X-ray crystallography practical

Oleron 2018

Last update: 21 Mai 2018

Table of Contents

Schedule of the X-ray crystallography practical	L
Data collection on a Gd derivative on beamline FIP)
Data reduction with XDS package	3
Introduction	3
Before lauching xdsgui:	5
Data reduction with XDS in graphic/automated mode	5
Graphic mode, using xdsgui	5
First look at the data statistics)
Optimization of data processing)
Final data scaling and merging with XSCALE11	L
Changing data format for ccp411	L
Quick SAD phasing with Phaser in ccp4i13	3
Phaser SAD Pipeline to determine experimental phases	3
What does this pipeline ?	ł
COOT for model building	5
Refinement / construction	7
Using Refmac5 to refine the (uncomplete) model	7
Completing the model with COOT)
Other useful information)

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 ($\lambda = 1.711$ Å) f'' = 28 e-, at $\lambda = 1.54$, 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 &

ŧ			XD	SGUI 2017	-01-25 runnin	g in /home/da	ita/Oleron20:	17/lyso-Gd_15N	/lay2013/xds_	kdsGUI			~
nu Help													
Projects	Frame	XDS.INP	XYCORR	INIT	COLSPOT	IDXREF	DEFPIX	INTEGRATE	CORRECT	tools	statistics	XDSCONV	XSCALE
older wi	ith XDS c	onfigurat	ion and out	tput files	5								
fault is the	e current dire	ectory. The tit	le bar of the XI	OSGUI wind	ow shows the	currently use	d folder.						
		Lo	ad recent proje	ct									
Choose or	create new	folder											
encose of	create new	- Court											

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"

ojects	Frame	XDS.INP	XYCORR	INIT	COLSPOT	IDXREF	DEFPIX	INTEGRATE	CORRECT	tools	statistics	XDSCONV	XSCALE
					Save			Run	XDS			stop xds	
IEDEI ARTIN ARTIN ATA_F ACKGF POT_F	J'S_LAW= IG_ANGLE REMPLATE RANGE=1 ROUND_RA RANGE=1 _ORIGIN=	FALSE =0.00000 COF_DATX 300 ANGE=1 5 20 = 0 0 0	00 2 SELECTIO FRAMES= INumbe INumbe IFirst = INDEXI	STARTING ON OF DAT =/img/e ers of fi ers of fi t and las ING PARAM	FRAME=1 FA IMAGE 2000_pre irst and irst and st data METERS =	(XPLA) (S ======= (fix_1_?) (last da (last da image n)	????.imd ata imad imber fo	g ge collect ge for bac pr finding	ted ckground g spots				
		===== CF	RITERIA H	FOR ACCEN	PTING RE	FLECTIO							

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.

0	XDSGUI 2017-01-25 running in /home/dhousset/Documents/TP-RX-2018/d	ta_MX/lyso-Gd_15May201	3/xds_2018 _ 🗆 🗙
Menu He	p		
Projects	Frame XDS.INP XYCORR INIT COLSPOT IDXREF DEFPIX INTEGRATE CORRECT tools statistics XDSCONV XSCALE		
cI	[197, I23] [199, I2(1)3] [211, I432] [214, I4(1)32]	-	x=-748,y=-187 res=1.50238A
			×
Max	imum oscillation range to prevent angular overlap at high resolution limit		
ass	uming zero (!) mosaicity.		
Max	imum oscillation range High resolution limit		
	(degrees) (Angstrom)		
	3.63 4.00		
	2.75 3.00		
	0.91 1.00		
			Contraction of the second s
cpu	time used 4.6 sec		
ela	psed wall-clock time 1.4 sec		
111	ERROR !!! INSUFFICIENT PERCENTAGE (< 50%) OF INDEXED REFLECTIONS		
AUT	MATIC DATA PROCESSING STOPPED. AS THE CRITERIA FOR A GOOD		
SOL	JIION ARE RATHER SIRICI, YOU MAY CHOOSE TO CONTINUE DATA		
".TO	ASSING AFIER CRANGING ING SODE -CARD IN SADS.INF IO		
IF	THE BEST SOLUTION IS REALLY NONSENSE YOU SHOULD FIRST HAVE		
A L	DOK AT THE ASCII-FILE "SPOT.XDS". THIS FILE CONTAINS THE		
INI	FIAL SPOT LIST SORTED IN DECREASING SPOT INTENSITY. SPOTS		
NEA	R THE END OF THE FILE MAY BE ARTEFACTS THAT CAN BE ERASED.		
ALT	ERNATIVELY YOU MAY TRY DIFFERENT VALUES FOR "INDEX_ORIGIN"		
AS	SUGGESTED IN THE ABOVE LISTING.		 The second state of the second st
YOU	MAY CHOOSE TO SKIP SOME OF THE FIRST FRAMES BY CHANGING		
THE	DATA_RANGE=" IN FILE "XDS.INP" AND START ALL OVER AGAIN.		<u>-</u>
		3	<u>Z</u>
		Range	
		Start	

