

REFINEMENT OF THE CRYSTALLOGRAPHIC STRUCTURE.

Starting from the protein model (in the example: lysozyme) introduced in the unit cell using Molecular Replacement, the refinement of the structure will be conducted through the manual inspection of the electron density in the real space and the minimization of a target function of Maximum Likelihood in the reciprocal space. At the end of the refinement the quality of the final model will be evaluated using geometric parameters as well as R-factors. For the reciprocal space refinement, the software Refmac [1] of the crystallographic suite CCP4 [2] will be used. The manual inspection of the electron density in real space, its comparison with the molecular model, and the modification of the latter to improve its congruence with the electron density will be performed using the software Coot [3].

Analysis of the electron density and its congruence with the initial model.

The software Coot can be opened directly from the CCP4 suite (1). In the following window, the graphical interface allows to select the file containing the molecular model (*pdb* file, the initial structure of the protein to be refined) (2) and the file containing structure factors (*mtz* files), from which the software will calculate the electron density map (3) and the difference electron density map (4). When the Coot software is started from the previous “job”, the one in which the phasing was carried on using Molecular Replacement, values in selection menus 2, 3 e 4 are those relative to model from MR and corresponding electron density maps. Once files have been selected, Coot can be opened using the “Run” button (5).

Coot can be run also outside the CCP4 interface, by launching the Coot software. In this case, the coordinates of the molecular model can be uploaded through *File > Open Coordinates* (6). Electron density maps are calculated by Coot once the user provides the diffraction intensities and the calculated phases, information contained in the *mtz* file. This file can be uploaded through *File > Auto Open MTZ* (7). The automatic opening option calculates a *2Fo-Fc* electron density map and a difference electron density map.

Coot usually visualizes the *2Fo-Fc* map as blue contour level map, while the difference map is visualized in green for the positive values (i.e., when electron density of the model, obtained from F_c , is lower than the experimental electron density, obtained from the experimental intensities and the calculated phases) and in red for the negative values (i.e., when the model would predict higher values of electron density compared to those measured through diffraction).

The screenshot displays the CCP4-7.1.000 Project Viewer interface for the Lysozyme project. The main window shows the results of a refinement job titled "Job 9: Molecular Replacement and refinement - MOLREP". The job is marked as "The job is Finished".

Job List (Left Panel):

Job/File	Evaluation
10 REFMACS	R=0.24 RFree=0.26
9 MOLREP	R=0.35 RFree=0.36
8 MOLREP	R=0.49 RFree=0.50
7 MOLREP	R=0.47 RFree=0.48
6 Data reduction	Spp=43 21 2 res=1.00 Rmeas=...
4 MOLREP	R=0.55 RFree=0.57
3 Data reduction	Spp=41 21 2 res=1.00 Rmeas=...
2 Integrate images with Mosflm	
1 Integrate images with Mosflm	

Refinement Results (Right Panel):

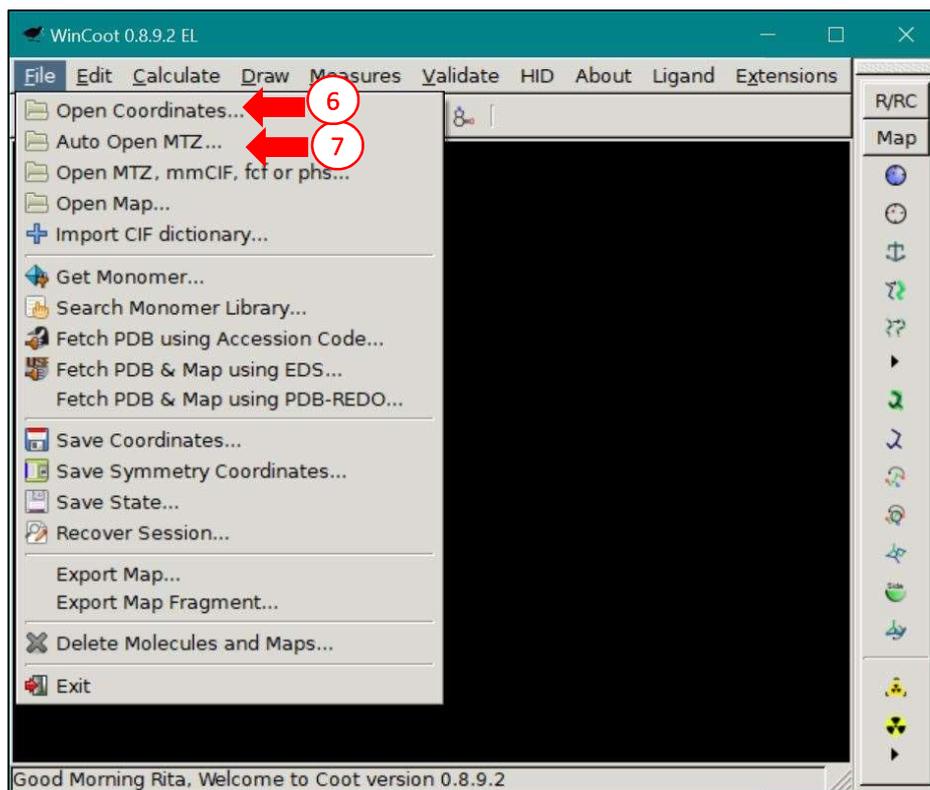
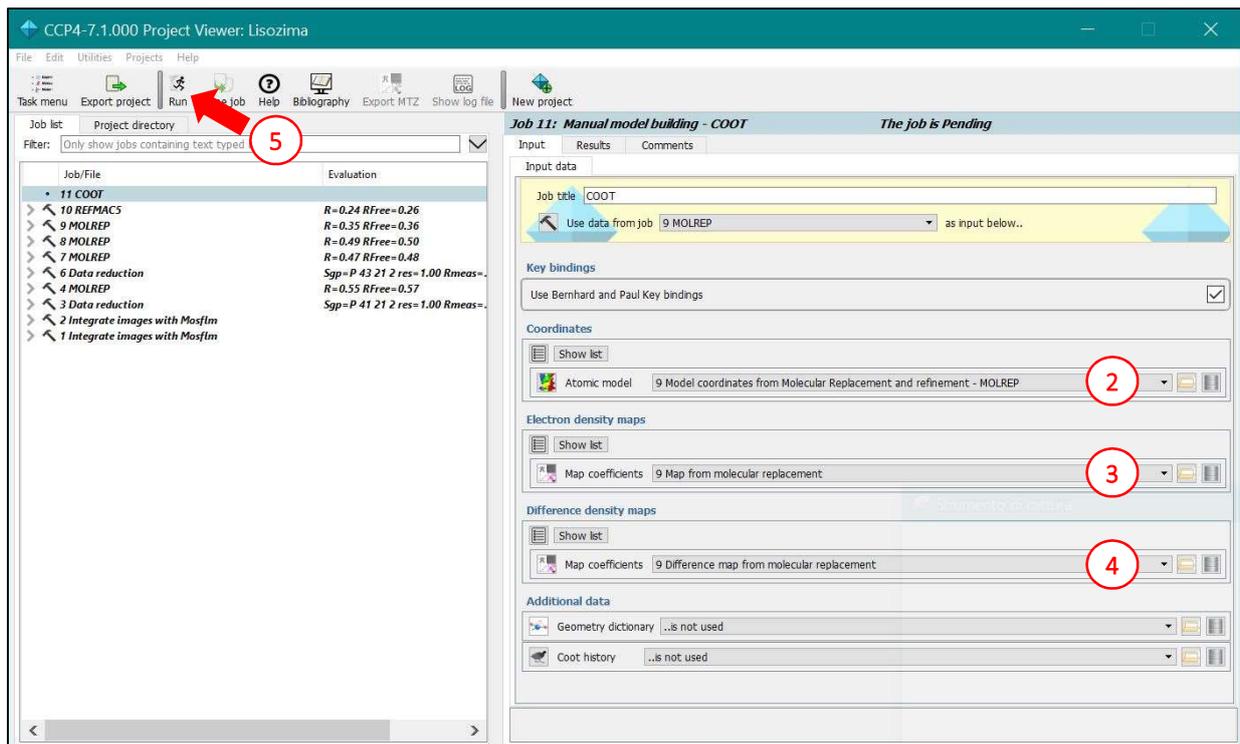
Statistic	Value
Resolution	22.25 - 1.00
No. reflections all/free	62753 / 3066
R-factor/R-free	0.352 / 0.365
RMS Deviations	
Bonds	0.0106
Angles	2.094
Chain mean B (No. atoms)	
AAA	11.4(1688)

Refmac R-factors Graph:

The graph shows the R-factor and R-free values over 5 cycles. The R-factor (blue line) starts at approximately 0.45 and decreases to about 0.35. The R-free (yellow line) starts at approximately 0.50 and decreases to about 0.36. The rmsBONDx100 (red line) starts at approximately 0.01 and decreases to about 0.005. The rmsANGLE (green line) starts at approximately 2.1 and decreases to about 2.0.

Bottom Toolbar:

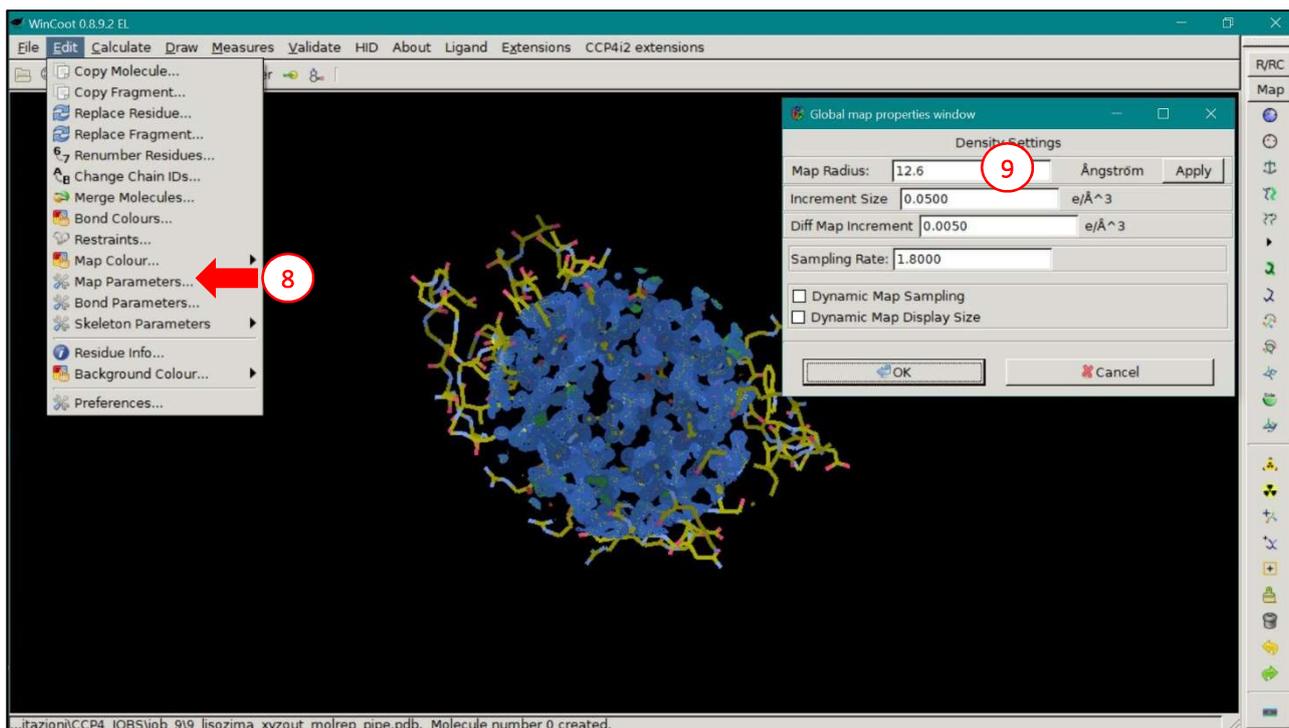
The bottom toolbar contains buttons for REFMACS and COOT. The COOT button is highlighted with a red circle and the number 1.



