X-ray crystallography practical

Oleron 2017

Last update: 19 June 2017

Table of Contents

Schedule of the X-ray crystallography practical

Tuesday 20 June, 9:15AM-12:45AM (1h45 + 1h30)

9h00-9h15 : sponsors

Présentation du TP, constitution des groupes (~15 min)

Data reduction (~30 min)

Données : BM30A-2014-11-19 (native, 164 frames, 1.1 A) lyso poule
15-05-13-lyso/lyso-Gd_SAD (anomales Gd, 300 frames, 1.65 A)
Soft : xds, xdsGUI ou mosflm
Support : tutorial_xds ou tutorial_iMOSFLM

Phasage par MR (~30 min)

Données : BM30A-2014-11-19 , divers modèles : lyso incomplet de poule ou modèle complet “basse” homologie (boeuf)
Soft : phaser
Support : ?

Phasage SAD (~30 min)

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

Pause café (11h00-11h15)

Construction des parties manquantes / affinement (~45 min)

Données : BM30A-2014-11-19, modèle lyso incomplet
Soft : coot, refmac
Support : ?

Construction des parties manquantes / affinement (~45 min)

Données : lyso-Gd_SAD, modèle lyso incomplet, modèle ligand+Gd
Soft : coot, refmac
Support : tutorial_SAD, tutorial_SAD_bis

Repas (12h45)

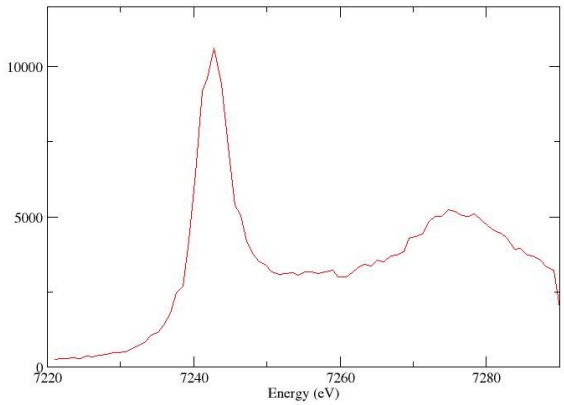
Data collection on a Gd derivative on beamline FIP

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)`

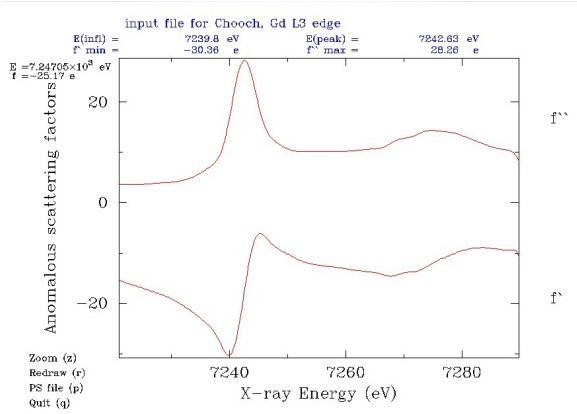
and the plot vs Energy in

`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 iMosflm

Start iMosflm

First open a terminal and go to the folder *andrea*:

```
cd Oleron2017_TP-MX/lyso-Gd_15May2013
```

Create a new folder

```
mkdir mosflm
```

and then move into this folder

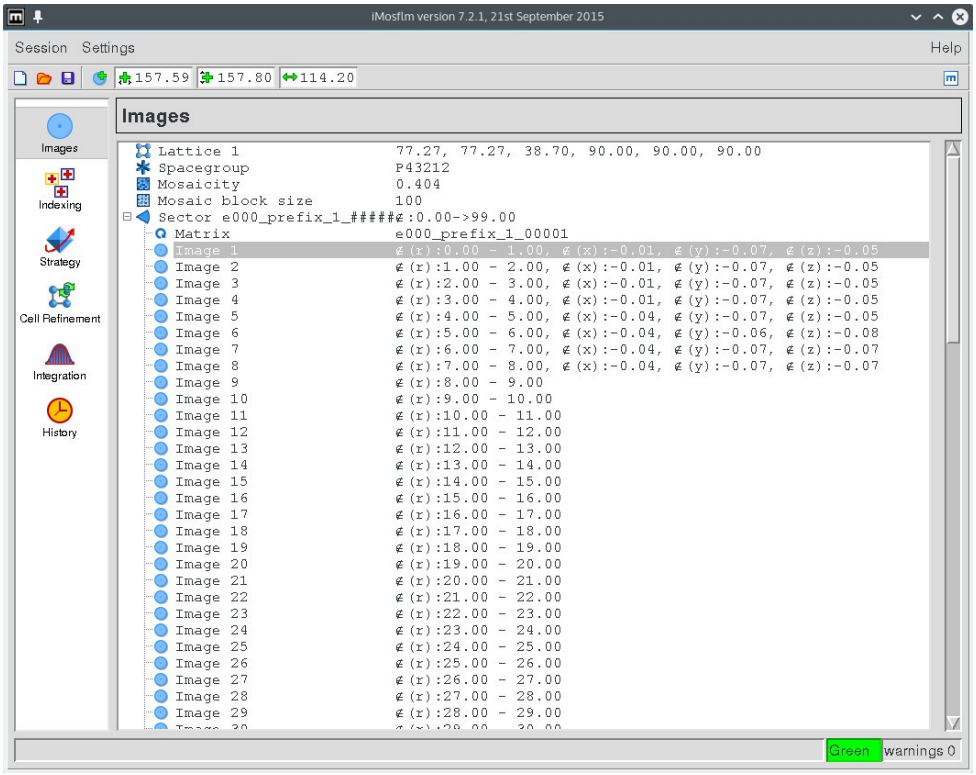
```
cd mosflm
```

Then run imosflm

```
imosflm
```

The main steps are:

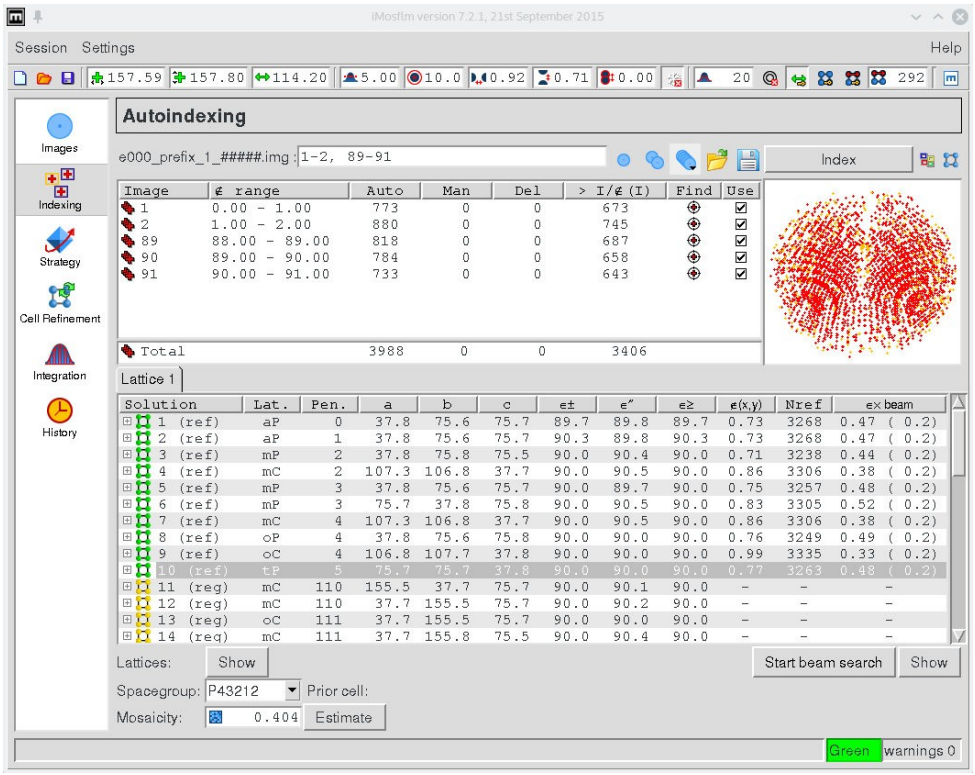
In “Settings → Experiment settings”, change select “Reverse direction of spindle rotation”
In “Session → Add images...” open the images (uncompress images) first.



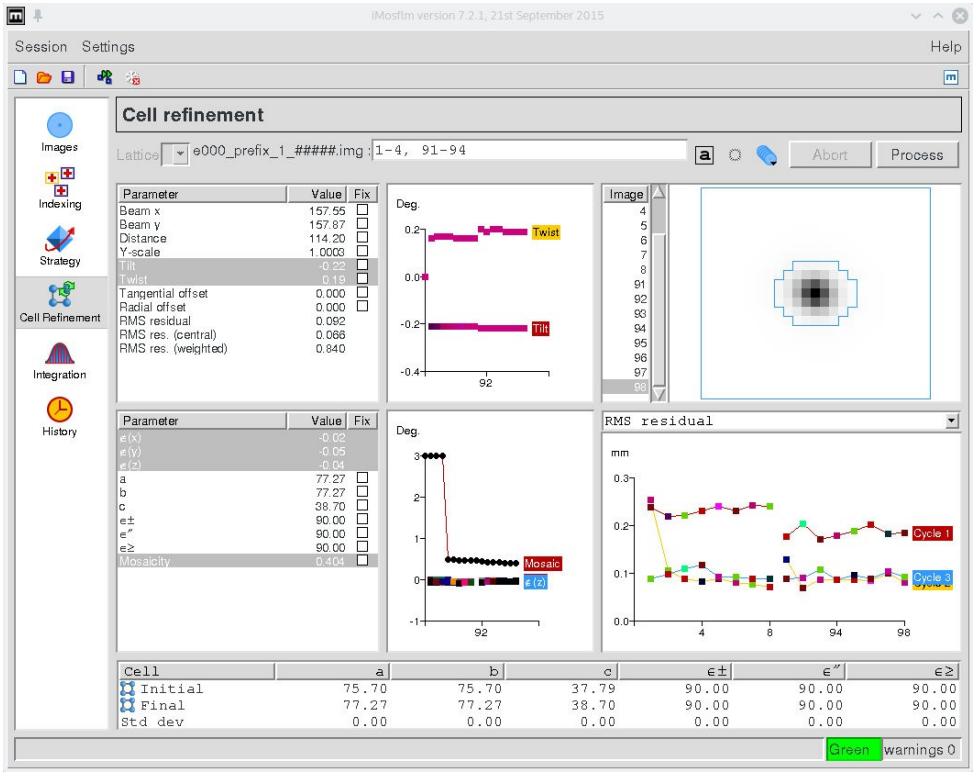
Indexation

Run “Indexing”. By default images 2 and 91 are used. You can use more by replacing “2, 91” by “2-6, 91-95”, for an exemple. Click on “Index” to redo the indexing.

