

XChemExplorer

Manual

Current version

Date

09/03/2017

Author

Tobias Krojer

tobias.krojer@sgc.ox.ac.uk

Table of Contents

Introduction	3
Reference	3
Getting started	4
Prerequisites	4
Installation	4
Usage	5
Refinement	6
Refinement stages	6
Overview	7
Working with PanDDA models.....	10
Frequently asked questions.....	14
What happens when you step through the models?	14
What happens when I make changes to the model?	14
What happens when I refine the model?	15

Introduction

XChemExplorer (XCE) is a data management and workflow tool which supports large scale analysis of protein-ligand structures by X-ray crystallography. It is not an actual algorithm, but serves as a launch pad for batch submission and analysis of the essential steps in the structure determination of protein-ligand structures.

Reference

Krojer, T., Talon, R., Pearce, N., Collins, P., Douangamath, A., Brandao-Neto, J., Dias, A., Marsden, B., and von Delft, F. (2017). The XChemExplorer graphical workflow tool for routine or large-scale protein–ligand structure determination. *Acta Cryst D* 73, 267–278.

XChemExplorer is a data management and workflow tool but not an algorithm. It makes extensive use of other people's software, therefore please cite their work accordingly:

- XIA2
- DIMPLE
- ACEDRG
- PanDDA
- COOT
- REFMAC
- PHENIX

Getting started

Prerequisites

It is essential that CCP4 (<http://www ccp4.ac.uk>) version 7.0 or higher is installed and correctly setup. XCE uses the python version that comes with it and will therefore not work if it does not exist. Additionally, it may be useful to also install PHENIX, since XCE uses several of its tools for validation purposes.

Installation

Download XChemExplorer from <http://tkrojer.github.io/XChemExplorer>

Put the gzipped tar archive to wherever you want XCE to be installed. In case you have no root privileges, put it somewhere into your home directory, e.g.:

```
/home/tkrojer/software
```

Then change to the respective directory and unpack the archive, e.g.:

```
cd /home/tkrojer/software  
gunzip XChemExplorer-1.0-beta.3.4.tar.gz  
tar -xvf XChemExplorer-1.0-beta.3.4.tar
```

This will create a new directory, i.e. from now on your XChemExplorer directory. Change into this directory, e.g.:

```
cd XChemExplorer-1.0-beta.3.4
```

The contents of the directory should look something like this when you type '**ls -l**':

```
-rwxr-xr-x 1 tobiaskrojer staff 185B 2 Mar 09:53 XChemExplorer.csh  
-rwxr-xr-x 1 tobiaskrojer staff 309K 2 Mar 09:53 XChemExplorer.py  
-rwxr-xr-x 1 tobiaskrojer staff 186B 2 Mar 09:53 XChemExplorer.sh  
drwxr-xr-x 11 tobiaskrojer staff 374B 2 Mar 09:53 helpers  
drwxr-xr-x 12 tobiaskrojer staff 408B 26 Jan 10:53 image  
drwxr-xr-x 34 tobiaskrojer staff 1.1K 9 Mar 14:41 lib  
drwxr-xr-x 4 tobiaskrojer staff 136B 2 Mar 09:53 setup-scripts  
drwxr-xr-x 7 tobiaskrojer staff 238B 9 Mar 14:41 web
```

The only thing left to do is to edit the XChemExplorer.sh or XChemExplorer.csh file, depending on which shell you are using. Open XChemExplorer.sh for bash shells or XChemExplorer.csh for C-shells with your editor of choice and edit the line

```
export XChemExplorer_DIR='/usr/local/scripts/tobias/XChemExplorer'
```

to where you XChemExplorer is installed. In our example this would be

```
export XChemExplorer_DIR='/home/tkrojer/software/XChemExplorer-1.0-beta.3.4'
```

That's it!

