“Coot for SHELXL” Tutorial

Paul Emsley & Judit É Debreczeni

August 27, 2022

Contents

1 Introduction

This is an extended description of the presentations of Coot at the ACA meeting in Salt Lake City, 2007 and 2014 IUCr meeting in Montreal.

It should also serve as a tutorial for Coot specific for use with SHELXL. As such, it is intended for those already familiar with the basics of SHELXL refinement[1], and would like to use Coot as a model building and validation tool between SHELXL refinement cycles. This is, however, no substitute for the standard Coot tutorial[2].

2 A SHELXL Project

Refinement in SHELXL requires two files:

- **ins** – instruction file containing atomic coordinates, restraints, constraints and other refinement instructions. Atomic coordinates are fractional coordinates and protein molecules are represented as one single chain.
- **hkl** – SHELX format reflection file (typically HKLF4).

and will produce the following:

- **res** – similar to `ins` with updated values of refined parameters,
- **fcf** – a small molecule style cif file containing phases,
- **lst** – listing file containing diagnostic information, useful for spotting errors in the model or refinement parameters,
- **pdb** – pdb style coordinate output if WPDB card was given.

All extensions are appended to a common root, e.g. *my-project*.`ins`, *my-project*.`fcf` etc.

2.1 Reading and writing SHELX coordinate files

Coordinate files in SHELX format (`ins` and `res`) are valid coordinate formats in Coot, i.e. they can be read and written. However, not all SHELX commands are interpreted.

- **CELL, LATT and SYMM cards** are used to determine the cell and space group when reading `ins/res`.
- **Atomic coordinates** are reformated upon as they are read: the single protein chain is split up based on gaps in residue number. It is assumed that a gap larger than 21 indicates the beginning of a new chain, where a new chain identifier is assigned. Residues are not renumbered, *i.e.* residue numbers correspond to those in the SHELX coordinate file.

  By default, hydrogens are displayed for protein molecules if they are present in `ins` or `res`. They can be set to be invisible in Edit→Bond Parameters...
- **PART cards** are converted to alternate conformation identifiers (up to four alternate conformations). Negative PART numbers translate as lower case identifiers.

[1] If this is not the case you might want to read the following: http://shelx.uni-ac.gwdg.de/SHELX/shelx.pdf and http://shelx.uni-ac.gwdg.de/SHELX/pn1a.htm

[2] http://www2.mrc-lmb.cam.ac.uk/Personal/pemsley/coot/tutorials/tutorial.pdf

- **Anisotropic displacement parameters** are read and converted to B-factors. Coot displays thermal ellipsoids, too.
- **Free variables** are mostly used in macromolecular refinement for co-restraining occupancies of atoms in alternate conformations. Coot keeps free variables read from `res` files and writes an updated FVAR list to `ins`. It assigns a new free variable on splitting a residue and dissociates an already existing fvar from the residue when its alternate conformation is deleted, but will not renumber or otherwise change fvars. SHELX style occupancies are not unencoded, they are used to reconstruct the SHELX format when coordinates are written to `ins`. (They should also serve as a potential mechanism of changing free variables within Coot in the future.)
- **AFIX cards** are handled as attributes of atoms.
- **Most other** pre-coordinate SHELXL instructions are not interpreted[3] and are written to `ins` unchanged. (In other words, there are certain limitations of Coot's ability to create a fully valid `ins` file, and will definitely fail if we try to create `ins` from scratch, e.g. from a `pdb` file.[4]) There is, however, a mechanism to insert SHELX instructions into the the new `ins` file: with the scripting command `(add-shelx-string-to-molecule` *`imol string`*`)`
It is also accessible through the SHELX menu (after running the script `shelx-extension.scm`).

To make SHELXL coordinate formats filterable in the file selection window, `(add-coordinates-glob-extension ".res")` and `(add-coordinates-glob-extension ".ins")` should be included in the `.coot` file.