Select the correct space-group: nb 96, $\boldsymbol{P}4_32_12$

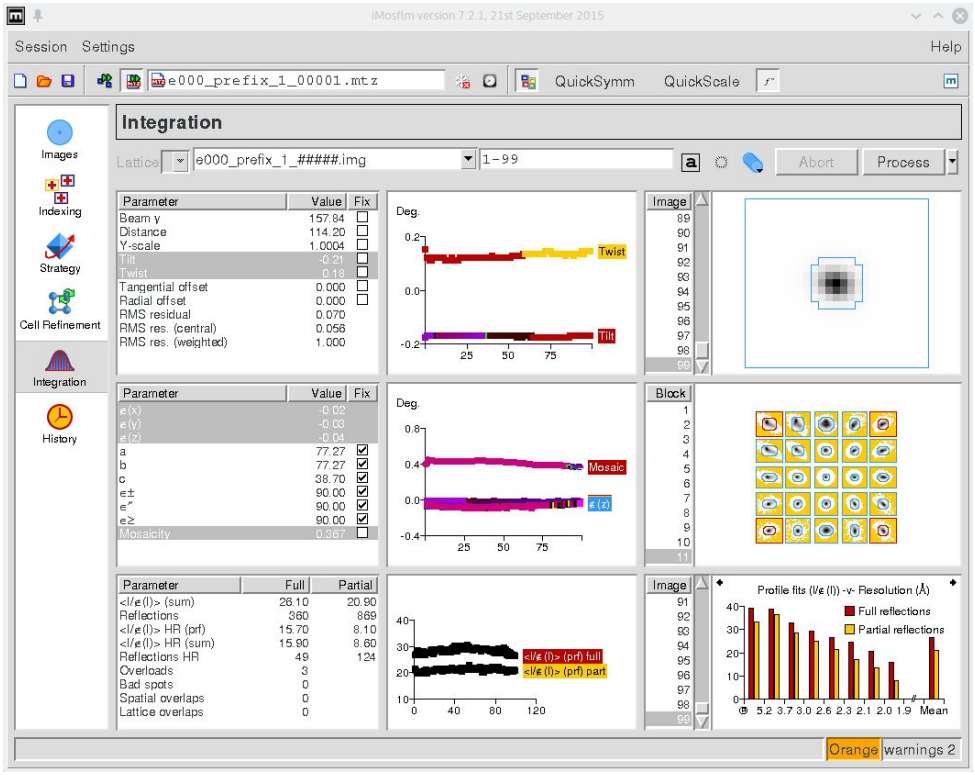


then run “Cell Refinement”



Integration and scaling

Select “Integration”



Then run “QuickSymm” (quick scaling to check the symmetry)
and “QuickScale” (scaling)
Check the Rmerge and Rmeas

Notably, compare Rmerge or Rmeas calculated either on:

all I+ and I- (includes the centrosymmetry (Fridel's law))

or

within I+ or I- (excludes the centrosymmetry (Fridel's law))

DeltaAnom correlation between half-sets tells you whether the anomalous signal is significant:
You divide your data into two sets, and compare the anomalous difference I+ - I- in the two sets. If the correlation is high, the signal is significant.

You can look also at the other figures on the table. All the integrated and scaled data are stored in the file **ctruncate_A5_1__00002-unique.mtz**
This file contains h, k, l, FreeRflag, F, SigF, Dano, SigDano, F(+), SigF(+), F(-), SigF(-), Isym, Imean, SigImean I(+), SigI(+), I(-), SigI(-)

For the following steps, you can use the **ctruncate_A5_1__00002-unique.mtz** file.

Data reduction with XDS package

Introduction

XDS is a suite of programs dedicated to the reduction of macromolecular crystallography 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.
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 your 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 generate a log file named with the .LP suffix.

XYCORR: computes a table of spatial correction values for each pixel: allow 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 inital 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 occuring 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, 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 modification of the input files and running the XDS package programs at the command line.

Default xds parameter files

Go to the directory **xds_step0_default-XDSINP**

You will find there the initial parameter file for xds, as created automatically by the beamline control software (simplified version)

```
xds_step0/XDS.INP
```

as well as a fully commented one

```
xds_step0/XDS.INP_sav
```

Run all step in a row

Go to directory **xds_step1_all** (783 sec elapsed time on a HP ElitBook 840 i5 notebook, with bzipped frames; 252 sec elapsed time with uncompressed frames)

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

```
FRIEDEL'S_LAW=FALSE
```

in XDS.INP. Then, run

```
xds
```

at the command line (or `xds_par` for the paralleled version).

Re-run scaling

Go to the directory **xds_step2_CORRECT** (60 sec elapsed time)

At the previous step, xds automatically figures out the Bravais symmetry, and picked up space group P422 as a representative. To check for extinctions (helices), just select space group 96 (P4(3)2(1)2) and enter refined cell parameters in XDS.INP

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

and select only the final scaling step of the processing (CORRECT)

```
JOB= CORRECT
```

Then run

```
xds
```

In CORRECT.LP, check for low intensity of reflections that should be absent (marked with "*") in the list above lines

```
AVERAGE INTENSITY FOR     207 REFLECTIONS WHICH SHOULD
BE SYSTEMATICALLY ABSENT IS     0.2% OF MEAN INTENSITY
```

xscale and xdsconv

Go to the directory **xds_step3_XSCALE-XDSCONV** (5.6 sec elapsed time)

Optional: use XSCALE for final scaling, merging of several dataset. Create the XSCALE.INP input file with the following lines. Not necessary for a single dataset, apart to redefine boundaries of resolution shells.

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

and run

```
xscale
```

at the command line (alternatively: `xscale_par`).

Use XDSCONV to generate reflection files in CCP4 FP/DANO format (F, SigF, Dano, SigDano):

```
INPUT_FILE=XSCALE.HKL  XDS_ASCII
OUTPUT_FILE=temp_ccp4.hkl  CCP4
FRIEDEL'S_LAW=FALSE
```

and run

```
xdsconv
```

at the command line.

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
```

Use XDSCONV again to generate reflection files in CCP4 F+/F- format (F, SigF, F+, SigF+, F-, SigF-):

```
INPUT_FILE=XSCALE.HKL  XDS_ASCII
OUTPUT_FILE=temp_ccp4_f.hkl CCP4_F
FRIEDEL'S_LAW= FALSE
GENERATE_FRACTION_OF_TEST_REFLECTIONS=0.05
```

and run again

```
xdsconv
```

at the command line

Then run CCP4 program f2mtz

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

Then, to run 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
```

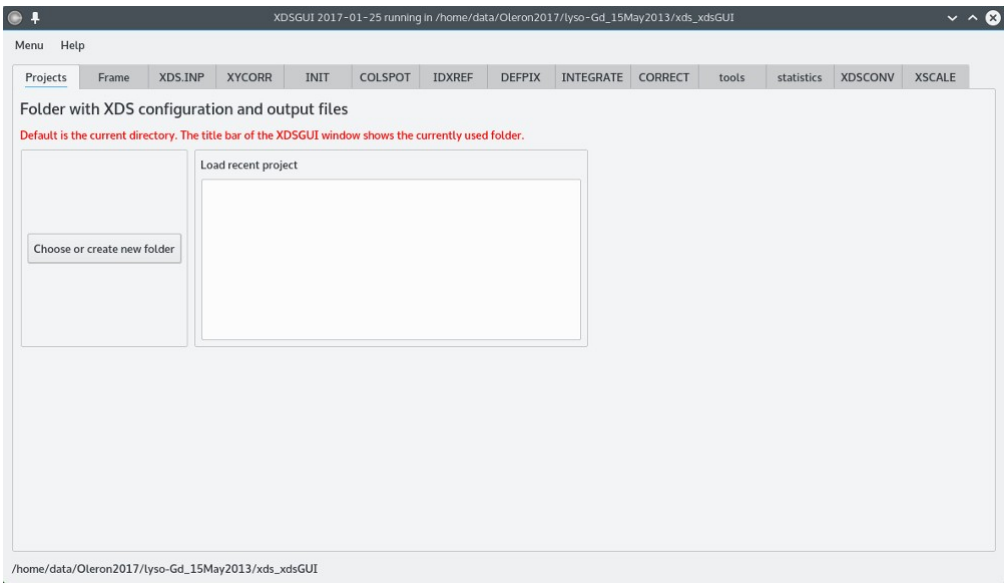
Data reduction with XDS in graphic/automated mode

Graphic mode, using xdsGUI

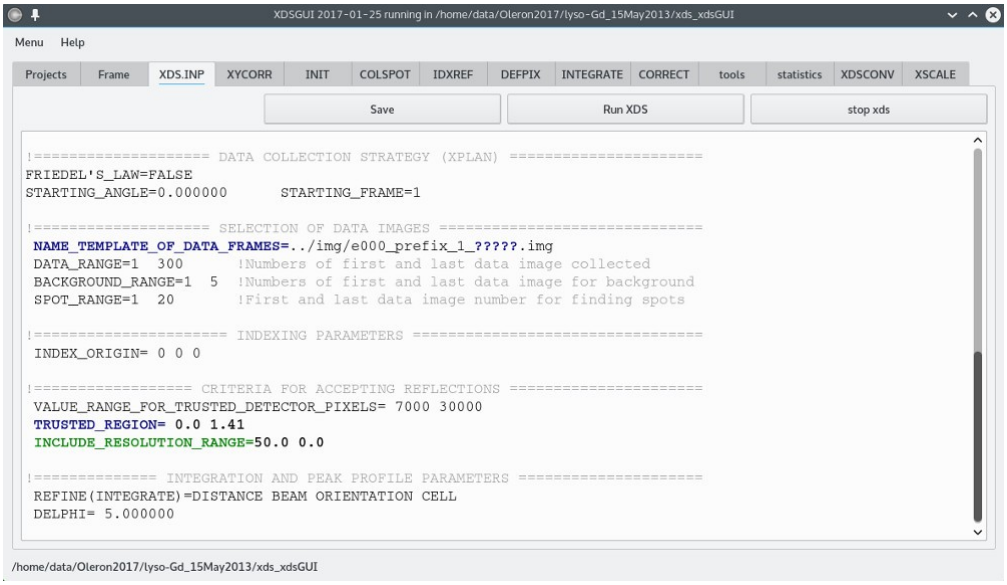
In xds_xdsGUI
Run

```
xdsgui
```

at the command line

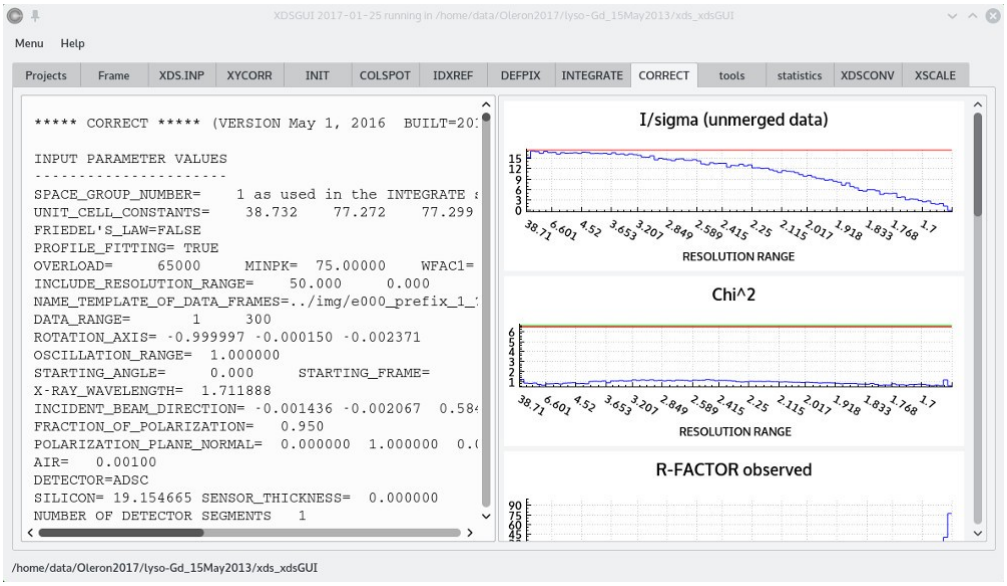


From there, all steps described above can be performed, starting with the edition of the XDS.INP parameter file

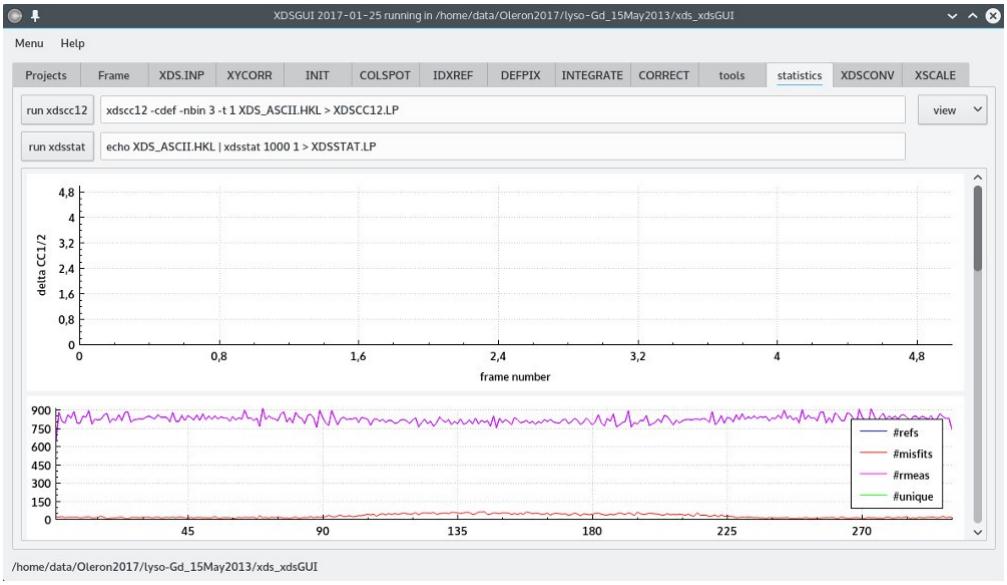


and then running xds (the “Run XDS” button).
Log files are available in the “XYCORR”, “INIT”, … windows, with graphical display of the

statistics:



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



Access to xscale and xdsconv is provided through the corresponding windows.