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:

ojects	Frame	XDS.INP	XYCORR	INIT	COLSPOT	IDXREF	DEFPIX	INTEGRATE	CORRECT	tools	statistics	XDSCONV	XSCALE	
****	CORRECT	****	(VERSION	May 1,	2016 BU	JILT=20:			I/sigma	(unmer	ged data)			
INPUT SPACE_ JNIT_C FRIEDH PROFII	PARAMET _GROUP_N CELL_CON EL'S_LAW LE_FITTI DAD=	ER VALU UMBER= STANTS= FALSE NG= TRU 65000	ES 1 as 1 38.7 E MINP	used in 32 7' K= 75.	the INTE 7.272 00000	GRATE : 77.299 WFAC1=	15 12 9 6 3 0 	^{6,602} ^{4,52} ^{3,65}	3 ^{3,2} 07 ^{2,8} 49 ² RES	2.589 ^{2.425} SOLUTION	25 ² .125 ^{.0} 2, RANGE	, ^{2,} 9 ₁₈ ^{2,833²,}	7 ₆₈ 2.7	
INCLUI	DE_RESOL	UTION_R. _OF_DAT.	ANGE= A_FRAMES:	50.000 =/img	0.00 /e000_pre	00 efix_1_:	Chi^2							
ROTATI SCILI	RANGE- ION_AXIS LATION_R ING_ANGI	= -0.99 ANGE= E=	9997 -0.0 1.000000 0.000	000150 START	-0.002371 ING_FRAME	L 3=	654 32 1							
X-RAY_NAVELENGTH= 1.711888 INCIDENT_BEAM_DIRECTION= -0.001436 -0.002067 0.584 FRACTION_OF_POLARIZATION= 0.950 DOLARIZATION_PULNER_NORMAL= 0.000000 1.000000 0.0								$\frac{3q_{2}}{2} \frac{3q_{2}}{2} \frac{q_{3}}{2} \frac{q_{3}}{2} \frac{3}{2} \frac{q_{3}}{2} \frac{2}{2} \frac{q_{4}}{2} \frac{2}{2} \frac{q_{5}}{2} \frac{q_{1}}{2} \frac{2}{2} \frac{q_{1}}{2} \frac{1}{2} \frac{q_{1}}{2} \frac{1}{2} \frac{q_{3}}{2} \frac{1}{2} \frac{q_{6}}{2} \frac{1}{2} \frac{1}{2} \frac{1}{2} \frac{q_{6}}{2} \frac{1}{2} $						
AIR= DETECT	0.0010 FOR=ADSC	0							R-FAC	TOR ob	oserved			
SILICO	ON= 19.1 R OF DET	54665 S ECTOR S	ENSOR_TH EGMENTS	ICKNESS	= 0.0000	000	<u><u>?</u></u>							