Usage

You can now run XCE by typing

```
/home/tkrojer/software/XChemExplorer-1.0-beta.3.4/XChemExplorer.sh
```

It may however be easier if you insert an alias into your .bashrc or .cshrc file:

```
alias XChemExplorer='/home/tkrojer/software/XChemExplorer-1.0-beta.3.4/XChemExplorer.sh'
```

Refinement

Refinement stages



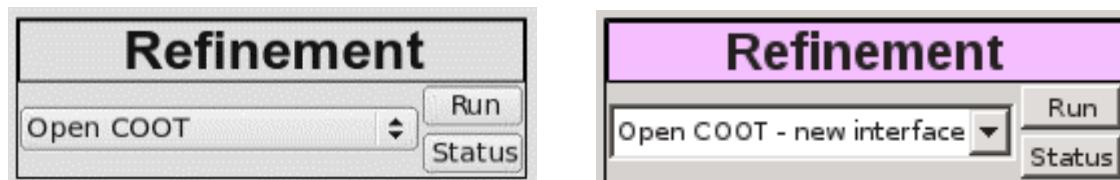
The figure summarises the refinement stage model that XCE uses to track the progress of each sample and which is also used to triage samples. You need at least do some initial refinement to be able to look at samples in the refinement interface.

Overview

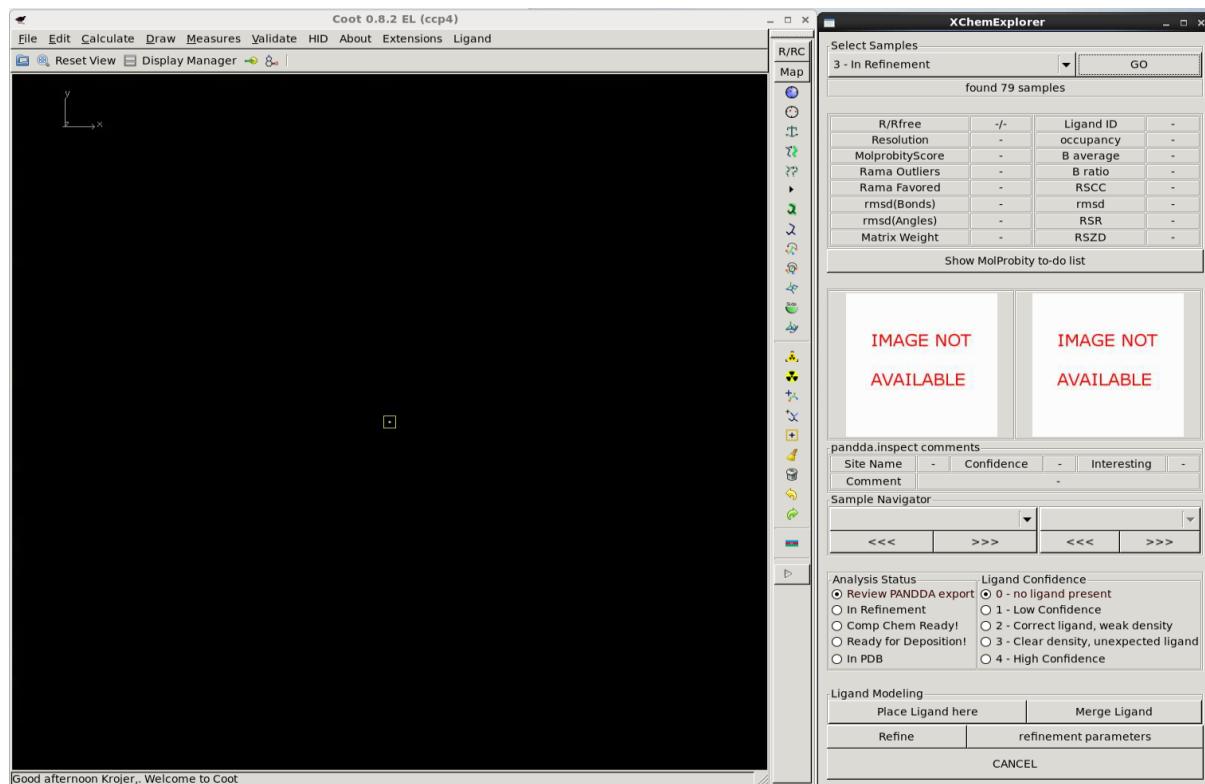
A list of all models that are currently ‘in refinement’ can be viewed in the ‘Refinement’ tab.

