REFINEMENT AND VALIDATION OF PROTEIN- GLYCAN COMPLEXES

In the previous practical lesson, by using **Molecular Replacement (MR)**, you have obtained the first **electron density maps** of human galectin-3 CRD, revealing the contour of the atoms in the crystal.

The electron density maps were calculated from Eq. 1:

Eq. 1

$$\rho(x,y,z) = (1/V)\sum_{hkl}|F_{hkl}| \cdot e^{2\pi i\alpha_{hkl}} \cdot e^{[-2\pi i(hx+ky+lz)]}$$

α_{hkl} is the phase angle of reflection *hkl*,
$|F_{hkl}|$ is the structure factor amplitude,
(x,y,z) are the fractional atomic coordinates in the unit cell,
V is the volume of the unit cell

In the two different maps you obtained from PhaserMR, the values of $|F_{hkl}|$ are calculated from two distinct Fourier differences:

$$\mathbf{2|F_{obs}|-|F_{calc}|}$$

and

$$\mathbf{|F_{obs}|-|F_{calc}|}$$

where the $\mathbf{|F_{obs}|}$ values correspond to the intensities ($\sim \sqrt{I_{hkl}}$) measured in the X-ray diffraction experiment;
on the other hand, the $\mathbf{|F_{calc}|}$ values were calculated after structure solution by MR from the atomic coordinates (x,y,z) of a similar model available from the Protein Data Bank (PDB) (or, very common nowadays, produced by artificial intelligence methods using AlphaFold). This model was placed in the unit cell of your crystal using the rotation and translation functions implemented in program PhaserMR.

In this lesson, you will start from this preliminary model and correct it, in a series of iterative steps, to bring it closer to the X-ray diffraction information. This requires calculation and inspection of $\mathbf{2F_{obs}-F_{calc}}$ and $\mathbf{F_{obs}-F_{calc}}$ difference electron density maps, **model building**,

refinement of the corrected model and calculation of new (hopefully, more informative) electron density maps.

Validation tools will be used to produce a final model that best explains the measured X-ray diffraction data. One of the validation parameters is known as the $\boldsymbol{R}$ **factor** and is calculated from Eq. 2. The R factor compares calculated structure factors $|F_{calc}|$, from the model, with observed structure factors $|F_{obs}|$, measured in the diffraction experiment:

Eq. 2

$$R = \frac{\sum(|F_{obs}| - |F_{calc}|)}{\sum|F_{obs}|}$$

By itself, this parameter is not sufficient to detect over-fitting, because any random set of atoms added to the model will approximate $|F_{calc}|$ to $|F_{obs}|$ and lower the R factor. This is overcome by associating a cross-validation parameter, calculated in the same way, but using as $|F_{obs}|$ only 5 to 10% of the unique reflections that are arbitrarily chosen and set apart from the normal refinement process. This is adequately known as the $\boldsymbol{R_{free}}$ **factor** and should not differ from the R factor by more than 5-6%. Agreement with the limits of stereochemical restraints is also a validation criterion that accompanies model building and refinement cycles and is given in the form of **root mean standard deviations (*rmsd*)** for bond lengths and bond angles.

Once refinement is taken to the best possible convergence, global validation takes place, evaluating several aspects, such as the distribution of amino acid residues in the energetically allowed regions of the Ramachandran plot, distribution of temperature factors, correctness of side-chain torsion angles, analysis of close contacts, water network contacts and other analysis, almost all implemented in software packages for validation. This global validation will help correcting and finalizing the best structural model.

