

Virtual Classroom: Intro to Model Building & Refinement. December 7 & 10, 2020.

Please complete this pre-tutorial checklist <u>before</u> the tutorial. Some of these steps may be time consuming:

- Download tutorial files (models and maps) here: <u>https://drive.google.com/file/d/1hIDNsOJJuDmQ4Y3pFIaHjQNiYxPUguGP/view?usp=sharing</u>
- □ Install ChimeraX (downloads are here: <u>https://www.cgl.ucsf.edu/chimerax/download.html</u>)
- □ Launch ChimeraX and open a model file (Menu: File>Open then choose any model (.pdb) in the tutorial files you downloaded)
- □ Open a map file in ChimeraX (Menu: FIle>Open then choose the map (.map) in the tutorial files)
- Install Coot
 - □ for Linux download from <u>https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/</u>.
 - □ For Mac download from <u>http://scottlab.ucsc.edu/xtal/wiki/index.php/Installing_Coot_on_OS_X</u>.
 - For windows download from <u>https://bernhardcl.github.io/coot/wincoot-download.html</u>.
 - You can use any version of Coot for this tutorial. I recommend version 0.9. It is significantly updated for lower resolution maps and is the future! Experienced coot users have been known to not like it, so if you must (or are using WinCoot) you can use version 0.8).
- □ Launch Coot and open a model file (Menu: File>Open Coordinates then choose any model (.pdb) in the tutorial files)
- Open a map file in Coot (Menu: File>Open Map then choose the map (.map) in the tutorial files).
- □ Install Phenix (downloads are here: <u>https://www.phenix-online.org/</u>)
- □ Launch Phenix and create a new project that points to a directory where you will save output files. (We won't use Phenix the first day and can address installation problems then).

If you can check all of these boxes, you're ready to start!

This tutorial will take place over two days. On **day one** we will use **ChimeraX** and **Coot** to assemble a starting model from existing publicly available structures and fit it into an experimental CryoEM map.

On **day two** we will use **Phenix** to perform real space refinements to improve both the fit of the model into the map and the chemical bond geometries of the model. We will use Coot to visually assess the output of the refinements and do some minor model adjustments by hand.

The aim of this tutorial is to provide familiarity with available software tools and techniques used for model generation and refinement. We will cover basic features and uses. This will not provide a comprehensive overview or advanced use of any particular software (this would require a much longer course!). We hope this introduction prepares you to work with your own data or attend more advanced tutorials in the future.

Our model system

In this tutorial, we will build a model of the human enzyme, ribonucleotide reductase (RNR). RNR converts ribonucelotides to deoxyribonucelotides and is the source for all de novo DNA building blocks in the cell, making it a key player in nucleotide biosynthesis. All living organisms use an RNR to generate dNTPs and because of this central role in cell division, human RNR is the target of existing anti-cancer drugs such as hydroxyurea and gemcitibine. The importance of this enzyme to human health is clear but high resolution structures of human RNR have been elusive because of difficulties in obtaining stable and homogeneous purified samples.

RNRs in aerobic organisms (including humans) require multiple subunits to perform catalysis. These subunits are named α and β . The architectures of α and β are conserved across species and the catalytic form is an $\alpha_2\beta_2$ dimer of dimers. Regulation of the enzymatic activity is controlled by changes in the oligomerization state of these two proteins with different assemblies reported for different species, including α_2 , α_4 , α_6 , β_2 , $\alpha_2\beta_2$, $\alpha_4\beta_4$, and $\alpha_6\beta_2$ forms. Cartoon representations of some of these states are shown in the figure on the right. The first structure of an α_2 from E. coli was solved by X-ray crystallography in 1994. In 2009 the first human α_2 structure was



reported. The possible assemblies of an α_6 were not obvious from the α_2 structure alone so the 3.3 Å resolution cryoEM structure of human α_6 reported in 2018 was an exciting advancement in the field. This α_6 is an inhibited state, obtained in the presence of dATP & CDP [reference:<u>https://elifesciences.org/articles/31502</u>; EMDB 7006, PDB 6AUI].

Here we will use existing α_2 models to build and refine a model of α_6 into the deposited map.

You can learn more about RNR structures from the PDB Molecule of the Month in Oct. 2019: <u>http://pdb101.rcsb.org/motm/238</u>.

Tutorial outline:

Day one:

- Part 1 Use Uniprot & PDB to become familiar with the sequence and make a plan
- Part 2 Collect model files, evaluate and fit to a map in ChimeraX

Part 3 - Using Coot, stitch together a starting α_2 model from multiple pieces and do some limited manual refinement

- Part 4 Using ChimeraX again, build a full α_6 from three α_2 models
- Part 5 Using Coot again, generate a final α_6 model

Day two:

- Part 6 Use Phenix to manipulate model files and perform a rigid body refinement
- Part 7 Use Coot to evaluate the refinement result and do some manual adjustments
- Part 8 Use Phenix to do a real space minimization refinement

1.1 Use Uniprot to get familiar with the protein

The hexamer is a trimer of α_2 dimers. D3 symmetry was imposed during 3D reconstruction. First, we will get familiar with the protein from existing knowledge.

