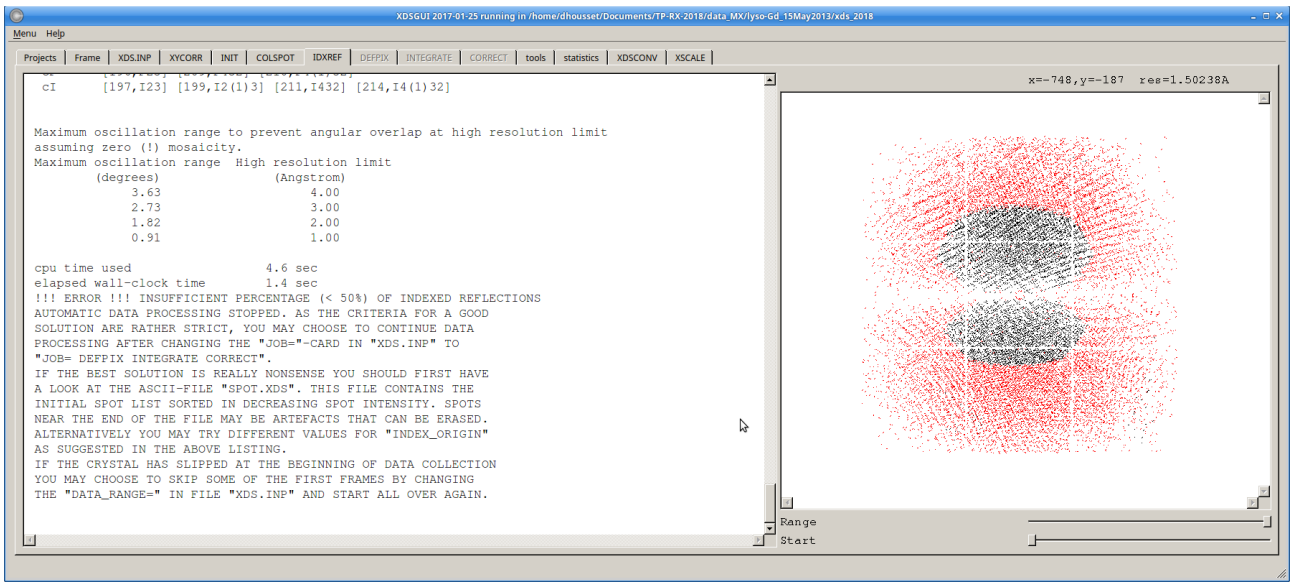
Once all the desired changes are made, click “Save” and “Run XDS”



tabs written in grey turn black once the task corresponding to the tab is completed.

When JOB= ALL is specified, all the tasks will be performed:
XYCORR INIT COLSPOT IDXREF DEFPIX XPLAN INTEGRATE CORRECT

However, if the indexation step (IDXREF) does not satisfy some criteria (less than 50% of the spots indexed, for example: check the terminal window or the IDXREF tab to see IDXREF log), the job will stop here.



It does not necessarily mean that the indexation fails (XDS criteria are known to be very strict), but may need inspection before continuing by replacing:

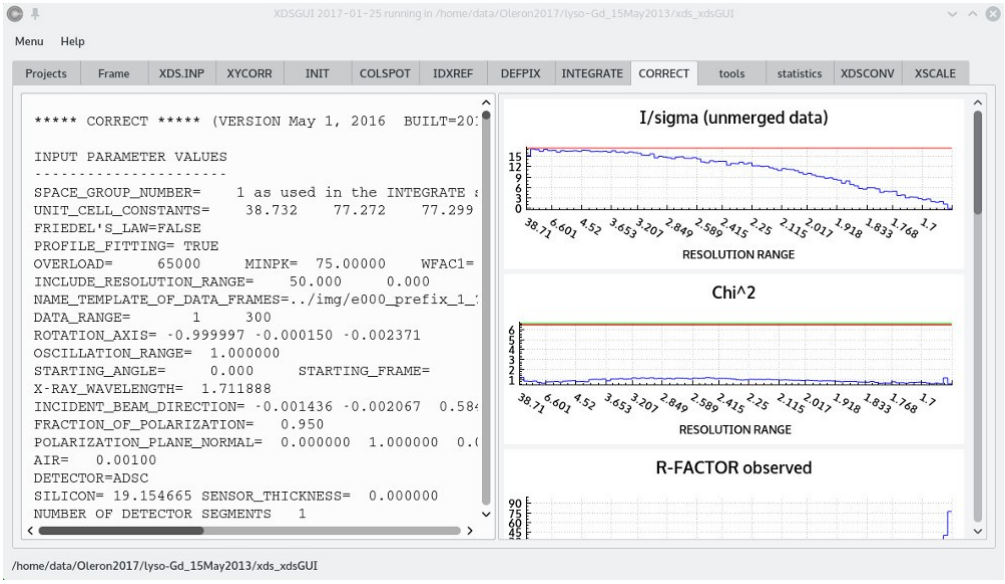
```
JOB= ALL
```

by :

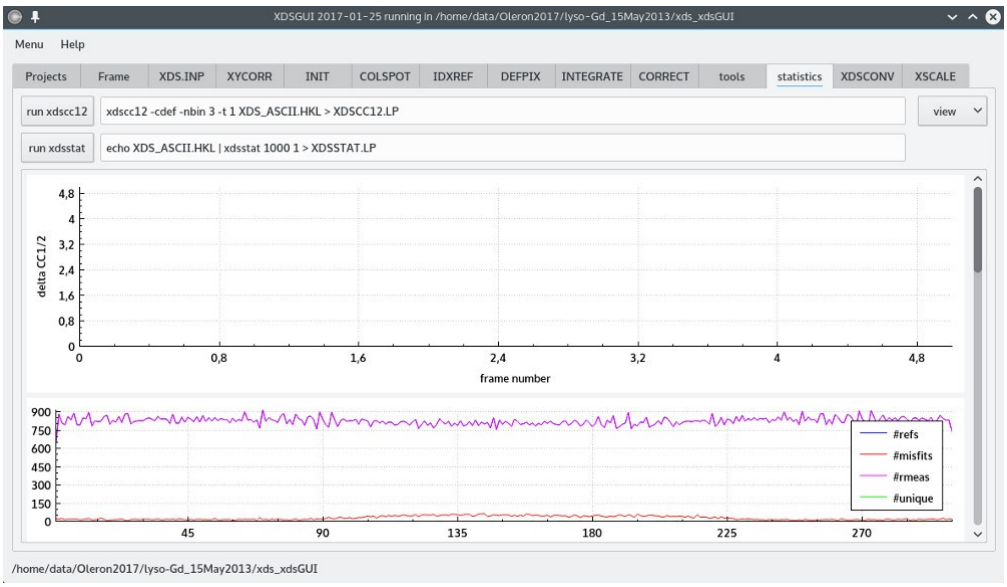
```
JOB= DEFPIX XPLAN INTEGRATE CORRECT
```

If this occurs, just update XDS.INP, save, and run XDS again.