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</pre>

```
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)





Help Job title [No title given] Wode for experimental phasing Single-wavelength anomalous dispersion (SAD) Phaser SAD pipeline Run Parrot (density modification) after Phaser Run Buccaneer (model building) after Parrot Define data Image: Start Sta	🔷 💿 Maximum Likelihood Experiment	al Phasing Initial parameters fro	om /home/data/15-05-13-lysc	o/lyso-Gd_S ⊙ ⊘ (
Job title [No title given] dode for experimental phasing Single-wavelength anomalous dispersion (SAD) Phaser SAD pipeline Run SHELXD before Phaser Run Parrot (density modification) after Phaser I Run Buccaneer (model building) after Parrot Define data MTZ in Full path Thome/data/15-05-13-lyso/lyso-Gd_SAD/xda_step3/Lyso-Gd_rowse View Crystal unknown belonging to Project Unknown Dataset name Unknown140515 T(+) F(+) SIGF(+) SIGF(+) P FP SIGFP SIGFP REER FreeRllag Resolution 38.635 A to 15.648 A; Wavelength (17.119 Space group read from mt file 'P 43 21 2'; Enantiomorph choice Both Enter scattering from fullorescence scan (default is to calculate 1' and 1" from wavelength) Define atoms tom sites run ShelxC/D Crystal contains cluster compound LLG-map completion on: all atom types Shek parameters 20.mposition of the asymmetric unit Total scattering determined by components in asymmetric unit SEQ file Full path Thome/data/15-05-13-lyso/lyso-Gd_SAD/ccp4_MR/193L.fasi Broke View Edit list Define another component Accessory parameters 20.mposition of the asymmetric unit Component #1 protein sequence file Number in asymmetric unit 1 SEQ file Full path Thome/data/15-05-13-lyso/lyso-Gd_SAD/ccp4_MR/193L.fasi Broke View Edit list Define another component Close				Help
Mode for experimental phasing Single-wavelength anomalous dispersion (SAD) Phaser SAD pipeline Run SHELXD before Phaser Run Parrot (density modification) after Phaser Run Buccaneer (model building) after Parrot Define data MTZ In Full path. Thome/data/15.05-13-lyso/lyso-Gd_SAD/xds_step3/Lyso-Gd_rowse View Crystal unknown belonging to Project Unknown Dataset name unknown140515 "(+) F(+) SIGF(+) SIGF(+) SIGF(+) SIGF(-) Chisa cattering from fluorescence	Job title [No title given]			
Phaser SAD pipeline Run SHELXD before Phaser Run Parrot (density modification) after Phaser Run Buccaneer (model building) after Parrot Define data • WTZ in Full path Thome/data/15-05-13-lyso/lyso-Gd_SAD/xds_step3/Lyso-Gd_orwse View Crystal unknown beforging to Project unknown Define data • Stataset name unknown140515 F(+) SIGF(+) SIGF(-) IGF(+) F(-) F(+) F(-) SIGF(-) SIGF(-) IGF(-) F(-) Stataset name unknown140515 F(+) F(-) SIGF(-) SIGF(-) F(-) F(-) SIGF(-) F(-) F(-) SIGF(-) F(-)	Mode for experimental phasing	Single-wavelength anomal	lous dispersion (SAD)	-
Run SHELXD	Phaser SAD pipeline			
Run Parrot (density modification) after Phaser ■ Run Buccaneer (model building) after Parrot Define data MTZ in Full path. Thome/data/15-05-13-lyso/lyso-Gd_SAD/xds_step3/Lyso-Gd_owse View Crystal unknown belonging to Project unknown Define data Dataset name Unknown belonging to Project unknown Dataset name Unknown140515 SIGF(+) SIGF(-) C(-) F(-) SIGF(-) SIGF(-) F(-) P P SIGF(-) SIGF(-) F(-) P FP SIGF(-) SIGF(-) F(-) Resolution 38.635 A to 16.648 A; Wavelength 1.7119 Space group read from mtz file 'P 43 21 2'; Enantiomorph choice Both Enter scattering from fluorescence scan (default is to calculate t' and f' from wavelength) Define atoms Atom sites run sheix C/D Image: State compound ILLG-map completion on: all atom types Crystal contains cluster compound LLG-map completion on: all atom types Felx parameters Image: State component in asymmetric unit Component #1 protein sequence file Number in asymmetric unit Image: State component Compo	Run SHELXD — before Phase	iser		