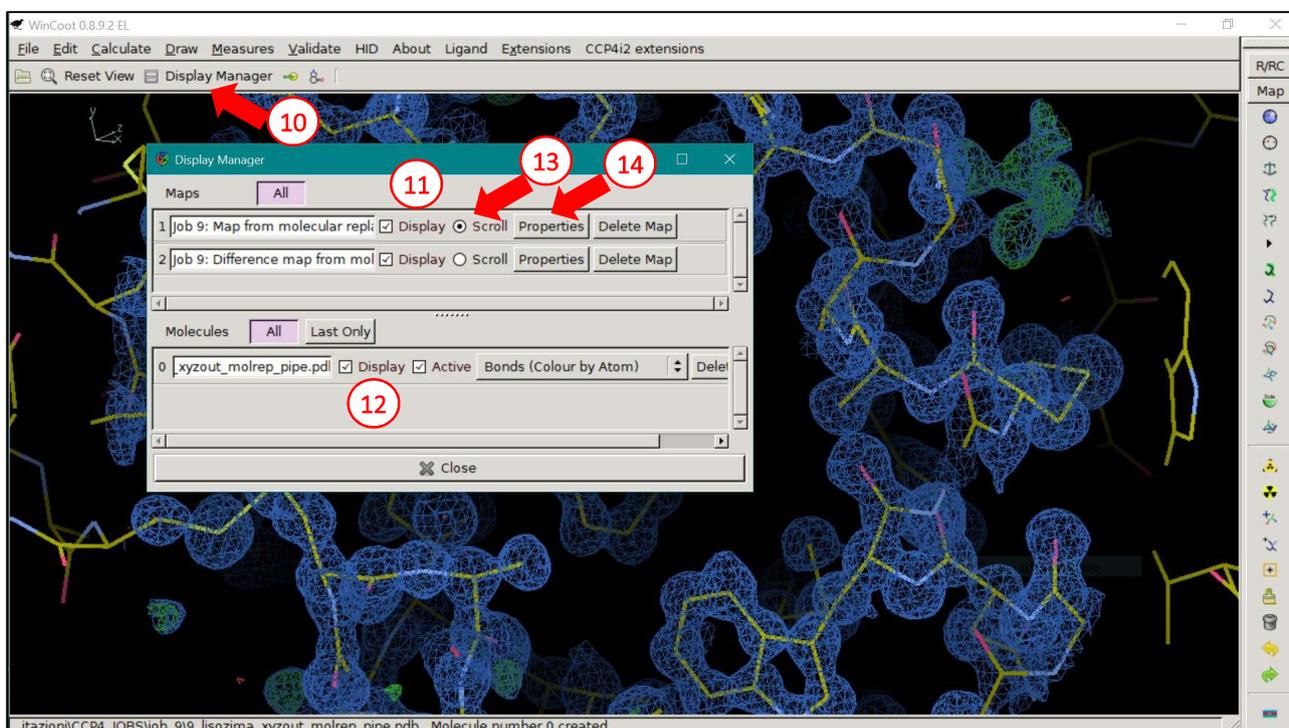
The electron density map is visualized up to a defined radius surrounding the center of the image. The visualization radius can be modified through *Edit > Map Parameters* (8), by selecting a suitable value (9).

In addition, mouse functions allow to:

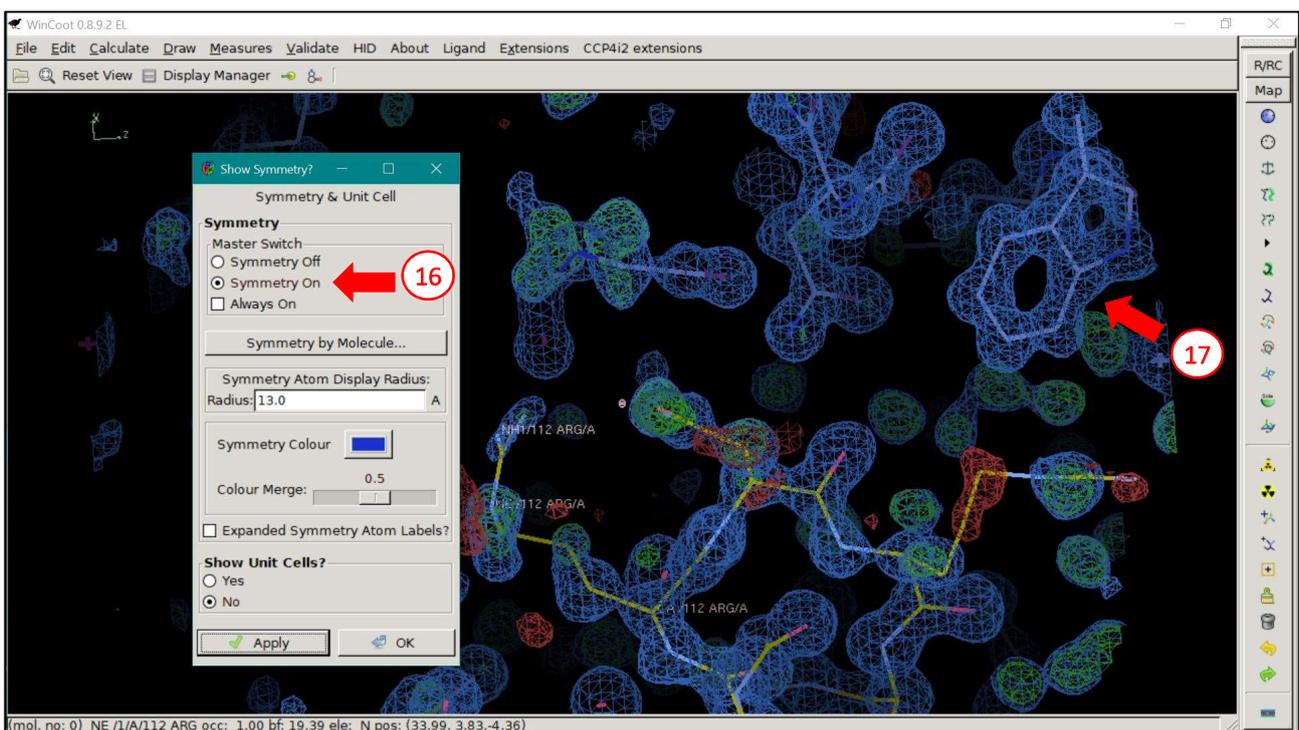
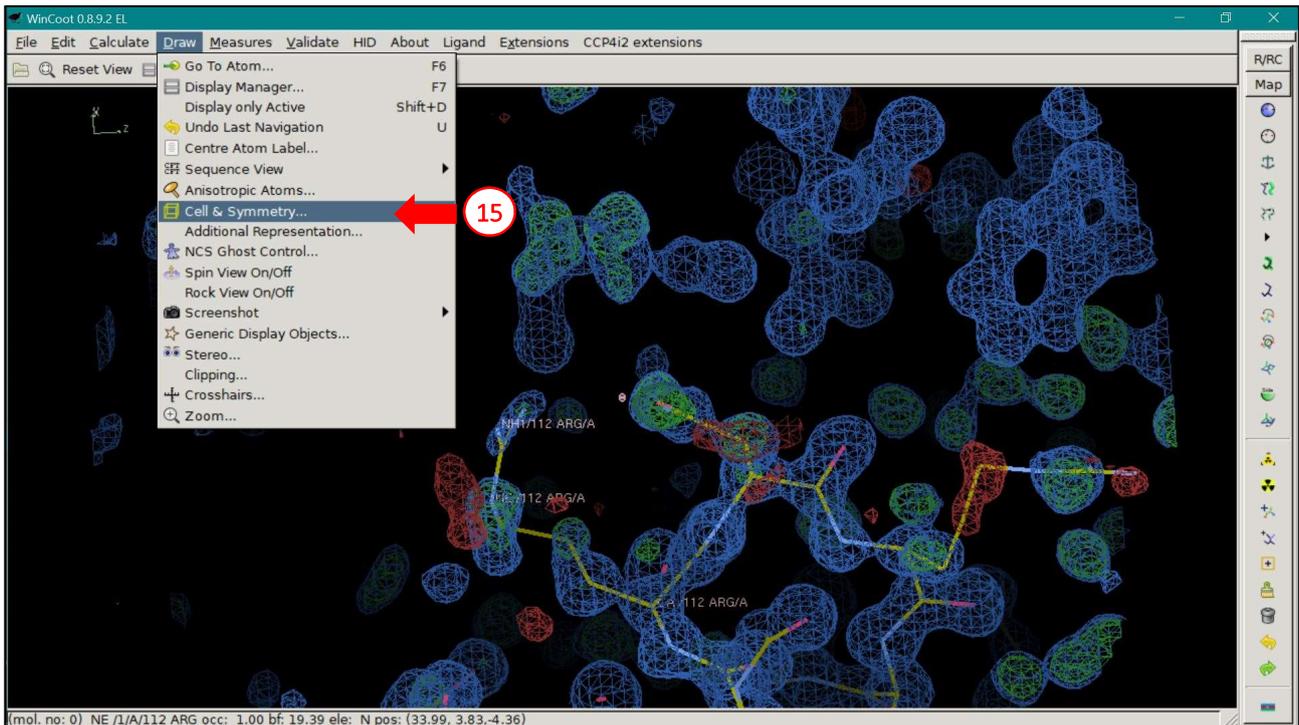
- (a) rotate the map with respect to the center of the image using the mouse left button,
- (b) zoom in or out the map using the right button,
- (c) increase or reduce the contour level of a map by scrolling the central mouse wheel,
- (d) move the center of the image onto a specific atom by clicking on the atom with the central mouse button.