Log files are available in the “XYCORR”, “INIT”, … windows, with graphical display of the statistics:



extra statistics are available running xdsstat in the “statistics” window:



Once this first run of data processing is completed, check the following output:

IDXREF for indexation
INTEGRATE for integration
CORRECT for scaling and merging

First look at the data statistics

Reminder: here, the data processing has been performed with no prior knowledge of the crystal symmetry. For IDXREF and INTEGRATE steps, the refinement of parameters was one assuming P1 space group.

Questions:
For all these steps, find out the relevant statistics and make your own opinion the data.
What are the possible Bravais lattices for your crystal and possible related space groups?
When no space group information is given, CORRECT is testing different Laue class.
What is the Laue class selected by CORRECT?
Is this choice OK?
What are the space groups compatible with the selected Laue class?
Check the information about possible systematic extinction in CORRECT output.
What are the possible space groups for these data?

If you wish to save this data processing, go on the "tools" tab, "Saving and comparing good results" and click on:
"backup files to ./save"

Optimization of data processing

You may improve data processing by introducing information on the crystal symmetry and refined cells and experimental setup parameters determined by CORRECT for the INTEGRATE step.

a) update cell and experimental setup parameters by:

go on the “tools” tab, “Optimizing data quality” and click on:
“copy latest geometry description over previous one”
“copy BEAM_DIVERGENCE, ...”

b) Go to XDS.INP tab, and replace

```
JOB= ALL
```

by :

```
JOB= DEFPIX XPLAN INTEGRATE CORRECT
```

also update space group and unit cell information:

After checking for extinctions (helices), you should have found out that you may either have space group 92 (P4(1)2(1)2) or space group 96 (P4(3)2(1)2): just enter the one you want in XDS.INP:

```
SPACE_GROUP_NUMBER=96
UNIT_CELL_CONSTANTS= 77.268   77.268   38.704  90.000  90.000  90.000
```

Then, click on “Save” and “Run XDS”

Questions:
Compare present and previous CORRECT output.
Has the data processing actually been improved?

Final data scaling and merging with XSCALE

Once the data are processed, XSCALE is used to scale and merge (to produce a file with unique reflections being the results of averaging all measurements equivalent by symmetry)
Go to XSCALE tab, update XSCALE.INP by adding below “OUTPUT_FILE=...” command:

```
FRIEDEL'S_LAW=FALSE
```

```
MERGE=TRUE
```

save & run xscale

Changing data format for ccp4

We need to provide ccp4 with data in a specific file format (named MTZ format, that is binary), while the file create by XSCALE is an ascii file. Moreover, for historical reason, the anomalous information may be stored in two ways:
a) explicitly providing F^+ and F^-
b) providing the anomalous difference F^+ - F^- (named Dano)
Since different programs within the ccp4 suite use either F^+ and F^- or Dano, we should have both in our MTZ file.

a) Use XDSCONV to generate reflection files in CCP4 FP/DANO format (F, SigF, Dano, SigDano): goto XDSCONV tab and update XDSCONV.INP:

```
INPUT_FILE=lyso-Gd.ahkl XDS_ASCII
OUTPUT_FILE=temp_ccp4.hkl  CCP4
FRIEDEL'S_LAW=FALSE
```

save and run XDSCONV

b) XDSCONV generates the input file F2MTZ.INP needed by f2mtz (CCP4 package) for the final conversion to binary mtz format. To run the CCP4 programs f2mtz just type the command:

```
f2mtz HKLOUT temp_ccp4.mtz < F2MTZ.INP
```

c) Use XDSCONV again to generate reflection files in CCP4 F+/F- format (F, SigF, F+, SigF+, F-, SigF-):
goto XDSCONV tab and update XDSCONV.INP:

```
INPUT_FILE=lyso-Gd.ahkl   XDS_ASCII
OUTPUT_FILE=temp_ccp4_f.hkl CCP4_F
FRIEDEL'S_LAW= FALSE
GENERATE_FRACTION_OF_TEST_REFLECTIONS=0.05
```

save and run XDSCONV

d) Then run CCP4 program f2mtz

```
f2mtz HKLOUT temp_ccp4_f.mtz < F2MTZ.INP
```

e) Then, to run CCP4 programm cad (to convert indices to the CCP4-asymmetric unit),

```
cad HKLIN1 temp_ccp4.mtz HKLIN2 temp_ccp4_f.mtz HKLOUT Lyso-
Gd_SAD.mtz <<EOF
  LABIN FILE 1 E1=FP   E2=SIGFP   E3=DANO E4=SIGDANO E5=ISYM
  LABIN FILE 2 E1=F(+) E2=SIGF(+) E3=F(-) E4=SIGF(-) E5=FreeRflag
  END
EOF
```

Lyso-Gd_SAD.mtz is the file you will be using in ccp4

Quick SAD phasing with Phaser in ccp4i

Go back to *Data/RX/lyso-Gd_15May2013:*
```
cd ..
```

Create a new directory ccp4_2018, move there
```
mkdir ccp4_2018
cd ccp4_2018
```

Copy **Lyso-Gd_SAD.mtz:**
```
cp ../xds_2018/Lyso-Gd_SAD.mtz .
```

Important Warning: SHELX does not like too long path and may fail is the ccp4 folder is too far in the directory tree. If this happens, you may have to create a symbolic link closer (ie in the $HOME (/home/tp) directory, with such a command:
```
cd /home/tp
ln -s /home/tp/Data/RX/lyso-Gd_15May2013/ccp4_2018 ccp4
```