Define data Image: Start S	Run Parrot (density modification)	n) after Phaser 🔳 Run Bu	ccaneer (model building)	after Parrot
MTZ in Full path Cthome/data/15-05-13-lyso/lyso-Gd_SAD/xds_step3/Lyso-Gd_rowse View Crystal unknown belonging to Project unknown Dataset name Unknown Dataset name unknown140515 SIGF(+) SIGF(-) SIGF(-) Crystal F(+) F(+) SIGF(-) SIGF(-) SIGF(-) Crystal FreeRflag SIGFP SIGFP SIGFP P FP SIGF (-) SIGFP SIGFP REER FreeRflag Store group read from mtz file 'P 43 21 2'; Enantiomorph choice Both Enter scattering from fluorescence scan (default is to calculate 1' and 1" from wavelength) Define atoms Atom sites Crystal contains cluster compound LLG-map completion on: all atom types Shelx parameters Component #1 protein sequence file Number in asymmetric unit 1 Stee of SAD/ccp4_MR/193Lfast Browse View Edit list Define another component Edit list Define another component Stores Stee of SAD/ccp4_MR/193Lfast Browse Yiew Close Run Save or Restore Close Save or Restore Close </td <td>Define data</td> <td></td> <td></td> <td>×.</td>	Define data			×.
Crystal unknown belonging to Project unknown Dataset name unknown140515	MTZ in Full path = //home/c	ata/15-05-13-lyso/lyso-Gd_S	AD/xds_step3/Lyso-Gd	Towse View
Dataset name unknown140515 F(+) F(+) SIGF(+) SIGF(-) F(-) F(-) SIGF(-) SIGF(-) FP FP SIGFP SIGFP REER FreeRflag FreeRflag Bace group read from mtz file 'P 43 21 2'; Enantiomorph choice Both Enter scattering from fluorescence scan (default is to calculate f' and f" from wavelength) Define atoms Define atoms Image: Component is cluster compound Image: Component is cluster compound LLG-map completion on: all atom types Image: Components in asymmetric unit Shelx parameters Image: Components in asymmetric unit Image: Component is cluster component Component #1 protein sequence file Number in asymmetric unit SEQ file Full path. Image: Component is component Image: Component Accessory parameters Image: Component Image: Component Image: Component Upput control Image: Component Image: Component Image: Component Run Save or Restore Close Close	Crystal unknown	pelonging to Project unkno	own	
F(+) F(+) SIGF(+) SIGF(-) F(-) F(-) SIGF(-) SIGF(-) FP FP SIGFP SIGFP FREER FreeRflag FreeRflag Space group read from mtz file 'P 43 21 2' ; Enantiomorph choice Both Enter scattering from fluorescence scan (default is to calculate 1' and 1" from wavelength) Define atoms Atom sites run ShelxC/D Find 4 heavy atoms of type GD Crystal contains cluster compound LLG-map completion on: all atom types Find 4 Shelx parameters Formation of the asymmetric unit Formation of the asymmetric unit Component #1 protein sequence file Number in asymmetric unit SEQ file Full path. whome/data/15-05-13-lyso/lyso-Gd_SAD/ccp4_MR/193L.fast Browse View Edit list Define another component file Suppert parameters Fourier of Save or Restore Fourier of Save or Restore	Dataset name unknown140515			
F(-) F(-) SIGF(-) SIGF(-) FP FP SIGFP SIGFP FREER FreeRflag FreeRflag Resolution [38.635] A to [1.648] A; Wavelength [1.7119] Space group read from mtz file 'P 43 21 2'; Enantiomorph choice Both Enter scattering from fluorescence scan (default is to calculate f' and f" from wavelength) Define atoms Atom sites run ShelxC/D Find 4 Ind 4 heavy atoms of type GD Crystal contains cluster compound LLG-map completion on: all atom types Find 4 Shelx parameters For a symmetric unit For a symmetric unit Component #1 protein sequence file Number in asymmetric unit 1 SEQ file Full path. whome/data/15-05-13-lyso/lyso-Gd_SAD/ccp4_MR/193L.fast Browse View Edit list Define another component file file File file Luccessory parameters For a scatterial For a scatterial file file Luctorial Close File Save or Restore Close	F(+) F(+)	SIGF(+)	SIGF(+)	-