Like in previous lessons, we remind you the **primary structure** (amino acid sequence) of human galectin-3 CRD:

```
>human_galectin-3_CRD
```

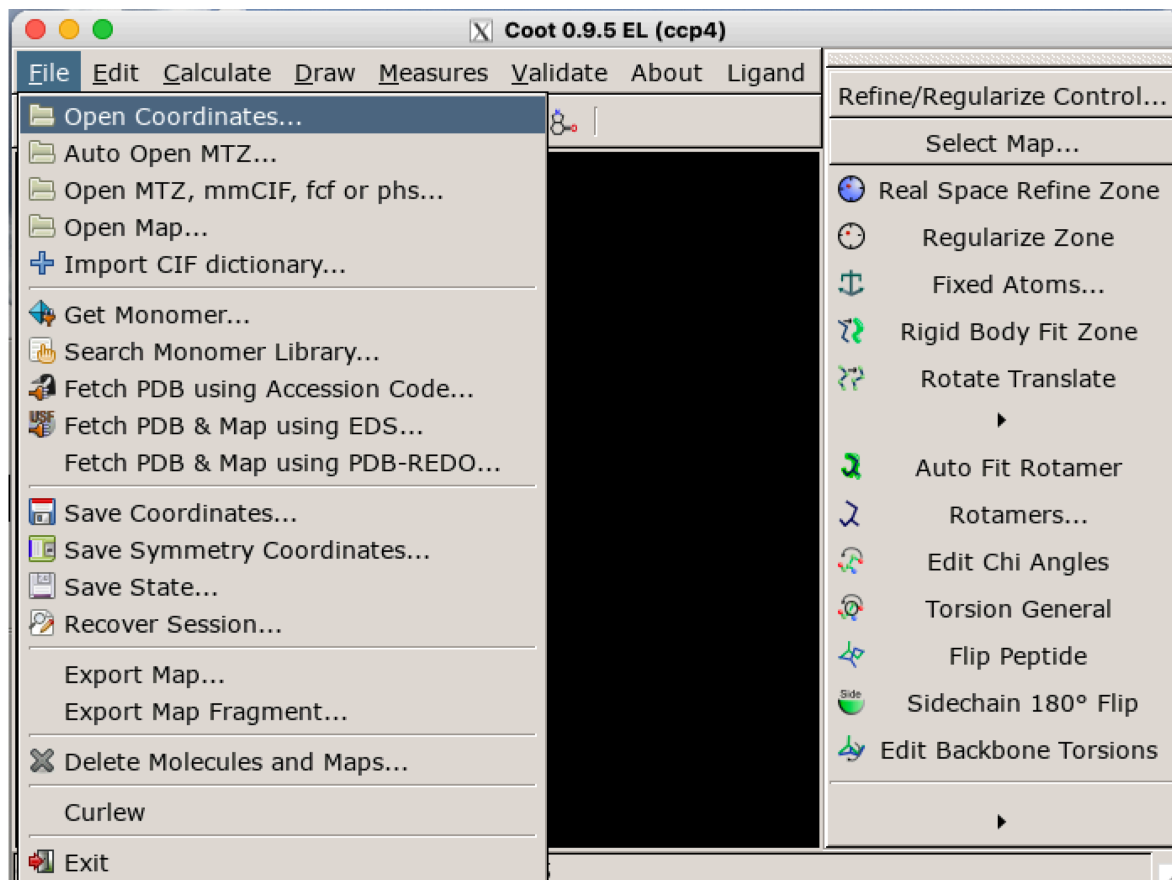
MLIVPYNLPLPGGVVPRMLITILGTVKPNANRIALDFQRGNDVAFHFNPRFNENNRRVIVCNTK
LDNNWGREERQSVFPFESGKPFKIQVLVEPDHFKVAVNDAHLLQYNHRVKKLNEISKLGISGDI
DLTSASYTMI

Dedicated software you will need: PHENIX (https://phenix-online.org/download/nightly_builds.cgi), CCP4 (https://www.ccp4.ac.uk/download/#os=windows) and Coot/WinCoot (http://bernhardcl.github.io/coot/wincoot-download.html)

***VERY IMPORTANT**: to access your results from any computer in the DQ-FCT-NOVA network, you must **use the working directory H:\\cdgeral** for creating projects and storing files.*

A. Electron density map inspection and model building

- Use the **Coot** program to open the .pdb and .mtz files obtained after solving the structure by MR. To do this, in the **File** menu, choose the **Open Coordinates** option, followed by the .pdb or **Auto Open mtz file**, followed by the .mtz file. Check that the model is within the contours of the electron density map.