And use this link in the ccp4 project (see below)

Launch
```
ccp4i &
```
and define a new project with
```
lyso-Gd_SAD/ccp4_SAD
```
as working directory

Phaser SAD Pipeline to determine experimental phases

In “Experimental Phasing” tab, select “Phaser SAD Pipeline” (button highlighted in blue in Figure 1) (~330 sec elapsed time).
Then enter the following parameters:
- reflexion file: `Lyso-Gd_SAD.mtz` (field highlighted in blue in Figure 2)
- sequence in fasta format directory ccp4_MR (field highlighted in green in Figure 2)
- heavy atom type: GD (field highlighted in red in Figure 2)
- wavelength: 1.7119 (field highlighted in orange in Figure 2)

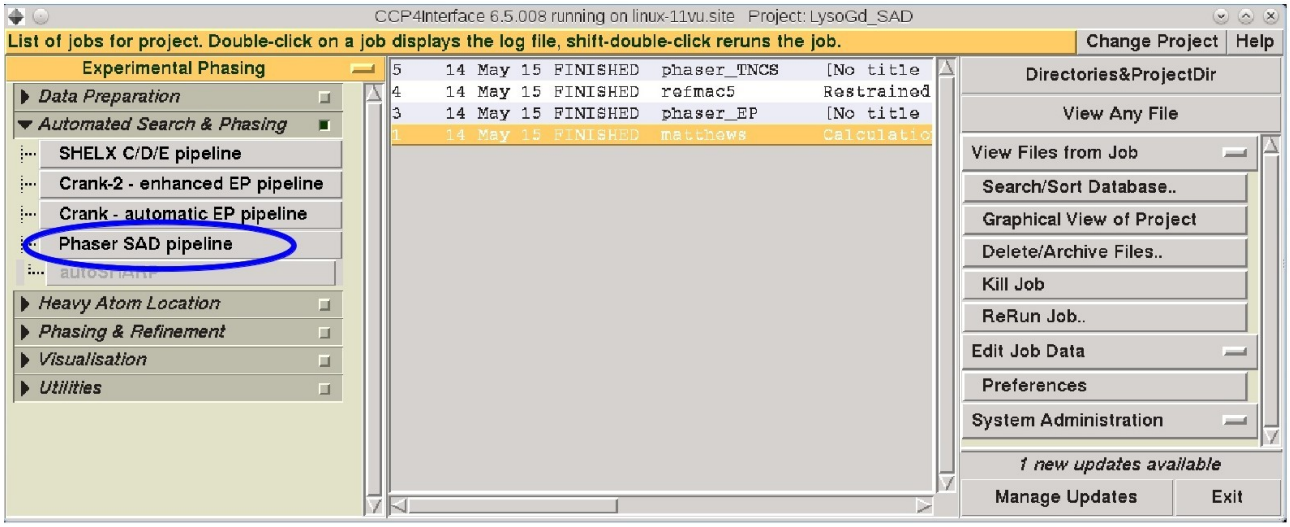


Figure 1

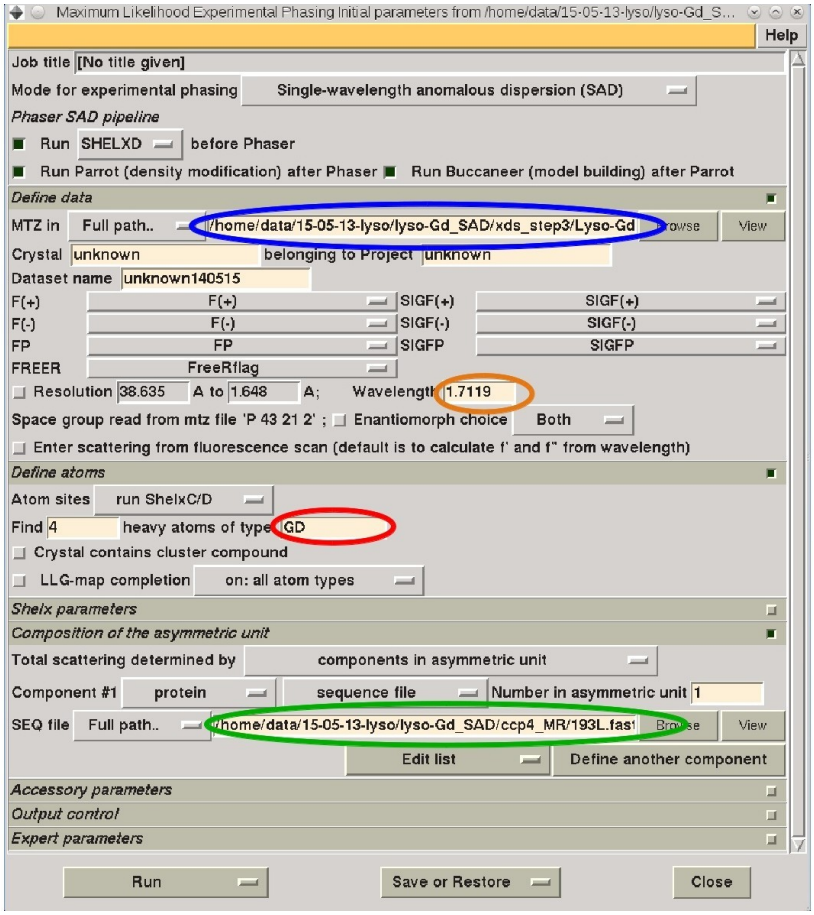


Figure 2

What does this pipeline ?

a) SHELXD will use the anomalous difference Patterson map and check the presence of peaks.

$$P_H(\vec{u}) = \sum_{h,k,l} \left(|F_{PH}(\vec{s})| - |F_{PH}(-\vec{s})|\right)^2 \exp[-2i\pi \vec{u}.\vec{s}]$$

