3 – Refining with Coot

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

INTRODUCTION

PHENIX and Coot are important computer programs for determining the structures of proteins and their protein complexes.

After a protein of interest has been crystallized and diffracted with a diffractometer at a synchrotron, the diffraction patterns, or the little spots that were captured by the detector, are scaled, integrated, and processed into an electron density map. Such an electron density map is essentially a detailed electron cloud of your protein of interest.

(a) (b)

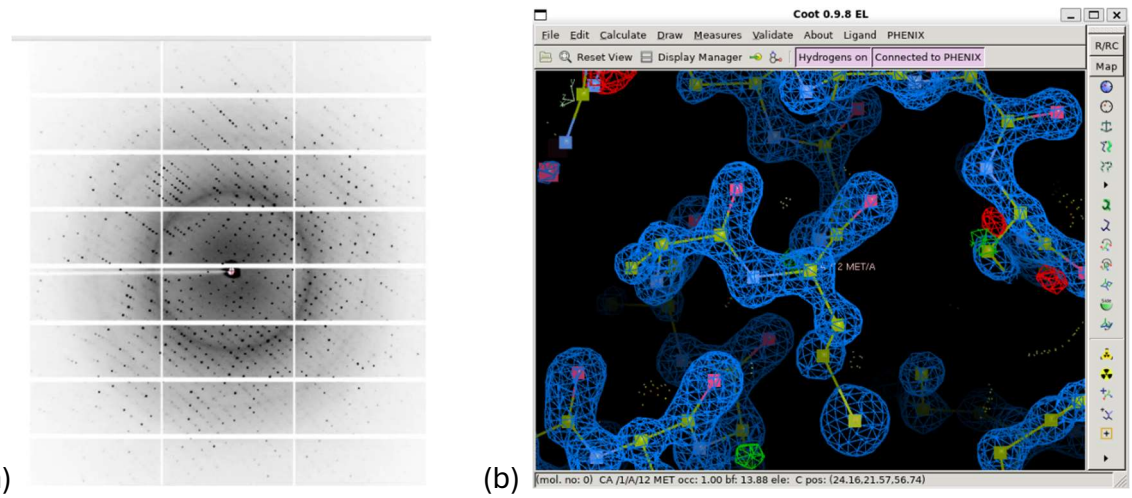

Figure 1. Diffraction patterns and solved models. (a) The diffraction patterns captured by a detector. (b) The blue electron density map that was processed from the diffraction patterns, accompanied by the protein model that was built from the density map. Photo credit: Susanna Huang.

The higher the resolution of the reflection file (.mtz file) (the processed electron cloud), the more clearly defined the electron cloud is. With a high-resolution reflection file, you can build and refine your protein model with high precision and accuracy.

Why does high precision and accuracy matter? If you were interested in observing the active site of a cancer protein and wanted to determine if a small molecule inhibitor is correctly inhibiting it, you would need to investigate how the inhibitor compound is interacting, at the atomic level, with the catalytic side chains of the target protein. If the reflection file has a low resolution (such as 4 Å or higher), then the electron cloud is only suggestive of the specific positions of the amino acid residues and the inhibitor molecule. The electron density is "blurry," and not much definitive information can be extracted from it (for the purposes of

analyzing the catalytic residues). If the reflection file has a high resolution (such as 1-2 Å), or also known as atomic resolution for this specific range, then it is very clear from the electron density where each of the amino acid side chains and the inhibitor should go. Having such high-resolution data would provide higher confidence on the placement of the atoms during model-building and refinement, leading to a more accurate model, with which the interactions between the protein and the inhibitor can be analyzed with high accuracy and detail.

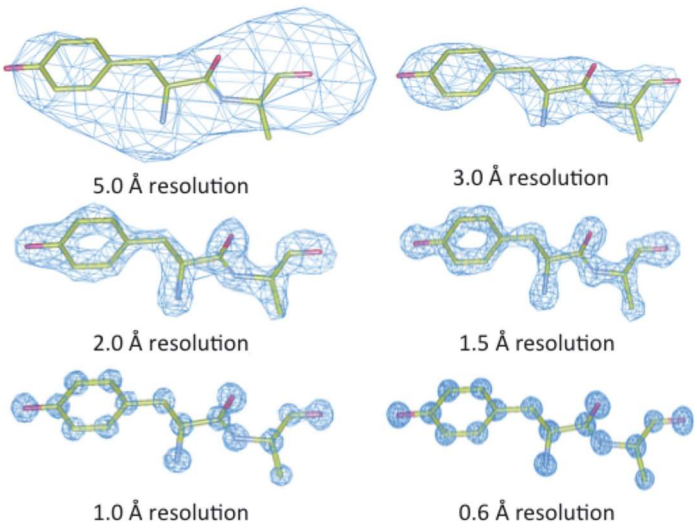

Figure 2. Electron density maps for a tyrosine-alanine peptide segment according to resolution. Higher resolution leads to higher confidence in the specific positions of amino acid residues, which can provide a higher resolution model and clearer insight on the catalytic site of the target protein. (Modified from Kuster D.J., Liu C., Fang Z., Ponder J.W., Marshall G.R. High-resolution crystal structures of protein helices reconciled with three-centered hydrogen bonds and multipole electrostatics. *PLOS ONE,* 2015, *10*. doi: 10.1371/journal.pone.0123146 under Creative Commons Attribution 4.0 International https://www.researchgate.net/figure/The-appearance-of-electron-density-as-a-function-of-the-nominal-resolution-of-the_fig2_276087260)