P P SlorP SlorP FREER FreeRilag Resolution [38.635] A to [1.648 A; Wavelength [1.7119] Space group read from mtz file 'P 43 21 2'; Enantiomorph choice Both Enter scattering from fluorescence scan (default is to calculate f' and f" from wavelength) Define atoms Atom sites run SheixC/D Crystal contains cluster compound LLG-map completion on: all atom types Sheix parameters Composition of the asymmetric unit Total scattering determined by components in asymmetric unit Component #1 protein SEQ file Full path. Upput control group Edit list Define another component tccessory parameters group Dutput control group Run Save or Restore Run Save or Restore	F(-) F(-)	SIGF(-)	SIGF(-)	
Intering Frommag Resolution 38.635 A to 1648 A; Wavelengti 1.7119 Space group read from mtz file 'P 43 21 2'; Enantiomorph choice Both	EPEEP FreeBflag		SIGPP	
Space group read from mtz file 'P 43 21 2'; Enantiomorph choice Both	Besolution 38.635 A to 1.648	A· Wavelength 17	119	
Enter scattering from fluorescence scan (default is to calculate 1' and 1" from wavelength) Define atoms Atom sites run ShelxC/D Crystal contains cluster compound LLG-map completion on: all atom types Shelx parameters Component #1 protein sequence file Number in asymmetric unit 1 SEQ file Full path. Informe/data/15-05-13-lyso/lyso-Gd_SAD/ccp4_MR/193L.fast Broke View Edit list Define another component Accessory parameters Run Save or Restore Close	Space group read from mtz file 'P (3 21 2' : Fnantiomorph c	thoice Both	
Certifier stattering from indorescence scan (denomination is to calculate if and if from wavelength) Define atoms Atom sites run ShelxC/D Find 4 heavy atoms of type GD Crystal contains cluster compound Crystal contains cluster compound LLG-map completion on: all atom types Shelx parameters Image: Components in asymmetric unit Component #1 protein SEQ file Full path Chome/data/15-05-13-lyso/lyso-Gd_SAD/ccp4_MR/193L.fast Brokes View Edit list Define another component Accessory parameters Image: Close Run Save or Restore Close	Enter scattering from fluorescel	ce scan (default is to calcu	late f and f" from wavele	anath)
Atom sites run ShelxC/D Find 4 heavy atoms of type GD Crystal contains cluster compound LLG-map completion on: all atom types Shelx parameters Composition of the asymmetric unit Fotal scattering determined by components in asymmetric unit Component #1 protein sequence file Number in asymmetric unit 1 SEQ file Full path	Define stoms	ice scan (denault is to calcu	nate i and i nom wavere	angun)
Find 4 heavy atoms of type GD Crystal contains cluster compound LLLG-map completion on: all atom types Shelx parameters Composition of the asymmetric unit Fotal scattering determined by components in asymmetric unit Component #1 protein sequence file Number in asymmetric unit 1 SEQ file Full path	Atom sites run ShelvC/D	1		-
Crystal contains cluster compound Crystal contains cluster compound LLG-map completion on: all atom types Shelx parameters Composition of the asymmetric unit Fotal scattering determined by components in asymmetric unit Component #1 protein sequence file Number in asymmetric unit SEQ file Full path. Full pa	Find 4 beaut atoms of two	(CD)		