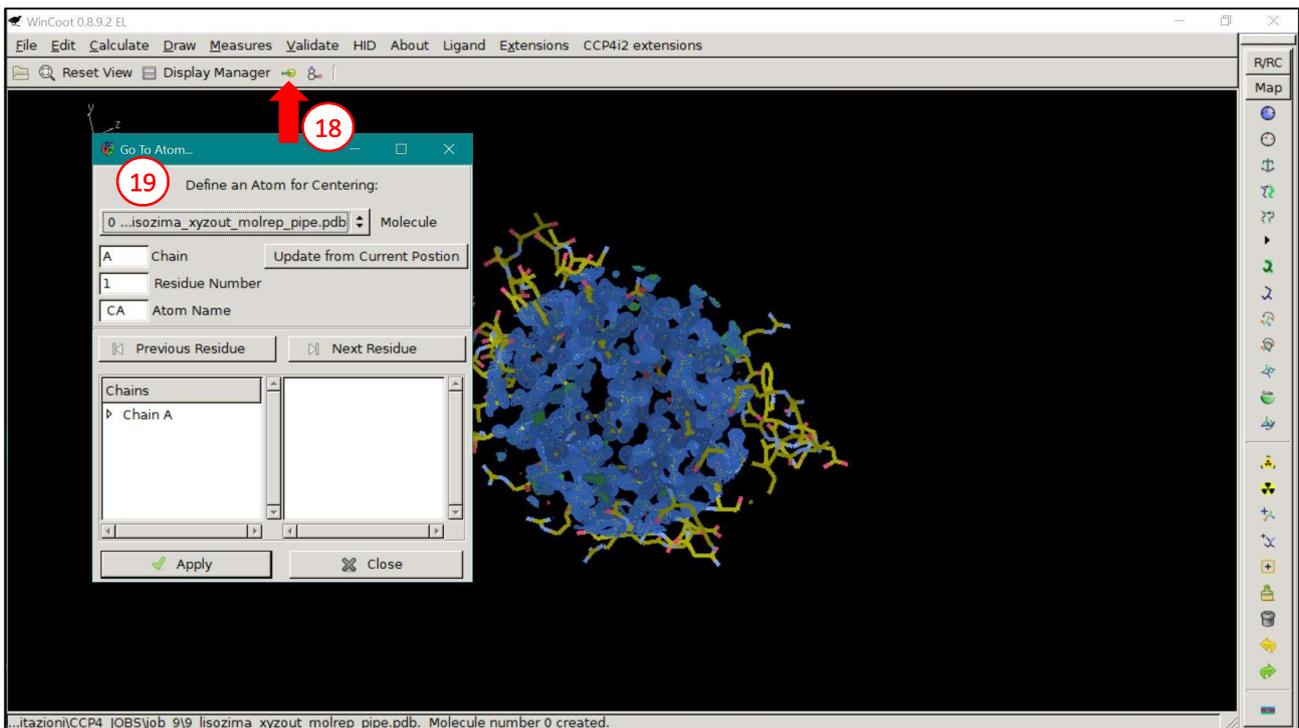
On the upper bar of the window, the option “Display manager” (10) opens another window that allows to select maps (11) and models (12) to be visualized. In the same window, the user can specify the map for which the scroll mouse function changes the contour level (“Scroll”, 13). Using the “Properties” button (14), the user can manually select map colors and levels.



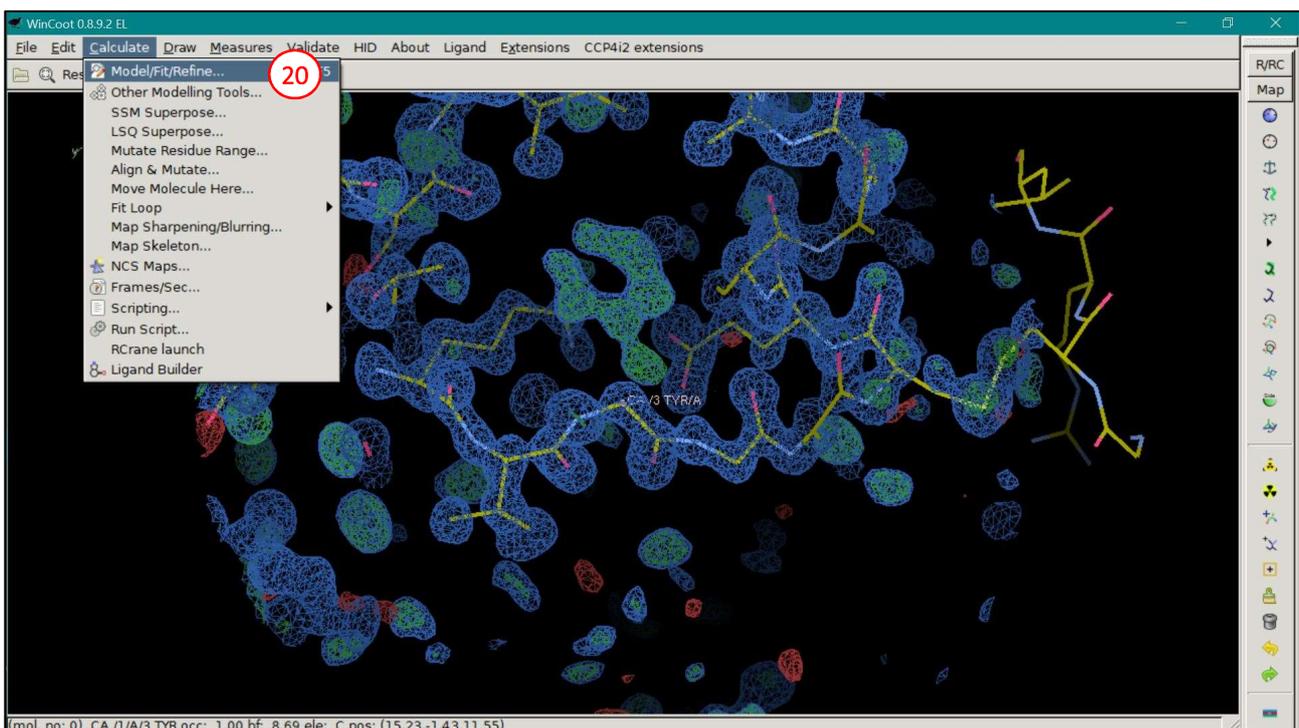
In many cases, it is useful to visualize symmetric molecules in order to understand which residual electron density can be attributed to molecules outside the asymmetric unit. To visualize symmetric molecules, the user can select *Draw > Cell & Symmetry* (15). In the specific window that opens (16), the option “Symmetry on” depicts symmetric molecules with a paler shade (17) compared to the independent model.

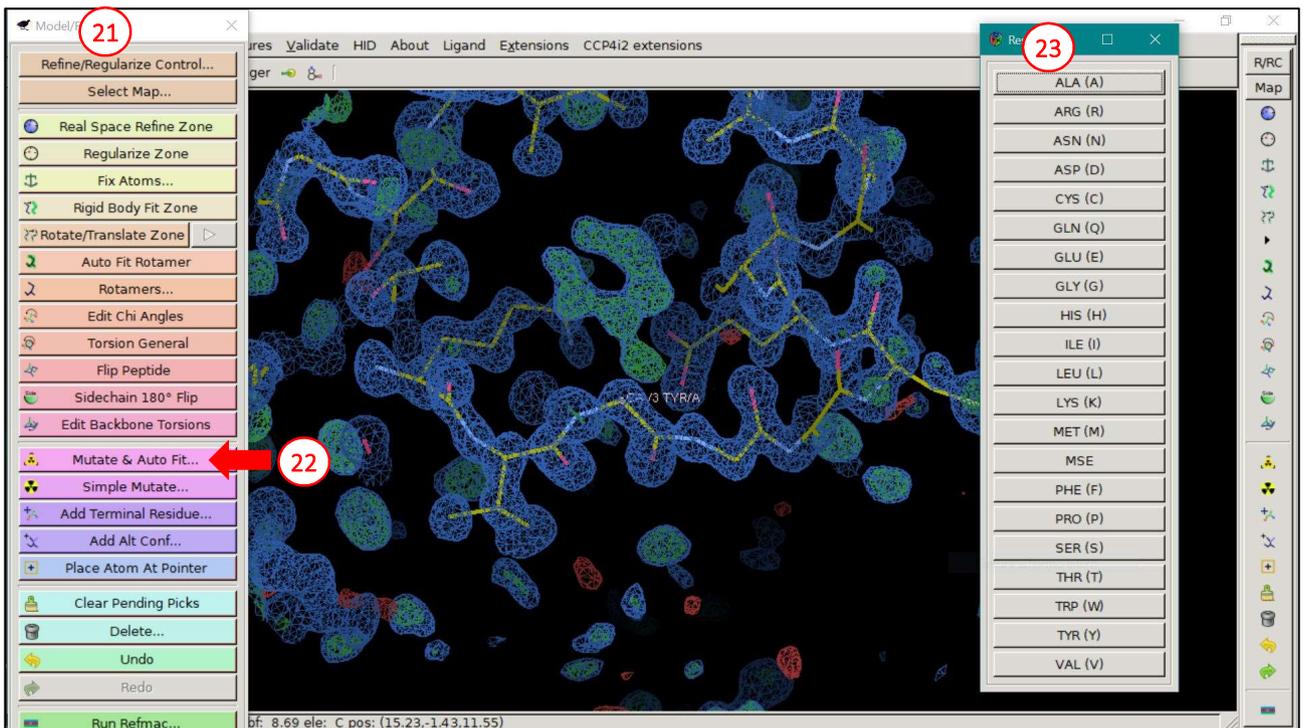


The software Coot allows for a direct modification of the model structure by the user, aiming at increasing congruence between model and observed electron density (blue map). When refining a new model, it is advisable to follow the main chain of the protein and check each residue starting from the first. The “Go to Atom” button (yellow circle with a green arrow, **18**) brings the center of the image to the location of a specific atom/residue (**19**), by selecting chain number, residue number and atom name (note that the atom name should be preceded by a space, e.g. “_CA”, with _ symbolizing the space).

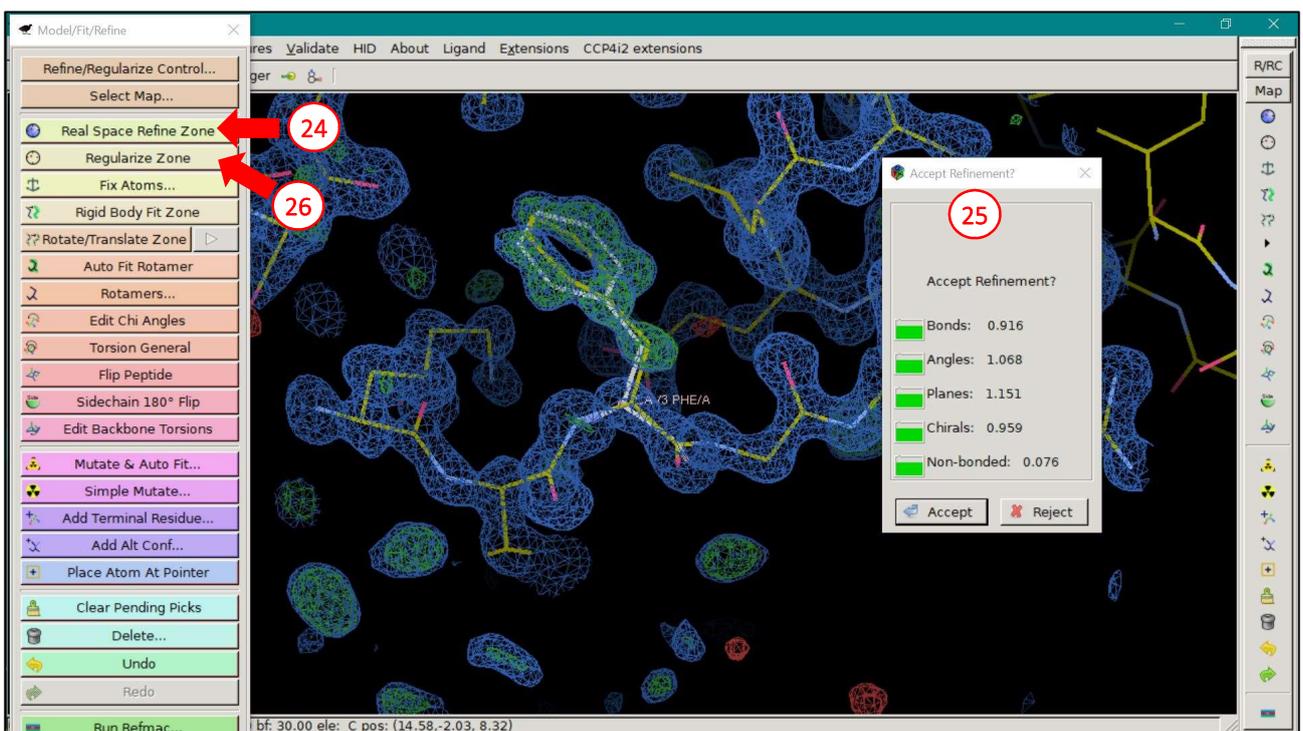


The Coot software provides many different tools that can be used to modify the model, all selectable from a window menu that can be open through *Calculate > Model/Fit/Refine* (20 and 21). The same options are also present as icons on the right side of the main window. One of the modifications usually required at the early stages of model refinement from MR solution is the mutation of residues that are different from the probe used in MR. Side chains of these residues may have been removed from the molecular probe, but the name of the residue is left in the *pdb* file and visualized by Coot. In order to mutate single residues, the user can select the “Mutate & Autofit” button (22) and click on the residue to be mutated. The software automatically selects the rotamer of the residue that better fits the electron density in real space.

