

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 $\sqrt{I_{hkl}}$), (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.



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

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 (GalNAcβ1-4GlcNAc). 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 (<u>http://bernhardcl.github.io/coot/wincoot-download.html</u>)

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.

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

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

	~ < >		0 (🔒 rcsb.org			ି ସେଦ		\checkmark	₾ +
тов С	<u>et</u> <u>fet</u>	O) (G)	s 😁	w	• 88	88 ₩	88 0	•	PDB PDB: Se
RCS	SB PDB	eposit -	Search -	Visualize 🗸	Analyze 🗸 🛛	Download 🗸	Learn -	About - [Documentation •	MyPDB - Careers C	Contact us	•
KCSB			<u> </u>	,347 Structures f	rom the PDB	▼ 3D	Structures 😧	human	galectin-3 CRD	Include CSM		
T RO	ILIN DAI	A DAAK	Mod	lels (CSM)				in Addition Keywords	nal Structure	otations	н	elp
RCC59	PDB-101	© PC	B S	EMDataResource	SAKB	WWPDB Foundation	OPDB-E	Comple 3 CRD, SUGAR	x, Human Galectin Type 1 LacNAc, BINDING PROTEIN	-	Fy D (2
S	earch Q	Jery History	Bro	owse Annotatio	ons MyPl	DB		in Structu	re Title			
QUE	RY: Full Text = "h	uman galecti	n-3 crd*					Crystal galectin with LN	structure of Human n-3 CRD in complex T	MyPDB Login	Search A	2
- Ac	dvanced Se	arch Que	v Builde	er 🕜				Crystal	structure of Human		He	qle
	▲ Full Text	0	,					galectin with LN	n-3 CRD in complex nT	K		
			human gale	ctin-3 crd				Crystal	Structure of Humar	1	Coun	+
			Add Term	Add Subguerv				Galectin with TD	n-3 CRD in Comple 139	Re	move Subauen	,
		Add Subque	ery					Crystal galectir	structure of Human -3 CRD in complex	(
	- Structure	Attributes										
	0											
	 Chemical 	Attributes										
	0											
	▼ Sequence	e Similarity										
	0											
	- Soguono	Matif O										

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



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



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.

				PHENIX ho	me		
0	☆? □	المعدود	${\cal K}$	Se			
Quit Pr	references Help Citations	Coot PyMO	L KING	Other tools	Ask for help		
Actions	Job history						
Project	S			Fa	vorites		
Show gr	oup: All groups	≎ Man	age	4	Analysis of data quality and crystal defects		
🛷 Selec	ct 🥹 Delete 🍦 New pro	oject 😣 Setting	S		Reflection file editor		
ID	Last modified	# of jobs	R-free		Utility for merging and converting reflections		
Bc16	100c Aug 31 2023 0	2:26 82	0.2071		PDB Tools Utility for simple modifications of models		
BE2	5_with_La Aug 15 2023 0	5.02 0		Cottings for n	reject hCol 2 with LoodiNAs	1	
BE				Settings for pr	roject nGai-3_with_LacdiNAc		
Phi	Project settings						
Bc(These settings will be applied	to the current proje	ect (hGal-3	_with_LacdiNAc)	only. Other application-specific settings can		
BE-	be changed via the Preference	es dialog, including	additional o	default parameter	files for individual programs.		
CtC	Project ID:		hGal-3_w	ith_LacdiNAc			
тві	Directory:		/Users/An	naLuisa/Documer	nts/TWINNING_GlycoNET/SummerSchool_S	Set2023/Hands-on_lessons_X-ray	/Cryst/phenix
BE-	Last modified:		Aug 15 20	023 06:02 PM			
Bt3	Groups:		None	Manage			
Sus							
C1:	Sequence file :		/Users/A	naLuisa/Docume	ents/TWINNING_GlycoNET/SummerSchoo	Browse	
BtO							
RNi	User comments :						
CT:							
Bt3	Always add hydrogens						
CBI	Use fast search in Autol	MR and Phaser					
Curre	Manage default files						
PHENIX							
2						Cancel	ОК

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)

