2 – PHENIX Refine

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

INTRODUCTION

As you may know, there are four stages of crystallography:

1. Protein expression and purification
2. Crystallization and crystal-growth optimization
3. Harvest crystals and diffraction of crystals
4. Building, solving, and refining molecular models

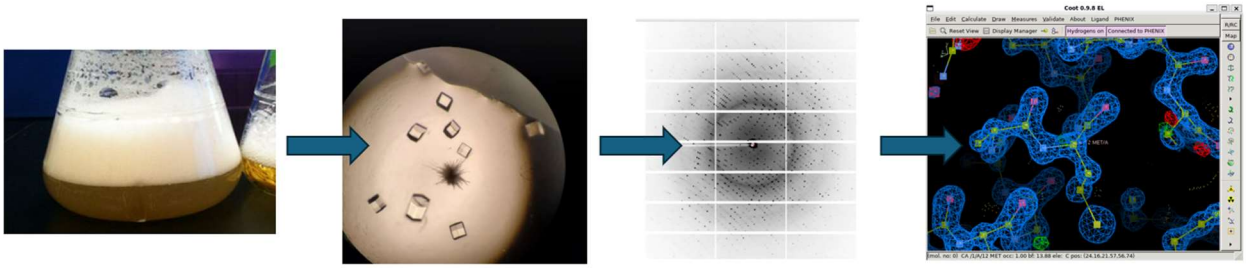


Figure 1. The four steps to crystallography, from left to right: (1) protein expression, (2) crystallization, (3) diffraction data collection, (4) determining the structure.

This model building step is how the diffraction data is turned into the structure, which can be useful for therapeutic drug discovery.

For instance, without the crystal structure of DNA helix, hemoglobin, and CRISPR/Cas9, we would have never known that DNA was made up of two strands and might have a copying and replication mechanism. We would have never known that hemoglobin was made of four units and was allosterically regulated with increasingly stronger binding affinities for oxygen with each bounded oxygen. We would have never known how Cas9 performs its job and how it could be used for creating useful genetic scissors to treat diseases. X-ray crystallography can reveal so much about the function of macromolecular complexes, and that is why learning this interpretation step from diffraction data to protein model is so important.

AIMS FOR THIS HANDOUT

This handout aims to:

- Provide background information on the importance of structure determination (INTRODUCTION)
- Provide alternative way to access PHENIX and Coot through *NoMachine!*
- Provide a source to learn how to use PHENIX Refine
- Provide Protein topics and sources to about proteins and their functions

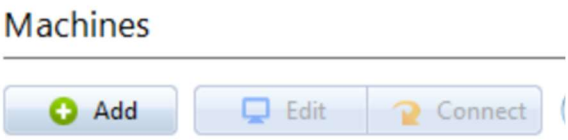
ALTERNATIVE WAY TO ACCESS PHENIX AND COOT

An alternative method to access PHENIX and Coot is by using *NoMachine!* to access a remote computer for using PHENIX and Coot.

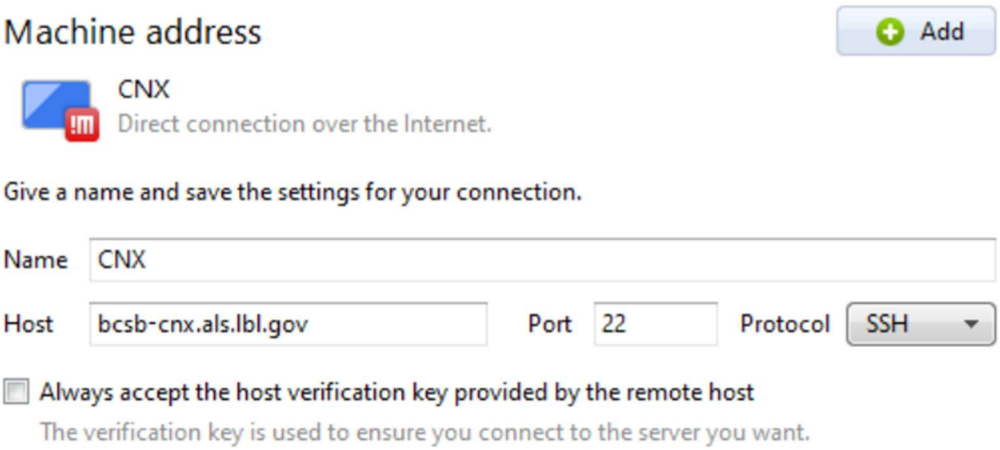
This remote computer is kindly provided by the Lawerence Berkeley National Laboratory and can be accessed through a previous user's login information.

To access this remote computer:

1. Download NoMachine! (https://www.nomachine.com/)
2. Upon installing and opening the program, you will see an Add button at the top left-hand corner

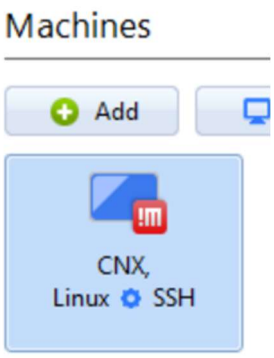


3. Click on this button, and it will go to a window where you can input information about the remote server you will be connecting to:
    a. Name: CNX
    b. Host: bcsb-cnx.als.lbl.gov
    c. Port: 22
    d. Protocol: SSH



e.