Get to know the program by trying out some shortcut key combinations:

Left-mouse Drag: Rotate view

Ctrl Left-Mouse Drag: Translate view

Shift Left-Mouse: Label Atom

Right-Mouse Drag: Zoom in and out

Ctrl Shift Right-Mouse Drag: Rotate View around Screen Z axis

Middle-mouse: Centre on atom

Scroll-wheel Forward: Increase map contour level

Scroll-wheel Backward: Decrease map contour level

Space: Next Residue

Shift Space: Previous Residue

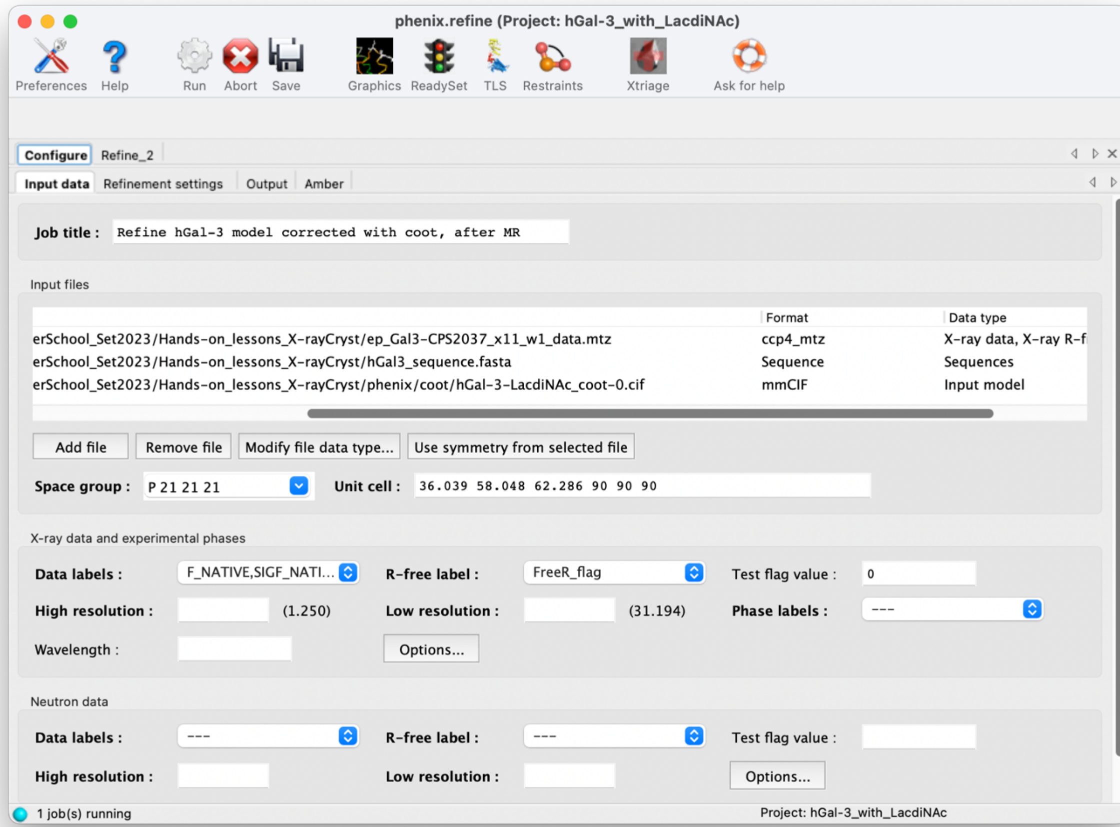
- Go to the N-terminus (**Draw -> Go To Atom**) and follow the polypeptide chain to perform any model adjustments that may be suggested by the $2F_{obs}$-F_{calc} and F_{obs}-F_{calc} electron density maps.

