This tutorial will provide a general outline on how to optimize a molecular model. Validation will follow in the next tutorial.

The goal of this tutorial will be to optimize a published crystal structure (PDBid: 1YAR) with the 2.8Å Thermoplasma acidophilum 20S proteasome density map (EMD-6287). Ideally, our real-space refinement method of the model will produce a better model (both stereochemistry and fit-to-density), however the crystal structure was solved to 1.9Å, and thus a model fit to our 2.8Å map may have slightly reduced stereochemistry. Moreover, we look to accomplish this within the 1 hour tutorial. The validation tutorial that follows will allow us to determine if an improvement was made.

Before we start, here is some helpful information:

```
phenix.real_space_refine: tool for extensive real-space refinement
of atomic coordinates against provided map

Usage:
  phenix.real_space_refine model.pdb ccp4_formatted_map.map
  or
  phenix.real_space_refine model.pdb map.mtz
  or
  phenix.real_space_refine model.pdb map.mtz
label='2FOFCWT,PH2FOFCWT'
```

or use the following to obtain all parameters.

```
phenix.real_space_refine --h
```

Download / Preprocess the Data

1. Through Chimera (**File > Fetch by ID…**), download the PDB: 1YAR, and the EMDB map: 6287. OR open the following files with Chimera (**File>Open…**), in the model optimization folder (files include 1YAR.pdb and emd_6287.map).
2. Map fitting process before optimization
   ***** The following steps (Step 2 and 3) are not needed during the tutorial since the data is provided. This is a good outline of exactly what I did to prepare the data. *****
   2.1. As one can see the file are not oriented the same due to being derived from different data sets. We will need to first align the data sets before optimizing. To do this open the model panel (**Favorites>Model Panel**), select the model from the list on the left and click the **activate only button** on the right. This allows the user to manually move the model into the density map. Once moved into the map press activate all.

2.2. Open the command line (**Favorites>Command Line**) and type "**fit #0 #1**", where #0 is the model and #1 is the map. The model should now be fit, however this is not the optimal model for the density map.

2.3. The user can then adjust the map in volume viewer to improve visualizing the data. To do this open the volume viewer (**Tools> Volume Data> Volume Viewer**) and change the step box to 1. In addition the user can adjust the threshold.

2.4. At this point we want to select the protein chains that need to be optimized. Ideally, we would like to quickly optimize a portion of the complex with a large amount of flexibility, and quickly. To do this click (**Select>Chains A-N**) followed by (**Select>Invert (Selected Models)**). This will allow us to select everything but Chains A-N. It should be noted that this can be done though the command line.

2.5. Remove the selected atoms by pressing (**Actions>Atoms/Bonds>Delete**).

2.6. Next, we want to remove non-protein atoms in the model (Waters for instance). To do this press (**Select>Residue>all nonstandard**) and remove with (**Actions>Atoms/Bonds>Delete**).

2.7. The resulting model needs to be saved. To do this, press (**File>Save PDB…**). Save the pdb as **1YAR_fit.pdb** and be sure to **save relative to the map** (checkbox below). *VERY IMPORTANT*

3. We now have a multi-subunit complex to operate on. So we open a terminal and navigate to the directory with the map file and the PDB file.

   Once there we will run the following:

   ```
   Optimize hryc$ phenix.real_space_refine 1YAR_fit.pdb emd_5995.map resolution=2.8
   ```

   3.1. and press enter.

   3.2. An **error** should occur. But this is an informative error:

   ```
   Sorry: Crystal symmetry mismatch between different files.
   ```

   This indicates that we will need to change the **CRYST** entry in the pdb to match the new density map.

   3.3. To do this we will open the pdb file 1YAR_fit.pdb in a text editor.

   1.3.1. Remove the CRYST1 line(s) completely.

   1.3.2. In addition, remove all non-atom, and END lines in the PDB file. Ideally, all that you would like to remain are atom lines.