4. Click “Add”
5. Now, when you return to the main page, you can see this new remote connection you can enter:



6. Click onto it, now you will be brought to a login page. You can input this username and password:
    a. Username: shuang6
    b. Password: LBLxrayDSTARS3#
7. Now, at the new window, along the top, it will say “New desktop,” click on this
8. Then it will say, “Create new virtual desktop,” click on this icon.
9. Now you will enter into the virtual desktop, where you can access the terminal to access PHENIX and Coot. If you want to use PHENIX, search for the terminal and type in the terminal: **phenix**
10. While in the virtual desktop, you can also open the internet and download protein files from the PDB, so that you can work on it on the PHENIX in the virtual desktop.
11. When you are saving your PHENIX projects onto this virtual desktop, navigate to the “2024-2025 STARS” subfolder under the “shuang6/Documents” folder. Create a new folder with your name in the STARS folder. You can save and access all your individual PHENIX projects in this folder.

HOW TO USE PHENIX REFINE

We will be following this tutorial created by the PHENIX development team to learn how to use PHENIX refine: https://youtu.be/6m9dX8gA5o0?si=nqJNW8jhHbgJrU1o

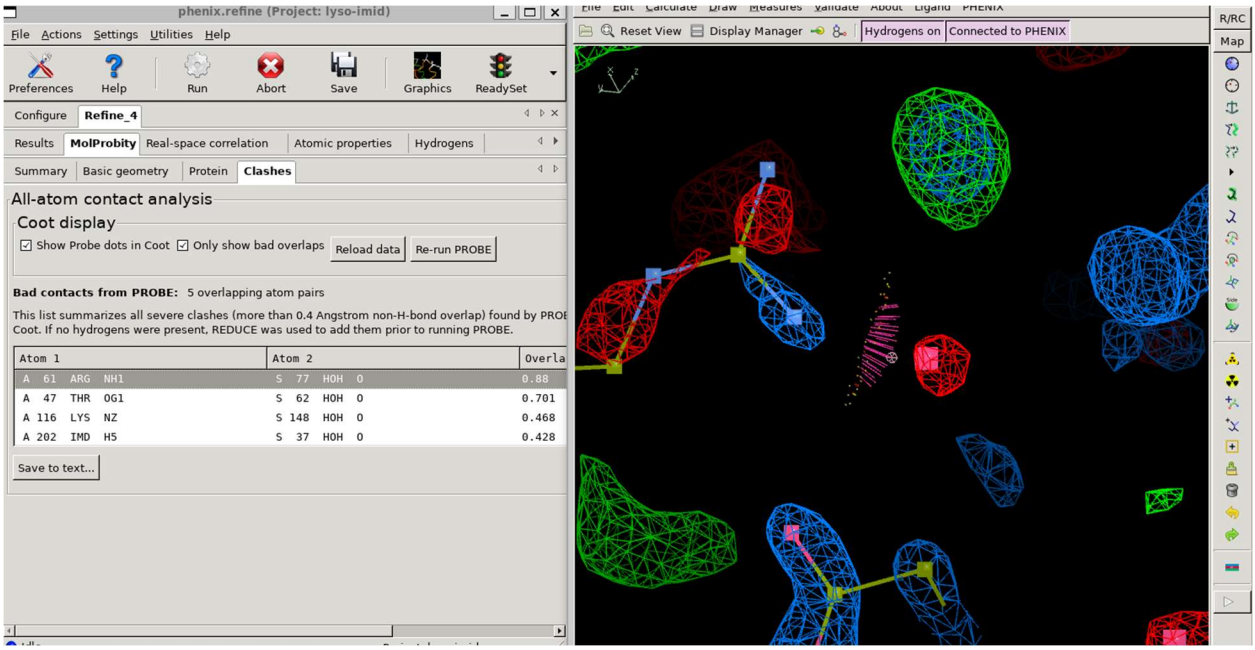
PHENIX refine is one of the important PHENIX commands that can be used to determine and refine crystal structures. PHENIX Refine essentially takes (1) the electron density map, which was produced by pre-processing of the diffraction spots, and (2) the protein structure to modify or tweak the structure to make it better fit the electron density map, making the current structure reflective of the experimental data. Hence, refine.