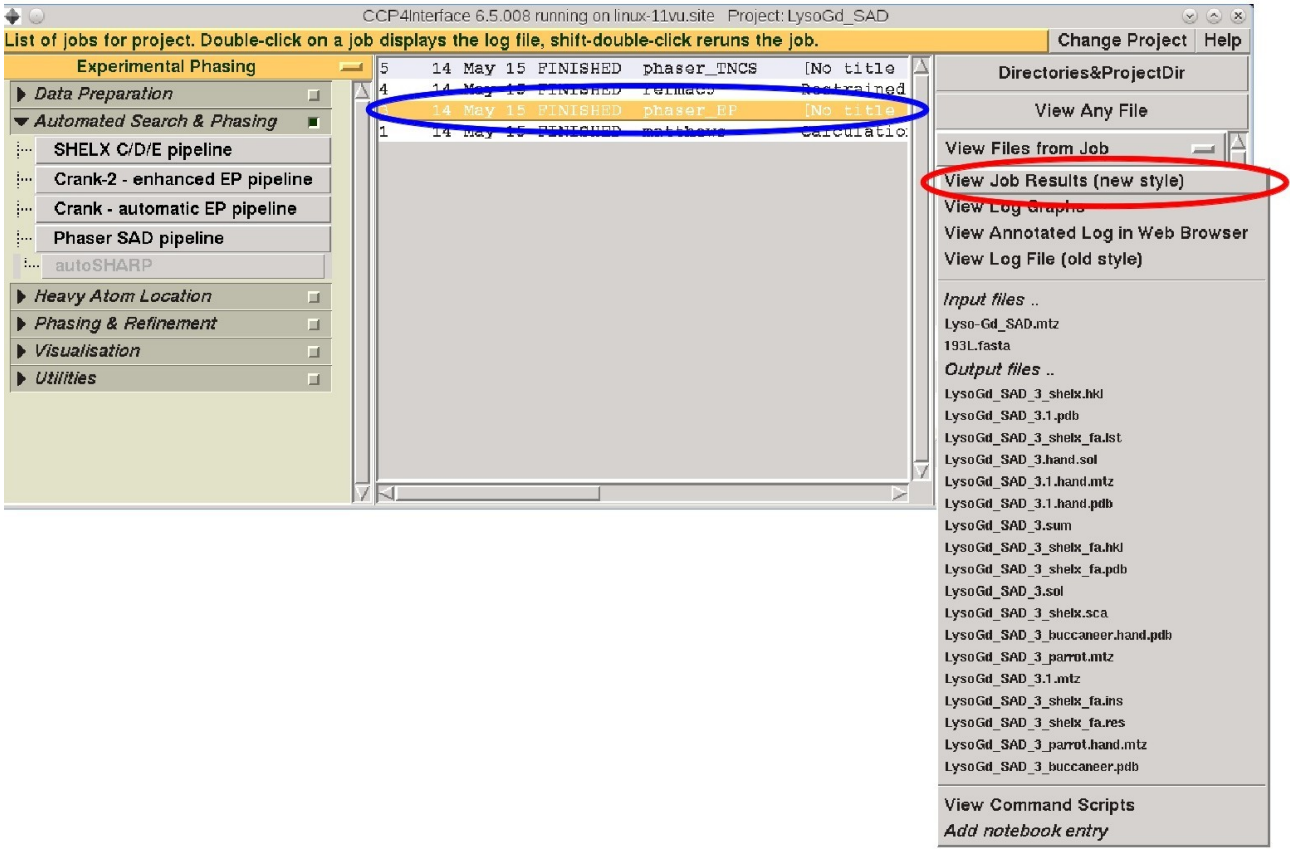
from the position of these peaks in the anomalous difference Patterson map, the position of Gd atoms in the asymmetric unit will be calculated (deconvolution of the Patterson map).

b) PHASER will calculate the experimental phases thanks to the Gd atoms located by SHELXD

c) PARROT will improve the phases through several density modification methods (solvent flattening, …)

d) BUCCANEER proceed to automated model building from the electron density map generated by PARROT and the amino-acid sequence sequence provided in input)

=> ~80% of residues built automatically
Upon completion of the job, and to analyze the log file, select the "Phaser_EP" job in the list (button highlighted in blue in Figure 3). Then, from the "View Files from Job", select "View Job Results (new style)" (button highlighted in red below)



COOT for model building

Experimental map, sub-structure of anomalous atoms and model can be displayed with Coot:
Run

`coot`

at the command line, and load pdb files (button highlighted in blue in Figure 4) and mtz files (button highlighted in red in Figure 4) as listed below:

Sub structure of Gd atoms is in

`LysoGd_SAD_3.1.pdb`

Experimental map coefficients are in

`LysoGd_SAD_3.1.mtz`

and after automated density modification with `parrot`

`LysoGd_SAD_3_parrot.mtz`

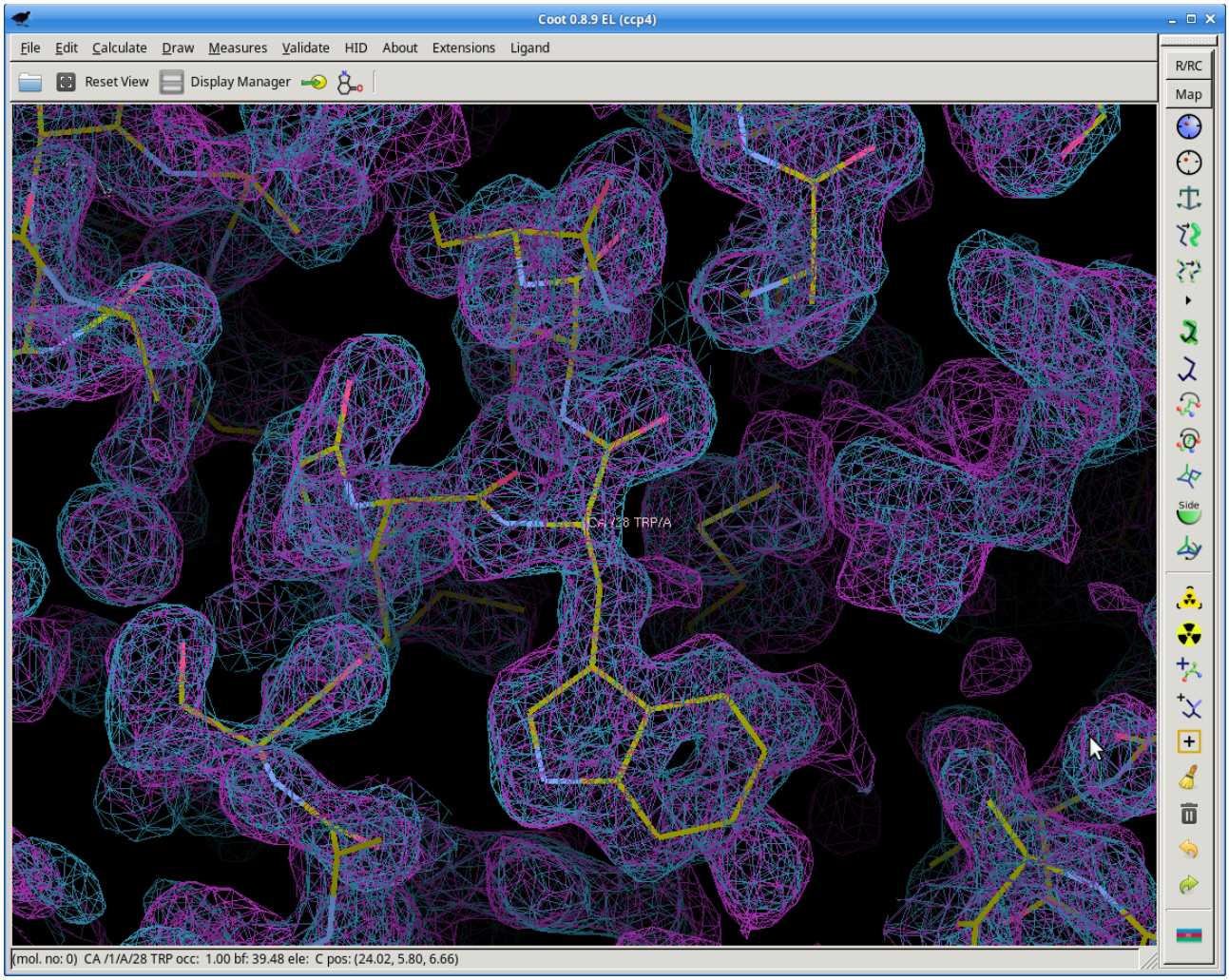
The model built automatically (80% of the residues) is available in

`LysoGd_SAD_3_buccaneer.pdb`

When running COOT with `LysoGd_SAD_3_parrot.mtz and LysoGd_SAD_3_buccaneer.pdb`, two maps are shown:

Map coefficients FWT and PHWT essentially correspond to F_{obs} and φ_{exp} , as calculated by PHASER (shown in purple below). A standard contour level is +1σ.

Map coefficients parrot.F_phi.F and parrot.F_phi.phi correspond to the improved experimental map, as calculated by PARROT (shown in cian below). A standard contour level is +1σ.



Questions:

compare both maps: which is the one that seems the easiest to build a model in it?

Start model building in the experimental map as model and the experimental map are good enough to start manual building.

Alternatively, run Refmac for a first refinement and manual rebuilt with

```
LysoGd_SAD_3_buccaneer.pdb
```

as pdb input file.

Warning: Depending on the initial choice of space group (92 or 96), the files above may be the one corresponding to correct or the wrong hand. You have to look also at

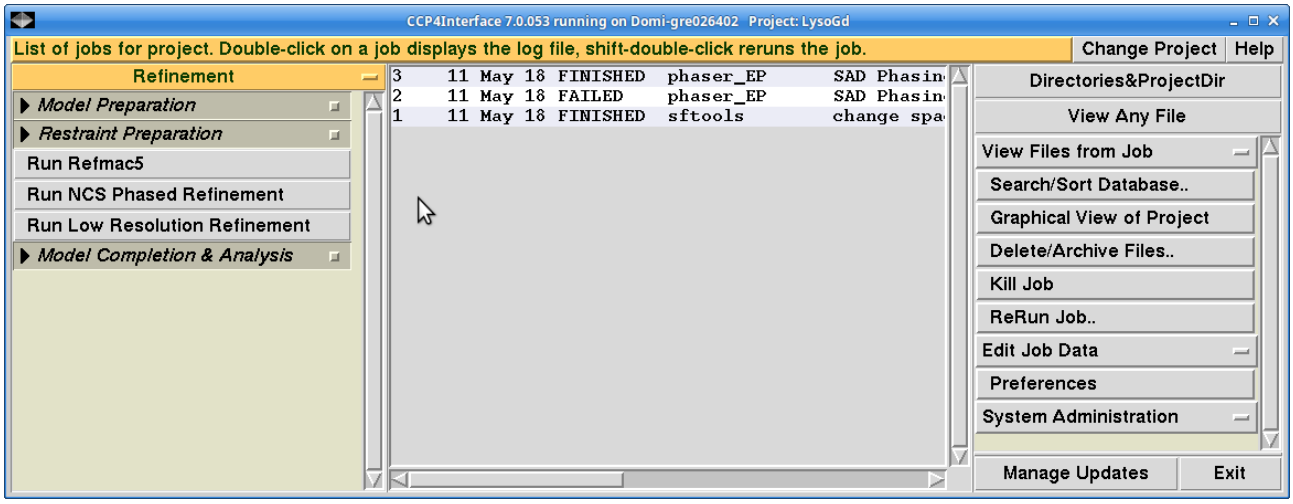
```
LysoGd_SAD_3_parrot.hand.mtz
LysoGd_SAD_3_buccaneer.hand.pdb
```

and check which is the correct hand and finalize space group determination.