	Sample ID	Compound ID	Refinement Space Group	Resolution	Rcryst	Rfree	Outcome	PandDA site details	Refinement Status												
1	NUDT22A-x0106	N13421a	P 21 21 21	1.39	0.20663	0.22948	3 - In Refinement	<table border="1"> <tr><td>Index</td><td>Name</td><td>Status</td></tr> <tr><td>2</td><td>Crystal contact</td><td>3 - In Refinement</td></tr> </table>	Index	Name	Status	2	Crystal contact	3 - In Refinement	finished						
Index	Name	Status																			
2	Crystal contact	3 - In Refinement																			
2	NUDT22A-x0161	N14124a	P 21 21 21	1.43	0.20613	0.23017	3 - In Refinement	<table border="1"> <tr><td>Index</td><td>Name</td><td>Status</td></tr> <tr><td>1</td><td>Mg site & putasteric site 1</td><td>4 - CompChem ready</td></tr> <tr><td>4</td><td>near xtal contact</td><td>3 - In Refinement</td></tr> </table>	Index	Name	Status	1	Mg site & putasteric site 1	4 - CompChem ready	4	near xtal contact	3 - In Refinement	finished			
Index	Name	Status																			
1	Mg site & putasteric site 1	4 - CompChem ready																			
4	near xtal contact	3 - In Refinement																			
3	NUDT22A-x0182	N14004a	P 21 21 21	1.48	0.21242	0.23591	3 - In Refinement	<table border="1"> <tr><td>Index</td><td>Name</td><td>Status</td></tr> <tr><td>1</td><td>Mg site & putasteric site 1</td><td>4 - CompChem ready</td></tr> </table>	Index	Name	Status	1	Mg site & putasteric site 1	4 - CompChem ready	finished						
Index	Name	Status																			
1	Mg site & putasteric site 1	4 - CompChem ready																			
4	NUDT22A-x0196	N14099a	P 21 21 21	1.75	0.21716	0.25772	3 - In Refinement	<table border="1"> <tr><td>Index</td><td>Name</td><td>Status</td></tr> <tr><td>1</td><td>Mg site & putasteric site 1</td><td>4 - CompChem ready</td></tr> </table>	Index	Name	Status	1	Mg site & putasteric site 1	4 - CompChem ready	finished						
Index	Name	Status																			
1	Mg site & putasteric site 1	4 - CompChem ready																			
5	NUDT22A-x0202	N13854a	P 21 21 21	1.40	0.21239	0.23580	3 - In Refinement	<table border="1"> <tr><td>Index</td><td>Name</td><td>Status</td></tr> <tr><td>1</td><td>Mg site & putasteric site 1</td><td>4 - CompChem ready</td></tr> </table>	Index	Name	Status	1	Mg site & putasteric site 1	4 - CompChem ready	finished						
Index	Name	Status																			
1	Mg site & putasteric site 1	4 - CompChem ready																			
6	NUDT22A-x0215	N13708a	P 21 21 21	1.56	0.21362	0.24442	3 - In Refinement	<table border="1"> <tr><td>Index</td><td>Name</td><td>Status</td></tr> <tr><td>1</td><td>Mg site & putasteric site 1</td><td>4 - CompChem ready</td></tr> <tr><td>2</td><td>Crystal contact</td><td>3 - In Refinement</td></tr> <tr><td>5</td><td>Met1 and other missing resl</td><td>3 - In Refinement</td></tr> </table>	Index	Name	Status	1	Mg site & putasteric site 1	4 - CompChem ready	2	Crystal contact	3 - In Refinement	5	Met1 and other missing resl	3 - In Refinement	finished
Index	Name	Status																			
1	Mg site & putasteric site 1	4 - CompChem ready																			
2	Crystal contact	3 - In Refinement																			
5	Met1 and other missing resl	3 - In Refinement																			