2.2 Reading `fcf` files

Coot can open `fcf` files and calculatesσ_A-weighted electron density and difference maps.

2.3 Reading `lst` files

Listing files corresponding to `ins/res` files can be opened with Coot and it will filter them for interesting features in the model and refinement flagged by SHELXL. Typically, these are potentially split atoms and disagreeable restraints, which are then displayed in an interesting-things-GUI. Such visualization of errors in the model provides a great validation tool in the context of a SHELXL project. The scripting function `(read-shelx-lst-file` *`lst-file imol`*`)` can be used to read `lst` files; alternatively, `shelx-extensions.scm` provides a GUI option: SHELX→Read LST File...

2.4 Opening SHELXL projects

As a consequence of SHELXL's file naming convention, it is straight-forward to open all relevant files of a project (`res`, `fcf` and `lst`) in one go. This can be done using `(read-shelx-project` *`file-name`*`)` where filename's root is the SHELX project's name. There is also a GUI option in the SHELX menu (after running the script `shelx-extension.scm`).

[3] This includes SHELX style restraints. As a consequence, unconventional residues, e.g. ligand molecules, require mmcif dictionaries (such as the Refmac dictionary) for refinement/regularization in Coot.

[4] You should be using SHELXPRO for that.

2.5 Running SHELXL from Coot

SHELXL can be run from Coot using either the function `(shelxl-refine` *`imol hkl-file`*`)` or SHELX→SHELXL refine... Specifying a `hkl` file is optional, to be used though if the `hkl` file does not correspond to the `res/ins` file.

Before running SHELXL, Coot creates a time-stamped `ins` file and a sym-link to the `hkl` file. These, together with the resulting files are stored in the coot-shelxl directory. `res` and `fcf` files generated by SHELXL will be read in automatically after refinement has finished.

3 03srv

In the small molecular world, it is generally not necessary to display electron density maps, as they can be represented by a simple peaklist quite precisely. However, visual examination of especially difference density maps can reveal subtle errors in refinement and can prove quite educational. So, suppose we have a small molecule project in some middle stage of refinement and wish to have a glance of the electron density map and maybe fix problematic parts of the structure.

3.1 Reading the project

- We'll start Coot:
  ```
  $ coot
  ```
  Now load the SHELX Module

- Calculate → Modules → SHELX

  [*An additional menu item, SHELX appears in the menu bar.*]

- Open the project 03srv with
  SHELX→Read SHELX project...
  and select any of the `03srv` files, i.e. `res`, `lst` or `hkl`.

  [*Coot now displays the 03srv molecule and the corresponding electron density maps. Note that it has also created `03srv.fcf.cif`, a macromolecular equivalent of `03srv.fcf`.*][5]

- Note that the fcf file from SHELXL should be make using "LIST 6".

- Note that Coot has read `03srv.lst` as well but it has decided that there was "nothing interesting" listed there.

- Now, just for speed and convenience, let's make the map radius a bit smaller. Select Edit→Map Parameters... and set the Map Radius to 5Å and click OK.

3.2 Hydrogens

- Let's examine the structure by moving around a bit in the density. We see large positive and negative difference density peaks around some of the atoms. Let's centre on C2BA (the methyl group with negative density on the hydrogen atoms and positive between them) by middle-clicking on it (Figure 1).

  "Hydrogens in wrong place!" — you might say.

[5] Alternatively, we could have opened the coordinates and the `fcf` file using File→Open coordinates... and File→Open MTZ, mmCIF, fcf and phs...

- Indeed, these hydrogens are incorrectly set: they are exactly 60 degrees off. Let's get rid of them. Select Calculate→Model/Fit/Refine,

  [*Coot displays the Model/Fit/Refine window*]

  click Delete,

  [*Coot displays the Delete item window*]

  click on the Atom and Keep Delete Active buttons and click on the hydrogen atoms to delete them. Finally, close the Delete window.