Automated data reduction

A series of automated scripted layers are available for automated data reduction with xds, such as

xdsme

xdsapp
Just run from the top directory

```
xdsapp -a -cmd --fried=false
```

The dataset will be automatically detected and processed (parameters from the header of images). Alternatively, xdsapp is also a graphic interface for xds.

Xia

Quick SAD phasing with Phaser in ccp4i

Launch

```
ccp4i
```

and define a new project with

```
lyso-Gd_SAD/ccp4_SAD
```

as working directory

Select “Phaser SAD Pipeline” (button highlighted in blue in Figure 1) (~330 sec elapsed time).
Then enter the following parameters:
- reflexion file: `xds_step3_XSCALE-XDSCONV/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)

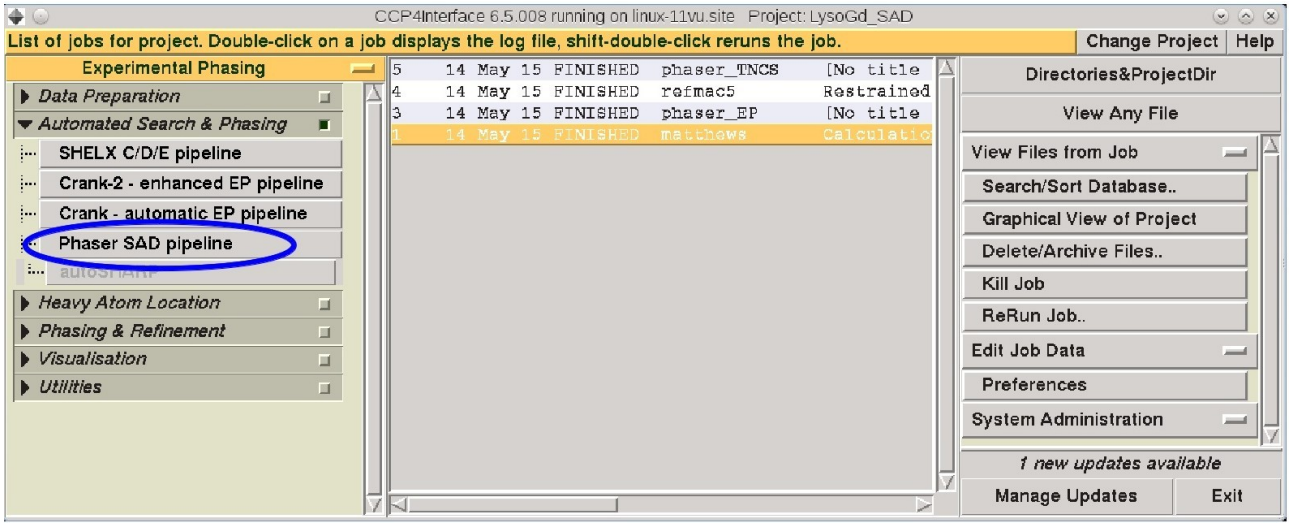


Figure 1

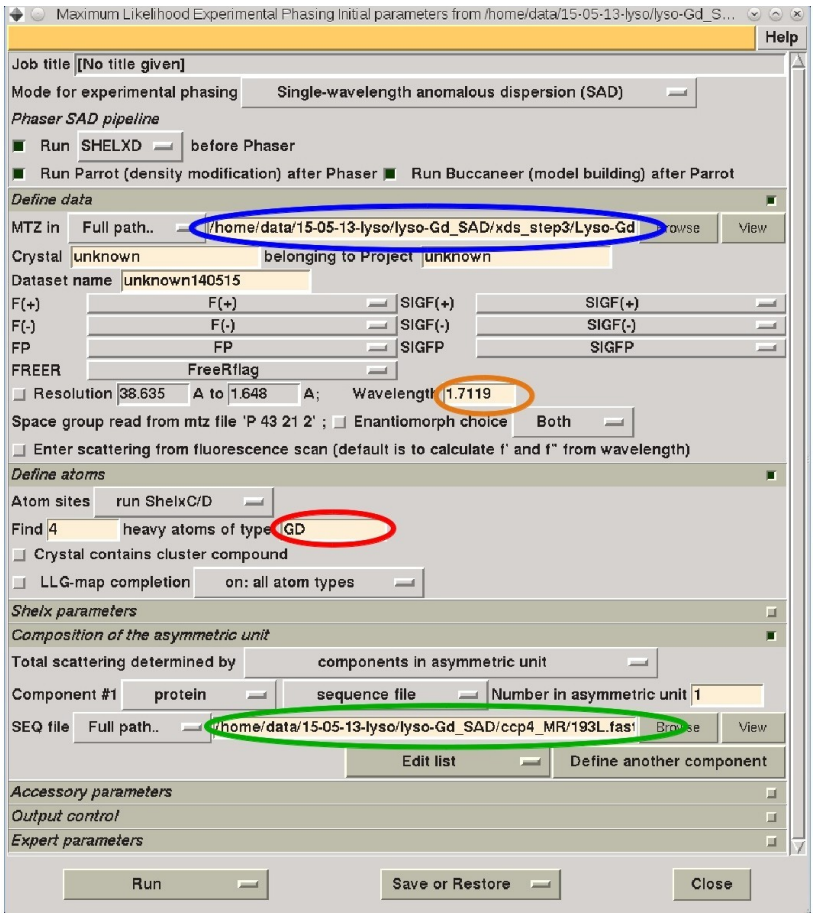


Figure 2

=> ~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 in Figure 3)

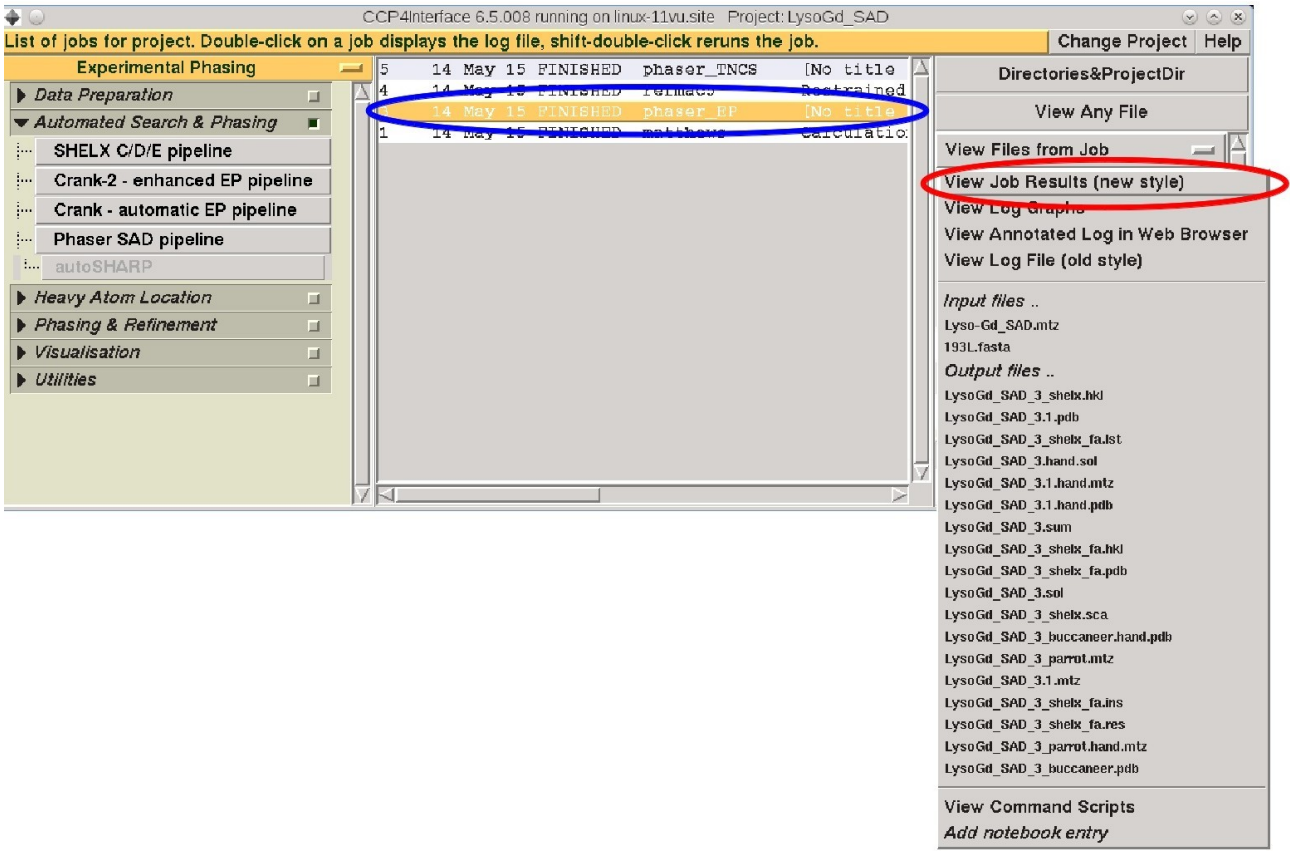


Figure 3

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:

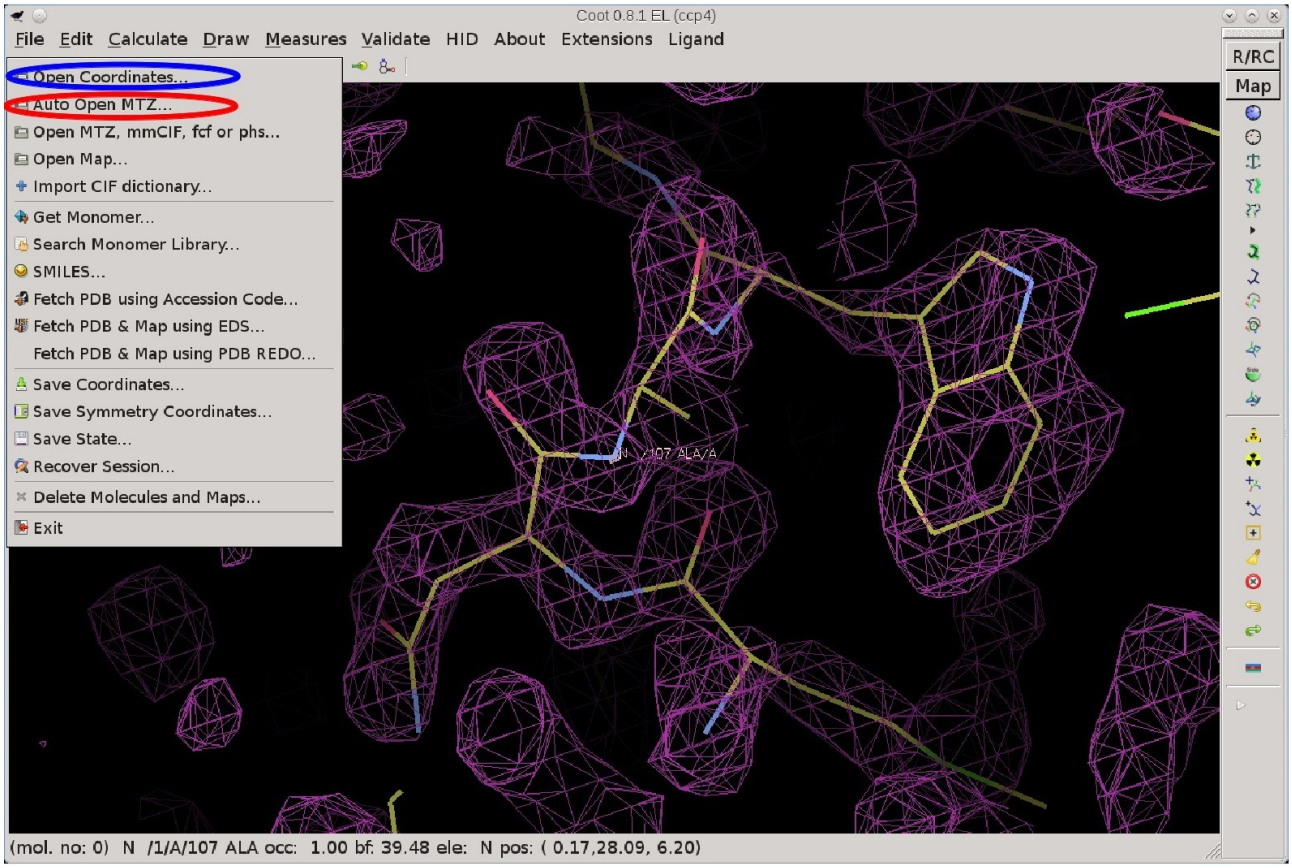


Figure 4

Sub structure of Gd atoms is in

`lyso-Gd_SAD/LysoGd_SAD_3.1.pdb`

Experimental map is in

`lyso-Gd_SAD/LysoGd_SAD_3.1.mtz`

and after automated density modification with `parrot`

`lyso-Gd_SAD/LysoGd_SAD_3_parrot.mtz`

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

`lyso-Gd_SAD/LysoGd_SAD_3_buccaneer.pdb`

This 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`

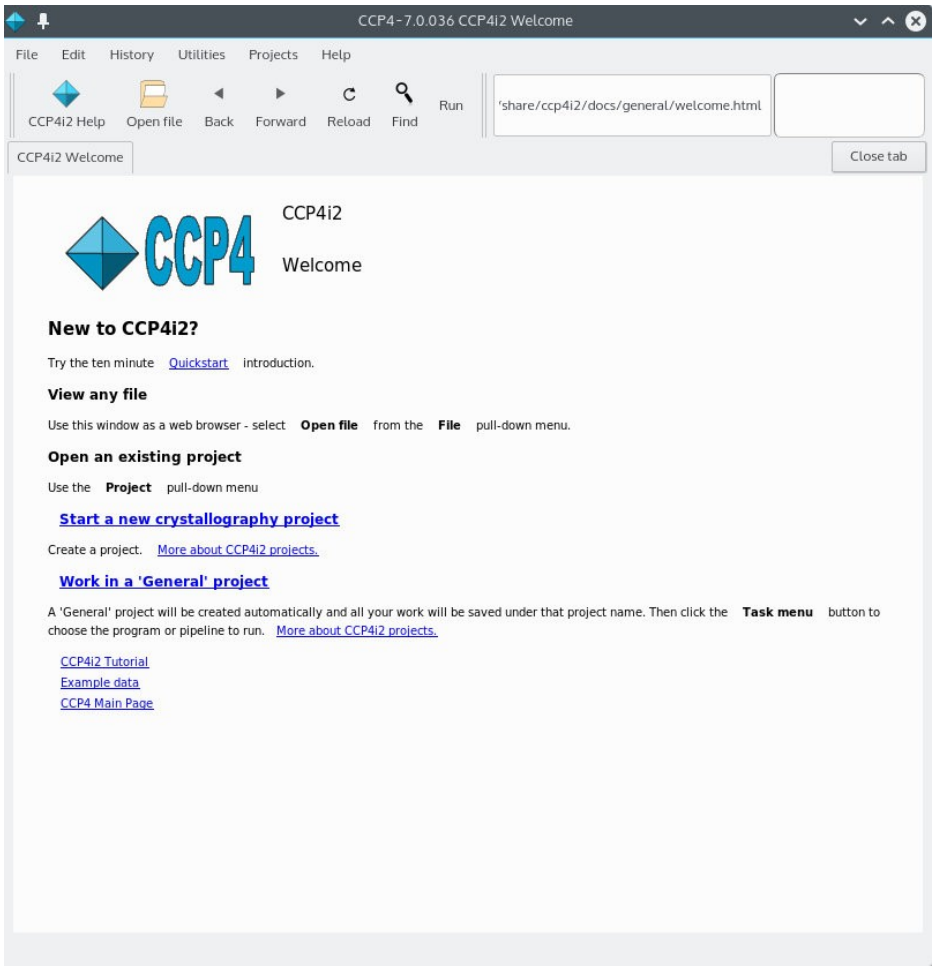
as pdb input file.

Quick SAD phasing in ccp4i2

From the directory ccp4i2_SAD, launch ccp4i2 at the command line

```
ccp4i2
```

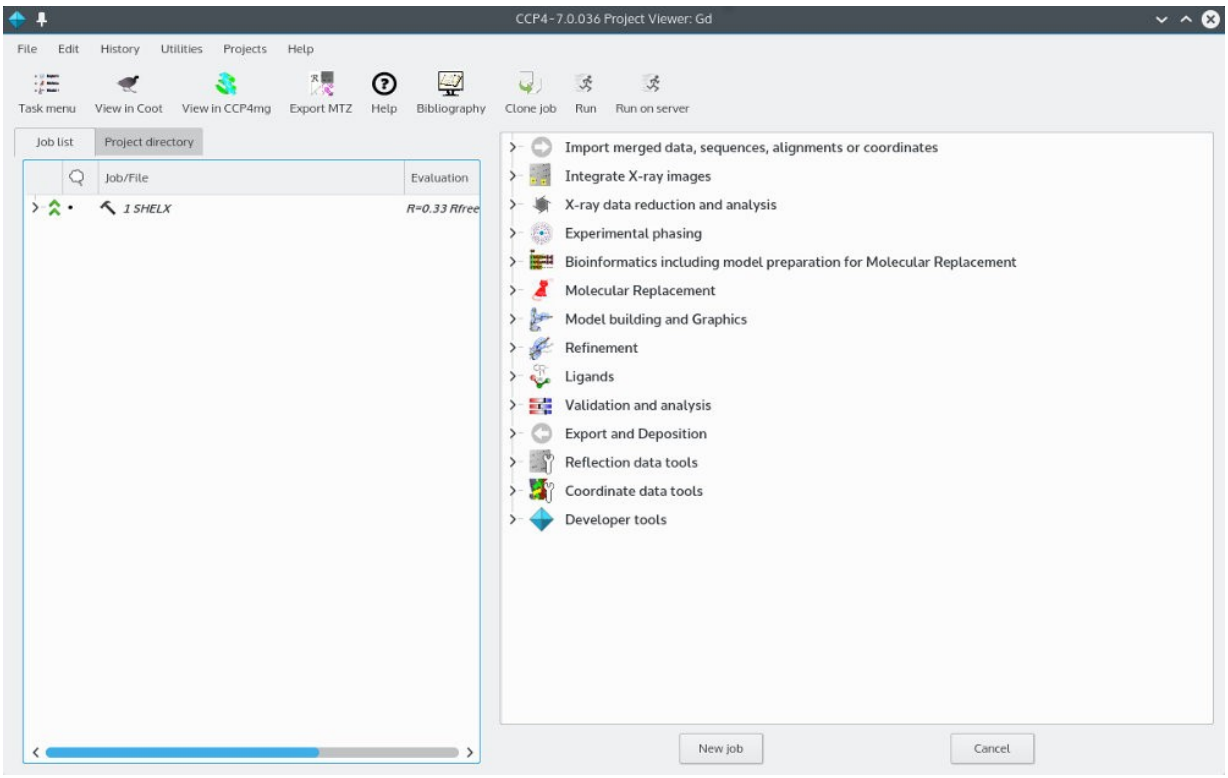
Then **Start a new crystallography project** and define a project directory.



Run the SHELX automated pipeline

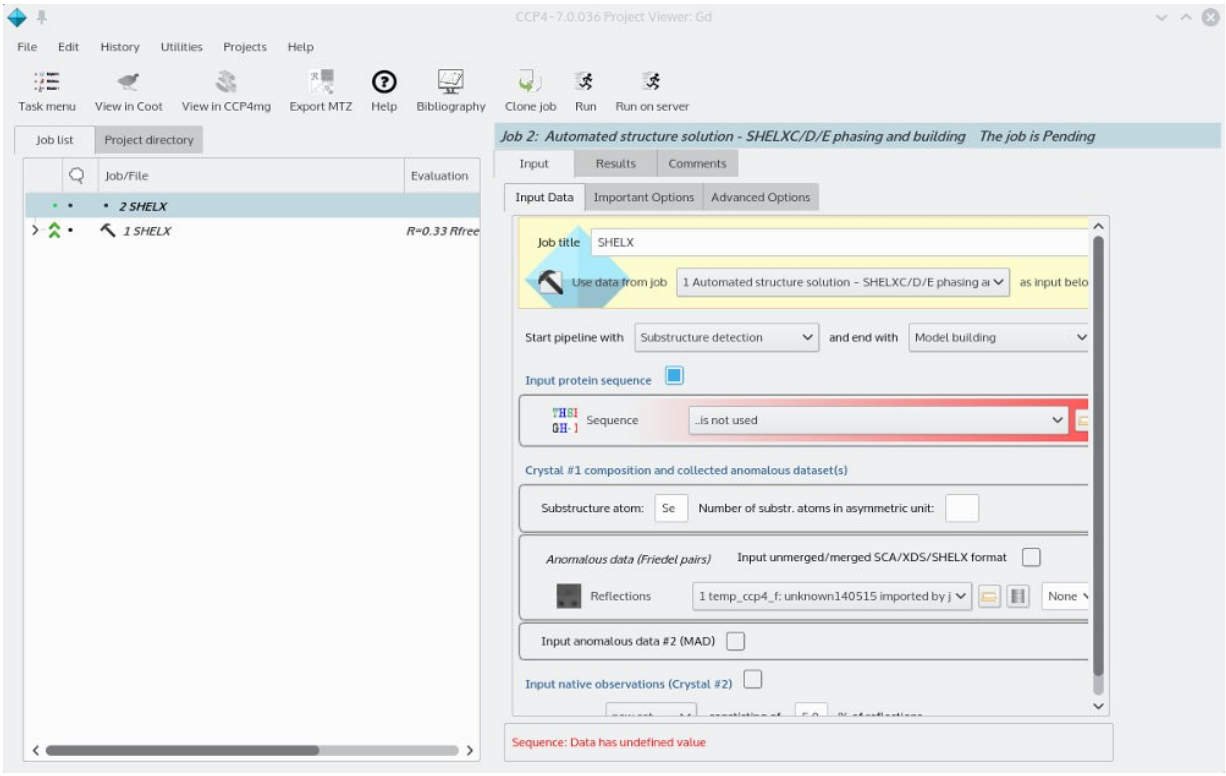
In the present section, the automated pipeline for SAD phasing is described. This pipeline is based on SHELX and Refmac5/Buccaneer.