*To adjust the model, you will need several tools from **Model/Fit/Refine** options in menu **Calculate**. Atom positions should be adjusted in the electron density using option **Real Space Refine Zone**.*

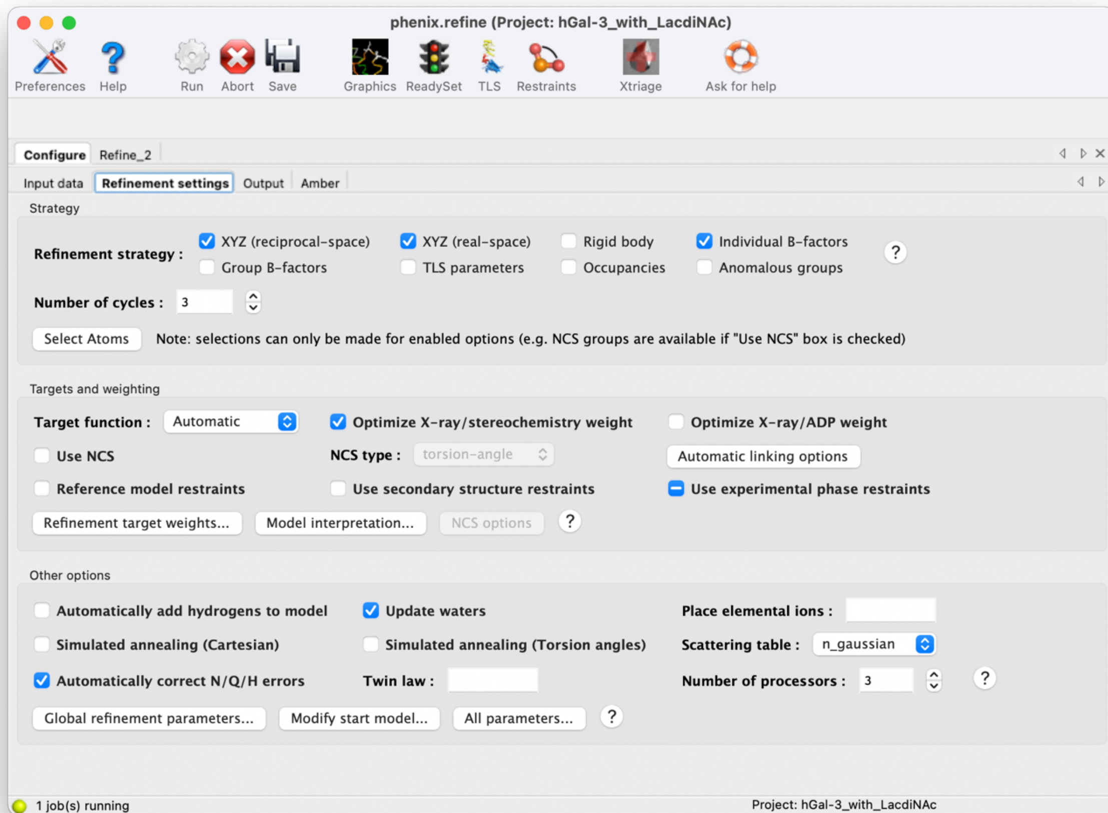
- In menu File, use option **Save Coordinates** to save your work.

B. Model refinement

- Use program **phenix.refine** to refine the model adjusted in **Coot** and calculate new electron density maps.