Conjourne control Interference Shelx parameters Interference Composition of the asymmetric unit Interference Fotal scattering determined by components in asymmetric unit Component #1 protein SEQ file Full path Component #1 protein SEQ file Full path Chome/data/15-05-13-lyso/lyso-Gd_SAD/ccp4_MR/193L.tas! Browse Edit list Define another component Accessory parameters Interference Dutput control Interference Expert parameters Interference Run Save or Restore Close	Crystal contains cluster compo	ind		
Composition of the asymmetric unit Composition of the asymmetric unit Fotal scattering determined by component #1 protein sequence file Number in asymmetric unit SEQ file Full path. Composition Edit list Define another component Accessory parameters Support parameters Support parameters Run Save or Restore Close	LL G man completion on; al	stom types		
Composition of the asymmetric unit Component and the asymmetric unit Fotal scattering determined by components in asymmetric unit Component #1 protein Sequence file Number in asymmetric unit Sequence file Number in asymmetric unit Edit list Define another component	Shelv parameters			
Fotal scattering determined by components in asymmetric unit Component #1 protein sequence file Number in asymmetric unit SEQ file Full path. Immode/data/15-05-13-lyso/lyso-Gd_SAD/ccp4_MR/193L.fast Broke View Edit list Define another component Accessory parameters Immode/data/15-05-13-lyso/lyso-Gd_SAD/ccp4_MR/193L.fast Broke View Edit list Define another component Immode/data/15-05-13-lyso/lyso-Gd_SAD/ccp4_MR/193L.fast Broke View Edit list Define another component Immode/data/15-05-13-lyso/lyso-Gd_SAD/ccp4_MR/193L.fast Immode/data/15-05-13-lyso/	Composition of the asymmetric uni	•		
Component #1 protein a sequence file Number in asymmetric unit 1 SEQ file Full path. In the final sequence file Define another component Edit list Define another component Accessory parameters Sutput control Run Save or Restore Close	Total scattering determined by	components in asym	metric unit	- -
Component #1 protein sequence file Number in asymmetric unit 1 SEQ file Full path. Image: Sequence file Image: Sequence file View Edit list Define another component Accessory parameters Image: Sequence file Image: Sequence file Dutput control Image: Sequence file Image: Sequence file Run Save or Restore Close		Components in asym		·
SEQ file Full path Chome/data/15-05-13-lyso/lyso-Gd_SAD/ccp4_MR/193L.tast Broke View Edit list Define another component Accessory parameters Image: Close Supprt parameters Image: Close	Component #1 protein	sequence file	Number in asymmetri	c unit 1
Edit list Define another component Accessory parameters	SEQ file Full path	data/15-05-13-lyso/lyso-Gd_5	SAD/ccp4_MR/193L.fas1	Brovise View
Accessory parameters		Edit list	Define and	other component
Cutput control	Accessory parameters			
Expert parameters	Output control			
Run — Save or Restore — Close	Expert parameters			
Run — Save or Restore — Close			1	,
	Run —	Save or Re	estore 🔤	Close

Figure 2

What does this pipeline ?

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

$$\mathbf{P}_{\rm H}(\vec{u}) = \sum_{\rm h,k,l} (|F_{\rm PH}(\vec{s})| - |F_{\rm PH}(-\vec{s})|)^2 \exp[-2i\pi \vec{u} \cdot \vec{s}]$$

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 $+1\sigma$.

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 $+1\sigma$.



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:

				CCP4In	terfa	:e 7.0	0.053	running on Dor	ni-gre026402 Project: Ly	/soGd				- 🗆 🗙
List of jobs for project. Double-clic	k on	ıaj	ob	displa	ys t	he l	log	file, shift-do	uble-click reruns th	ne job.			Change Proje	ct Help
