

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

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

 $\begin{array}{l} \alpha_{hkl} \text{ is the phase angle of reflection } hkl, \\ |F_{hkl}| \text{ is the structure factor amplitude,} \\ (x,y,z) \text{ are the fractional atomic coordinates in the unit cell,} \\ V \text{ is the volume of the unit cell} \end{array}$

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

2|Fobs|-|Fcalc| and |Fobs|-|Fcalc|

where the $|\mathbf{F}_{obs}|$ values correspond to the intensities (~ $\sqrt{I_{hkl}}$) 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 **2F**_{obs}-**F**_{calc} and **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 Xray diffraction data. One of the validation parameters is known as the *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 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 (<u>https://phenix-online.org/download/nightly_builds.cgi</u>), CCP4 (https://www.ccp4.ac.uk/download/#os=windows) and Coot/WinCoot (<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.

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.



•				X	Coot 0.9	5 EL (ccp4	4)		
<u>F</u> ile	<u>E</u> dit	<u>C</u> alculate	<u>D</u> raw	<u>M</u> easures	<u>V</u> alidat	e About	Ligand	Det	fina/Dagulariza Control
	Open C	coordinates			8-			Rei	nne/Regularize Control
	Auto O	pen MTZ			0~				Select Map
🖹 (Open M	ITZ, mmCI	F, fcf or	phs				\odot	Real Space Refine Zone
8	Open M	1ap						\odot	Regularize Zone
	Import	CIF dictior	nary					Ŧ	Fixed Atoms
💠 (Get Mo	nomer						33	Rigid Body Fit Zone
00 ×	Search Fetch P	Monomer L DB using A	_ibrary.	 n Code				55	Rotate Translate
- 19 F	Fetch P	DB & Map (using E	DS					•
F	Fetch P	DB & Map (using Pl	DB-REDO				2	Auto Fit Rotamer
	Save C	oordinates.						2	Rotamers
	Save S	ymmetry C	oordina	ites				\mathbf{R}	Edit Chi Angles
	Save S Recove	tate r Session						@	Torsion General
	Evport	Man						R	Flip Peptide
	Export	Map Fragm	ent					Side	Sidechain 180° Flip
8	Delete	Molecules a	nd Map)S				4y	Edit Backbone Torsions
- (Curlew								•
- 	Exit								

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.

				PHENIX home
🕑 🔀 🦿		🛫 💉	$\boldsymbol{\mathscr{K}}$	🤹 🗘
Quit Preferences He	lp Citations	Coot PyMOL	KING	Other tools Ask for help
Actions Job history				
Projects				Xtriage
Show group: All group	oups	O Mana	ae	Analysis of data quality and crystal defects
Select O Dele	te New project	Settings	J	Reflection file editor
		octangs		
D	Last modified	# of jobs	R-free	PDB 100IS
BC16100C	Aug 31 2023 02:26.	. 82	0.2071	· outry for simple mountations of models
rGal-3_with_La	Sep 05 2023 10:13.	. 1		PDB file editor
BE2023_aula1P4	May 25 2023 01:13.	4	0.3200	Simple hierarchy-based editing of PDB files
BE2023	May 15 2023 11:21.	3	0.3318	phenix.refine
Phils_Lysozyme	NOV UZ ZUZZ U3:04.	. 19	0.2208	Automated refinement using diffraction data (X-ray, neutron,)
BE-2022	Jun 02 2022 11:51.	/0	0.3703	AlphaFold (predicted models)
NaCCP	May 05 2022 01:01.	35	0.2037	Crystals: Data analysis and manipulation
CtCBM50	Jan 27 2022 02:01.	. 19	0.1983	
TBI-Carolina	Oct 29 2021 04:56	. 5	0.2066	Xtriage
BE-2021	Jun 15 2021 11:21		0.2233	
Bt3015	May 07 2021 01:50	13	0.1748	Reflection file editor
SusD_BACOVA2	Apr 08 2021 04:34 .	63	0.1836	Utility for merging and converting reflections
Bt0996C	Jan 16 2021 10:50 .	74	0.2991	Validation and map-based comparisons
C13D19	Jan 07 2021 07:20 .	3	0.3398	Experimental phasing
Bt0866	Jan 01 2021 09:33 .	29	0.3807	
RNaseA	Nov 02 2020 02:43.	3		Molecular replacement
CT28	Aug 30 2020 02:50	. 48	0.4254	Phaser-MR (simple one-component interface)
Bt3013	Jul 14 2020 12:28	. 4	0.2413	Automated molecular replacement with Phaser – use only for structures
CBM11_hexa	Oct 03 2019 02:34	. 1		with a single type of component
	0 07 0010 00:11	47	0.0004	
CBM11_hexa	Oct 03 2019 02:34 07 0010 00:11 /Users/AnaLuisa/Docu	. 1 ments/TWINN	ING_Glyco	DNET/SummerSchot Browse
PHENIX version 1.20-44	59-000			Project: hGal-3_with_LacdiNAc



