



MR STRUCTURE SOLUTION AND MODEL BUILDING

As seen in the previous practical lesson, each reflection spot in a diffraction pattern results from a monochromatic wave, constructively scattered by all equivalent lattice points. The intensity of the reflection corresponds to the intensity of this wave and is recorded by the detector.

One other crucial piece of information to reconstruct the image of the macromolecule inside the crystal is the **phase of the scattered wave** when intercepted by the detector. This information cannot be retrieved and is in the loss of this phase angle that resides what is known as the **Phase Problem in crystallography**.

Solving the phase problem consists of **correctly estimating the phase angle values**, introduce them in the **electron density equation** (Eq. 1) and retrieve the **electron density map** that surrounds all the atoms of the macromolecule in the crystal.

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 (proportional to $\sqrt{|hkl|}$),
 (x, y, z) are the fractional atomic coordinates in the unit cell,
 V is the volume of the unit cell

One of the main methods to overcome the loss of the phase information is **Molecular Replacement (MR)**. This method requires that the atomic coordinates of at least one **similar protein/domain** are available (previously determined by X-ray crystallography, NMR or even cryo-EM methods). Similarity is evaluated by alignment and comparison of the primary sequences of both proteins.

In this method, the **potentially similar model** (sequence identity higher than 30%) is used to “find the place” of the protein of interest in the unit cell of the unknown structure. This is computationally performed in two steps (guided by the calculation of particular electron density maps where phase angles are assumed to be equal to 0, known as **Patterson maps**): first, a **rotation search** finds the orientation of the molecule in the unit cell; second, a **translation search** throughout the unit cell finds the position of the previously rotated model.



MR is probably the most used method to solve a protein's 3D structure, a consequence of the increasing number of structures available in the PDB. Very recently, artificial intelligence (AI) algorithms are being used to predict 3D protein structures based on primary sequence information (e.g. AlphaFold2 or RoseTTAFold software). The potentially similar models (either from the PDB or calculated with AI methods) can provide the necessary phasing information in MR structure solution strategies. MR is particularly important for **ligand-binding studies**, as the free structure can be used as a search model on an X-ray diffraction dataset from a protein crystallized with a ligand of interest.

In this practical lesson you will experimentally address one of the methods of solving the Phase Problem, **Molecular Replacement**, by going through the steps to obtain the three-dimensional structure of the carbohydrate-recognition domain of human galectin-3 (hGal-3 CRD), crystallized in the presence of the tumour and pathogenic glycobiomarker LacdiNAc (GalNAc β 1-4GlcNAc). X-ray diffraction data (list of measured intensities) used in this tutorial were obtained in a synchrotron source.

In the **Protein Data Bank** (PDB), you will find already known structures of hGal-3 CRD that will allow us to use this method to solve the Phase Problem.

Dedicated software you will need: PHENIX (https://phenix-online.org/download/nightly_builds.cgi) and Coot (<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:\lcdgeral** for creating projects and storing files.

Besides the diffraction data collected previously (provided by the teacher), you will need the **primary structure** (amino acid sequence) of human galectin-3 CRD:

```
>human_galectin-3_CRD
MLIVPYNLPLPGGVVPRMLITILGTVKPNANRIALDFQRGNDVAFHFNPRFNENNRRVIVCNTK
LDNNWGREERQSVFPFESGKPFKIQVLVEPDHFKVAVNDAHLLQYNHRVKKLNEISKLGISGDI
DLTSASYTMI
```



A. Find a potentially similar model (search model) in the PDB and prepare it for MR

1. Search the PDB for human galectin-3 CRD structures (<https://www.rcsb.org> or <https://www.ebi.ac.uk/pdbe/>)

The screenshot shows the RCSB PDB website interface. The search bar contains the query "human galectin-3 CRD". The search results are displayed in a table with columns for "in Additional Structure Keywords", "in Structure Title", and "Count". The results list several entries, including "Complex, Human Galectin-3 CRD, Type 1 LacNAc, SUGAR BINDING PROTEIN" and "Crystal structure of Human galectin-3 CRD in complex with LNT". The "Advanced Search Query Builder" is visible on the left side of the page, showing the search query "Full Text = *human galectin-3 crd".