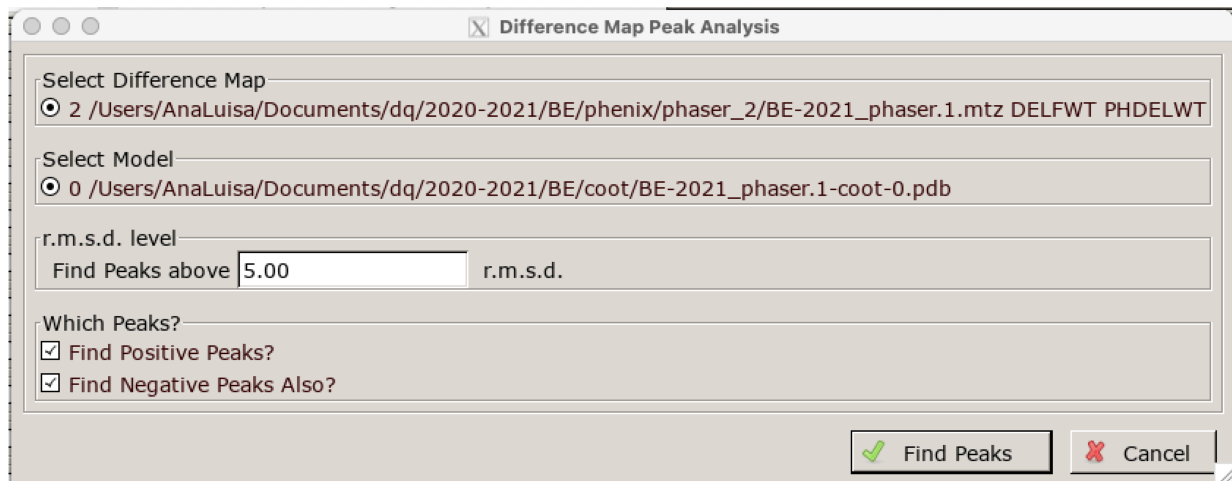
- In Refinement settings, choose to refine **XYZ** coordinates and **individual B-factors**. Choose to **Optimize X-ray/stereochemistry weight** and **Update waters**.



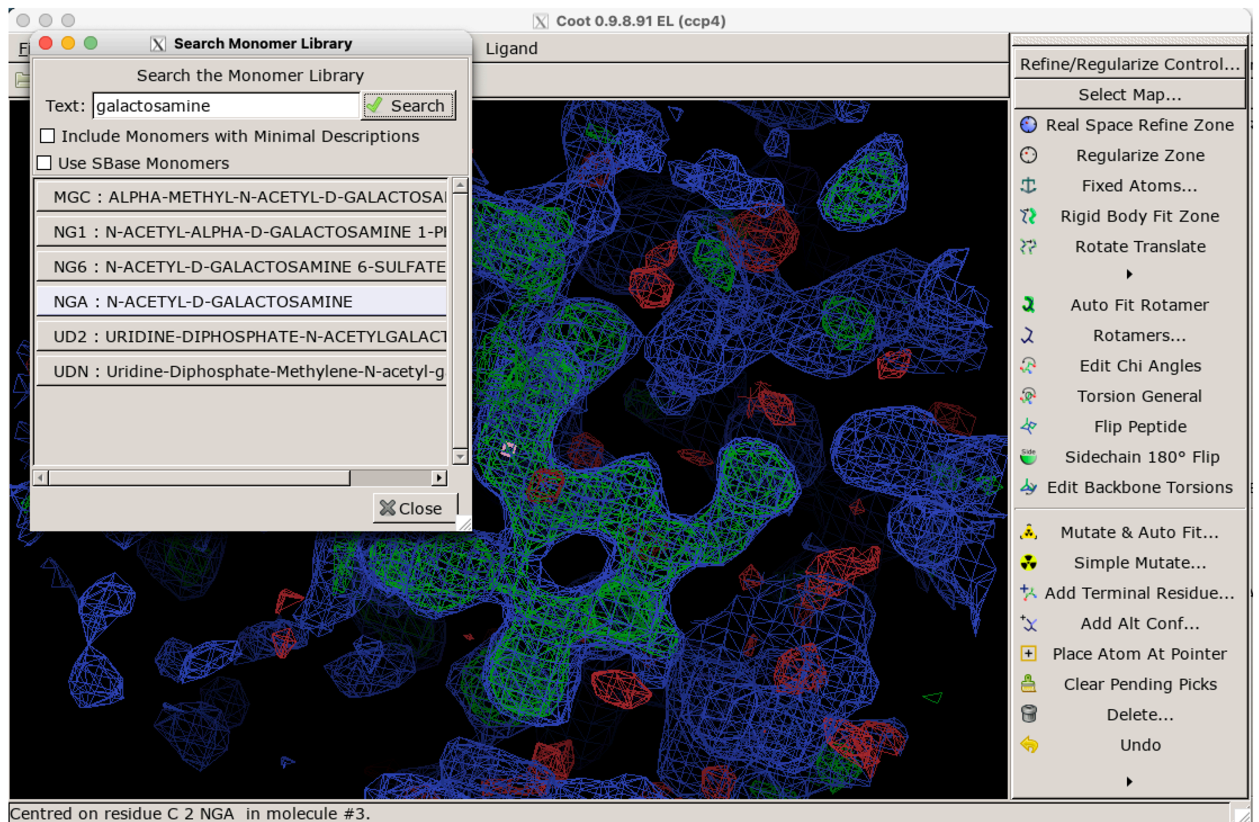
- After refinement, confirm the improvement of refinement parameters like R factors and rms deviations for bonds and angles.
- Inspect the new electron density maps in Coot.