- Now we need to tell SHELXL to set the hydrogens properly. As seasoned SHELXers, we know that methyl groups require the "HFIX 137" instruction. Select SHELX→Add SHELXL instruction... from the menu bar.

  [*A window appears where we can select our molecule and type the new command.*]

  Select 03srv.res for model and type `HFIX 137 C2BA` in the instruction entry box. Click OK.

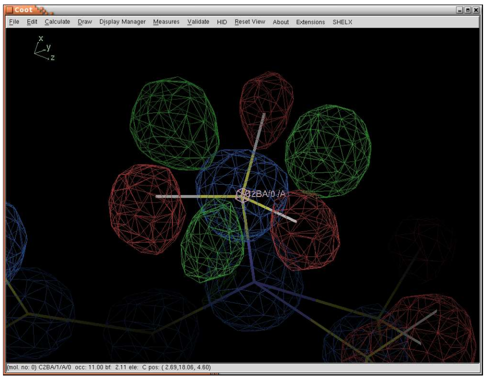
- Let's undisplay the difference density — just to see the improvement after refinement. Click Display Manager in the menu bar

  [*The Display Control window appears*]

  and click the Display button next to 03srv.fcf.cif Difference SigmaA.

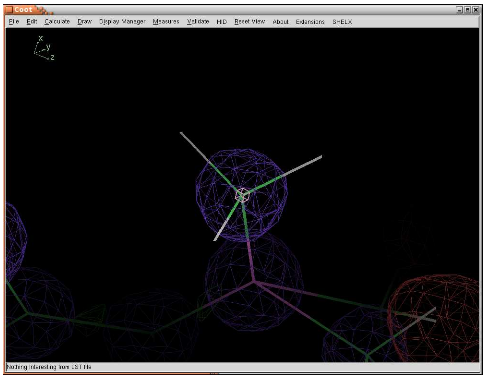
- To rerun SHELXL on the model with the new instruction for hydrogens, click SHELX→SHELXL Refine... and select 03srv.res as molecule to be refined. We do not need to set the `hkl` file as we already have a `03srv.hkl` to refine against. Click Refine.

  [*Coot now runs SHELXL in the background and displays the new structure and electron density maps when the refinement has finished.* (Figure 2)]



(mol. no: 0) C2BA/1/A/0 occ: 11.00 bf: 2.11 ele: C pos: (2.69,18.06, 4.60)

Figure 1: Difference density peaks around C2BA indicating that hydrogen atoms were incorrectly set



Nothing Interesting from LST file

Figure 2: No difference density around the same region after SHELXL refinement with corrected hydrogens

SHELXL has set the hydrogens correctly and we see no difference density around `C2BA`. Well done.

- Can you fix any other hydrogens? (Sure you can. C9AA has similar issues, for instance.) This will be your homework.

3.3 Symmetry

It might be quite interesting to look at crystal packing and intermolecular interactions by displaying symmetry equivalents. This is how it is done in Coot:

- First, for clarity, we will switch off maps: in the Display Control window click on the All toggle button above the maps.
- Select Draw→Cell and Symmetry... and click Yes in the master symmetry switch. Finally, click Apply.
- You can switch on the unit cell too, if you like.

We are done with 03srv, you can close Coot now.

4 Viscotoxin-B2

Let's take Viscotoxin-B2 (an atomic resolution structure of a small peptide toxin, PDB-ID: 2V9B) as an example to show how Coot can be used for validation and model building of a protein refined with SHELXL.

4.1 Analysing `lst` files

- Start Coot with
  `$ coot`
- Extensions → Modules → SHELX

  [*An additional menu item, SHELX appears in the menu bar.*]
- Open the project viscotoxin-b2 with SHELX→Read SHELX project...
  and select any of the `viscotoxin-b2` files, i.e. `res`, `lst` or `hkl`.