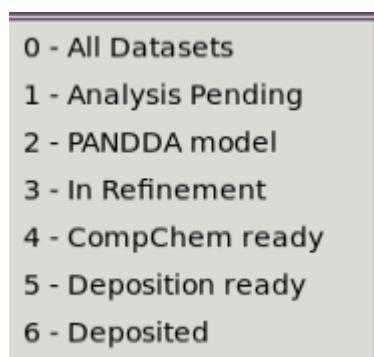
If you want to inspect them in COOT and if necessary want to refine them further, choose ‘Open COOT’ or ‘Open COOT – new interface’ from the magenta action box and press RUN. From v1.0-beta.3.4 (21/02/20160 onwards, XCE provides two different interfaces for inspection and refinement of structures and it is recommended to use the new interface! The remainder of the manual will only refer to the new version!



This will launch COOT and the respective XCE interface.



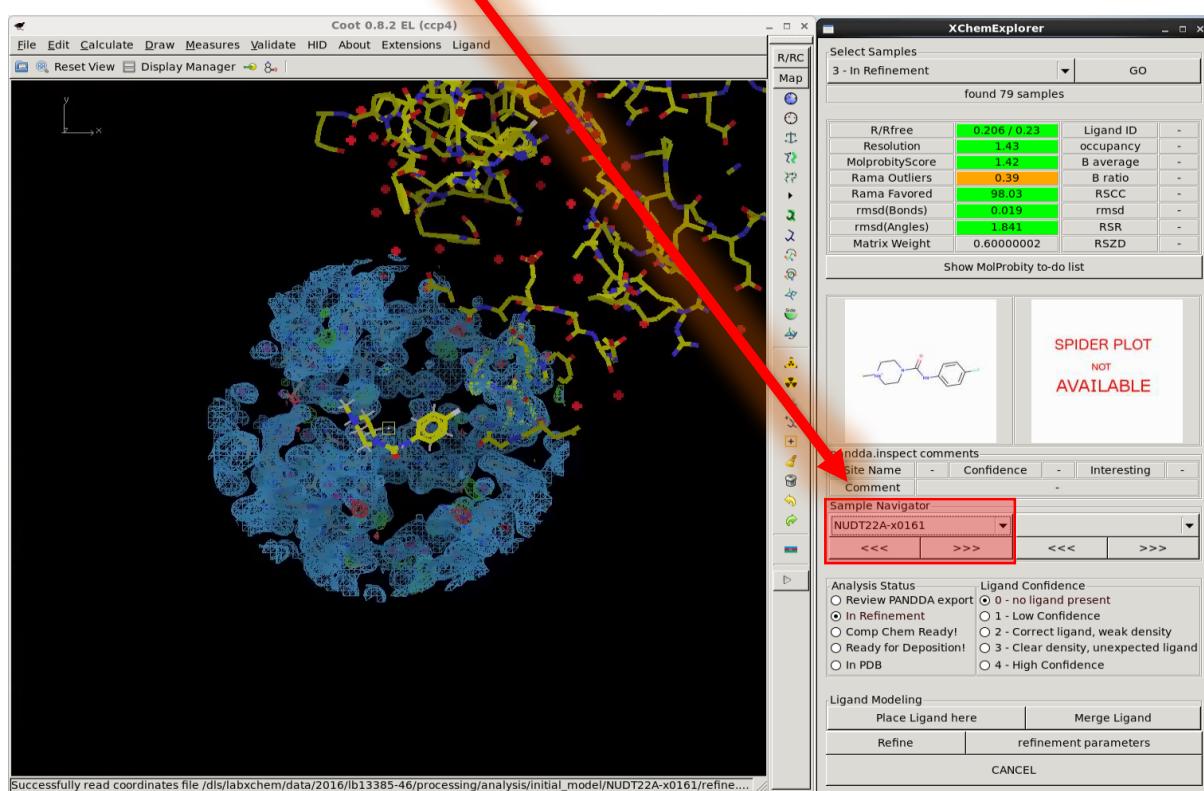
First you need to select the subset of samples you want to review/ refine from drop-down menu at the top of the XCE interface (1). The drop-down lets you choose the Refinement Stage and once you press GO will load all samples that are in the respective Refinement stage. The image below shows you all the available categories



For example: if you want to browse through all the models that are already in the PDB, then select category 6 and press GO. And 40 structures are already in the PDB in this example.