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

			phenix.refine (Project: hGal-3_wi	ith_LacdiN	Ac)		
iferences Pelp	Run Abort	Save Gra	phics ReadySet TLS	Restraints Xt	triage A	Ask for help		
onfigure Refine_2	2							4 1
put data Refinem	ent settings Ou	utput Amber						4
Job title : Refin	ne hGal-3 mode	el corrected w	with coot, after	MR				
put files						Format		Data type
rSchool_Set2023	/Hands-on_less	ons_X-rayCryst/	ep_Gal3-CPS2037_x	11_w1_data.mtz		ccp4_mtz		X-ray data, X-ray R-f
rSchool_Set2023	/Hands-on_less	ons_X-rayCryst/l	hGal3_sequence.fast	a		Sequence		Sequences
rSchool_Set2023	/Hands-on_less	ons_X-rayCryst/	phenix/coot/hGal-3	-LacdiNAc coot-0.ci	if	mmCIF		Input model
Add file	Remove file M	odify file data tyr	De Use symmetry	from selected file	_			
Add file R	Remove file M	odify file data typ	De Use symmetry	y from selected file	00			_
Add file R	Remove file Me	odify file data typ	cell: 36.039 58.0	y from selected file	90			
Add file R Space group : F X-ray data and expen	Remove file Mo P 21 21 21 rimental phases	odify file data typ	Use symmetry	y from selected file	90			
Add file R Space group : F X-ray data and exper Data labels :	Remove file Me P 21 21 21 rimental phases F_NATIVE,S	Unit oligf_NATI 📀	Cell : 36.039 58.0 R-free label :	y from selected file 048 62.286 90 90 FreeR_flag	90	Test flag value :	0	
Add file R Space group : F X-ray data and exper Data labels : High resolution :	Remove file Mr P 21 21 21 rimental phases F_NATIVE,S	Unit of IGF_NATI (>)	Cell : 36.039 58.0 R-free label : Low resolution :	y from selected file 048 62.286 90 90 FreeR_flag (3	90 ? 31.194)	Test flag value : Phase labels :	0	
Add file R Space group : F X-ray data and exper Data labels : High resolution :	Remove file Mr P 21 21 21 rimental phases F_NATIVE,S :	Unit of GF_NATI () (1.250)	cell : 36.039 58.0 R-free label : Low resolution :	y from selected file 048 62.286 90 90 FreeR_flag (3	90 C 31.194)	Test flag value : Phase labels :	0	
Add file R Space group : F X-ray data and exper Data labels : High resolution : Wavelength :	Remove file Mr P 21 21 21 rimental phases F_NATIVE,S	IGF_NATI C	cell : 36.039 58.0 R-free label : Low resolution : Options	y from selected file 048 62.286 90 90 FreeR_flag	90 3 1.194)	Test flag value : Phase labels :	0	e
Add file R Space group : F X-ray data and exper Data labels : High resolution : Wavelength : Neutron data	Remove file Mr P 21 21 21 rimental phases F_NATIVE,S :	iGF_NATI 3	Cell : 36.039 58.0 R-free label : Low resolution : Options	y from selected file 048 62.286 90 90 FreeR_flag (3	90 © 31.194)	Test flag value : Phase labels :	0	
Add file R Space group : F X-ray data and exper Data labels : High resolution : Wavelength : Neutron data Data labels :	Remove file Mr P 21 21 21 rimental phases F_NATIVE,S :	CIL250	cell : 36.039 58.0 R-free label : Low resolution : Options R-free label :	y from selected file 048 62.286 90 90 FreeR_flag (3	90 31.194)	Test flag value : Phase labels : Test flag value :	0	6
Add file R Space group : F X-ray data and exper Data labels : High resolution : Wavelength : Veutron data Data labels : High resolution : High resolution :	Remove file Mr P 21 21 21 rimental phases F_NATIVE,S :	IGF_NATI C	cell : 36.039 58.0 R-free label : Low resolution : Options R-free label : Low resolution :	y from selected file 048 62.286 90 90 FreeR_flag 	90 © 31.194)	Test flag value : Phase labels : Test flag value : Options	0	C