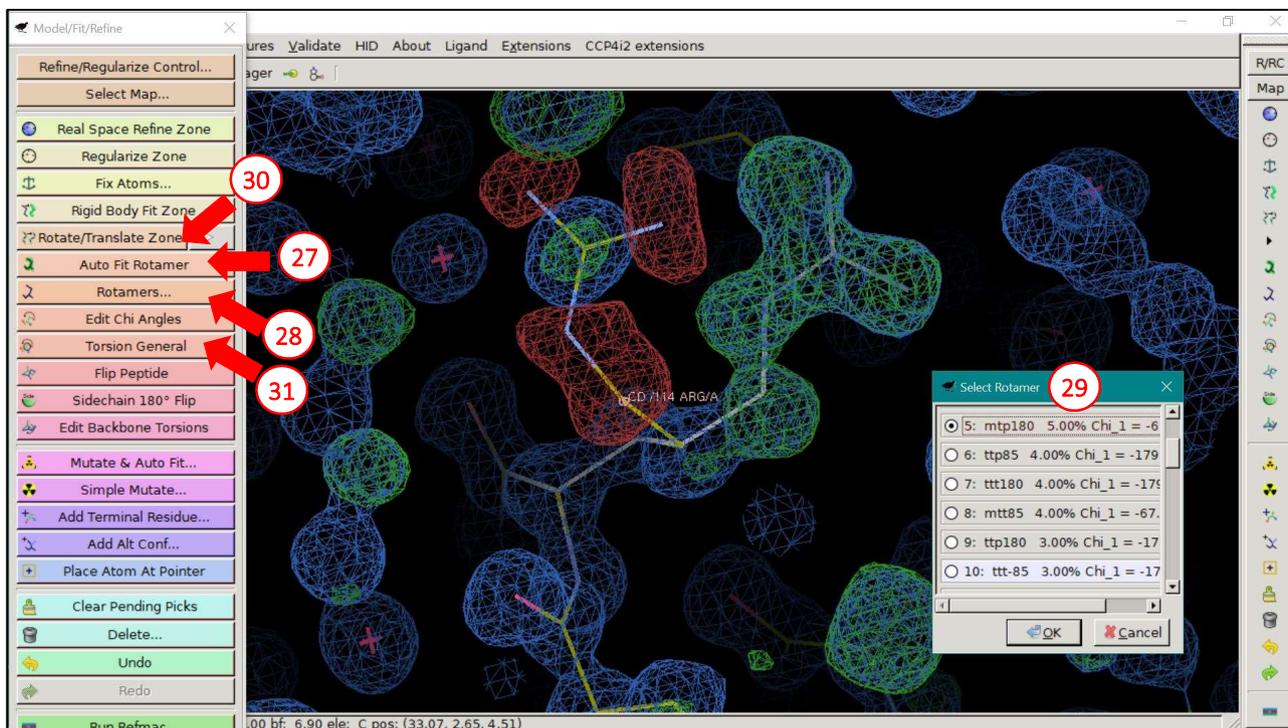




When the residue to be mutated has been selected, a new window (23) appears with a list of residues from which the user can choose the correct mutation. After the mutation, it is advisable to perform a further real space refinement to correctly locate the residue in the electron density, using the “Real Space Refine Zone” function of Coot (24). This function requires the selection of a segment of the protein chain, by clicking on two residues/atoms. After the optimization, the software allows the user to retain or reject the modifications, considering also the geometric parameters of the real space-refined model (25). If geometrical values are not satisfactory, the user can optimize the geometry with the “Regularize zone” function (26), that changes angles and distances in a defined protein segment.



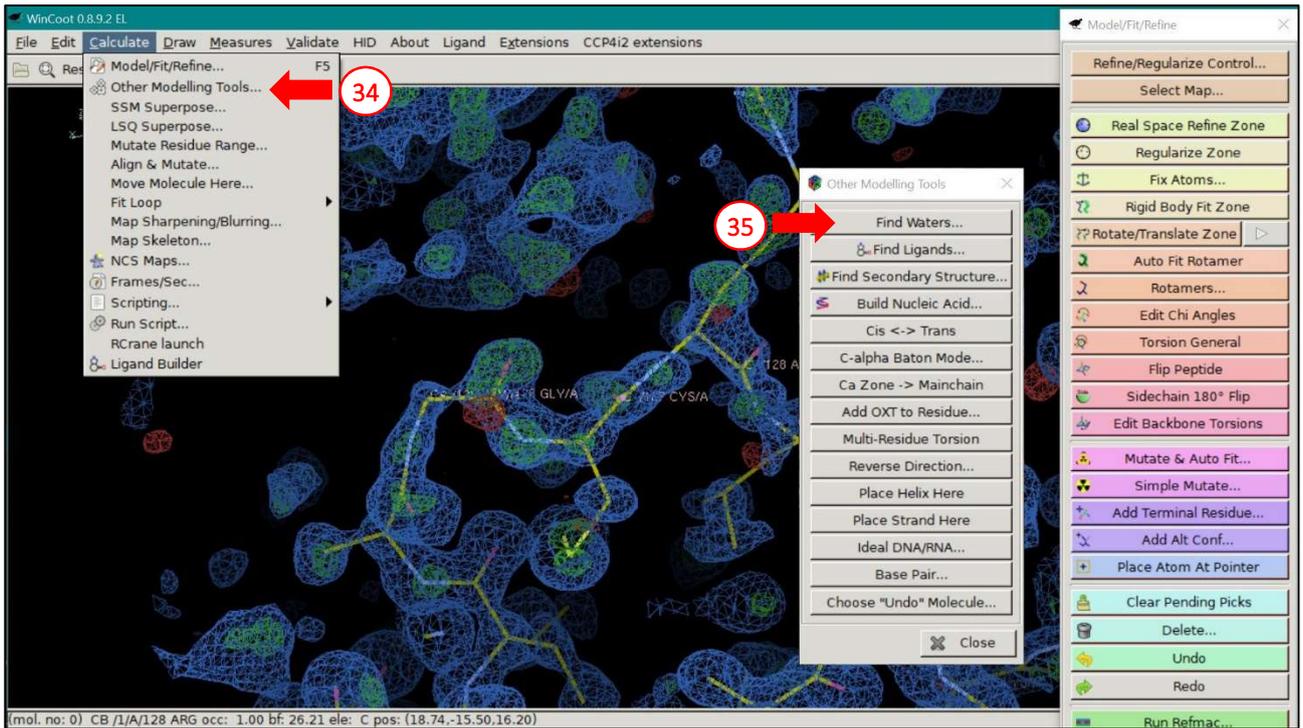
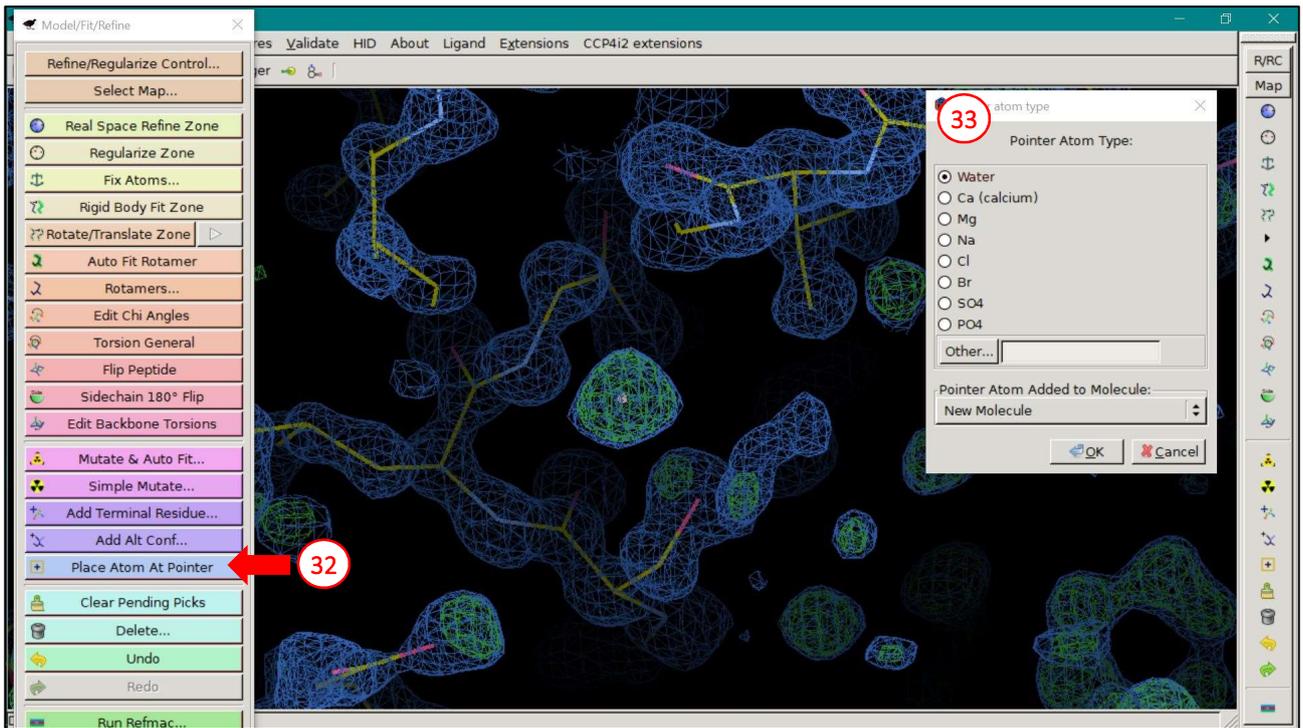
Another common occurrence that the user can encounter while refining is the presence of residues whose side chains are in positions that do not match the observed electron density. In this case, the user can select the “Auto Fit Rotamer” function (27), that allows to automatically refine the rotamer conformation considering the real space electron density, or the “Rotamers” function (28), that allows the user to manually choose the rotamer conformation that better fits the observed electron density (29).