C. Analysis of the carbohydrate recognition site and ligand interactions

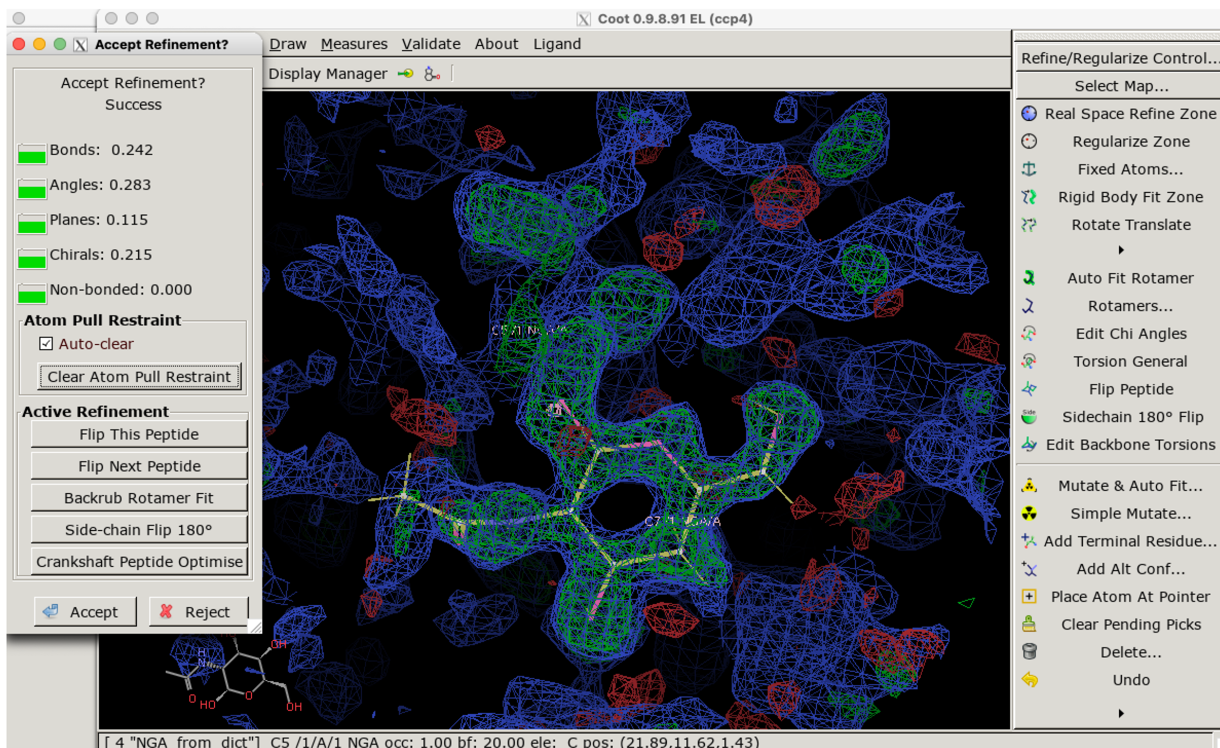
- In Coot's Validate menu, use the **Difference Map Peak Analysis** option to investigate positive and negative peaks of $F_{obs} – F_{calc}$ electron density



- Check if there is electron density that can correspond to the ligand with which the protein was crystallized.
- In Coot's File Menu, search the Coot database for the structure of the ligand used (to facilitate ligand construction, the tutors will provide the coordinates for LacdiNAc).