After a protein has been diffracted and after you have created a model based on the experimental data, you will need to refine your model, or essentially make your model reflect the experimental data as closely as possible.

One way to refine your model is to iterate over several cycles of (1) running PHENIX.refine and (2) manually refining the model result on Coot until something that is known as the R-value and R-free of your model are as low as you can make them (usually around 0.2 to 0.25).

For now, we will need to learn how to manually refine models in Coot first.

AIMS FOR THIS HANDOUT

This handout aims to:

- Provide background information on the importance of high-resolution diffraction data and the introduction to how a model can be refined (INTRODUCTION)
- Provide a source to learn how to use Coot to manually refine a model
- Provide a source for learning more about the PDB and cancer/therapeutics topics

HOW TO REFINE WITH COOT

From the previous handout, you were able to run a PHENIX.refine command. It required the reflections file (the electron density map, .mtz file format) and an approximate model of the protein (the proposed protein structure, .pdb file format). The PHENIX.refine command produced a newly refined .pdb file. In the near future, you will manually refine this new protein model in Coot.

For now, you will need to learn how to use Coot first.

We will be using a video tutorial created by one of the lead developers Dr. Paul Emsley for Coot: https://youtu.be/YUHaSUX32Ic?si=KlUpwjeipkY_lUFF

There are a couple of key things needed to manually refine a .pdb structure in Coot:

- The .mtz file (the electron density map)
- The .pdb file (the structure file that you are refining)

Open PHENIX. Directly open Coot on the PHENIX GUI home page.

Open the tutorial video and follow along. The tutorial data in Coot will provide both the .mtz file and the .pdb file.

After following along with the tutorial video, you should be able to open any .mtz file and its accompanying .pdb file, check the validation charts to determine which residues in the model need to be fixed, fix those amino acid residues, and add in ligands at their electron density locations.

These are the topics the tutorial goes through:

- Load tutorial model (.pdb file) and data (.mtz file)
    - Click “Calculate” tab
    - Click “Load Tutorial Model and Data”

- How to navigate Coot (some key ones)
    - Left button, click and drag left-to-right, rotates the view
    - Right button, click and drag up-to-down, zooms in and out
    - Middle button (on mouse), click on an atom, brings atom to center of screen
    - Left button, double-click on an atom, turns on/off atom labels
    - Middle button (on mouse) or two fingers on mouse pad, scroll, changes contour level (it is good to stay at around 1 angstroms when looking at the model and the data)
    - Middle button (on mouse), click and hold and drag, pans the view
    - Ctrl + left button, click and hold and drag, pans the view
    - Ctrl + right button, click and drag up-to-down, shifts view on z-axis
    - Ctrl + right button, click and drag left-to-right, changes depth of field
- How to change the electron density color
    - Click “Display Manager”
    - Click “Properties” on the selected map
    - Click “Colour”
    - Change color to your desired color (or light steel grey in the example)
- Preparing for the refinement of a model
    - Click “Map” on the right panel
    - Click “estimate” to estimate the weight for the electron density map
- Determining which amino acids need refinement
    - Click “Validate” tab
    - Click “Density Fit Analysis” for your model
    - Change scale to 1.4 or 2.4 to find outliers (these are the amino acid residues that need your attention)
    - Click on the red/orange blocks to bring the Coot window to that specific amino acid residue, which needs to be fixed
        - In the tutorial example, it is residue 89 (phenylalanine)
- How to turn on the text for the icons on the right panel
    - Click on the last triangle button icon on the right panel
    - Click on “Icons and text”
- Refining amino acid positions
    - (Make sure to scroll through electron density so that you can find a setting to more easily view the model and the electron density)
    - If the amino acid chain only needs to be rotated from a red region to a green region, use “Rotamers” on the right panel to rotate the side chain to the correct location

