4 – PHENIX AutoBuild, PHENIX Refine, and Coot

(handout created by Susanna Huang for STARS collegiate branch located at the Georgia Tech campus)

INTRODUCTION

Therapeutic drug discovery is important for the treatment of diseases. Diseases can be treated at their symptoms or at their protein interaction root causes. How can the root causes, such as the mechanism of action of a target protein, be identified? One important way is through structure determination, such as X-ray crystallography.

Structure determination provides information about the target protein of interest, specifically the molecular basis of the protein or protein complex, especially as it relates to how it interacts with and influences other proteins or other cellular signaling.

Being able to see the protein structure, along with its cofactors, ligands, and catalytic residues, at the atomic resolution, enables you to specifically create therapeutics that can target the protein. For example, if a protein has a specific active site that is important for the recognition of an overexpressed signal that is leading to cancerous growth, through structure determination, you can understand which amino acid residues or which cofactors are important at the active site and thus design molecules to block those residues or cofactors at the active site to stop the protein signaling. Another way: you can also design a compound that binds at the active site and that recruits a protein degrader to the target protein, degrading the target protein and preventing the signal from being transduced through the cell.

These are only two examples, one of protein inhibition, one of protein degradation (design of PROTACS), of the many possible examples on how therapeutic drugs can be designed to inhibit target proteins to treat diseases.

Several key steps with this type of medicinal chemistry research includes: Analyzing your protein structure to understand the active site, designing and synthesizing your molecule to target the protein, and testing the molecule in cancer cell lines to see if they have any activity. The molecules that do have activity and that inhibit cancer cell growth can be then co-crystallized with your protein target to see how your designed therapeutic molecule binds and inhibits your protein. This structure determination information can be used to see how your therapeutic molecule can be improved.

One way to solve protein-inhibitor structures for structure determination is through X-ray crystallography.

AIMS FOR THIS HANDOUT

This handout aims to:

- Provide background information on the importance of structure determination as it relates to therapeutic drug design (INTRODUCTION)
- Provide a source to learn the whole workflow from PHENIX autobuild to PHENIX refine to Coot, to improve the refinement of a model
- Provide a source to download PyMol to visualize how well the refinement process is going, in comparison to the original structure
- Provide a source for learning more about PHENIX terminology and example literature that uses structural biology

PHENIX AUTOBUILD AND THE WORKFLOW

From the previous two handouts, you were able to learn how to perform the PHENIX.refine command as well as begin refining structures in Coot.

The question is: Where do you get your original structure to refine with? There are several ways you can tackle this question. For this handout, the method that will be used is PHENIX.autobuild.

Just as it sounds, PHENIX AutoBuild essentially is capable of taking in:

- The .mtz file (with the phase information)
- The .dat / .seq / .fasta file (with the protein sequence information)

to create a brand new model of the protein from scratch. Because it is of course from scratch, that is why the produced .pdb file needs to be refined.

The refinement process looks like this:

- Take the produced .pdb structure file from PHENIX.autobuild
- Take the .mtz file
- Take the .dat / .seq / .fasta file
- Put all of this into PHENIX.refine

PHENIX.refine will then produce an updated, refined .pdb structure file, now this file needs to be modified and fixed up in Coot.

- Take the produced .pdb structure file from PHENIX.refine
- Take the .mtz file
- Refine the .pdb file against the .mtz file in Coot

- Plug the newly saved .pdb file from Coot back into PHENIX.refine
- Repeat the PHENIX.refine, Coot, PHENIX.refine, Coot, etc. cycle until the R-free and R-work are as low as possible (while still having R-free and R-work as close together as possible)

The modified file in Coot is saved under a new .pdb structure file name, and this new .pdb structure file undergoes another round of PHENIX.refine, and the cycle goes on and on until the R-work and R-free are less than 0.2 (or if the resolution of the experimental data is 1.5 angstroms, less than 0.15).

From this next tutorial, you will:

- Learn where to find and download .mtz and .fasta files on the PDB website
- Learn where to find the reported R-free, R-work, and resolution values on the PDB
- Learn how to run PHENIX.autobuild
- Run PHENIX.refine from a .pdb that was produced from PHENIX.autobuild
- Use Coot to perform further refinement
    - Ensure all the amino acids, which can be modeled, are present
    - Go through each amino acid residue and real-space refine it to ensure each residue is fitting inside its electron density as best as it can (this may require flipping residues or shifting them)
    - Eliminate waters that are causing clashes and add in small molecule inhibitors (which in this case is imidazole)
    - Save the updated .pdb file as a new structure file, which can then be used for further PHENIX.refine
- Use PyMol to compare the structure you refined against the originally deposited PDB structure and assess how well you are performing the refinement so far