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



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

pher	nix.refine (Project: hGal-3_with	_LacdiNAc)	
💥 🤈 🛞 😭 🛄 🔣 K	👲 🐛 🏡 🛛 🔏	N 🔿	
eferences Help Run Abort Save Graphics Re	adySet TLS Restraints Xtria	age Ask for help	
onfigure Refine 2			4 4
nout data Refinement settings Output Amber			4
Strategy			
Refinement strategy: VXYZ (reciprocal-space) Group B-factors	XYZ (real-space) Rigid bod TLS parameters Occupance	ly ♥ Individual B-factors ? cies Anomalous groups	
Select Atoms Note: selections can only be made for er	nabled options (e.g. NCS groups ar	re available if "Use NCS" box is checked)	
Targets and weighting			
Target function : Automatic 😒 🗹 Optimize	X-ray/stereochemistry weight	Optimize X-ray/ADP weight	
Use NCS NCS type :	torsion-angle 🗘	Automatic linking options	
Reference model restraints Use secon	ndary structure restraints	😑 Use experimental phase restraints	
Refinement target weights Model interpretation	NCS options ?		
Other options			
Automatically add hydrogens to model V Upda	ite waters	Place elemental ions :	
Simulated annealing (Cartesian)	lated annealing (Torsion angles)	Scattering table : n_gaussian ᅌ	
Automatically correct N/O/H errors Twin lay	w •	Number of processors : 3	2
Global refinement parameters Modify start model.	All parameters		
1 iah(a) suming		Project: bGal-3 with LoodiNAs	

- 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



0 0 0	X Difference Map Peak Analysis
Select Difference Map	
● 2 /Users/AnaLuisa/Documents/dq/202	20-2021/BE/phenix/phaser_2/BE-2021_phaser.1.mtz DELFWT PHDELWT
Select Model	
● 0 /Users/AnaLuisa/Documents/dq/202	20-2021/BE/coot/BE-2021_phaser.1-coot-0.pdb
r.m.s.d. level	
Find Peaks above 5.00	r.m.s.d.
Which Peaks?	
☑ Find Positive Peaks?	
☑ Find Negative Peaks Also?	
	✓ Find Peaks X Cancel

- 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).



Centred on residue C 2 NGA in molecule #3



 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.





•		1		-			CCP4-8.0.015	Project Vi	ewer: hGal-3			
Nolecu Molecu Model I menu I	Export proje	ct Ru	Clone	Help B	bliography	R X Export MT	LOG Z Show log file Sh	i2run ow i2run cor	nmand New	project		
1: Valida	ntion of ca	rbohydrat	e structures -	Privateer				The j	ob is Finished			
							Input	Results	Comments			
-01491				0								
				0	•	•						
				45								
				De leta								
				·								
				135								
				180	330 300	270 24	0 210 180 150	120 00	60 30 0			
				000	000 000	210 24	Phi	120 00				
▼ D	etailed	d mon	osacchai	ride valio	lation d	ata						
/alidati	ion resu	Its for p	oyranose s	ugars:								
Chain	Name	Q^1	Phi	Theta	Anomer	D/L ²	Conformation	RSCC ³	<bfactor></bfactor>	Diagnostic		
С	NDG	0.556	256.39	3.89368	alpha	D	⁴ C ₁	0.85	30.2228	Ok		
С	NGA	0.561	321.113	6.94636	beta	D	⁴ C ₁	0.90	19.3744	Ok		
				COOT				I			REFMAC5	
				5501								

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. <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.