In the left column of the ‘Sample Navigator’ Section, use the arrow buttons or the drop-down menu to select the structure of interest:



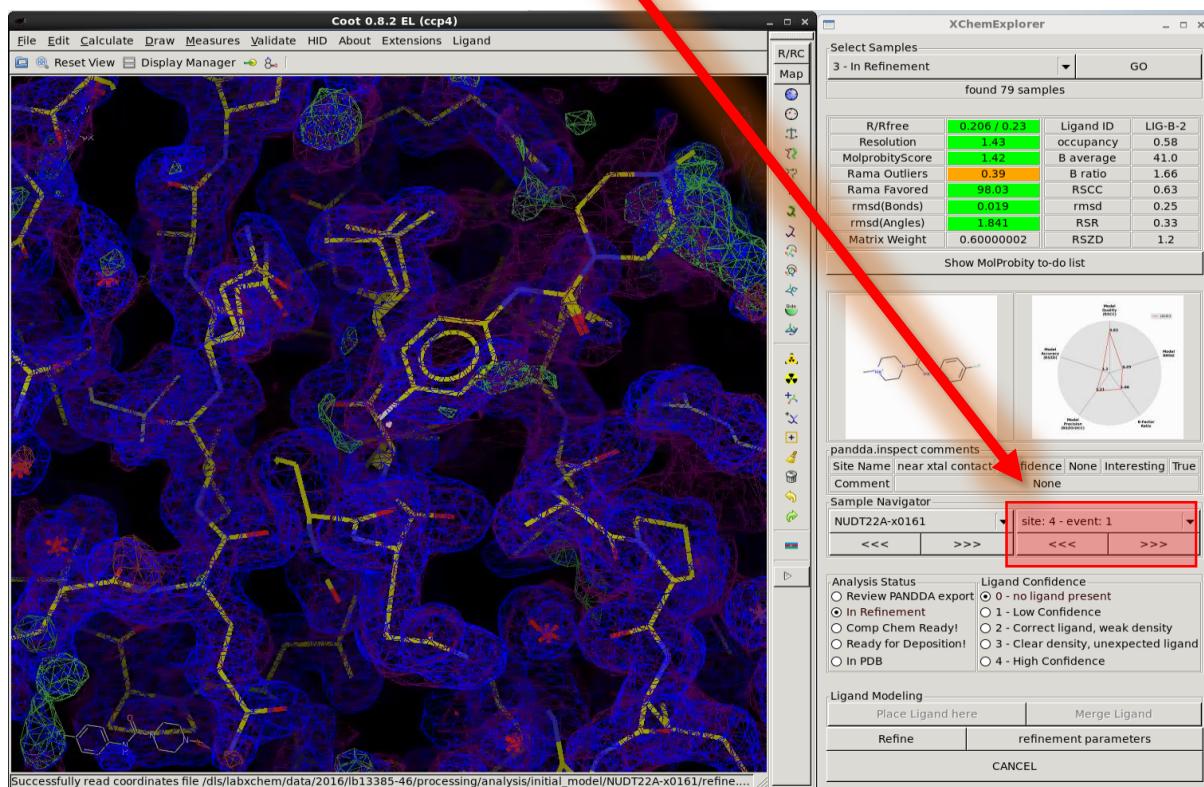
XCE will load the structure¹, 2fofc map (blue), fofc (green/red) map and a file (+ dictionary) of the ligand if the latter was created and specified in the database. This ligand molecule may be slightly confusing because it may seem to just float in space. However, the molecule is completely ignored as long as it is not merged into the main structure! It is only loaded to enable quick modelling if the ligand is not already part of the structure.

¹ Structure refers to the file called refine.pdb in the respective sample directory, or if no refinement has been carried out so far and category 0 or 1 are selected, then it will try to load dimple.pdb. Note: if you use XCE throughout the process, then refine.pdb as well as dimple.pdb are not actual files but symbolic links that point to the most recently refined file.