From the ccp4i2 window, in job list, select **Experimental phasing**, and then the **SHELX** pipeline.



Then provide

- the sequence
  - file **193L.fasta**, to be renamed as .seq
  - directory **lyso-Gd_15May2013/ccp4_MR**
- the crystallography data
  - file **temp_ccp4_f.mtz**
  - directory **lyso-Gd_15May2013/xds_step3_XSCALE-XDSCONV**



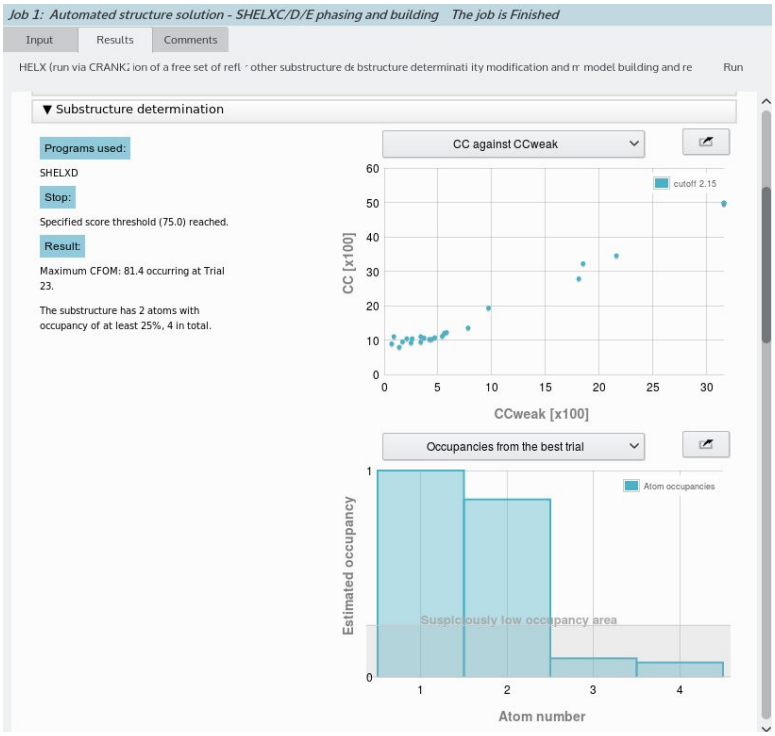
and mention that the data were collected at the peak of fluorescence.

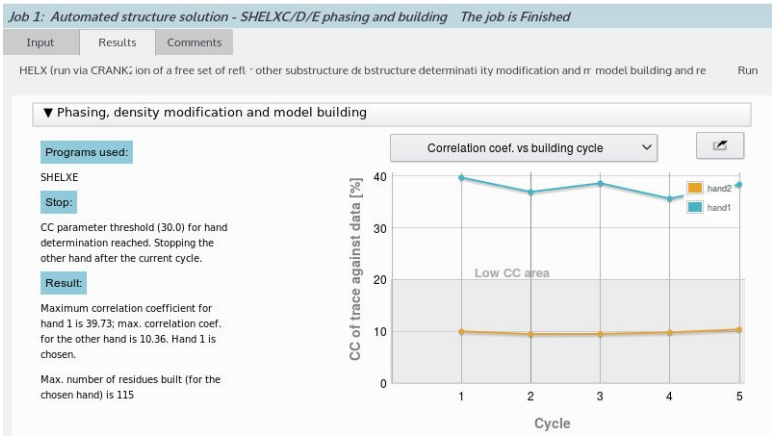
Then press **Run** to chain the following steps:

- Substructure determination (SHELXD)
- Phasing, Density Modification, first model (SHELXE)
- Refinement (Refmac5) / Model building (Buccaneer)

Results analysis

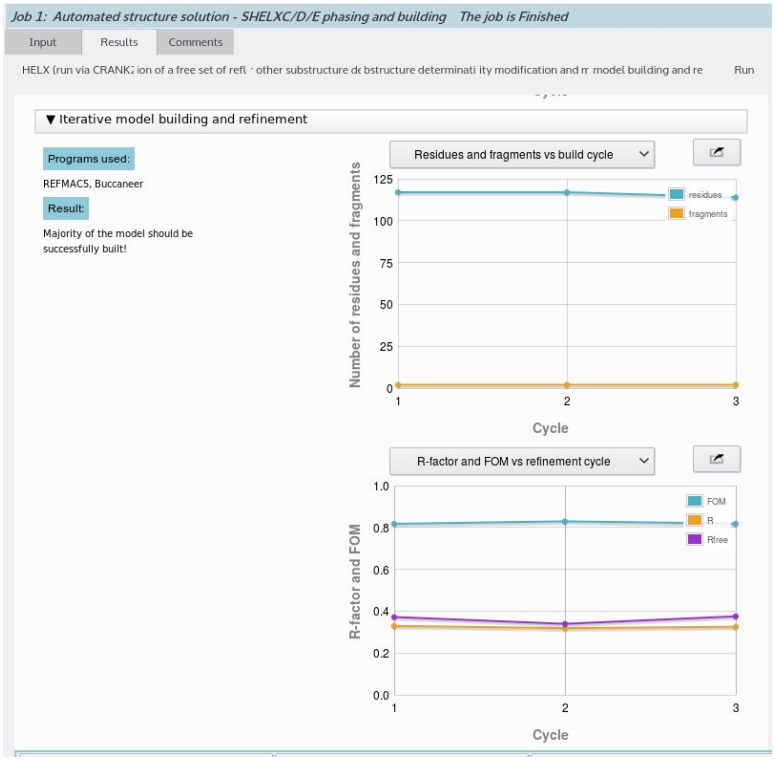
Substructure determination:
The substructure consists in 4 Gd atoms, 2 of them with >10 sigma density in the 2mfo-Dfc density map upon refinement, at the interface between 2 molecules, close to residues 62-63.





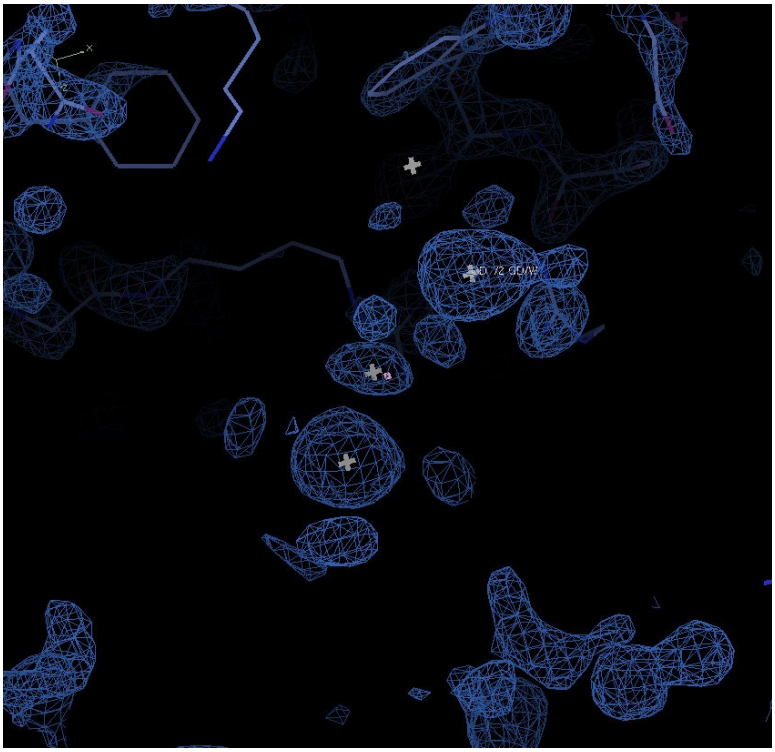
Model building and refinement:

- 3 cycles of refinement, with final FOM of 0.818, Rfree of 37.5% and Rfactor of 32.3%
- 114 residues built automatically.

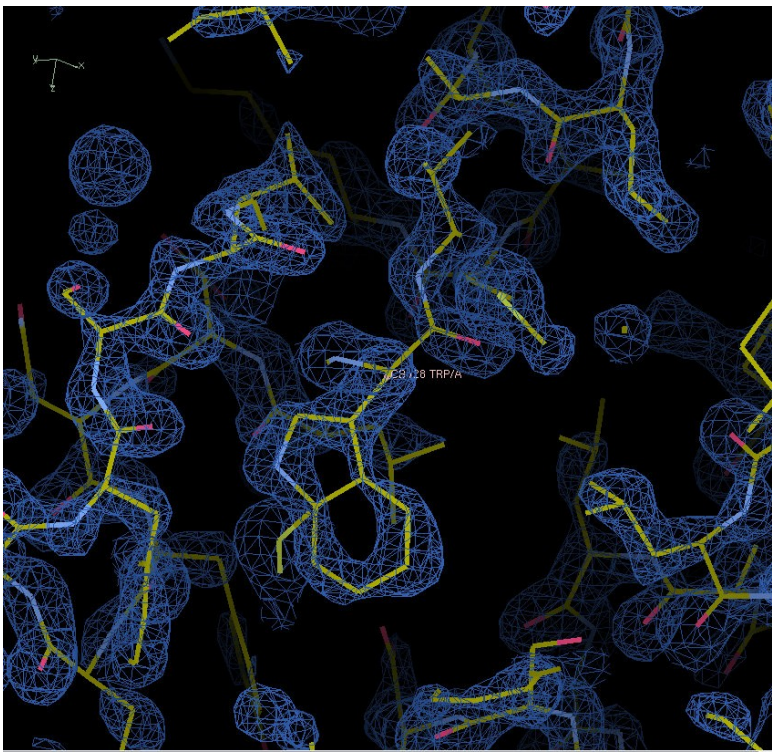


Results displayed in ccp4i2 for the model building/refinement step.

The 2mfo-Dfc map, sub-model and refined model can be displayed with coot (press the **Manual COOT** button).



Partial model displayed in the 2mfo-Dfc contoured at 1sigma.Extra density is visible around the chelating HPDO3A for manual building.