Refinement		-	3	3 1	1 M	ay	18	FINISHED	phaser_EP	SAD	$Phasin \Delta$	Dire	ctories&Project	Dir
Model Preparation				2 1	1 М 1 м	ay	18	FAILED	phaser_EP	SAD	Phasin nge sna		View Any File	
Restraint Preparation							10	THIOHD	5100015	onu	nge spu		The Any The	117
Run Refmac5												View Files	from Job	
Run NCS Phased Refinement												Search/S	ort Database	
Run Low Resolution Refinement				3								Graphica	l View of Projec	t
Model Completion & Analysis												Delete/A	rchive Files	
-												Kill Job		
												ReRun J	ob	
												Edit Job D	ata	_
												Preferen	ces	
												System A	dministration	_
			F	1		_						Manage	Updates	Exit

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.

Run Refmac5		_ 🗆 X									
Output MTZ File		Help									
Job title Refinement											
Do restrained refinement - using no prior phase information - input											
Input fixed TLS parameters											
no 🛁 twin refinement											
Use Prosmart: no - (low re:	Use Prosmart: no - (low resolution refinement)										
Run libg to generate external restraints (DNA/RNA) automatically											
Run Coot:findwaters to automatically add/remove waters to refined structure											
MTZ in LysoGd - LysoGd_3_parrot.hand.mtz	Browse	View									
FP FP Sigma SIGFP											
MTZ out LysoGd - LysoGd_3_parrot.hand_refmac1.mtz											
PDB in LysoGd - LysoGd_3_buccaneer.hand.pdb	Browse	View									
PDB out LysoGd - LysoGd_3_buccaneer.hand_refmac1.pdb	Browse	View									
LIB in LysoGd - Merge LIBIN	s Browse	View									
Output lib LysoGd - LysoGd_3_buccaneer.hand.cif	Browse	View									
Refmac keyword file LysoGd -	Browse	View									
Data Harvesting											
Refinement Parameters											
Setup Geometric Restraints		-									
Setup Non-Crystallographic Symmetry (NCS) Restraints											
External Restraints											
Monitoring and Output Options											
Run - Save or Restore -	Close										

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 +1 σ .

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*	Constraint	s 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 (alo	ong c)	$a = b; \alpha =$	$\beta = \gamma = 90$
Hexagonal 6-fold (alo	ong 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 ₁ 2 ₁ 2, P2 ₁ 2 ₁ 2 ₁ , C222 ₁ , C222, F222, I222, I2 ₁ 2 ₁ 2 ₁
Quadratic	4 422	P4, P4 ₁ , P4 ₂ , P4 ₃ , I4, I4 ₁ , P422, P42 ₁ 2, P4 ₁ 22, P4 ₁ 2 ₁ 2, P4 ₂ 22, P4 ₂ 2 ₁ 2, P4 ₃ 22, P4 ₃ 2 ₁ 2, I422, I4 ₁ 22
Trigonal	3 32	P3, P3 ₁ , P3 ₂ , R3, P312, P321, P3 ₁ 12, P3 ₁ 21, P3 ₂ 12, P3 ₂ 21, R32
Hexagonal	6 622	P6, P6 ₁ , P6 ₅ , P6 ₂ , P6 ₄ , P6 ₃ , P622, P6 ₁ 22, P6 ₅ 22, P622, P6 ₄ 22, P6 ₃ 22
Cubic	23 432	P23, F23, I23, P2 ₁ 3, I2 ₁ 3, P432, P4 ₂ 32, F432, F4 ₁ 32, I432, P4 ₃ 32, P4 ₁ 32, I4 ₁ 32

No. 96







422

Origin on 2[110] at $2_11(1,2)$

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

Po Mu Wy	sitio Itiplic ckoff	ns city, letter,		(Coordinates			Reflection conditions
on	e sym	incu y						General:
8	b	1	(1) x, y, z (5) $\bar{x} + \frac{1}{2}, y + \frac{1}{2}, \bar{z} + \frac{1}{2}$	(2) $-\frac{3}{4}$ (6) .	$ar{x}, ar{y}, z + rac{1}{2} \\ x + rac{1}{2}, ar{y} + rac{1}{2}, ar{z} + rac{1}{4}$	(3) $\bar{y} + \frac{1}{2}, x + \frac{1}{2}, z + \frac{3}{4}$ (7) y, x, \bar{z}	(4) $y + \frac{1}{2}, \bar{x} + \frac{1}{2}, z + \frac{1}{4}$ (8) $\bar{y}, \bar{x}, \bar{z} + \frac{1}{2}$	$\begin{array}{l} 00l: \ l=4n\\ h00: \ h=2n \end{array}$
								Special: as above, plus
4	а	2	<i>x</i> , <i>x</i> ,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 $2k + l = 4n$