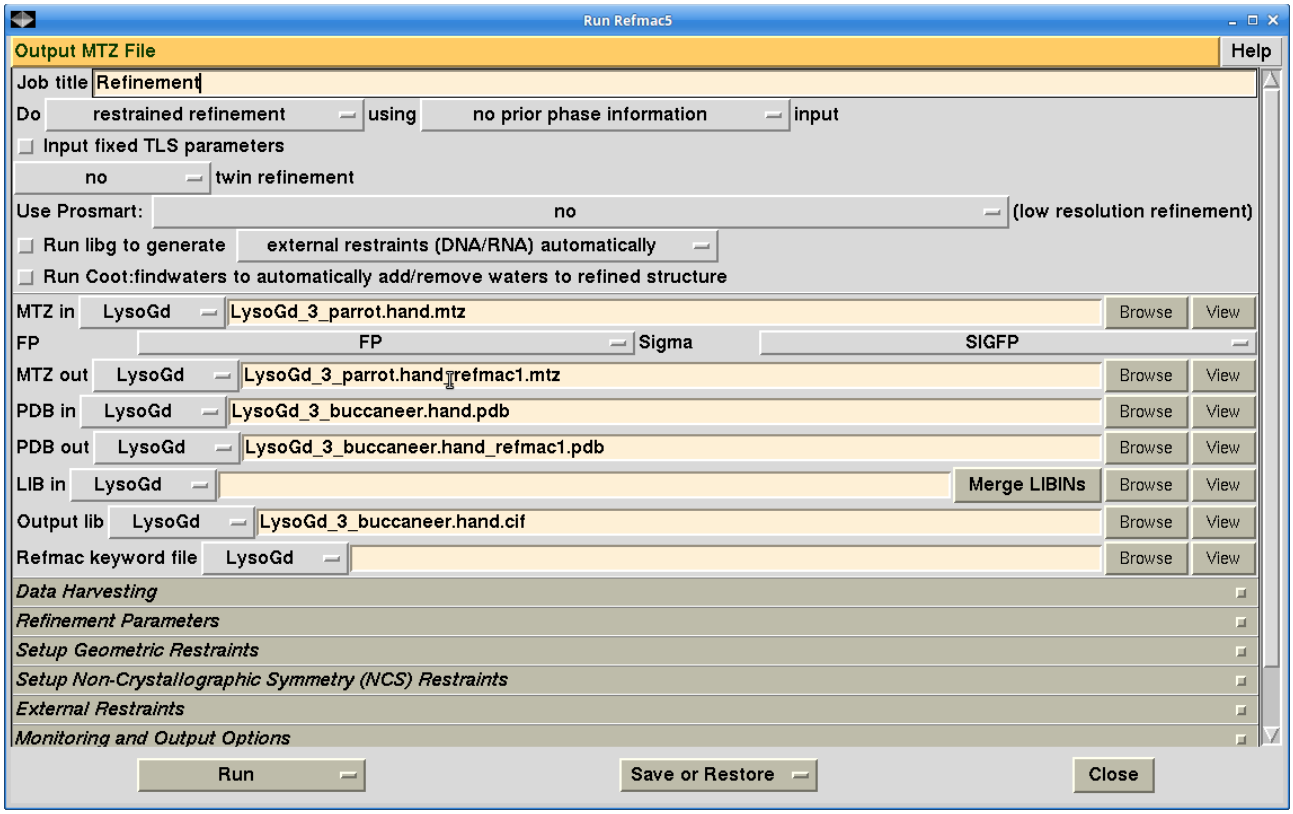
Refinement / construction

Using Refmac5 to refine the (uncomplete) model

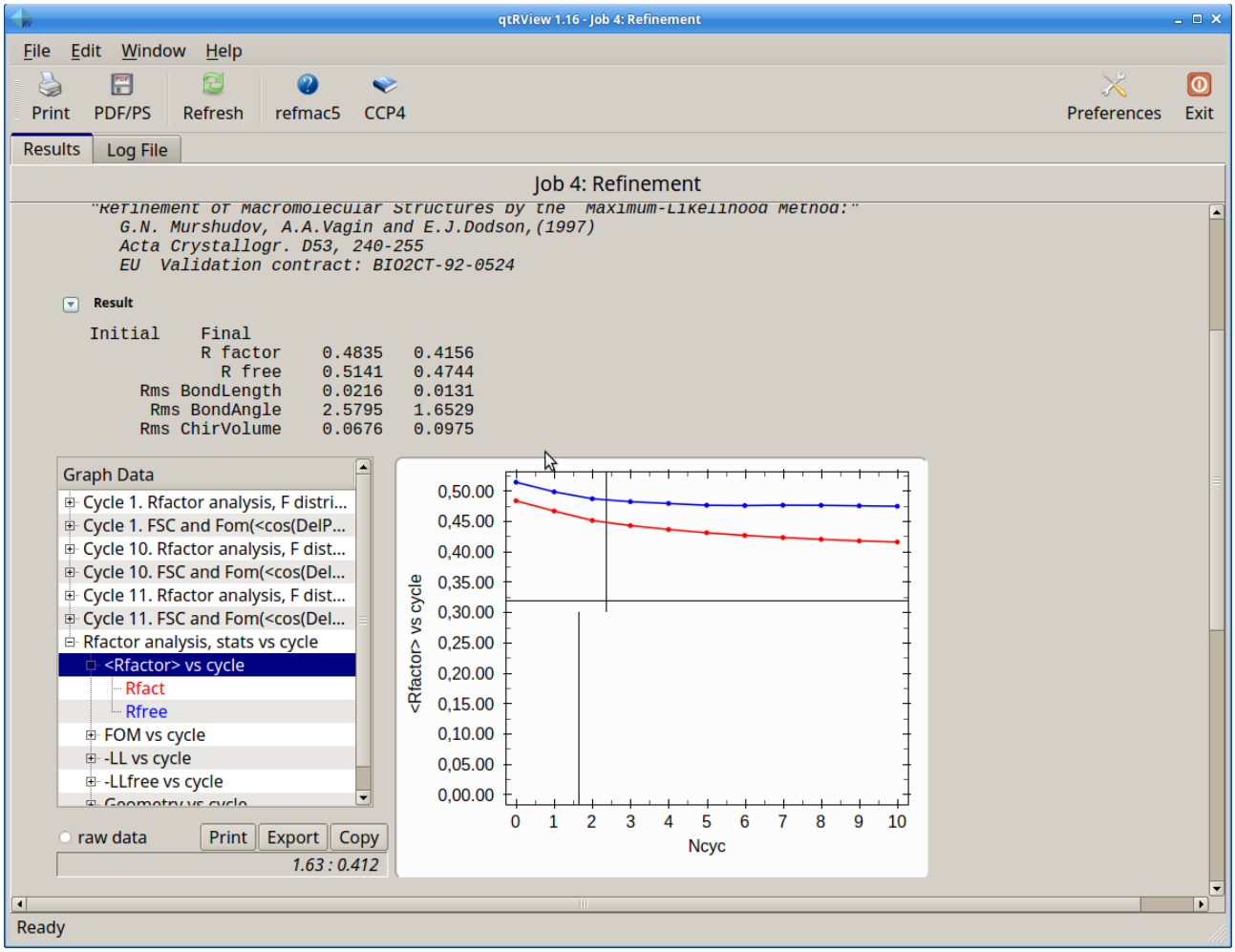
Based on the model built by BUCCANEER, and possibly completed by you (using COOT), you can initiate REFMAC to refine the structure of your model.
select Run Refmac5 tab in Refinement:



Refmac requires (i) an mtz file that contains structure factors or intensities and a free reflection set (FreeRflag): use the initial mtz file generated after XDS, and (ii) a pdb file of a model, that will be refined. After refinement, the refined model PDB file will be named as indicated on line « PDB out ». The amplitudes, and electron density map coefficient will be stored in the mtz binary file indicated on line MTZ out.



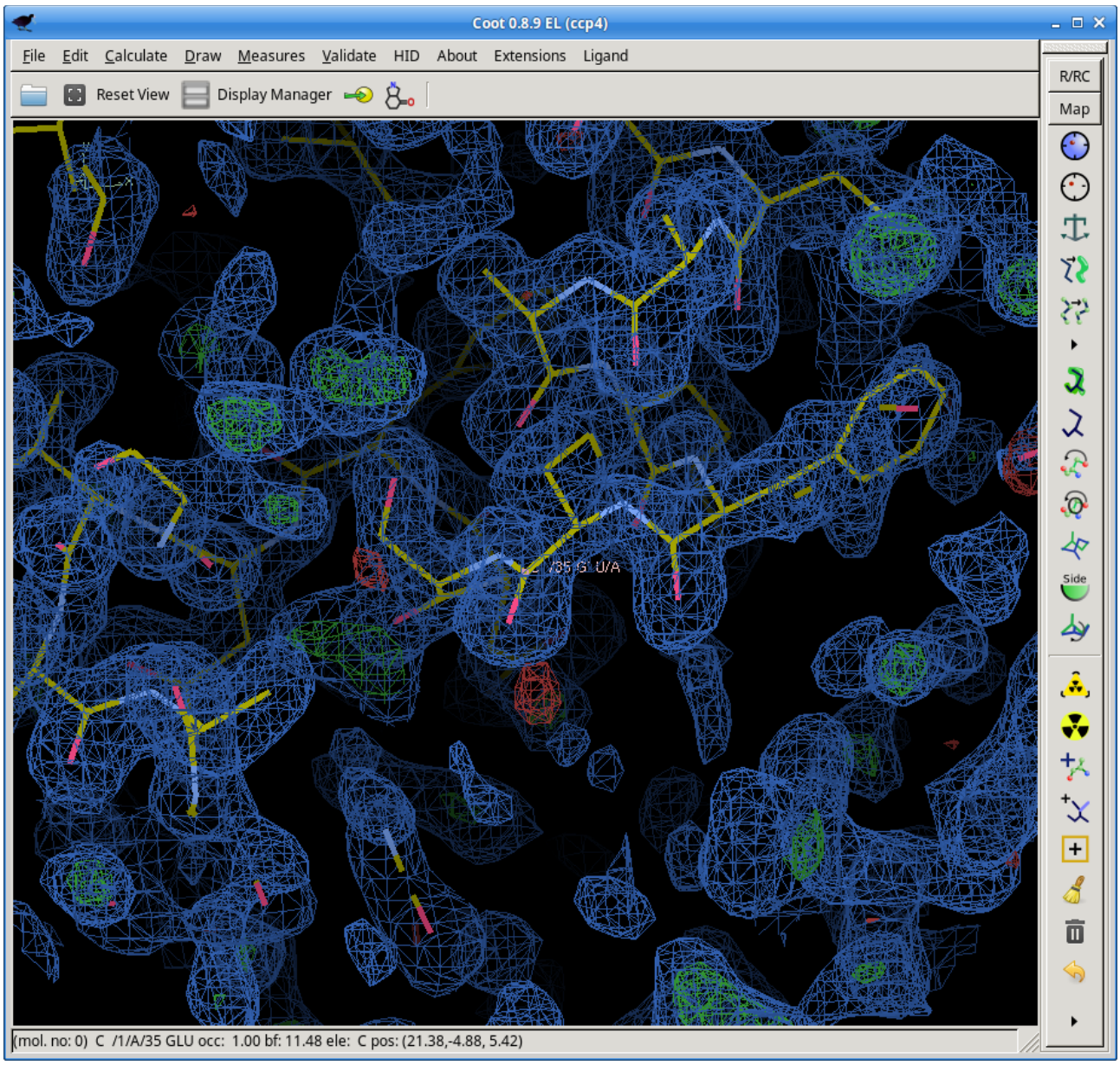
Press Run to start the calculation. As previously, you can open the output file by double click on the job line (or with the tab "View Files from Job"):



Questions:

What are the statistical criteria that provide information about the refinement behaviour?
What do you think about your refinement cycle?

Press the “Coot” tab in "Output files" section (bottom of the window) to open automatically the output files (pdb and mtz files) in coot.
Files automatically opened are the ones mentioned in MTZ out and PDB out



Important notice:
In this map, the phases are the ones derived from the model, and no longer the experimental ones.
Map coefficients FWT and PHWT essentially correspond to $2F_{obs} - F_{calc}$ and φ_{calc} (shown in blue above). A standard contour level is +1σ.
Map coefficients DELFWT and PHDELWT essentially correspond to $F_{obs} - F_{calc}$ and φ_{calc} (shown in green (positive) and red (negative) above). A standard contour level is +/- 3σ.
At some point in the refinement, the model phases become closer to the real phases than the experimental phases; It is up to you to decide when you think the model phases contains more information than the experimental phases.