Important: if you are working with models from PanDDA, then you must select one of the sites (see below) and continue. It is of utmost importance that in case you're editing PanDDA models, modelling of the respective sites should be done with **pandda.inspect**. The XCE interface for COOT does not take care of PanDDA specific issues like proper assignment of occupancies and alternative conformations. In this case, use the interface to check the binding site, annotate the structure, tweak refinement parameters and for correcting other parts of the structure. If you discover that there is an error in the way you modelled the bound conformation, please go back to **pandda.inspect** (<http://pandda.bitbucket.org/manual.html#manual>), improve your model and export it again!

Working with PanDDA models

Use the drop-down menu or the arrow button in the right part of the 'Sample Navigator' section. XCE will centre on the event and will load the respective PanDDA event maps (dark violet) as well as quality metrics , spider plots and annotations.



Once you selected one of the sites, XCE will delete the ligand PDB file and grey out the 'Place ligand here' button and the 'Merge ligand' button in the 'Ligand Modelling' section to avoid any unwanted placement of ligand. You can of course use all the other features in COOT to add ligands, but unless you really know what you're doing, this will almost certainly end up in a mess. You really need to decide at the beginning if you want to use PanDDA for ligand identification and modelling or if you model ligands in the 'traditional' way!!!

You can now step through the sites and evaluate how confident you are that the ligand pose/identity is correct:

Ligand Confidence
<input type="radio"/> 0 - no ligand present
<input type="radio"/> 1 - Low Confidence
<input checked="" type="radio"/> 2 - Correct ligand, weak density
<input type="radio"/> 3 - Clear density, unexpected ligand
<input type="radio"/> 4 - High Confidence

The 'Analysis Status' refers to the overall structures, whereas 'Ligand Confidence' refers to the specific ligands! Also, the database will be updated immediately when you change status or confidence.

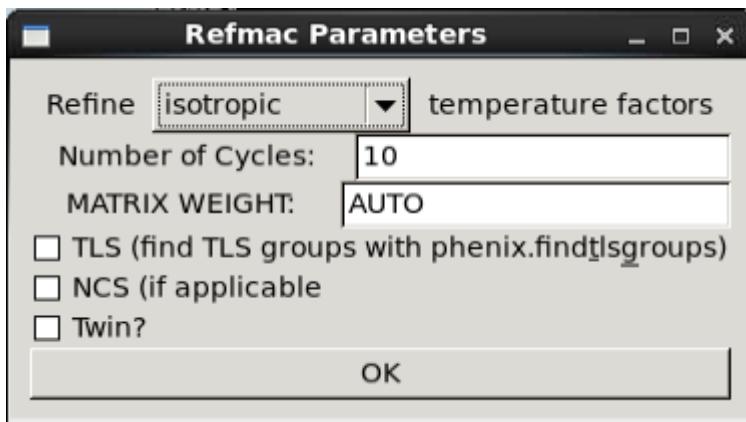
First review all the sites, then check the overall quality of the model:

R/Rfree	0.206 / 0.23	Ligand ID	LIG-B-1
Resolution	1.43	occupancy	0.61
MolprobityScore	1.42	B average	30.89
Rama Outliers	0.39	B ratio	1.21
Rama Favored	98.03	RSCC	0.8
rmsd(Bonds)	0.019	rmsd	0.26
rmsd(Angles)	1.841	RSR	0.23
Matrix Weight	0.60000002	RSZD	0.0
Show MolProbity to-do list			

Check Molprobity analysis:

MolProbity to-do list			
Ramachandran outliers			
Chain	Residue	Name	Score
A	132	ASP	0.023119
A	160	GLN	0.001898
Rotamer outliers			
Chain	Residue	Name	Score
A	1	MET	0.286132
A	5	VAL	0.012325
A	160	GLN	0.244349
A	162	LEU	0.000000
A	212	ARG	0.107241
Severe clashes			
<input type="checkbox"/> Show Probe dots	<input checked="" type="checkbox"/> Overlaps only		
Atom 1	Atom 2	Overlap	
A 261 ARG HB2	W 84 HOH O	-0.723000	
A 2 ASP OD1	A 4 GLU OE1	-0.717000	
A 4 GLU HB2	A 215 THR HG23	-0.643000	
A 212 BARG HD2	A 218 GLY O	-0.594000	
A 212 BARG CD	A 218 GLY O	-0.579000	
A 4 GLU HG23	A 274 PRO HG23	-0.551000	
<input type="button" value="Close"/>			