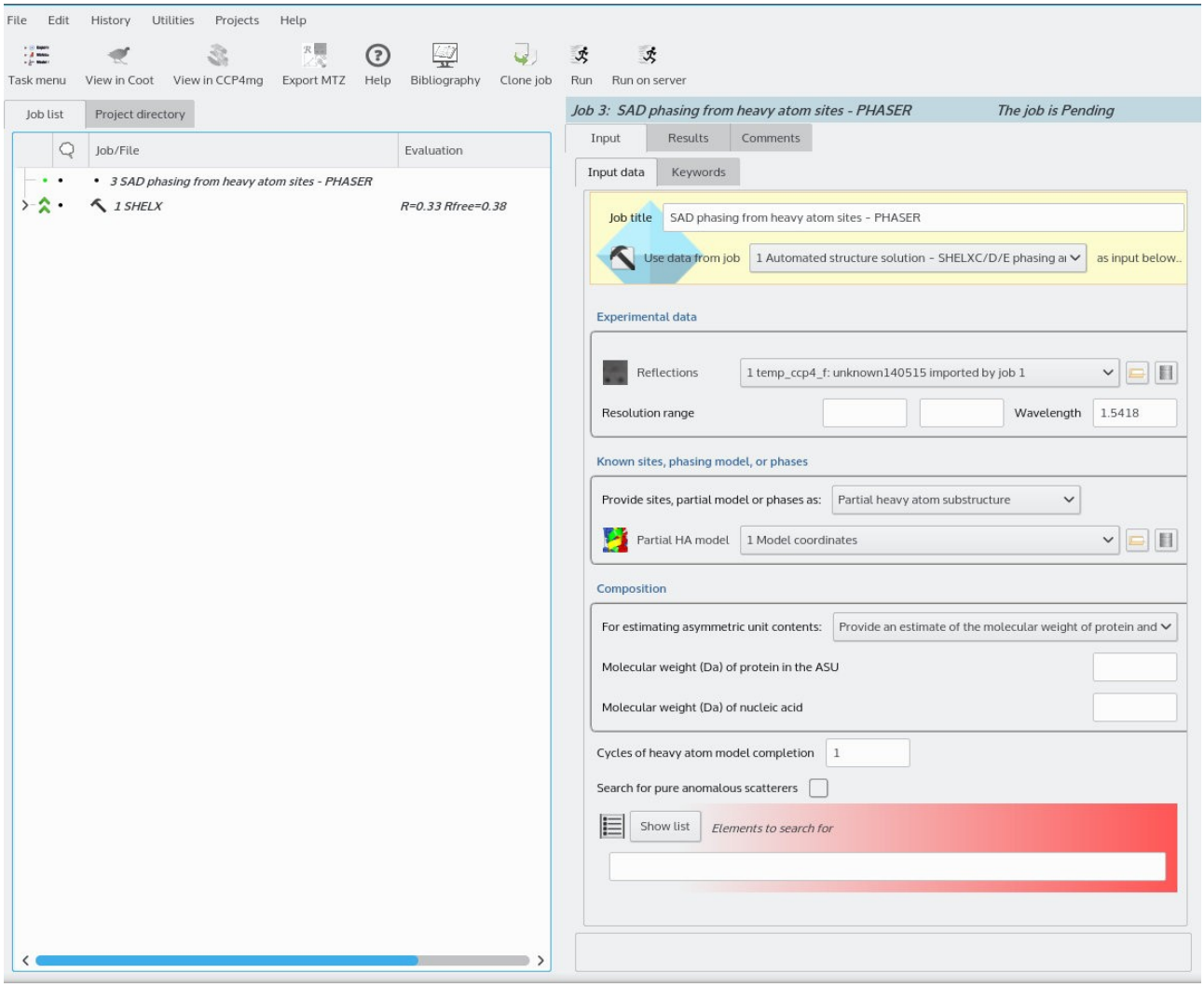
Refined model displayed in the 2mfo-Dfc contoured at 1sigma.

Automated SAD phasing with Phaser

Instead of SHELXE, Phaser can be used for phasing, using the substructure established by SHELXD.

To do so, in the **Experimental phasing** menu, select **SAD phasing from heavy atom sites – Phaser**.

Provide the requested information (wavelength: 1.711888 Ang; MW: 14300).


File
Edit
History
Utilities
Projects
Help
Task menu
View in Coot
View in CCP4mg
Export MTZ
Help
Bibliography
Clone job
Run
Run on server
Job list
Project directory
Job/File
Evaluation
3 SAD phasing from heavy atom sites - PHASER
1 SHELX
R=0.33 Rfree=0.38
Job 3: SAD phasing from heavy atom sites - PHASER
The job is Pending
Input
Results
Comments
Input data
Keywords
Job title
SAD phasing from heavy atom sites - PHASER
Use data from job
1 Automated structure solution - SHELXC/D/E phasing ai
as input below..
Experimental data
Reflections
1 temp_ccp4_f: unknown140515 imported by job 1
Resolution range
Wavelength
1.5418
Known sites, phasing model, or phases
Provide sites, partial model or phases as:
Partial heavy atom substructure
Partial HA model
1 Model coordinates
Composition
For estimating asymmetric unit contents:
Provide an estimate of the molecular weight of protein and
Molecular weight (Da) of protein in the ASU
Molecular weight (Da) of nucleic acid
Cycles of heavy atom model completion
1
Search for pure anomalous scatterers
Show list
Elements to search for

SAD phasing, step by step, with ccp4i

Launch

```
ccp4i
```

and define a new project with

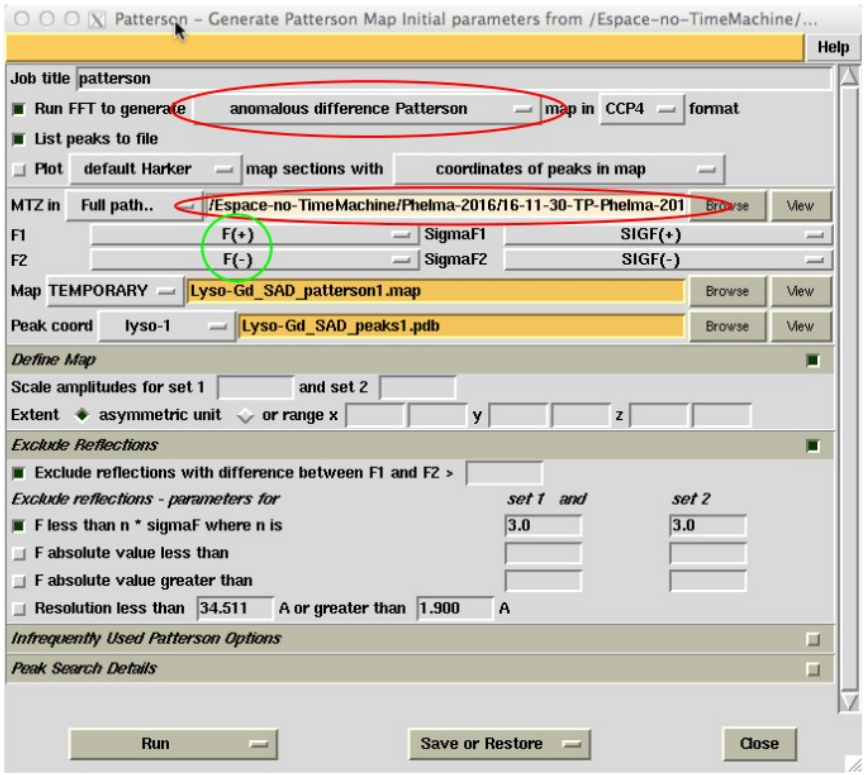
```
mosflm
```

(the directory in which you have processed the data) as working directory

Step 1: Calculate the anomalous difference Patterson map and check the presence of peaks.

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

Go to menu / sub-menus: Experimental Phasing / Heavy Atom Location / Generate Patterson Map



select “anomalous difference Patterson”
MTZ in: **ctruncate_A5_1__00002-unique.mtz** or **Lyso-Gd_SAD.mtz**
check that F1 stands for F+ and F2 for F-
In “Exclude Reflections” you can chose the resolution range, indicate a threshold in the ratio signal-to-noise, etc.

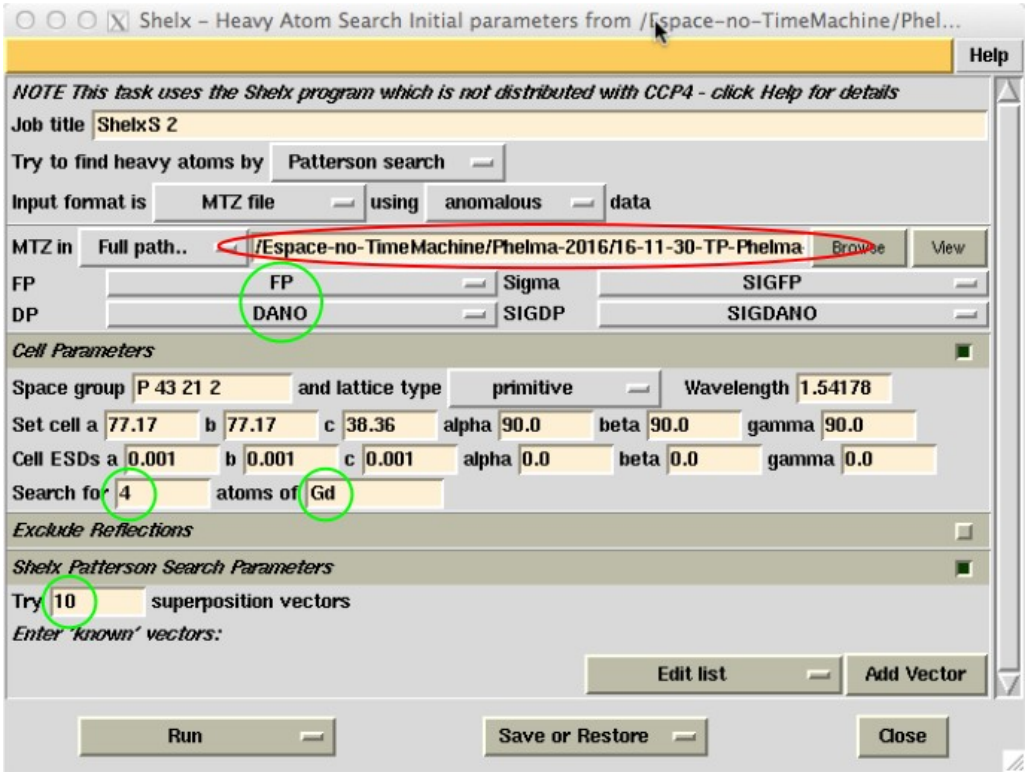
The list of peaks in the anomalous diffrence Patterson is stored in Lyso-Gd_SAD_peaks1.pdb (Peak coord).
Edit this file with any text editor:

```
CRYST1   77.170   77.170   38.360  90.00  90.00  90.00 P 4/m m m
SCALE1      0.012958 -0.000000 -0.000000        -0.00000
SCALE2     -0.000000  0.012958 -0.000000         0.00000
SCALE3      0.000000 -0.000000  0.026069        -0.00000
ATOM      1  OW  WAT X   1       0.000   0.000   0.000137.19137.19           O
ATOM     17  OW  WAT X   2      25.077   0.000  14.622  9.35  9.35           O
ATOM     13  OW  WAT X   3      38.585  35.227  10.790  9.16  9.16           O
ATOM     14  OW  WAT X   4      38.585  10.136  12.893  6.86  6.86           O
ATOM     12  OW  WAT X   5      15.234  13.308   9.606  6.63  6.63           O
ATOM     12  OW  WAT X   6      13.308  15.234   9.606  6.63  6.63           O
ATOM      6  OW  WAT X   7       2.715   3.095   4.514  6.40  6.40           O
ATOM     18  OW  WAT X   8      25.208  25.208  16.120  6.22  6.22           O
ATOM      4  OW  WAT X   9      23.342  23.342   3.180  5.87  5.87           O
ATOM      9  OW  WAT X  10      38.585  36.598   6.570  5.72  5.72           O
ATOM     20  OW  WAT X  11      28.584   1.994  19.180  5.71  5.71           O
ATOM      7  OW  WAT X  12      12.806   9.762   5.186  5.33  5.33           O
ATOM      7  OW  WAT X  13       9.762  12.806   5.186  5.33  5.33           O
ATOM     16  OW  WAT X  14      11.822  15.680  14.051  5.29  5.29           O
ATOM     16  OW  WAT X  15      15.680  11.822  14.051  5.29  5.29           O
ATOM      8  OW  WAT X  16      29.094  29.094   5.656  5.27  5.27           O
ATOM     21  OW  WAT X  17      21.835   3.112  19.180  5.26  5.26           O
ATOM      3  OW  WAT X  18      26.280  26.280   1.214  5.25  5.25           O
ATOM     11  OW  WAT X  19       9.264  12.258   9.493  4.64  4.64           O
ATOM     11  OW  WAT X  20      12.258   9.264   9.493  4.64  4.64           O
ATOM      5  OW  WAT X  21      38.585  16.886   3.847  4.64  4.64           O
ATOM     10  OW  WAT X  22      13.464  36.192   8.329  4.62  4.62           O
ATOM     10  OW  WAT X  23      36.192  13.464   8.329  4.62  4.62           O
ATOM     19  OW  WAT X  24      22.978  28.723  17.844  4.23  4.23           O
ATOM     19  OW  WAT X  25      28.723  22.978  17.844  4.23  4.23           O
ATOM     15  OW  WAT X  26      26.327  26.327  13.383  3.52  3.52           O
ATOM      2  OW  WAT X  27      15.437  38.585   0.000  3.18  3.18           O
END
```

You should have peaks above 4 sigmas.