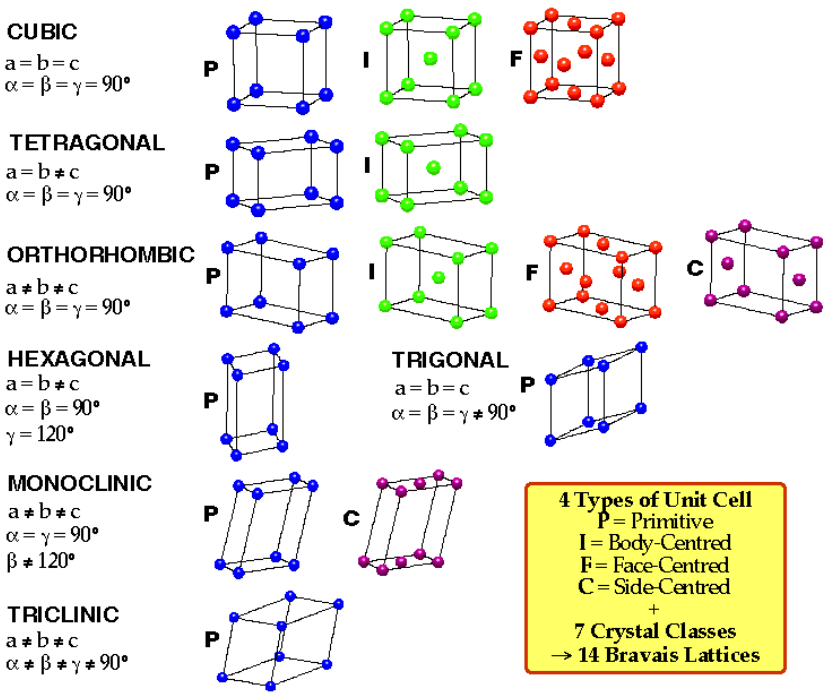
Questions:
What does a positive peak (green) in the difference Fourier map indicate?
What does a negative peak (red) in the difference Fourier map indicate?

Completing the model with COOT

Use COOT again to continue model building. Usually, numerous cycles of refinement and manual

model modifications are required to finalize the refinement process and obtain the most complete model, including solvent molecules, ligands, etc …

Other useful information

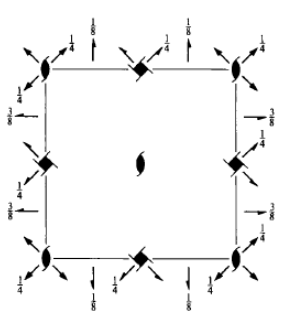


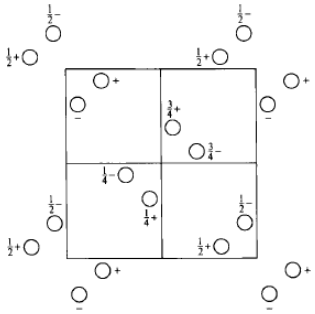
Crystal System	Minimum Symmetry*	Constraints on unit cell
Triclinic	None	None
Monoclinic	One 2-fold (along b)	$\alpha = \gamma = 90$
Orthorhombic	Three 2-folds (along a,b,c)	$\alpha = \beta = \gamma = 90$
Trigonal	3-fold (along c)	$a = b$; $\alpha = \beta = 90$; $\gamma = 120$
Tetragonal	4-fold (along c)	$a = b$; $\alpha = \beta = \gamma = 90$
Hexagonal	6-fold (along c)	$a = b$; $\alpha = \beta = 90$; $\gamma = 120$
Cubic	Four 3-fold axes (along body diagonal)	$a = b = c$; $\alpha = \beta = \gamma = 90$

System	Laue class	Space Groups
Triclinic	1	P1
Monoclinic	2	P2, $P2_1$, C2
Orthorhombic	222	P222, $P222_1$, $P2_12_12$, $P2_12_12_1$, $C222_1$, C222, F222, I222, $I2_12_12_1$
Quadratic	4 422	P4, $P4_1$, $P4_2$, $P4_3$, I4, $I4_1$, P422, $P42_12$, $P4_122$, $P4_12_12$, $P4_222$, $P4_22_12$, $P4_322$, $P4_32_12$, I422, $I4_122$
Trigonal	3 32	P3, $P3_1$, $P3_2$, R3, P312, P321, $P3_112$, $P3_121$, $P3_212$, $P3_221$, R32
Hexagonal	6 622	P6, $P6_1$, $P6_5$, $P6_2$, $P6_4$, $P6_3$, P622, $P6_122$, $P6_522$, P622, $P6_422$, $P6_322$
Cubic	23 432	P23, F23, I23, $P2_13$, $I2_13$, P432, $P4_232$, F432, $F4_132$, I432, $P4_332$, $P4_132$, $I4_132$

$P4_32_12$ D_4^8 422 Tetragonal

No. 96 $P4_32_12$ Patterson symmetry $P4/mmm$





Origin on $2\,[1\,1\,0]$ at $2_1\,1\,(1,2)$

Asymmetric unit $0 \le x \le 1;\quad 0 \le y \le 1;\quad 0 \le z \le \frac{1}{8}$

Positions

Multiplicity, Wyckoff letter, Site symmetry	Coordinates	Reflection conditions
8 b 1	(1) x,y,z (2) $\bar{x},\bar{y},z+\frac{1}{2}$ (3) $\bar{y}+\frac{1}{2},x+\frac{1}{2},z+\frac{3}{4}$ (4) $y+\frac{1}{2},\bar{x}+\frac{1}{2},z+\frac{1}{4}$ (5) $\bar{x}+\frac{1}{2},y+\frac{1}{2},\bar{z}+\frac{3}{4}$ (6) $x+\frac{1}{2},\bar{y}+\frac{1}{2},\bar{z}+\frac{1}{4}$ (7) $y,x,\bar{z}$ (8) $\bar{y},\bar{x},\bar{z}+\frac{1}{2}$	General: $00l:\ l=4n$ $h00:\ h=2n$
		Special: as above, plus
4 a . . 2	$x,x,0$ $\bar{x},\bar{x},\frac{1}{2}$ $\bar{x}+\frac{1}{2},x+\frac{1}{2},\frac{3}{4}$ $x+\frac{1}{2},\bar{x}+\frac{1}{2},\frac{1}{4}$	$0kl:\ l=2n+1$ or $2k+l=4n$