  [*Coot has read the `res` and `fcf` files. It has also found quite a few interesting features in the `lst` file and displays them in the Interesting things from SHELXL window.*]

  As we browse the Interesting things from SHELXL list, we see many disagreeable bond distance, angle distance and even chiral volume restrains (DFIX, DANG and CHIV) around residue 1024. Navigate there by clicking on, say, the first disagreeable DANG button.

  What a mess! — the structure around 1024 is strongly distorted and out of density.
- Other validation tools within Coot, such as the Ramachandran plot, B-factor Variance Graph and Density Fit Graph show 1024 A as an outlier (Figure 3). Click Validate→Temp. fact. variance analysis and select viscotoxin-b2.res. Examine the plot by clicking on outliers. Similarly, you can take a look at the Ramachandran plot and Density fit analysis as well (you will find a similar degree of badness there).

  Note that Coot has split the molecule into four chains: A and B are protein chains, whereas C is for sulfate ions and D for waters.
- Let's fix 1024 A now. The easiest way to do that is real space refinement. We select a map first: select Calculate→Model/Fit/Refine...

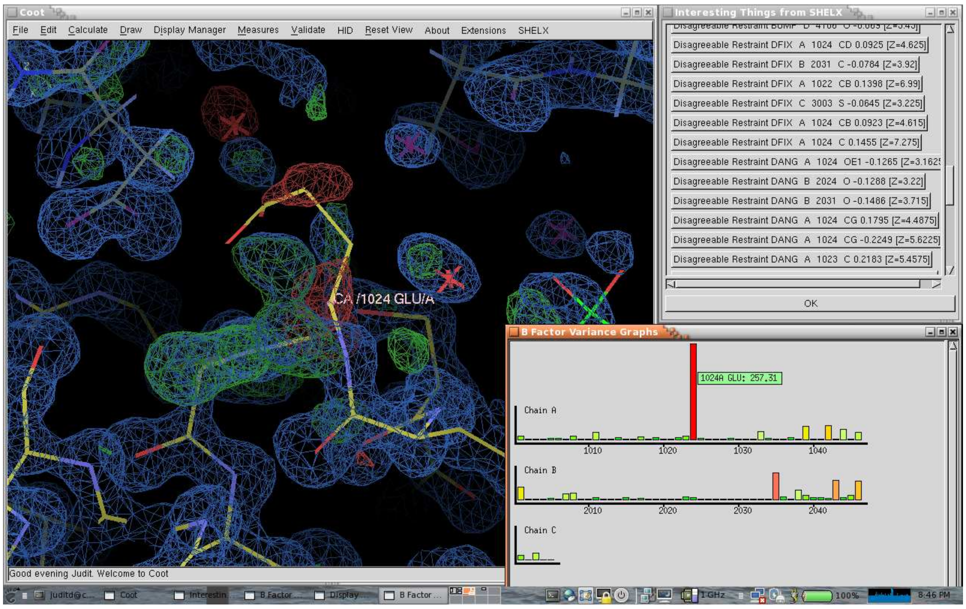
  [*The Model/Fit/Refine window pops up...*]

where we click on Select map... and select viscotoxin-b2.fcf.cif SigmaA. Now click Real Space Refine Zone in the Model/Fit/Refine window and select the region 1023-1025 by clicking on these residues in the graphics window.

[*Coot displays white intermediate atoms that migrate towards the density. The Accept Refinement window pops up.*]

Try to pull the atoms into the density. It might be a good idea to start with the carbonyl oxygen of 1024.[6] Click Accept when done.