Step 2: determine the position of Gd atoms using Shelxd program. Basically calculates the coordinates of Gd atoms from the anomalous difference Patterson map.

Go to menu / sub-menus: Experimental Phasing / Heavy Atom Location / ShelxS - Heavy Atom Search



MTZ in: **ctruncate_A5_1__00002-unique.mtz** or **Lyso-Gd_SAD.mtz**
check that FP stands for FP and DP for DANO
Guess how many Gd you could have bound to one lysozyme molecule (You can slightly overestimate), and specify that you are looking for Gd.
10 superpositions vectors (number of Patterson peaks that will analysed as Gd-Gd interatomic vectors, you can try from 4 to about 10, not critical in normal cases)

Look at the log file (old style): Two Gd found:

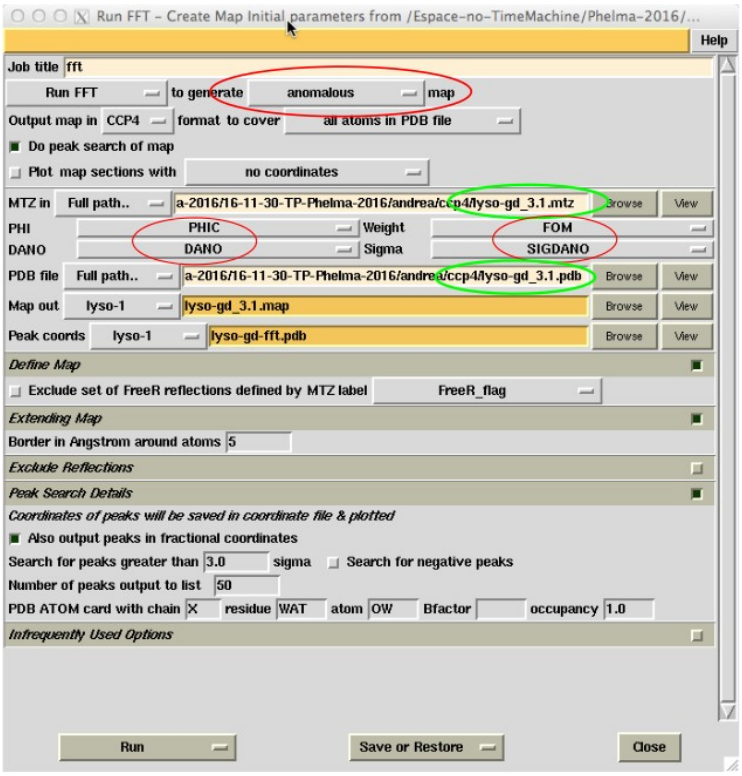
```
 Patterson vector superposition minimum function for E ShelxS 2

 Patt. sup. on vector 1     0.1222  0.1606  0.2497   Height   6.   Length 18.28

 Maximum =    8.96,  minimum =  -98.87       highest memory used = 15501 /673654

       0.1 seconds CPU time

   500 Superposition peaks employed, maximum height  75.1  and minimum height  19.1  on atomic
number scale


 Heavy-Atom Location for E ShelxS 2

   9562 reflections used for structure factor sums


 Solution   1    CFOM =  1.35    PATFOM = 99.9    Corr. Coeff. = 11.6    SYMFOM = 99.9

 Shift to be added to superposition coordinates:  0.4209  0.0618  0.0471

 Name  At.No.   x       y       z    s.o.f.  Minimum distances / PATSMF (self first)

GD1    68.4  0.1857  0.5136  0.7047  1.0000   22.47
                                                3.0
```

```
GD2    63.1   0.1403   0.4813   0.8238   1.0000    18.18     6.27
                                                      2.5      2.7


 Maximum memory for Patterson interpretation = 15161 /  4800

        0.1 seconds CPU time
```

Step 2 (alternative)

Use already known phases to locate Gd atoms.
advantages: more precise localization of anomalous scatterers (especially when you have several atoms to locate), no need for very accurate phases
drawbacks: need for phases (either from molecular replacement or from other heavy atoms derivatives)
Here, we can use the phases from the molecular replacement solution done with the bovine lysozyme model (lyso-gd_3.1.mtz contain phases information from the molecular replacement solution)

Go to menu / sub-menus: Map & mask Utilities / Run fft - Create Map



PHIC are the phases from molecular replacements
FOM is the Figure of Merit, i.e. a value between 0 and 1 that quantify the quality of the phase estimates (used here has a weight)
DANO is the anomalous difference F+ - F-, associated with the sigma (error) SIGANO
The pdb file is only used to determine where the map is calculated (around the model, easier for interpretation, but not mandatory, the map can be calculated on the asymmetric unit if no model is known)
The map calculated is :

$$\rho_H(\vec{r}) = \sum_{h,k,l} w.(|F_{PH}(\vec{s})| - |F_{PH}(-\vec{s})|)\exp[i(\varphi_{calc} - \frac{\pi}{2})].\exp[-2i\pi\vec{r}.\vec{s}]$$
$$\text{with}\,(|F_{PH}(\vec{s})| - |F_{PH}(-\vec{s})|) = \text{DANO}\,;\, w = \text{FOM}\,;\, \varphi_{calc} = \text{PHIC}$$

the shift of -π/2 come from the fact that we are looking at the f'' contribution (the one that breaks Friedel's law), which contributes with a phase shift of π/2 (if'' in the structure factor formula)

The highest peaks of this map should correspond to Gd atoms.
The file lyso-1_7_peaks.ha provides the peaks in fractional coordinates:

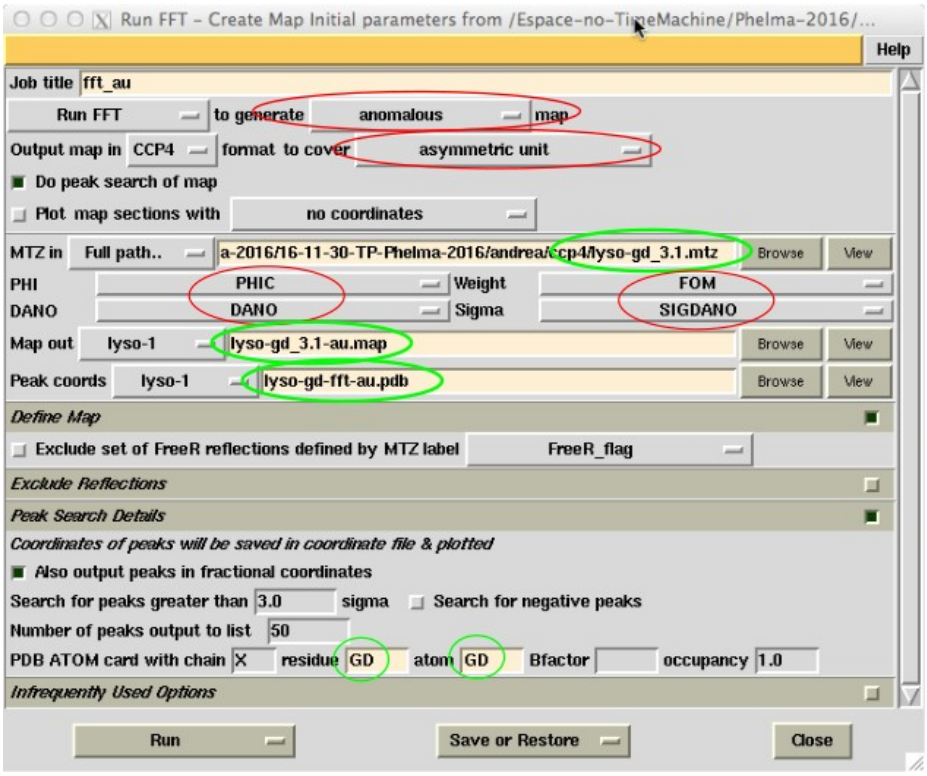
```
GRID 120 120  64
CELL   77.1700   77.1700   38.3600   90.0000   90.0000   90.0000
ATOM1   Ano  -0.1838  0.5129  0.0428       29.37  0.0 BFAC  20.0
ATOM2   Ano  -0.3162  0.0129 -0.2928       29.37  0.0 BFAC  20.0
ATOM3   Ano  -0.3162  0.0129  0.7072       29.37  0.0 BFAC  20.0
ATOM4   Ano  -0.0129  0.3162 -0.2072       29.37  0.0 BFAC  20.0
ATOM5   Ano  -0.0129  0.3162  0.7928       29.37  0.0 BFAC  20.0
ATOM6   Ano   0.1838  0.4871  0.5428       29.37  0.0 BFAC  20.0
ATOM7   Ano  -0.1403  0.4817 -0.0746       27.04  0.0 BFAC  20.0
ATOM8   Ano   0.1403  0.5183  0.4254       27.04  0.0 BFAC  20.0
ATOM9   Ano   0.0183  0.3597 -0.3246       27.04  0.0 BFAC  20.0
ATOM10  Ano   0.0183  0.3597  0.6754       27.04  0.0 BFAC  20.0
ATOM11  Ano  -0.1421  0.4820  0.8906        5.75  0.0 BFAC  20.0
ATOM12  Ano  -0.1315 -0.0038 -0.2092        4.02  0.0 BFAC  20.0
ATOM13  Ano  -0.1315 -0.0038  0.7908        4.02  0.0 BFAC  20.0
ATOM14  Ano   0.1315  0.0038  0.2908        4.02  0.0 BFAC  20.0
ATOM15  Ano   0.0038  0.1315 -0.2908        4.02  0.0 BFAC  20.0
ATOM16  Ano   0.0038  0.1315  0.7092        4.02  0.0 BFAC  20.0
ATOM17  Ano   0.2835  0.5146  0.2851        3.88  0.0 BFAC  20.0
ATOM18  Ano   0.0146  0.2165  0.5351        3.88  0.0 BFAC  20.0
```

Two independent sites (the first 6 ones are related by crystallographic symmetry and have the same heights (29.37), same for the next 4 ones (heights of 27.04)

the file lyso-gd-fft.pdb provides the peaks in cartesian coordinates:

```
CRYST1   77.170   77.170   38.360  90.00  90.00  90.00 P 43 21 2
SCALE1      0.012958 -0.000000 -0.000000        -0.00000
SCALE2     -0.000000  0.012958 -0.000000         0.00000
SCALE3      0.000000 -0.000000  0.026069        -0.00000
ATOM      5  OW  WAT X   1     -14.183  39.579   1.643 29.37 29.37           O
ATOM      5  OW  WAT X   2     -24.402   0.994 -11.233 29.37 29.37           O
ATOM      5  OW  WAT X   3     -24.402   0.994  27.127 29.37 29.37           O
ATOM      5  OW  WAT X   4      -0.994  24.402  -7.947 29.37 29.37           O
ATOM      5  OW  WAT X   5      -0.994  24.402  30.413 29.37 29.37           O
ATOM      5  OW  WAT X   6      14.183  37.591  20.823 29.37 29.37           O
ATOM      3  OW  WAT X   7     -10.831  37.174  -2.860 27.04 27.04           O
ATOM      3  OW  WAT X   8      10.831  39.996  16.320 27.04 27.04           O
ATOM      3  OW  WAT X   9       1.411  27.754 -12.450 27.04 27.04           O
ATOM      3  OW  WAT X  10       1.412  27.754  25.910 27.04 27.04           O
ATOM     23  OW  WAT X  11     -10.962  37.199  34.164  5.75  5.75           O
ATOM      6  OW  WAT X  12     -10.151  -0.292  -8.025  4.02  4.02           O
ATOM      6  OW  WAT X  13     -10.151  -0.292  30.335  4.02  4.02           O
ATOM      6  OW  WAT X  14      10.151   0.292  11.155  4.02  4.02           O
ATOM      6  OW  WAT X  15       0.292  10.151 -11.155  4.02  4.02           O
ATOM      6  OW  WAT X  16       0.292  10.151  27.205  4.02  4.02           O
ATOM     13  OW  WAT X  17      21.879  39.714  10.936  3.88  3.88           O
ATOM     13  OW  WAT X  18       1.129  16.706  20.526  3.88  3.88           O
```