2. Choose a suitable coordinate model and download it (in PDB format) to a folder in your work computer.



RCSB PDB

Deposit Search Visualize Analyze Download Learn About Documentation Careers COVID-19

Structure Summary 3D View Annotations Experiment Sequence Genome Versions

Biological Assembly 1

Display Files Download Files Data API

1A3K

X-RAY CRYSTAL STRUCTURE OF THE HUMAN GALECTIN-3 CARBOHYDRATE RECOGNITION DOMAIN (CRD) AT 2.1 ANGSTROM RESOLUTION

PDB DOI: <https://doi.org/10.2210/pdb1A3K/pdb>

Classification: **GALECTIN**
Organism(s): **Homo sapiens**
Expression System: **Escherichia coli**
Mutation(s): No

Deposited: 1998-01-22 Released: 1998-07-15
Deposition Author(s): **Seetharaman, J., Kanigsberg, A., Slaaby, R., Leffler, H., Barondes, S.H., Rini, J.M.**

Experimental Data Snapshot

Method: X-RAY DIFFRACTION
Resolution: 2.10 Å
R-Value Free: 0.240
R-Value Work: 0.170
R-Value Observed: 0.170

wwPDB Validation

Metric	Percentile Ranks	Value
Clashscore	5	5
Ramachandran outliers	0	0
Sidechain outliers	0.8%	0.8%

Biological assembly 1 assigned by authors.

Find Similar Assemblies

Validation Report

Global Symmetry: Asymmetric - C1
Global Stoichiometry: Monomer - A1

This is version 2.0 of the entry. See complete [history](#).

Reminder: you can inspect the structures and evaluate your choice using Mol (in the PDB website) or Chimera or Coot (in your computer).*

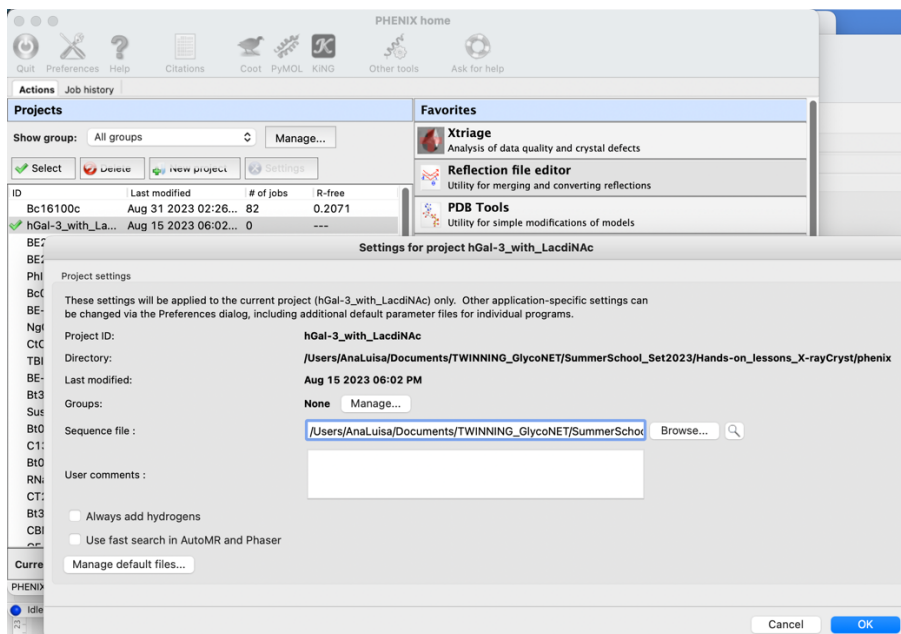
Tips for Coot: Use the left mouse button to rotate the molecule and the right button to enlarge / reduce the image. With the central button you can click and center the image on an atom of your choice. In the Display Manager you can manage and modify parameters of the represented objects (coordinates and electron density maps).