- Position it in the density using the commands in the **Rotate / Translate** and the **Real Space Refine Zone** options.



- In Coot's Edit Menu, merge N-acetylgalactosamine into the protein model, using "Merge Molecules" Tool.
- At this stage, you can refine this preliminary model, and see how the refinement statistics improve, or add the N-acetylglucosamine moiety, completing the ligand's model. Refine in phenix.refine and calculate new electron density maps.
- Check what type of interactions the ligand makes with the protein and which residues are involved.

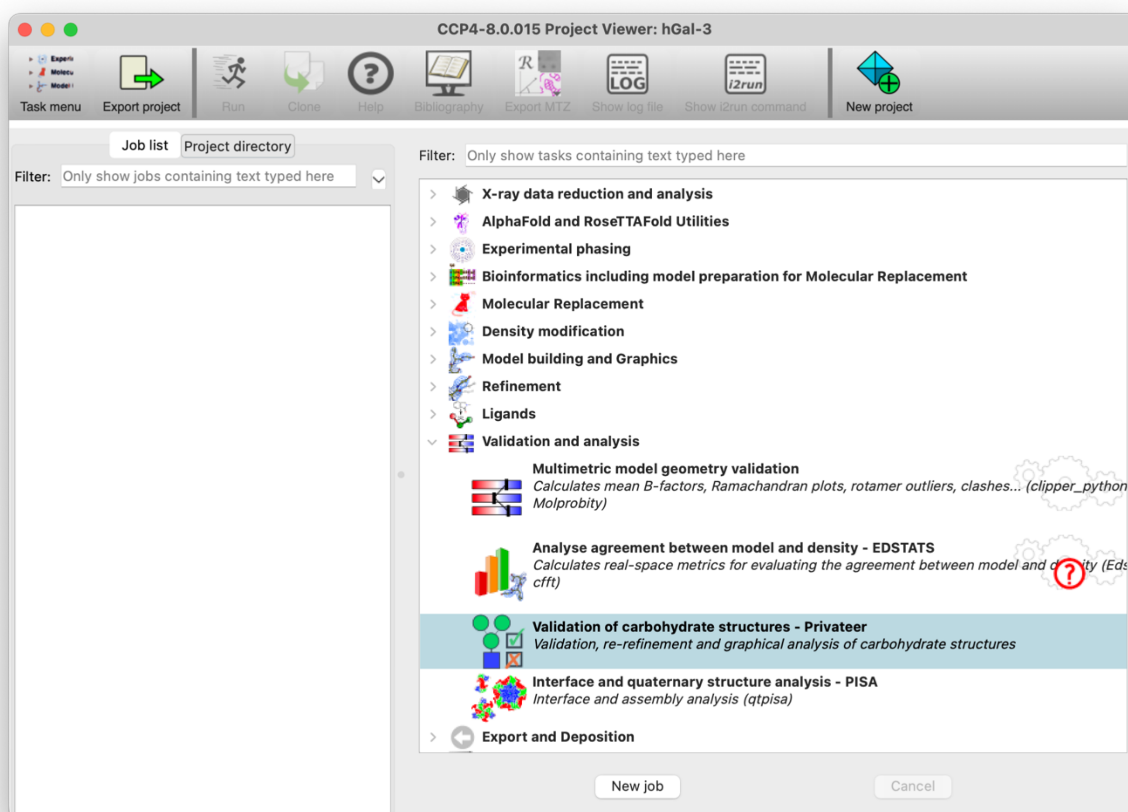
D. Model validation

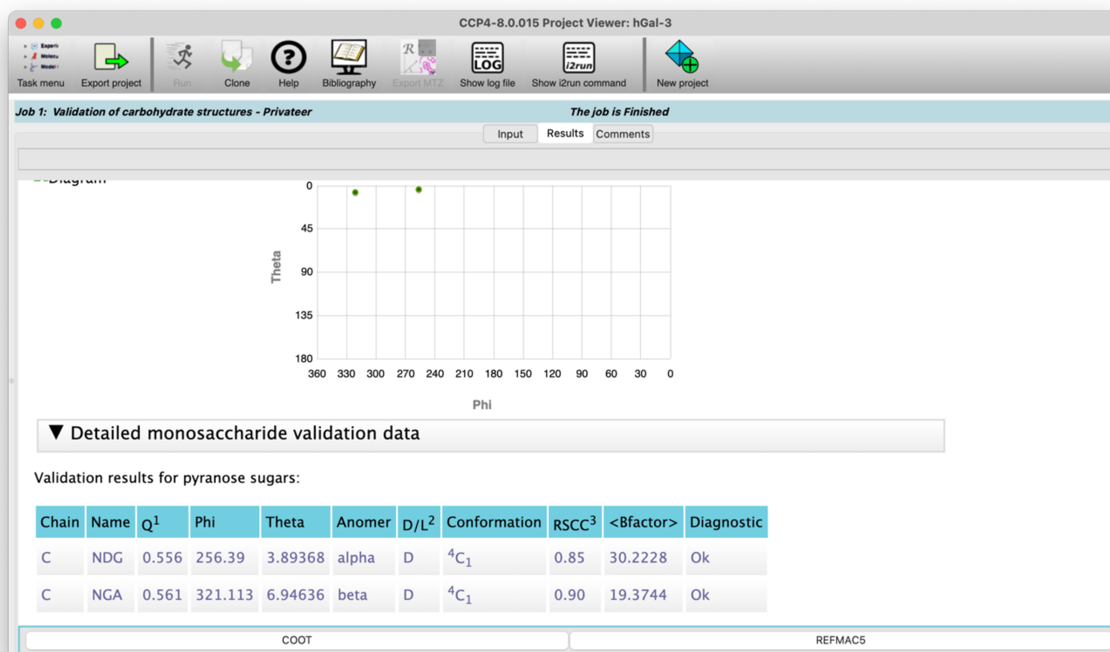
- In the **Validate menu** you can find several validation tools that you should use to correct possible errors in the model.
- Using the **Real Space Refine Zone** option you can adjust the model to the electron density, obeying to the validation criteria.

- Follow the various validation steps suggested until you get a final model, that best represents the X-ray diffraction and conforms with the stereochemical restraints.
- In the end, don't forget to save your work.

E. Validation of the LacdiNAc structure