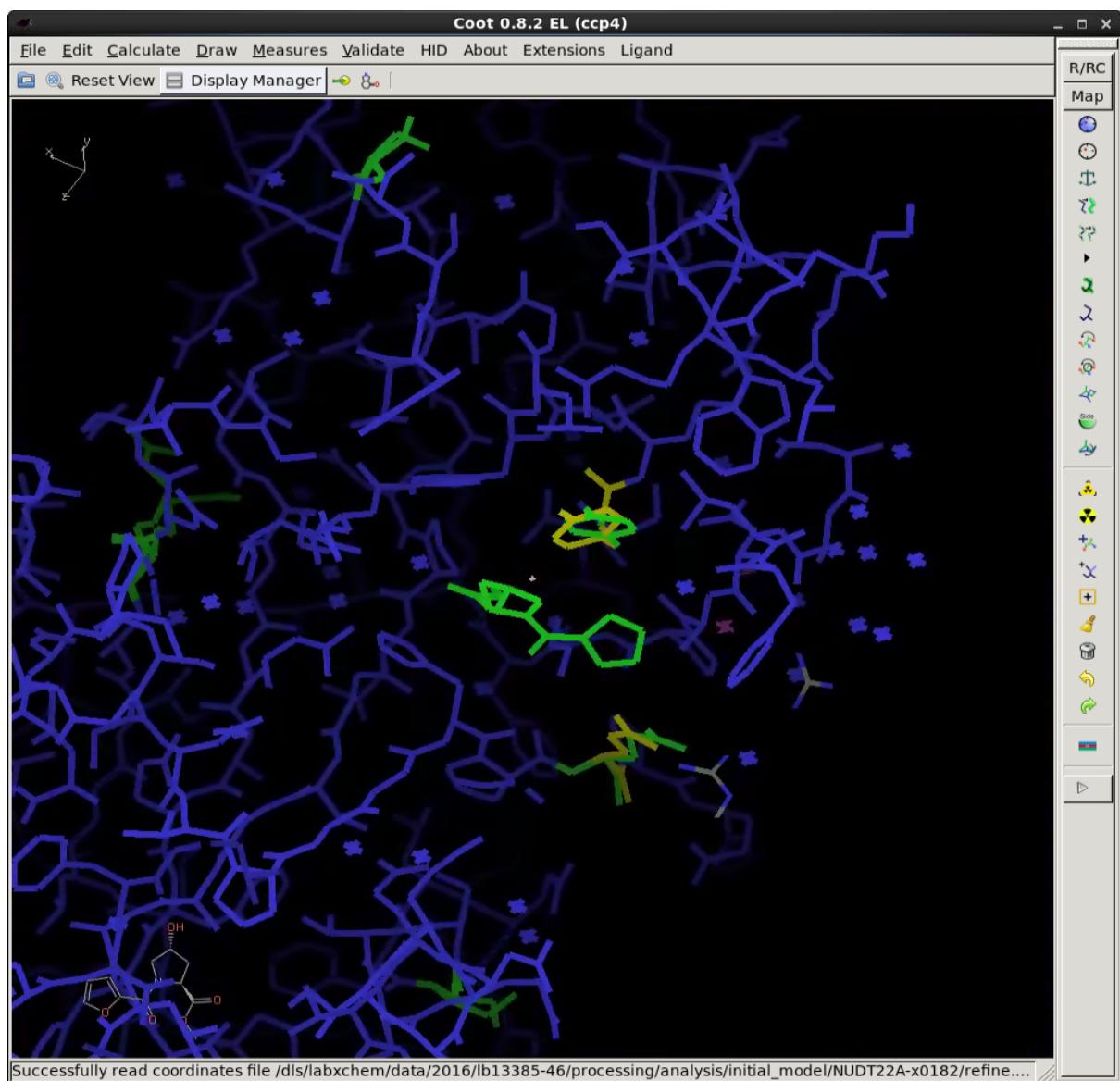
You can change some REFMAC parameters by clicking on ‘Refinement Parameters’ on the main panel. XCE will ‘remember’ your previous settings.