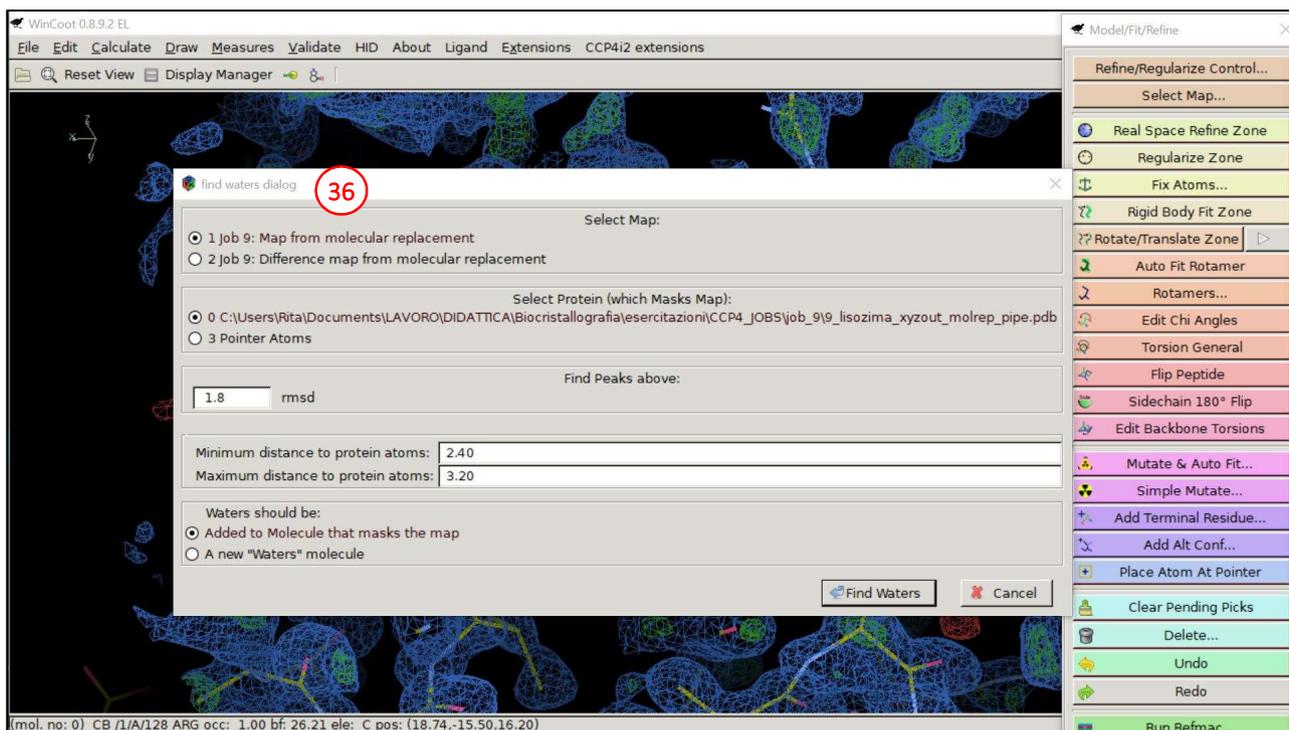


The “Rotate/Translate Zone” (30) and the “Torsion General” (31) functions are more rough instruments to modify the model to better fit the electron density. The former allows to select a protein segment, between two defined initial and final residues, and to move the entire region. The latter expects the user to input 4 atoms (by clicking on them) that define the torsion angle and then allows its manual change. These tools require more user intervention on the model and should, therefore, be used with some caution and considering the chemical analysis of the model.

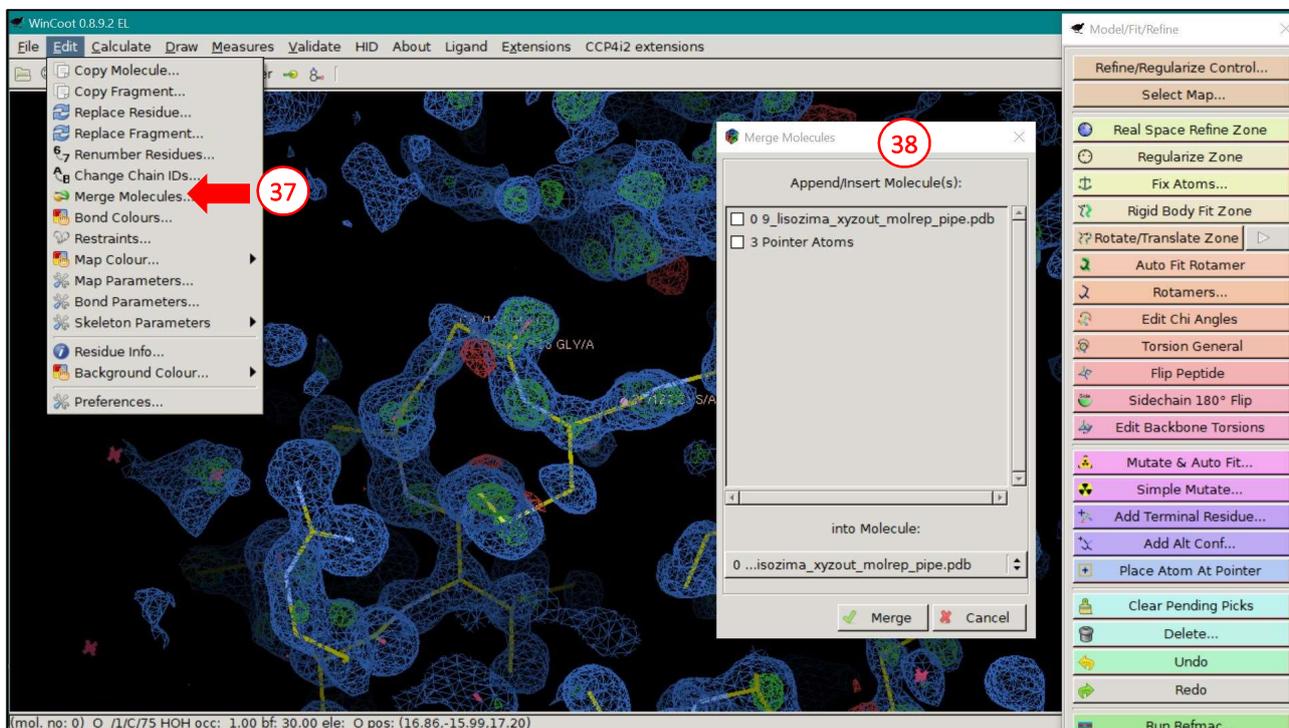
Once the user is satisfied with the modifications on the protein chain, the next step is the addition of water molecules in residual electron densities. Again, Coot offers to the user two different approaches: the manual identification of residual densities and introduction of the water molecules, or an automatic approach. In the first case, the user should move the center of the map (identified by the pink cube) onto the unexplained electron density and add a water molecule in this position using the “Place Atom at Pointer” function (32). This function opens an additional window (33) in which the user can choose the type of atom/molecule/ion to be added.

If the automatic approach is preferred, the procedure can be carried on by choosing the *Calculate > Other Modelling Tools* menu (34). In this case, the user runs a program to identify regions of the difference electron density that can be attributed to water molecules by selecting “Find waters” (35). In the further window, the threshold difference electron density for the automatic search can be set (36).



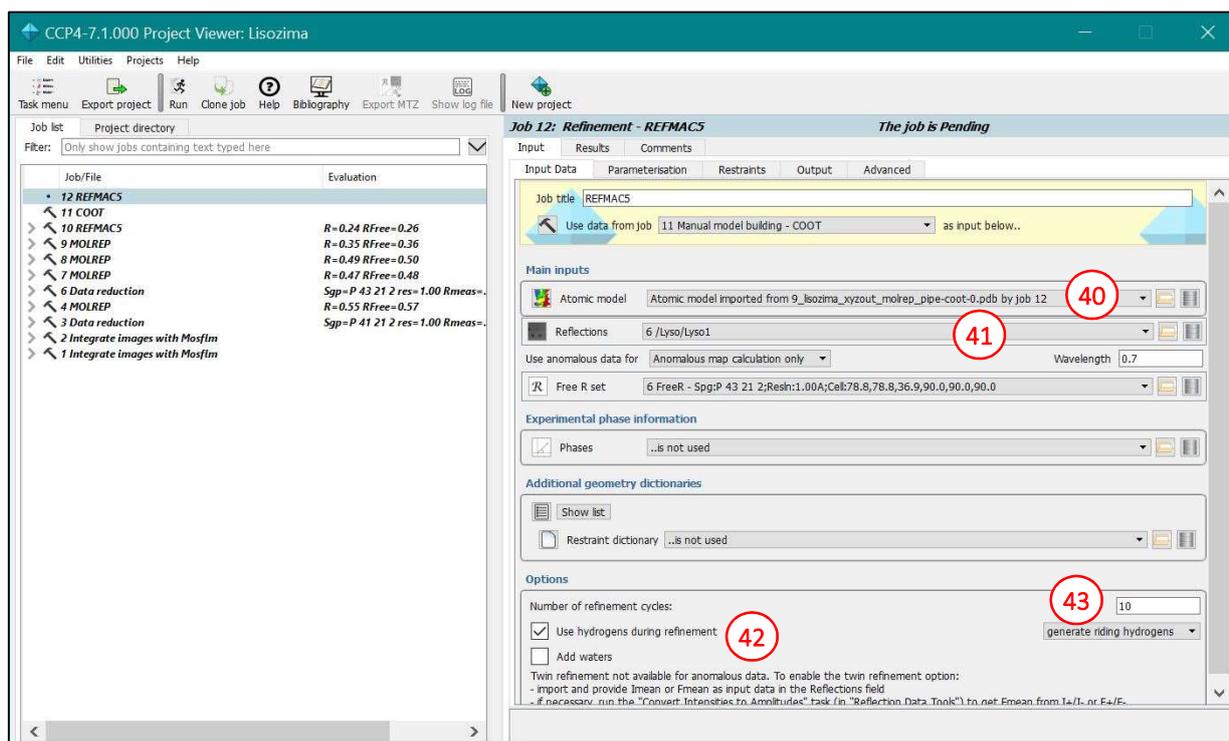
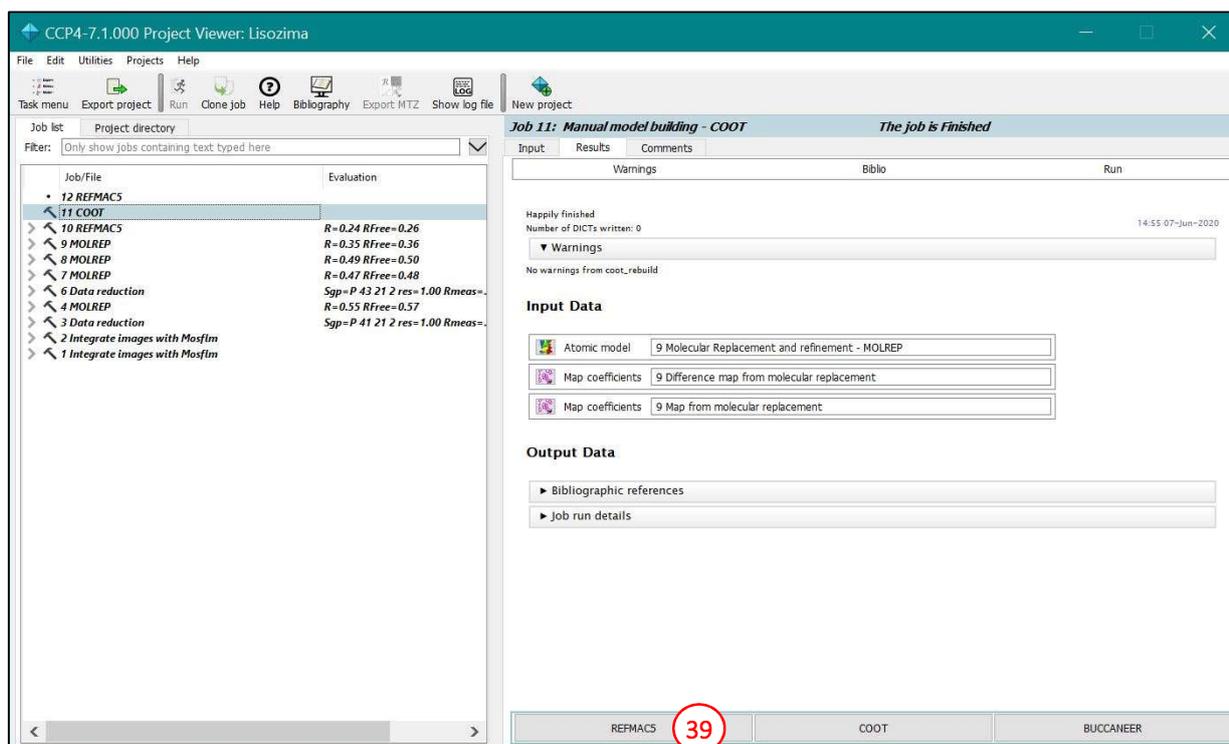