Use program Privateer, implemented in CCP4i2 interface, to check the correctness of your LacdiNAc model.





Chain	Name	Q[1]	Phi	Theta	Anomer	D/L[2]	Conformation	RSCC[3]	<Bfactor>	Diagnostic
C	NDG	0.556	256.39	3.89368	alpha	D	$^{4}C_{1}$	0.85	30.2228	Ok
C	NGA	0.561	321.113	6.94636	beta	D	$^{4}C_{1}$	0.90	19.3744	Ok

References

Lima, Carlos D L, Helena Coelho, Ana Gimeno, Filipa Trovão, Ana Diniz, Jorge S Dias, Jesús Jiménez-Barbero, et al. 2021. “Structural Insights into the Molecular Recognition Mechanism of the Cancer and Pathogenic Epitope, LacdiNAc by Immune-Related Lectins.” *Chemistry – A European Journal* n/a (n/a). https://doi.org/https://doi.org/10.1002/chem.202100800.

Carvalho, Ana Luísa, Teresa Santos-silva, Maria João Romão, Eurico J. Cabrita, and Filipa Marcelo. 2018. “Structural Elucidation of Macromolecules.” *Essential Techniques for Medical and Life Scientists*, September, 30–91. https://doi.org/10.2174/9781681087092118010005.

Carvalho, Ana Luísa, José Trincão, and Maria João Romão. 2010. “X-Ray Crystallography in Drug Discovery.” In *Methods in Molecular Biology (Clifton, N.J.)*, 572:31–56. https://doi.org/10.1007/978-1-60761-244-5_3.