We will be using this video tutorial created by Susanna Huang:
https://youtu.be/aiiBv5v89s8

Here is a zip folder with the three files you will be needing: https://gtvault-my.sharepoint.com/:u:/g/personal/shuang466_gatech_edu/EWJrkTE6L0VNmeiaoUvCQUQBVkeK6FOXYPVgwc2CzYeVMQ?e=OAgi7b

There are a couple of key things needed to run PHENIX.refine:

- The .pdb file (the structure file produced from PHENIX.autobuild)
- The .mtz file (the electron density map with the phase information)
- The .dat file (the sequence file for the protein)

Open PHENIX. At the bottom, create a new directory for this PHENIX project. Move PHENIX into this new directory, to tell PHENIX this is the folder to work in. Create a new project for this folder.

Copy the three files from the tutorial data folder into your specific 7ol6 PHENIX project folder:

- overall_best.pdb (the rough structure file produced by PHENIX.autobuild)
- overall_best_refine_data.mtz (the experimental data that also has R-free flags)
- rcsb_pdb_7OL6.dat (the .dat file type for the protein sequence)

Open the tutorial video and follow along.

After following along with the tutorial video, you should be able to find the source .mtz and .fasta files on the PDB as well as know where to find the original R-work, R-free, and resolution values for the electron density map. You should also be able to understand how PHENIX.autobuild works as well as be able to run PHENIX.refine and Coot iteratively until the R-work and R-free values begin decreasing. You should also be able to compare your refined structure results to that of the original PDB structure in PyMol.

PYMOL DOWNLOAD INFORMATION:

PyMOL is a gold standard program in the research realm for observing proteins from the Protein Data Bank. You can download it here: https://www.pymol.org/

Once you are in, you can open your most recently refined 7ol6 model (.pdb file). You can also type into the terminal to fetch the PDB deposited 7ol6 model: **fetch 7ol6**

To compare the two structures, align your refined model against the fetched 7ol6 model. You can type **show surface** to show the surface of the protein. This is how you can see where the imidazole inhibitor is binding to the protein. You can also see if there are any extra bumps in your refined model that you might need to go back and check.

You can type **hide surface** to remove the surface. You can also type **show sticks** to show all the individual atoms and **hide cartoon** to make the ribbon model of the protein go away (so you can see the atoms more clearly).

With this setting, you can compare and contrast how well your model is refined, as compared to the deposited PDB structure. This will help you understand what other refinements you might need to try or what other refinements you ought not perform, especially if it is something that the authors did or didn't do.

PHENIX TERMINOLOGY AND FURTHER READING

PHENIX can sometimes be confusing.

Thankfully, we have a handy dandy webpage that serves as a dictionary of sorts to help explain what some of the PHENIX terminology mean: https://phenix-online.org/documentation/dictionary.html

If there is ever a term that you don't quite understand or would like to learn more about, you can always look it up in the PHENIX dictionary or Google it. The PHENIX dictionary will give you more context in terms of the PHENIX program.

Literature that discuss how X-ray crystallography and structure determination play a role in medicinal chemistry research projects and the design and synthesis of small molecules for therapeutic drug discovery (Google search query: "drug design crystallography"):

- Bijak V., Szcygiel M., Lenkiewicz J., Gucwa M., Cooper D., Murzyn K., et al. The current role and evolution of X-ray crystallography in drug discovery and development. *Expert Opinion on Drug Discovery*. 2023, *18*(11):1221-1230. https://doi.org/10.1080/17460441.2023.2246881.
    - Use and importance of X-ray crystallography for designing therapeutics, especially in the context of cryoEM and AI
- Maveyraud L., Mourey L. Protein X-ray crystallography and drug discovery. *Molecules*. 2020, *25*(5): 1030. https://doi.org/10.3390/molecules25051030
    - Use of X-ray crystallography for therapeutic drug discovery, fragments of therapeutics and small molecules
- Zheng H., Hou J., Zimmerman M.D., Wlodawer A., Minor W. The future of crystallography in drug discovery. *Expert Opinion Drug Discovery*. 2014, *9*(2): 125-137. https://doi.org/10.1517/17460441.2014.872623
    - Use of X-ray crystallography for creating, understanding, and analyzing macromolecular models

CHALLENGE:

- Read through two of the listed three papers and determine a couple key points that the two papers both agree on as well as a couple key points they disagree on. Are there any factors that influence why their perspectives would be consistent or inconsistent?