At the end of the model analysis, the user should save the modified model and the water molecules (or ions) in a single *pdb* file. Coot proceeds in two steps: first, the coordinates of the water molecules are added to the structural model of the protein through *Edit > Merge Molecules* (37) and by selecting both the initial *pdb* file and the added molecules "Pointer Atoms" (38); second, the modified model including the added molecules should be saved in a *pdb* file by *File > Save Coordinates*.

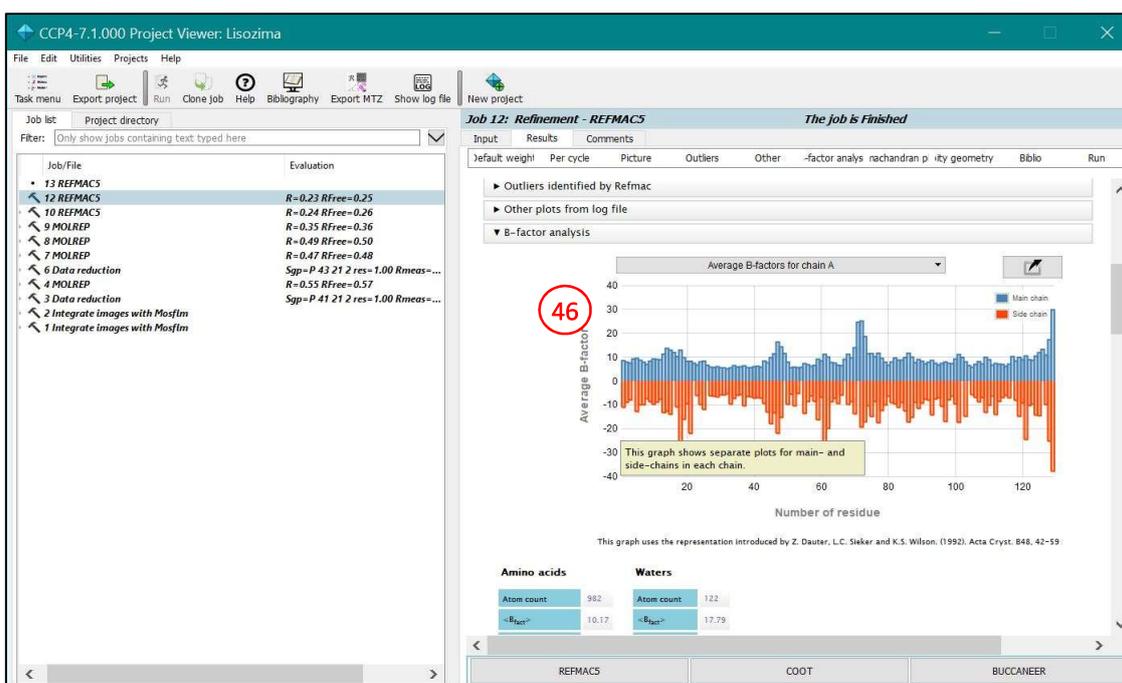
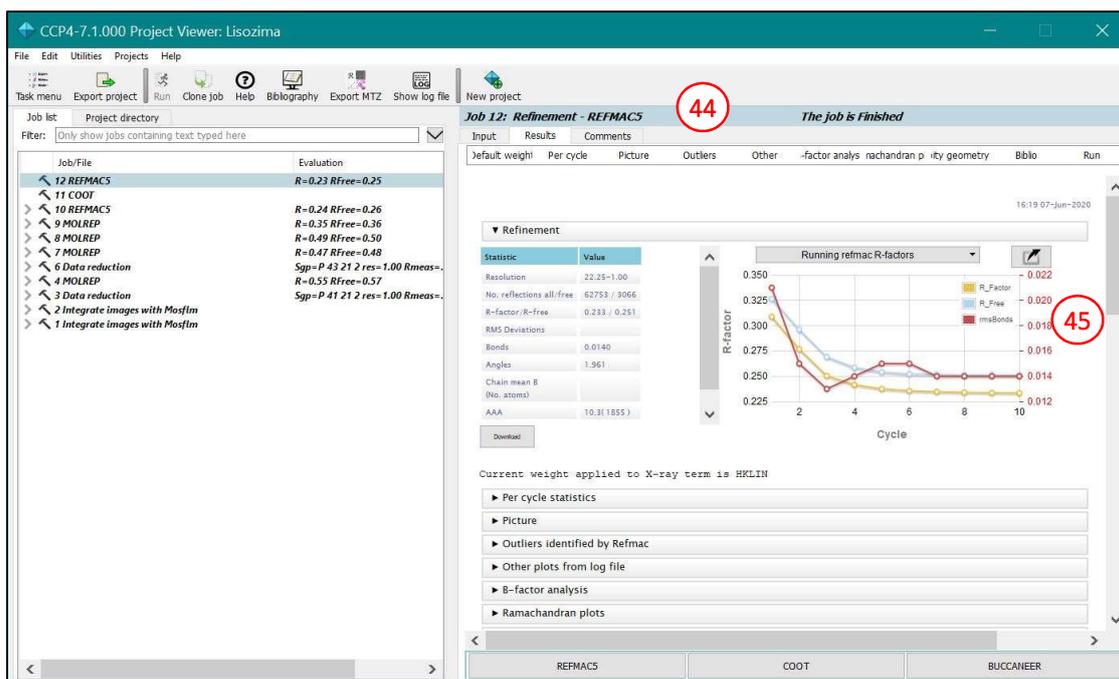


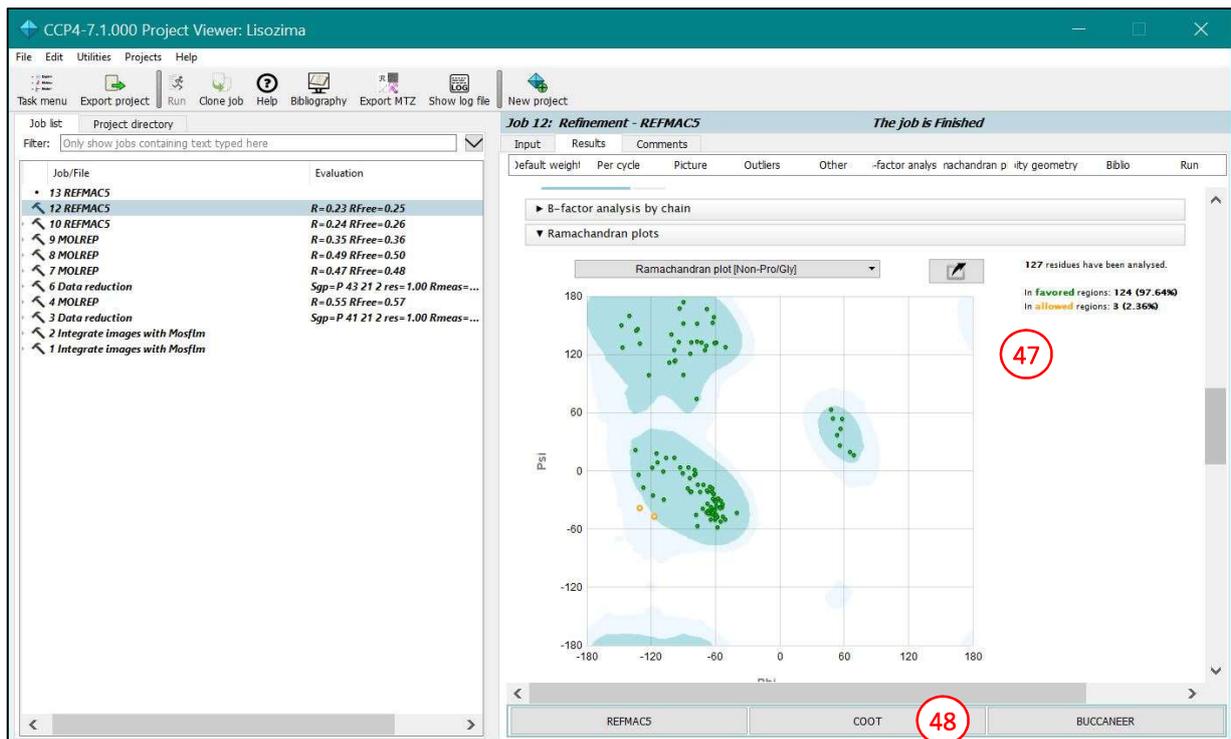
Fourier space refinement.

The software Refmac allows to carry on the Fourier space refinement of the model through the minimization of a Maximum Likelihood target function. Refmac can be run from the CCP4 graphical interface, after the program Coot has been closed (and after the coordinates of the new model have been saved in a *pdb* file!). A button in the bottom of the same window from which the Coot program has been run can start the Refmac software (39). In the following window, the user is asked to specify the name of the starting *pdb* file (40), e.g. the one saved from Coot and containing modified coordinates and solvent molecules, and the *mtz* file containing the experimental data (41).



Refmac runs a restrained refinement of the atomic coordinates and it is advisable to add the hydrogen atoms during this refinement (42). Positions of the hydrogen atoms will not be refined, but their electron density will be taken into account during the calculation. Usually, 10 refinement cycles (43) are sufficient for the optimization to converge. Once options have been set, the “Run” button starts the Fourier space refinement. At the end of the optimization, results are visualized in the same window (44). Graph 45 shows the variation of R_{work} and R_{free} values during refinement cycles and allows the user to check whether the refinement converged, i.e., values do not change significantly in the last refinement cycles, and the refinement lowered both values, not increasing their difference. Further graphs in the same window facilitate the evaluation of the quality of the protein model, such as the B factor values of the residues (46) and the Ramachandran plot to monitor the main chain torsion angles (47).