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

	Phaser-M	R (simple one-o	compone	ent interfa	ace) (Project:	hGal-3 v	vith Laco	liNAc)		
	🖎 🚫 🗖	1 🚺	1		<u>9</u> 9		A			
	👾 💽 🇤		N	Freembler		A alv f	🥩 ar haln			
Preferences Help	RUN ADORT Sav	ve xtriage	Sculptor	Ensempler	Advanced GUI	ASK T	or neip			
Configure									4	Þ ×
Input files Search option	IS								<	1 1
composition either by single search	adding a sequer	nce file, or enteri	ng the mo	olecular w	eight below. T	his interfa	ce only su	pports a		
Title : solve hGal-	3 structure b	y MR								
						For	mat	Data type		
ummerSchool_Set2023	/Hands-on less	ions X-ravCryst/	ep_Gal3-	CPS2037_	x11_w1_data.ı	ntz ccp-	4_mtz	X-ray data		
	,									
ummerSchool_Set2023	/Hands-on_less	ions_X-rayCryst/	phenix/1	a3k.pdb		PDB		Search model		
ummerSchool_Set2023 ummerSchool_Set2023	Hands-on_less Hands-on_less	cons_X-rayCryst/ cons_X-rayCryst/	phenix/1 hGal3_se	a3k.pdb quence.fas	sta	PDB Seq	uence	Search model Sequence		
ummerSchool_Set2023 ummerSchool_Set2023	3/Hands-on_less 3/Hands-on_less	sons_X-rayCryst/	phenix/1 hGal3_se	a3k.pdb quence.fas	sta	PDB Seq	uence	Search model Sequence		
ummerSchool_Set2023 ummerSchool_Set2023	;/Hands-on_less ;/Hands-on_less	sons_X-rayCryst/	phenix/1 hGal3_se	a3k.pdb quence.fas	sta	PDB Seq	uence	Search model Sequence		
ummerSchool_Set2023 ummerSchool_Set2023 Add file Rem	/Hands-on_less /Hands-on_less /Hands-on_less	ions_X-rayCryst/	phenix/1 hGal3_se	a3k.pdb quence.fas	sta	PDB Seq	uence	Search model Sequence		
ummerSchool_Set202: ummerSchool_Set202: Add file Remu Unit cell : 36.039 5	/Hands-on_less /Hands-on_less pove file 18.048 62.286	sons_X-rayCryst/ sons_X-rayCryst/ 90 90 90	phenix/1 hGal3_se	a3k.pdb quence.fa	sta Spac	PDB Seq e group :	uence P 21 21	Search model Sequence 21		
ummerSchool_Set202: ummerSchool_Set202: Add file Remu Unit cell : 36.039 5 Data labels :	y Hands-on_less /Hands-on_less pove file 18.048 62.286 F_NATIVE,SIG	sons_X-rayCryst/ sons_X-rayCryst/ 90 90 90 GF_NATI 📀	phenix/1 hGal3_se	a3k.pdb quence.fas pe :	sta Spac protein	PDB Seq e group :	uence P 21 21	Search model Sequence 21		
ummerSchool_Set202: ummerSchool_Set202: Add file Remain Unit cell : 36.039 Data labels : Sequence identities :	<pre>//Hands-on_less //Hands-on_less //Hands-on_less ave file ave file ave file F_NATIVE,SIC ave file ave file</pre>	sons_X-rayCryst/ sons_X-rayCryst/ 90 90 90 GF_NATI 📀	phenix/1 hGal3_se Chain tyj Model RI	a3k.pdb quence.fa: pe : MSDs :	sta Spac protein	PDB Seq e group :	uence	Search model Sequence		
ummerSchool_Set202: ummerSchool_Set202: Add file Remain Unit cell : 36.039 5 Data labels : Sequence identities : Molecular weight :	y Hands-on_less / Hands-on_less ove file 8.048 62.286 F_NATIVE,SIC 100	sons_X-rayCryst/ sons_X-rayCryst/ 90 90 90 GF_NATI 😒	phenix/1 hGal3_se Chain ty Model RI Compon	a3k.pdb quence.fa: pe : MSDs : ent copies	sta Spac protein : 1	PDB Seq e group :	P 21 21	Search model Sequence 21		
ummerSchool_Set202: ummerSchool_Set202: Add file Remu Unit cell : 36.039 5 Data labels : Sequence identities : Molecular weight : Either the RMSDs or se Molecular weight	y Hands-on_less / Hands-on_less i / Hands-on_less	sons_X-rayCryst/ sons_X-rayCryst/ 90 90 90 GF_NATI 📀	phenix/1 hGal3_se Chain tyj Model R! Compon	a3k.pdb quence.fa: pe : MSDs : ent copies odel, sepa	sta Spac protein : 1 :rated by space	PDB Seq e group : ©	P 21 21	Search model Sequence 21 💌		

- 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



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). <u>https://doi.org/https://doi.org/10.1002/chem.202100800</u>.
- 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. <u>https://doi.org/10.2174/9781681087092118010005</u>.
- 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.