There are a couple of key things needed to run such a command in PHENIX:

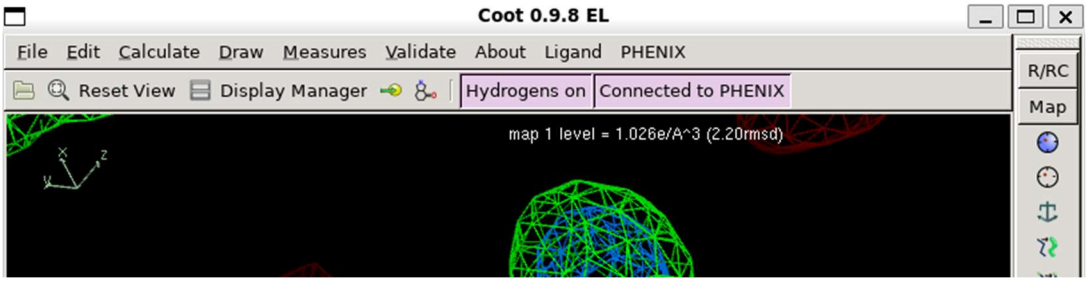
- The .mtz file (which is the electron density map, the experimental data, from the diffraction experiment)
- The .pdb file (which is the current protein structure that is used by the program to tweak)

Create a folder where you want to save your PHENIX projects into, create a specific folder for this specific PHENIX refine run, change to this directory (essentially put your files in here and work in this directory), and follow along with the tutorial video with the tutorial data while working in this folder.

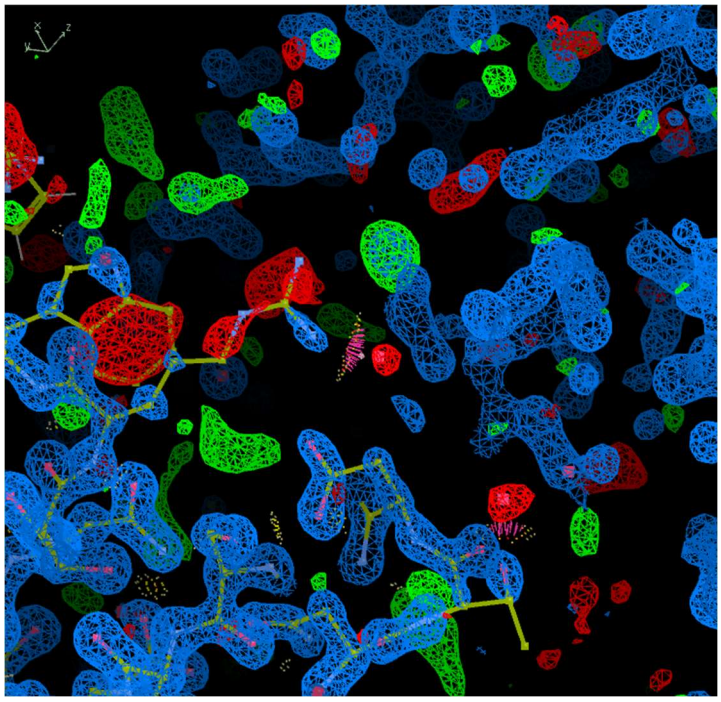
After following along with the tutorial video, you should be able to open the PHENIX refine results in Coot and examine the clashes.



Scroll through the Angstrom resolution on Coot until it says around 1 Angstrom. This will help with accurately determining how the model might need to be tweaked.



The blue regions are the experimental data, the electron cloud of the protein. The red regions are where there is structure taking up space, with no electron cloud to explain it. The green regions are where there is no structure taking up space, with electron cloud density present for some reason.



Our task is to go into Coot and minimize as many red and green regions as possible by manually tweaking the protein structure. After that is finished, subsequent rounds of PHENIX refine and tweaking in Coot will get your protein build closer and closer to the real structure to match it up with the electron density data.

PROTEIN TOPICS TO LEARN:

To understand what is going on in Coot, it is advantageous to have a solid understanding of proteins, their structures, and their amino acid residues.

Topics to understand include:

- The levels of protein structure (such as primary, secondary, tertiary, and quaternary structure, and the differences between those)
- The interactions that occur at each of the protein structural levels (such as hydrogen-bonding, London Dispersion forces, salt bridges/electrostatic interactions, disulfide bridges)
- The amino acid residues and their polypeptide backbone that enables for the interactions to occur (each of the 20 amino acid sidechains and their characteristic properties and their abreviations)
- The overall structure of the polypeptide backbone (N, C-alpha, C-prime) and its directionality (N-terminus vs C-terminus)

Consider looking into these topics so that you can understand more about how you are tweaking the proteins with PHENIX and Coot. This background knowledge will provide you with a richer experience with learning about structure determination.

A program that can assist you with visualizing proteins: PyMOL

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 type into the terminal: **fetch 5keg**

This will fetch you the 5keg protein from the PDB directly, without needing to download anything on your end.

You can type into the terminal: **remove solvent**

This will remove all the little red spots that are water molecules and give you a clearer picture of the protein at hand.

With **show surface**, you can now see the surface of the protein and potential binding pockets on the protein that can be leveraged for recognizing the protein or inhibiting it.

There are a lot of other cool tricks you can use with PyMOL, such as making a video showing the protein spinning, which you can include in presentations, or color-coding the protein to emphasize certain domains of the protein. Here is a website that has 10 key

commands that you can use for PyMol: https://medium.com/@snippetsbio/pymol-10-very-basic-commands-that-you-really-need-to-know-part-1-89eb77b61f7c