In order to avoid symmetry related peaks, you can calculated the map on the asymmetric unit:



and the files lyso-1_8_peaks.ha lyso-gd-fft-au.pdb only contains two peaks much higher than the rest (which is basically noise):

```
GRID 120 120  64
CELL   77.1700   77.1700   38.3600   90.0000   90.0000   90.0000
ATOM1   Ano   0.8162  0.5129  0.0428       30.70  0.0 BFAC  20.0
ATOM2   Ano   0.4817  0.8597  0.0746       28.26  0.0 BFAC  20.0
ATOM3   Ano   0.4962  0.6315  0.0408        4.20  0.0 BFAC  20.0
ATOM4   Ano   0.9854  0.7835  0.0351        4.05  0.0 BFAC  20.0
ATOM5   Ano   0.6936  0.3170  0.0308        4.01  0.0 BFAC  20.0
ATOM6   Ano   0.2161  0.3438  0.0380        3.96  0.0 BFAC  20.0

CRYST1   77.170   77.170   38.360  90.00  90.00  90.00 P 43 21 2
SCALE1      0.012958 -0.000000 -0.000000        -0.00000
SCALE2     -0.000000  0.012958 -0.000000         0.00000
SCALE3      0.000000 -0.000000  0.026069        -0.00000
ATOM     16  GD    GD X   1       62.987  39.579   1.643 30.70 30.70           O
ATOM     23  GD    GD X   2       37.174  66.339   2.860 28.26 28.26           O
ATOM     17  GD    GD X   3       38.293  48.736   1.565  4.20  4.20           O
ATOM     11  GD    GD X   4       76.041  60.464   1.346  4.05  4.05           O
ATOM      9  GD    GD X   5       53.523  24.464   1.182  4.01  4.01           O
ATOM     10  GD    GD X   6       16.678  26.533   1.457  3.96  3.96           O
```

Here, the peaks of the map gives you directly the coordinates of the Gd atoms.
Further analysis should show that these two peaks match the two peaks found by difference Patterson map, if you take into account the possible symmetries.

Step 3: calculating the experimental phases using the anomalous signal and the two gadolinium sites determined previously.

A) the lyso-gd-fft-au.pdb needs some editing to be interpreted properly (it will be saved in lyso-gd-fft-aus.pdb):
- just keep the two main sites
- put the occupancy to 1.0 (it will be refined anyway afterwards)
- replace o by GD in the last column (otherwise the program will believe you have oxygen instead of Gd, and will fail)

```
CRYST1   77.170   77.170   38.360  90.00  90.00  90.00 P 43 21 2
SCALE1      0.012958 -0.000000 -0.000000       -0.00000
SCALE2     -0.000000  0.012958 -0.000000        0.00000
SCALE3      0.000000 -0.000000  0.026069       -0.00000
ATOM     16  GD   GD X   1      62.987  39.579   1.643  1.00 30.70          GD
ATOM     23  GD   GD X   2      37.174  66.339   2.860  1.00 28.26          GD
END
```

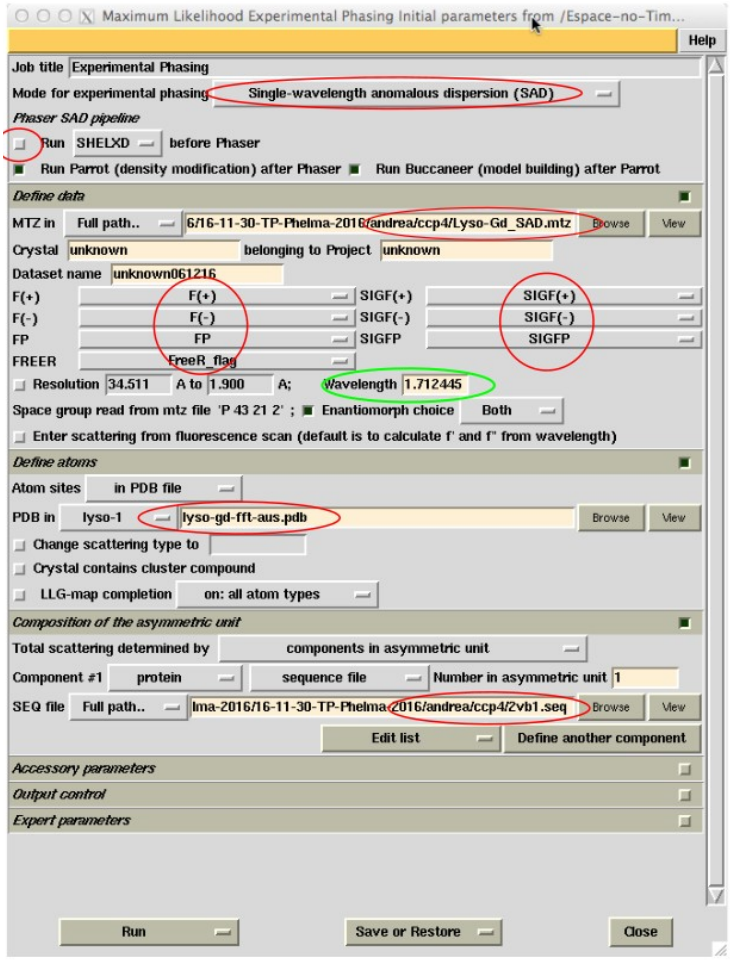
B)

Go to menu / sub-menus: Experimental Phasing / Automated Search & Phasing / Phaser SAD pipeline:

Pay attention to deselect run SHELXD before Phaser (you have already found where the GD atoms are)
some files are in the ccp4 directory and not in the mosflm one: just be sure where they are in you session.
Here you need F+ and F- and not DANO

The main 3 steps of the program will be:

1) Using Phaser to calculate phases from the data and positions of Gd atoms, using the principles described in the lectures. Positions, B-factor and occupancy of Gd atoms will be refined. Other sites will be searched and refined. (check log file)
2) Using Parrot for density modification procedure (solvent flatening, etc ...): should improve the quality of the phases
3) Using Buccaneer to build automatically a model in the resulting electron density map, using the primary sequence information provided.



the experimental phases are in lyso-1_10_parrot.mtz. The model automatically constructed by buccaneer is in lyso-1_10_buccaneer.pdb
Files with 'hand' are also created in cases where the correct hand is not known. Here, the space group was set properly, since it was already known.

Step 4: Use coot to look at maps and models

open coot

open the molecular replacement solution with bovine lysozyme in ccp4 directory: lyso-gd_3.1.pdb
menu File / Open coordinates

open molecular replacement maps in ccp4 directory: lyso-gd_3.1.mtz
menu File / Auto Open MTZ

open the automatically built model in mosflm directory: lyso-1_10_buccaneer.pdb
menu File / Open coordinates

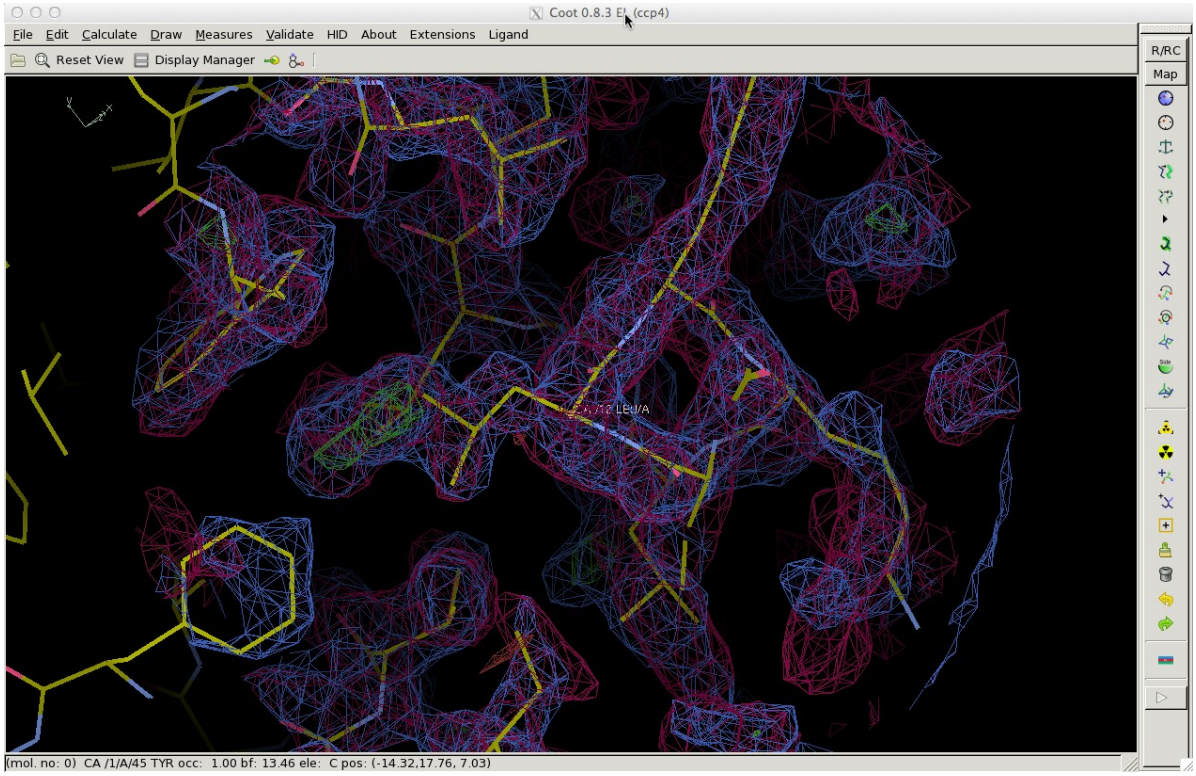
Open experimental phases map in mosflm directory: lyso-1_10_parrot.mtz
menu File / Open MTZ, mmCIF, fcf or phs... and chose
parrot.F_phi.F as amplitudes
parrot.F_phi.phi as phases

adjust the contour to 1 sigma

Compare the different maps. Focus on regions were the sequence of bovine and Hen lysozyme are different.

Hen and bovine lysozymes sequence alignment :

```
>2z2f_A Lysozyme C-2; stomach lysozyme, 1,4-beta-N-acetylmuramidase C, bacteriolytic enzyme,
hydrolase; 1.50A {Bos taurus} SCOP: d.2.1.2
Length = 129    Score =  154 bits (389), Expect = 1e-38,    Method: Composition-based stats.
Identities = 71/130 (54%), Positives = 96/130 (73%), Gaps = 2/130 (1%)

Query: 19  KVFGRCELAAAMKRHGLDNYRGYSLGNWVCAAKFESNFNTQATNRNTDG-STDYGILQIN 77
           KVF RCELA  +K+ GLD Y+G SL NW+C  K+ES++NT+ATN N    STDYGI QIN
Sbjct: 1   KVFERCELARTLKKLGLDGYKGVSLANWLCLTKWESSYNTKATNYNPSSESTDYGIFQIN 60

Query: 78  SRWWCNDGRTPGSRNLCNIPCSALLSSDITASVNCAKKIVSDGNGMNAWVAWRNRCKGTD 137
           S+WWCNDG+TP + + C++ CS L+ +DI  +V CAK IVS+  G+ AWVAW++ C+  D
Sbjct: 61  SKWWCNDGKTPNAVDGCHVSCSELMENDIAKAVACAKHIVSE-QGITAWVAWKSHCRDHD 119

Query: 138 VQAWIRGCRL 147
           V +++ GC L
Sbjct: 120 VSSYVEGCTL 129
```



MR

under construction

Complete model

Partial/remote model

Refinement / construction

under construction

Complete model

Partial/remote model