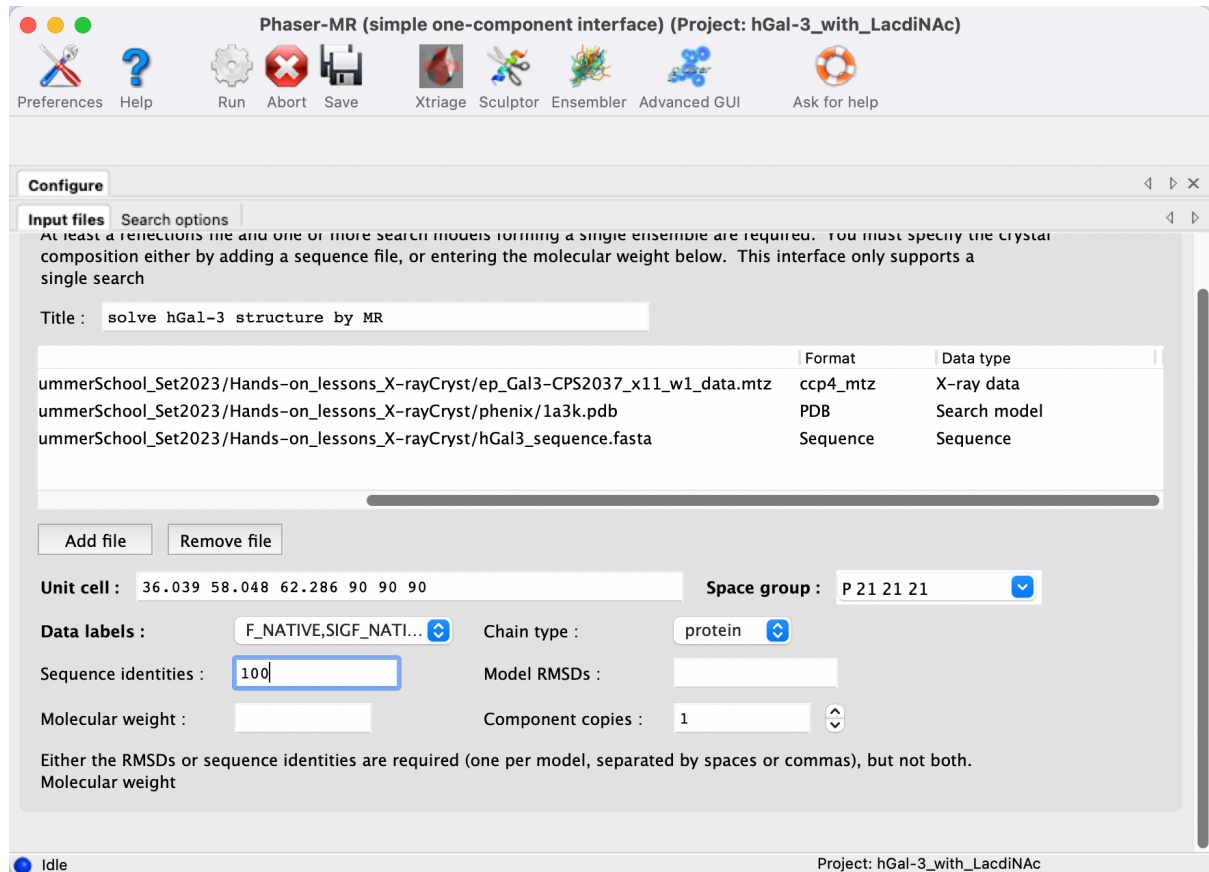
To solve the Phase Problem by Molecular Replacement you may want to “clean” the 3D model found. All pre-existing ligands, water molecules, as well as other solvent molecules, ions or cofactors can be removed from the model. For this, you can inspect and directly edit the coordinate file (it is a text file). Alternatively, you can use the **Model tools** options in Phenix. Or you can leave it as is since most MR programs can deal with the presence of non-protein molecules in the search model coordinate file.

B. Solve the Phase Problem by Molecular Replacement

1. Start the PHENIX software package and create a New Project with a name of your choice.
2. Choose a folder for Phenix to store all generated files.
3. Indicate to Phenix the location of the file containing the primary sequence of hGal-3 CRD.



4. In the PHENIX interface, find and click on the **Molecular replacement** menu on the right sidebar. Choose the option **Phaser-MR (simple one-component interface)** and add the required files (diffraction data, coordinates of search model and sequence file)



5. In **Search options**, fill in the estimated number of hGal-3 CRD molecules in the asymmetric unit (*number of copies to search for*). It should not be necessary to change the predefined parameters.
6. Run the program and analyze the results obtained - check the solution found.
7. Use the **Coot** program to open the obtained .pdb and .mtz files and confirm that the structure has been solved - the model must be contoured by the electron density map; there must be zones with no electron density (and no model) that will correspond to the solvent channels.

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



GLYCOTwinning Summer School
25th – 29th September 2023, FCT-NOVA, Caparica, Portugal

Space: Next Residue

Shift Space: Previous Residue

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.