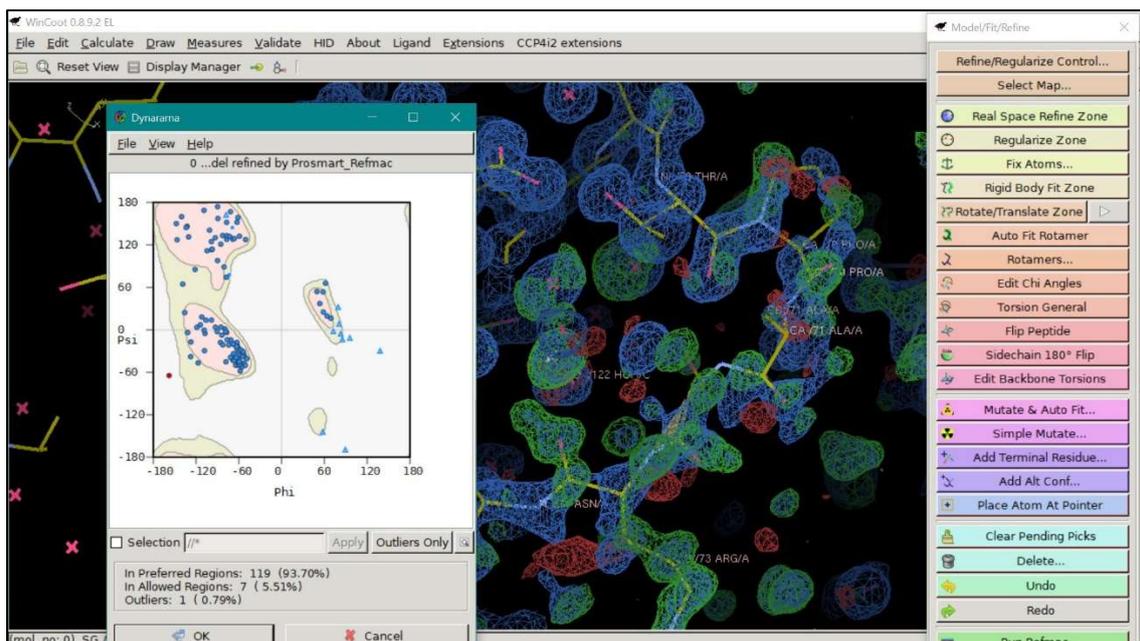


The reciprocal space refinement is usually followed by another visual inspection of the electron density, using the software Coot that can be started from the same window (48).

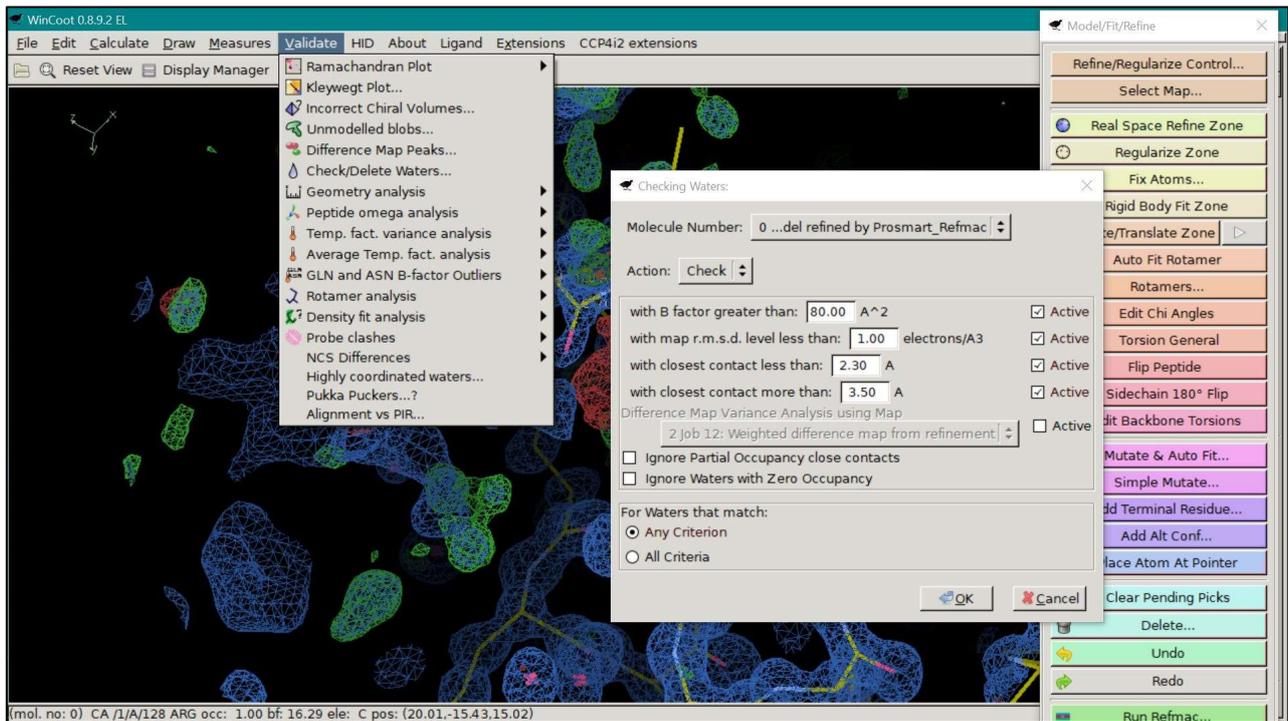
Analysis of the quality of the structural model.

Besides the analysis of R_{work} and R_{free} values (45), of the B factors along the protein chain (46), and the Ramachandran plot (47), other geometric parameters can be monitored from the Coot software, together with the analysis of the difference electron density. In particular, the *Validate* menu offers multiple tools:

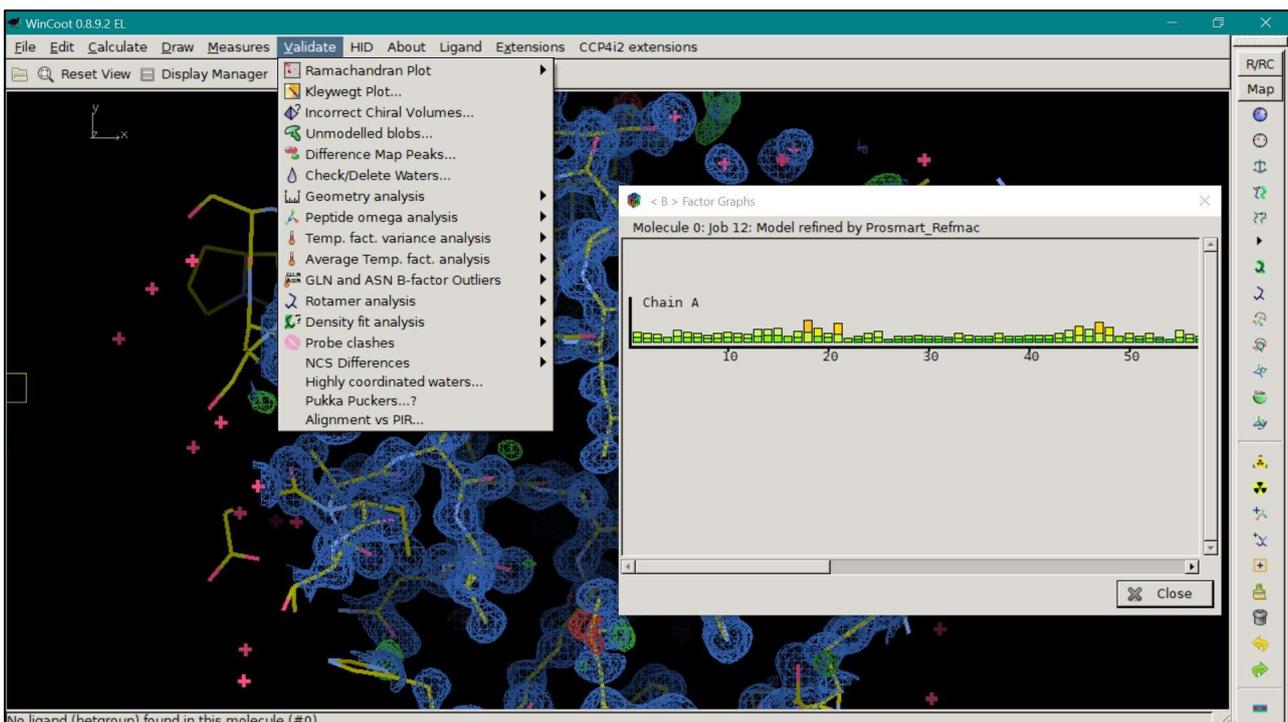
- Analysis of the Ramachandran plot through *Validate > Ramachandran plot* is more thorough than that offered from the main Refmac window as each residue can be inspected by clicking on the marker in the plot.



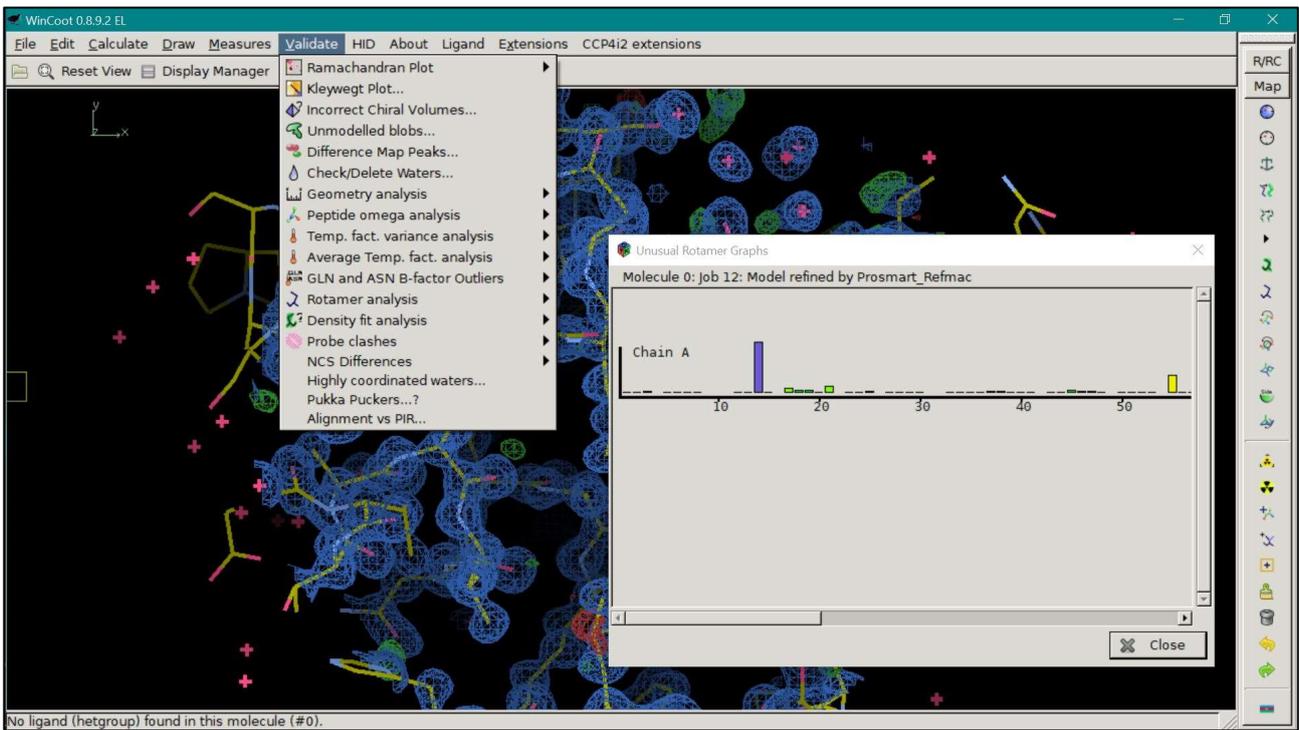
- Analysis of the solvent molecules introduced in the model (*Validate > Check/Delete Waters*): the positioning of some water molecules may not be supported by the electron density after Fourier space refinement, particularly when water molecules have been introduced automatically by the software. This function allows to inspect each water molecule, remove them, or convert them into ions (positively or negatively charged) according to chemical considerations regarding their environment. A further indication of misplaced solvent molecules may come from the B factor values. Sometimes the electron density has a more complex shape compared to spherical water molecules or ions, suggesting the presence of different species (multiatomic ions, additives, cryoprotectants, ...).