And if you think the model needs further refinement, then press ‘Refine’ on the main panel. If you refine structures at Diamond, each refinement job will automatically be sent to the cluster. In case no functional PBS queueing system is available, refinement will be carried out on the local machine. The latter will probably be the case when you work on your linux box or MAC at home.

Just to reiterate again: you can use the interface to tweak side-chains, place solvent molecules, ions or whatever you would usually do in COOT. However, do not touch the residues, ligands, solvent molecules and their surroundings that you modified in pandda.inspect. You can quickly find out which regions are affected by colouring your molecule by occupancy (see below). But again, if you find that

your model needs updating all the time in the same region than it is certainly quicker and better to modify the apo model accordingly and then repeat the map calculation and PanDDA analysis.



'Traditional' refinement

If you are not refining models built with pandda.inspect, then you can directly go to refinement once you've calculated initial maps. Technically of course, a cycle of initial refinement has been carried out already with REFMAC as part of the DIMPLE difference pipeline, but for XCE the Refinement Status is still '1 – Analysis Pending'.

Will read a symbolic link called dimple.pdb (and the corresponding dimple.mtz file) in the relevant sample directory.

In this case no ligands will have been built and there are two ways:

- use the default COOT protocol, i.e. first fit ligand to the electron density, then merge it into the protein pdb file
- use the XCE interface to put the protein into the screen centre, then merge it into the protein file

Same as before: ligand PDB and CIF files will be read automatically

Note: ligand is deleted from the list of models in COOT. This is done to avoid any confusion in case the 'merge' button was pressed twice, but can be a bit annoying if you want to build multiple instances of the ligand. Either launch a round of refinement and build the ligand during the second refinement cycle or load the ligand a second time and merge

Frequently asked questions

What happens when you step through the models?

XCE will load a file called refine.pdb (or dimple.pdb in case no refinement has been carried out so far) from the sample directory and if available a pdb file of the ligand and the respective restraints. Additionally, 2fofc as well as fofc maps are loaded (or they are calculated on the fly from refine.mtz/dimple.mtz if the map files were for whatever reason no pre-calculated). Note that refine.pdb/mtz and dimple.pdb/mtz are therefore reserved file names. Hence, if you wanted to manipulate your model outside the XCE environment you can easily do so and XCE will read the manipulated model in as long as it's called refine.pdb and present in the respective sample directory. One thing to keep in mind though: XCE carries out the actual refinement in a subfolder called Refine_<cycle number> and only links the resulting pdb/mtz files as refine.pdb/mtz into the project directory. Every time a new refinement is launched, it will first delete the respective symbolic links. But it will also delete it if it is an actual file. So if you want to keep the original, better create a symbolic link called refine.mtz. When you go through the models, it will remove all currently loaded models and load the aforementioned files from the next sample directory.

What happens when I make changes to the model?

All changes that you make to the model called refine.pdb will be preserved if you launch refinement (see next point). They will however be lost if you go to the next dataset. XCE will currently not ask if you want to keep the changes, the pdb file will be deleted from the list of molecules in COOT and it

will be lost forever! Also, be careful if you read in additional molecules with the same name, for example if you want to analyse something. COOT does not mind if molecules have the same name since every molecule that is read in gets a unique internal identifier. XCE however recognises molecules by filename it may get confused in case of duplicates.

What happens when I refine the model?

When you press refine, XCE will take the model called refine.pdb and save it to a subfolder called Refine_<cycle number + 1> as in.pdb together with the shell script that will be used for refinement. If something goes wrong and the reason for failure is not clear it is usually a good starting point to look at the logfiles in the respective folder. Keep in mind that if you added for example water or other solvent molecules and did not merge them into refine.pdb, then they will not be included. XCE does not do any automatic merging!