[*The Ramachandran plot and the Density fit graph look much better.*]



Disagreeable Restraint BUMP D 4106 O -0.085 [Z=3.43]
Disagreeable Restraint DFIX A 1024 CD 0.0925 [Z=4.625]
Disagreeable Restraint DFIX B 2031 C -0.0784 [Z=3.92]
Disagreeable Restraint DFIX A 1022 CB 0.1398 [Z=6.99]
Disagreeable Restraint DFIX C 3003 S -0.0645 [Z=3.225]
Disagreeable Restraint DFIX A 1024 CB 0.0923 [Z=4.615]
Disagreeable Restraint DFIX A 1024 C 0.1455 [Z=7.275]
Disagreeable Restraint DANG A 1024 OE1 -0.1265 [Z=3.162
Disagreeable Restraint DANG B 2024 O -0.1288 [Z=3.22]
Disagreeable Restraint DANG B 2031 O -0.1486 [Z=3.715]
Disagreeable Restraint DANG A 1024 CG 0.1795 [Z=4.4875]
Disagreeable Restraint DANG A 1024 CG -0.2249 [Z=5.6225]
Disagreeable Restraint DANG A 1023 C 0.2183 [Z=5.4575]

Good evening Judit. Welcome to Coot

Figure 3: Mistraced region in `viscotoxin-b2.res`. When reading the project, Coot has extracted potentially split atoms and disagreeable restraints from the listing file and displayed them in an interesting-things-GUI (upper-right corner). You can navigate to the problematic regions by clicking on the items in the list. The B-factor Variance Graph (bottom-right corner) shows a couple of outliers, one of which is the residue shown in the graphics window.

4.2 Adding Atoms

- Let's locate missing bits of the structure now. Select Validate→Difference Map Peaks... and hit Find blobs.

  [*The new pop-up contains a clickable list of difference density blobs. We click on the first one and Coot navigates to a large green cloud.*]

  Let's examine the first blob: it is clearly tetrahedral. We know that the crystals grew from ammonium-sulfate, so we will try to fit a sulfate ion here.

- Select Place Atom At Pointer in the Model/Fit/Refine window.

[6] Eagle-eyed cooters might have noticed that some of the hydrogen atoms have flown away upon real-space-refine. This is due to some library incompatibilities. Nothing to worry about, SHELXL will duly knock them into shape.

[*We are given a new window to pick what kind of atom/molecule we wish to place here.*]

Select SO4. Select viscotoxin-b2.res to put the sulfate into and hit OK.

[*A pretty sulfate ion appears, but it is a bit out of density. Let's fix it.*]

- Simply select Real Space Refine in the Model/Fit/Refine window and click twice on the new sulfate ion. And finally hit Accept in the Accept Refinement? window. Done.

4.3 Adding Alternate Conformations

SHELXL gives us the flexibility to refine occupancies of discretely disordered atoms, where free variables are used to control occupancies of alternate atom pairs. Coot, on the other hand, can set free variables and occupancies according to the SHELX convention.

- You might remember another outlier in the B-factor Variance Graph: Ile 2035 in chain B. Let's click on the corresponding bar in the graph (it is the salmon coloured bar in the second row).

  [*Coot centers on 2035. We see a side chain with some difference density blobs around it. It is a hint for an alternate conformation.*]

- To add an alternate conformation here, we click on the "Add Alternative Conformation to a Residue" button in the Coot toolbar (down the right-hand side).

  In the pop-up you can choose to split the whole residue or only atoms beyond Cα. Click on any atom of residue 2035.

  [*Coot has now split the residue, displays a set of white atoms. We can pick a rotamer in the rotamer selection window.*]

- In this case rotamer 3 fits the density best, so let's select that. We can set the occupancy of the new conformer to, say, 0.3. (We do not need to be very precise, it will be refined by SHELXL anyway.) Click Accept.

  [*Well, Cδ is still not perfect.*]

- We pull Cδ to the density with real space refinement.

  But how about the occupancies? Select Measures/Residue Info... and click on any atom of this residue to check what Coot did about the occupancies.

  [*In the Residue Info pop-up we see that occupancies are set to 161.00 and -161.00 (and a new free variable, nr. 16, has been assigned to them).*]

We can now save coordinates or perhaps re-run SHELXL to see how well we managed to fix things (this should take a couple of minutes).