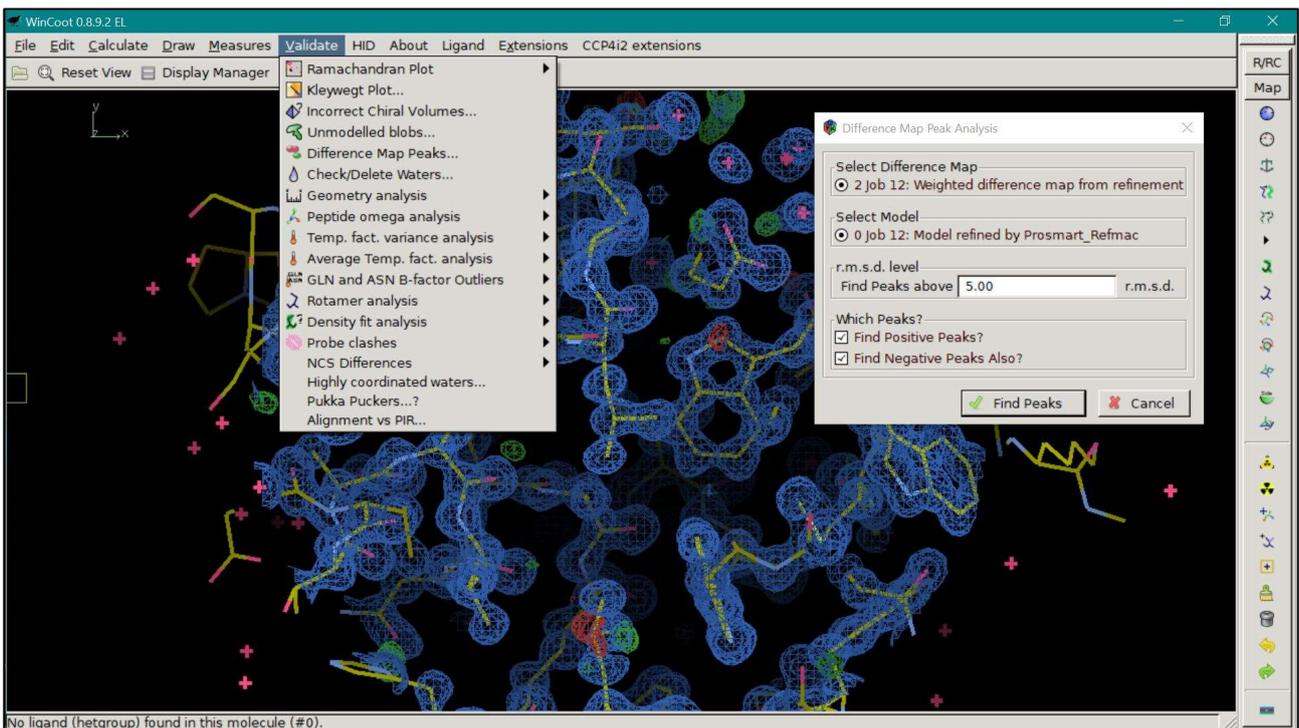
- Analysis of B factors (*Validate > Average Temp. fact. analysis*).



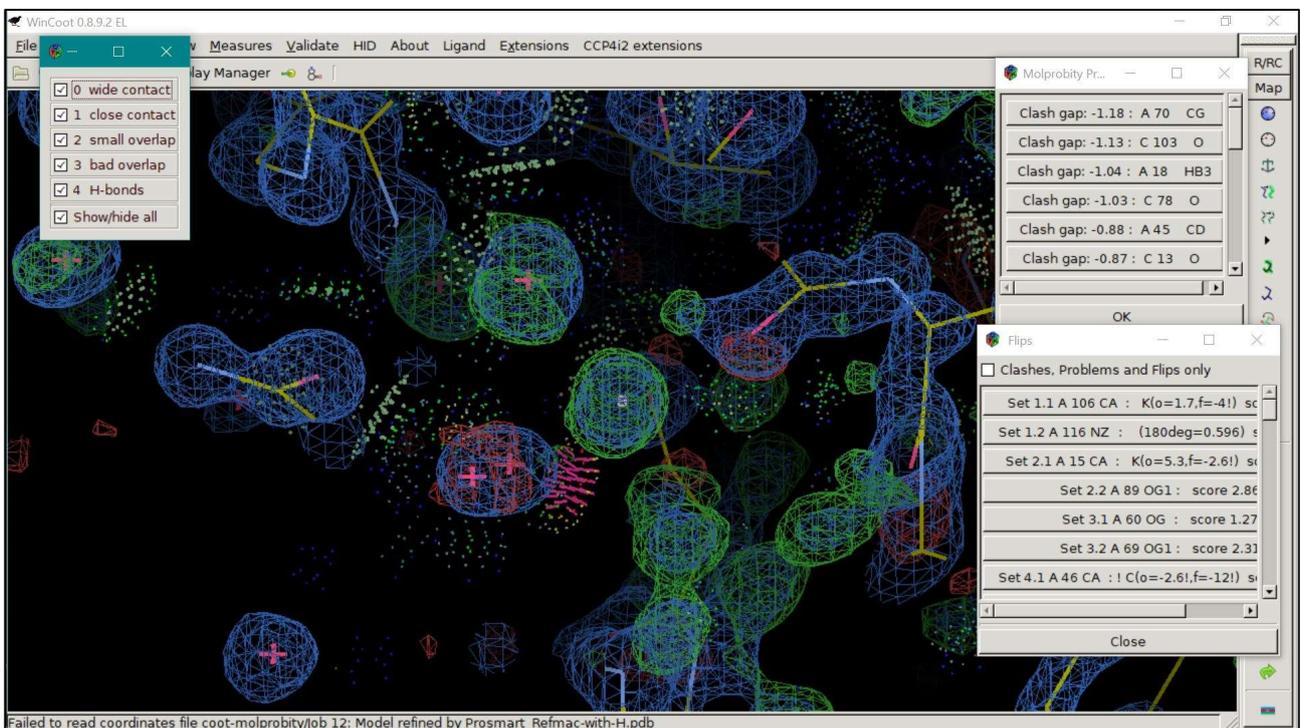
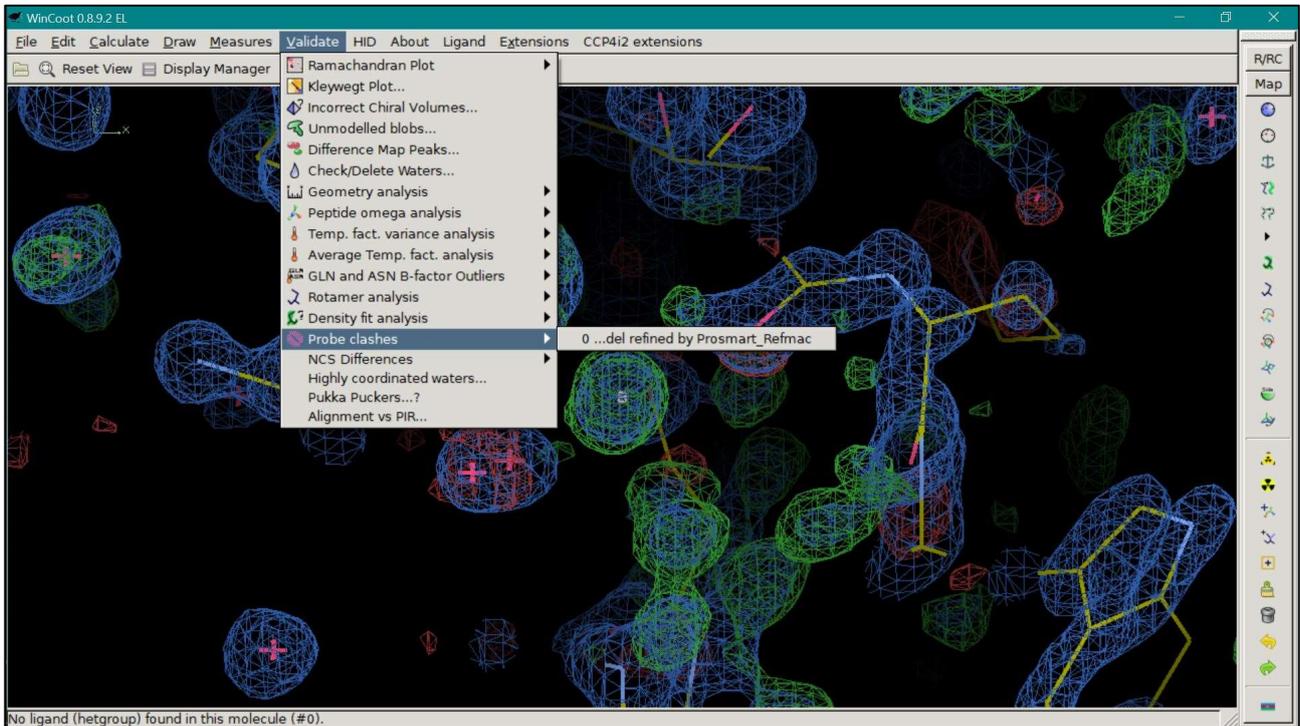
- Analysis of the side chain conformations (*Validate > Rotamer analysis*) to identify side chains in non-favored or less probable conformations. The purple bars indicate residues for which the side chain is not present in the model as it could not be located in the electron density.



- Analysis of the highest values of the difference electron density map (*Validate > Difference Map Peaks*): to visualize regions of the map with the highest discrepancies between model and experimental density, a possible indication of the presence of solvent molecules or of the wrong attribution of an electron density.



- Analysis of clashes present in the model (*Validate > Probe clashes*): this tool visualizes regions of the model in which the distance between non-bonded atoms is less than the sum of their van der Waals radii. Each clash should be manually inspected to assess whether a close distance is supported by the experimental electron density – and should, therefore, be considered for its biological role – or is the result of a misplacement of atoms or groups. Clashes are indicated by colored lines, with the pink/red lines being the most significant.



References.

- [1] G.N. Murshudov, A.A. Vagin, and E.J. Dodson, "*Refinement of Macromolecular Structures by the Maximum-Likelihood method*". **Acta Cryst.** **1997**; D53, 240-255.
- [2] M. D. Winn et al., "*Overview of the CCP4 suite and current developments*". **Acta Cryst.** **2011**; D67, 235-242.
- [3] P. Emsley, B. Lohkamp, W.G. Scott, and K. Cowtan, "*Features and Development of Coot*". **Acta Cryst.** **2010**; D66(Pt 4), 486-501.