   1.3.3. Save the file.

   3.4. Feel free to open the 1YAR_fit.pdb file in a text editor to see what the resulting PDB should look like.

4. Good to go!

Run Initial Phenix.real_space_refine

At this point we now have a single subunit ready for a quick optimization. At this step we like to see the most amount of movement in the optimization process. We can adjust parameters to provide this movement, however, more movement requires more time. Before running Phenix.real_space_refine we can see what is needed by typing in the following to the terminal:

```
Optimize hryc$ phenix.real_space_refine
```

To see additional parameters, type the following to the command line (this can also be done in the Phenix GUI, however sometimes there are less options when it comes to the model refinement routine):

```
Optimize hryc$ phenix.real_space_refine --h
```

1. Now let's run Phenix.real_space_refine on the single complex (with altered model) using default parameters.

   ```
   Optimize hryc$ phenix.real_space_refine 1YAR_fit.pdb
   emd_6287.map resolution=2.8
   ```

   Various start information should appear including some basic model statistics. Ideally, the model will converge to an improved model. With this complex, and default parameters, the model should converge to the result quickly. This should take ~6 minutes.
2. The result should include the following (and will be written into the same directory):
   a. 1YAR_fit_initial.geo - geometry restraint file
   b. 1YAR_fit_real_space_refined_all_states.pdb - all intermediate models including the converged model. Note - this file may not always be produced depending on refinement parameters.
   c. 1YAR_fit_real_space_refined.pdb - the resulting optimized model

Build Up Complex / Run Complex on Phenix.real_space_refine

We have now quickly optimized half the complex. Its stereochemistry and fit-to-density should have improved. Ideally, we want to take this one step further and optimized the interfaces. To do this we will build up the complex in Chimera.

1. Build up the complex in Chimera using **copy/combine…** in the model panel.
   a. Make a duplicate of the real-space refined model.
   b. Fit the complex into the other half of the map.

   c. Select all subunits in the model panel and again use **copy/combine…** to join all the subunits, creating one combined PDB file with 28 chains.
   d. Save the new combined PDB file as 1YAR_fit_rsr_complex.pdb. Be sure to save only the last combined model and save it relative to the map.
2. Run phenix.real_space_refine on the complex. Back to the terminal and type in the following command:

```
Optimize hryc$ phenix.real_space_refine
1YAR_fit_rsr_complex.pdb emd_6287.map resolution=2.8
run=minimization_global+adp ncs_constraints=True
```

   Here we are specifying the operations that will be run in the refinement (global minimization and compute atomic displacement parameters (ADP or B-factors per-atom)
3. During this run the user should note a few things:
   a. All chains are being optimized to the EM density map.
   b. If one looks at the Start info that is output, the cross correlation (CC) around atoms is very similar to that of the unit cell. This is because we are now modeling the complete complex. Ideally these values will improve throughout the refinement.
   c. We made sure NCS constraints (On as default) was on since we would like to treat all the subunits identically. This reduces time during the refinement. Typically, if the subunits are not identical we turn off NCS constraints (ncs_constraints=False) at this step. It should be noted that Phenix will identify the subunits are symmetrically similar.
4. The output will take slightly longer than that of the initial model.
5. Again, the output will include three files.
6. Typically another refinement is run with ramachandran_restraints=False. Following this COOT is used to adjust regions with poor dihedral angles, improving the Ramachandran plot. This step is then iterated with clashscore, rotamers, and fit-to-density.

Helpful References:

https://www.phenix-online.org/documentation/reference/real_space_refine.html

Campbell MG, Veesler D, Cheng A, Potter CS, Carragher B. 2.8 Angstrom resolution reconstruction of the Thermoplasma acidophilum 20S proteasome using cryo-electron microscopy. eLife (2015) 4, pp. E06380-e06380.

Goddard, T. D.; Huang, C. C.; Meng, E. C.; Pettersen, E. F.; Couch, G. S.; Morris, J. H.; Ferrin, T. E. (2017). "UCSF chimerax: meeting modern challenges in visualization and analysis". Protein Science. 27.

Hryc CF*, Chen D-H*, Afonine PV, Jakana J, et. al. Accurate Model Annotation of a Near-atomic Resolution Cryo-EM Map. Proceedings of the National Academy of Sciences of the United States of America doi:10.1073/pnas.1621152114 (2017). \

PHENIX: a comprehensive Python-based system for macromolecular structure solution. P. D. Adams, P. V. Afonine, G. Bunkóczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L.-W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger and P. H. Zwart. Acta Cryst. D66, 213-221 (2010).

Wang, Zhao, Corey F Hryc, Benjamin Bammes, Pavel V Afonine, Joanita Jakana, Dong-Hua Chen, Xiangan Liu, *and others.* "An Atomic Model of Brome Mosaic Virus Using Direct Electron Detection and Real-space Optimization." *Nature communications* 5 (2014): doi:10.1038/ncomms5808.