    - If after such a rotation it does not fit quite nicely into the electron density, use "Real Space Refine Zone" on the right panel to let the side chain adjust on its own
    - Alternatively, "Auto Fit Rotamer" can be used on the right panel to automatically rotate the side chain into its correct location while also adjusting nearby atoms
- Finding unmodeled blobs
    - Click "Validate" tab
    - Click "Unmodeled blobs"
    - This will result in a list of green blobs that have electron density with no structure to explain it
- Modeling unmodeled blobs: sulfate ion
    - Click "Calculate" tab
    - Click "Modeling"
    - Click "Add Other Solvent Molecules"
    - Click "SO4: SULFATE ION"
- Show the symmetry related structures
    - Sometimes there will be clear electron density with no model in it. These electron density may be due to a second identical unit of the model. If this is the case, you can try this:
    - Click "Draw" tab
    - Click "Cell and Symmetry"
    - Click "Symmetry On"
    - Change the model color, if you would like (the example was light purple)
- Showing interactions between the protein and the ligand: sulfate ion
    - Click "Measures" tab
    - Click "Environment Distances"
    - Click "Show Residue Environment?"
- How to fix a main chain issue
        - This example is for residue 41, glutamic acid
    - Click "Validate" tab
    - Click "Ramachandran Plot" (this is a phi psi angle map of the amino acids on the polypeptide. The red outliers are the ones with non-realistic phi psi angles, which would mean those outliers need to be checked on)
        - Kleywegt Plots can provide information on the relative distance between chains
    - Click the "Selection" check box and change it to "//A" for just the A chain of the protein

  - Clear the view – minimize the red/green density annotations by clicking "Display Manger" across the top, clicking "Scroll" for the second map, which is the red/green density map, and scroll back the red/green density
  - Clear the view – turn off the "Environment Distances" under the "Measures" tab
  - Use "Real Space Refine Zone" to make the model go into the electron density
  - Use "Flip peptide" to flip a peptide backbone direction into the electron density
- Modeling unmodeled blobs: blob 2
  - Bring ligand blob to center of screen
  - The name of the compound is known (this either comes from the research paper or your experimental data)
  - Click "File" tab
  - Click "Search Monomer Library"
  - Type the key word components for searching the library
    - In the example, it is "phosphate mono guanosine", for 3GP
  - To make the ligand fit the blob:
  - Undisplay the imported ligand by going to "Display Manager" and unchecking "Display" for the 3GP molecule
  - Click "Ligand" tab
  - Click "Find Ligands" (for ligand fitting)
  - Checkbox the specific desired ligand (in this case it is 3GP)
  - Specify that the search can be performed "Right here" or "Everywhere"
  - Check to see that the ligand fits, while scrolling through the angstrom resolution
  - In this specific case, the OH group was pointing in the wrong direction
  - Click "Real Space Refine Zone"
  - Click on the molecule that you want to refine
  - Click "Clear Atom Pull Restraint" in the pop-up window
  - Click and drag on an atom that you want to move across, which in this case would be the OH hydrogen. You can drag and let it stabilize before you let go and confirm that you accept the change
  - Merge the fitted ligand into the protein model by clicking "Edit" tab, "Merge Molecules," check on "Fitted ligand" to merge it with file 0, which is the protein model

PDB-101 AND CANCER TOPICS:

Learning how to use Coot and PHENIX to build and refine protein structures based on diffraction data is important, especially for structure determination of protein-inhibitor complexes.

A broader understanding on which proteins play key roles in the activation of cancer provides insight on why those proteins can serve as treatment targets.

PDB-101 is a website maintained by the Protein Data Bank that provides accessible and important information on key concepts and applications of structure determination, especially as they relate to diseases and their treatments. This is the PDB-101 website: https://pdb101.rcsb.org/

The PDB has all their previous Molecule of the Month displayed on the website. These entries are protein and/or nucleic acid structures that were determined through X-ray crystallography, cryoEM, or solution NMR and that have a significance in medicine.

The second subsection of Molecule of the Month entries on this website is specifically dedicated to cancer research: https://pdb101.rcsb.org/motm/motm-by-category

Two specific examples of structural research for cancer:

- ***Histone Deacetylases***
    - Histone Deacetylase 6 (HDAC6) has been identified as a protein target for prostate cancer treatment. If HDAC6 is inhibited, an inhibition pathway is activated, and the cancer cell growth is inhibited. For this reason, developing HDAC6 inhibitors can be one way to inhibit the growth of prostate cancer.
    - PDB-101 entry and HDACs structures: https://pdb101.rcsb.org/motm/285
- ***Cisplatin***
    - Cisplatin is a platinum atom bonded to two chlorine atoms and two amines. This therapeutic drug can bind at a DNA site and cause the DNA to bend, recruiting an HMG (high mobility group)-domain protein to protect the kinked DNA, but this situation only leads to apoptosis pathways and cancer cell death. Cisplatin is very effective at treating testicular cancer.
    - PDB-101 entry and cisplatin-DNA structure: https://pdb101.rcsb.org/motm/255

CHALLENGE:

- Read through three Molecule of the Month entries and understand how the structure of the protein and/or nucleic acid complex provides insight to the function and usage of the complex for the development of treatments