Uniprot will let us quickly familiarize ourselves with the domain architecture and sequence of α : <u>https://www.uniprot.org/uniprot/P23921</u>

Importantly we can easily find domain boundaries:

Family & Domain	S ⁱ				
Domains and Repeats					
Feature key					Length
Domain ⁱ	1 - 92	ATP-cone	🏦 Add 🔧 BLAST		92
Region Feature key	Position(s)	Description	Actions	Graphical view	Length
Region ⁱ	11 - 17	Allosteric activator binding 🧇 By similarity			7
Region ⁱ	217 - 218	Substrate binding 🕜 By similarity			2
Region ⁱ	285 - 288	Allosteric effector binding, determines substrate specificity			4
Region ⁱ	427 - 431	Substrate binding 🕜 By similarity			5
Region ⁱ	603 - 607	Substrate binding 🕜 By similarity			5

Structureⁱ 00 0 0 0 PDB Entry Method 2WGH 2.30 Å A/B 75-742 **PDBe** X-ray RCSB PDB PDBj PDBsum **3HNC** 2.41 Å A/B 1-792 PDBe X-ray RCSB PDB PDBj PDBsum **3HND** 3.21 Å A/B PDBe X-ray 1-792 RCSB PDB PDBj PDBsum **3HNE** 3.11 Å A/B 1-792 PDBe X-rav RCSB PDB PDBj PDBsum 3.16 Å 1-792 **PDBe 3HNF** X-ray A/B RCSB PDB PDBj PDBsum structure Legend: 🗧 Helix 🚦 Turn 🚪 Beta strand 📗 PDB Structure known for this area

And existing structures that are available in public repositories:

We will compare a few of these to decide which to use as our starting model. If we look up more details in the PDB entries (following the provided links), we find out that two of these were solved in the presence of dATP: 2WGH & 3HNF. 3HND was solved with a substrate, GDP. As we might expect dATP and CDP to be bound in our structure all of these may be useful for a starting model (note GDP and CDP bind to the same active site). We will consider these for building our starting model.

Advanced topics

What do I do if I can't find a structure with identical sequence? What if the exact pieces don't exist?

Making threading models or homology models:

Modeller: <u>https://salilab.org/modeller/</u> SWISS-MODEL: <u>https://swissmodel.expasy.org/</u> I-TASSER: <u>https://zhanglab.ccmb.med.umich.edu/I-TASSER/</u> Chainsaw (part of CCP4): <u>http://www.ccp4.ac.uk/html/chainsaw.html</u>

2.1 We can download the model files (.pdbs) from RCSB, or fetch them directly into ChimeraX from the database. Open ChimeraX and then open these three files, either from a local download (File>Open>coordinates.pdb) or on the command line (**open 2WGH**, etc.). Web-fetched data are stored locally in **~/Downloads/ChimeraX**/

https://www.cgl.ucsf.edu/chimerax/docs/user/commands/open.htm

After opening three models, it will look like a mess:



Each of these models is from a crystal structure of an α_2 dimer, with chains named A and B. [Note: Crystal structures do not always contain a biological unit! You may need to add or delete symmetry related chains. The **sym** command is helpful for this. Look at 1S5H and 6C3I for examples].

Let's spend a couple of minutes orienting in ChimeraX:

How to navigate: https://www.cgl.ucsf.edu/chimerax/docs/user/commands/ui.html#mousedefaults

Action	3-button mouse	Mac track pad	Windows track pad	
left click	Left button	click	click	
Right click Right button		Command-click	alt-click	
Middle click	Middle button	Control-click	Control-alt-click	
Rotate	Left mouse drag	Click drag	Click drag	
Translate	Middle mouse	alt-click	Control-alt-clic	
Zoom	Mouse scroll wheel	Pinch	2 finger drag	
Select	Left mouse + ctrl	Control-click	Control-click	

ChimeraX is great for making pretty pictures. Try different display modes in the Molecule Display Menu until you find something you like. Play with the lighting & effects in the Graphics menu. The Snapshot button in the Home menu will save a .png to your Desktop (or you can save as an image file format from the Save options). Take a pretty image and paste it into the shared doc.

2.2 The three structures we opened are in random orientations relative to each other. Let's align them using the matchmaker command (<u>https://www.cql.ucsf.edu/chimerax/docs/user/commands/matchmaker.html</u>)

mm #1-2 to #3 (The model ID numbers are in the Models panel, see purple arrow below)



Matchmaker does a sequence alignment, finds the best match and then does a structure based alignment on that best matched chain. In the log there is an RMSD value for the alignment. Put the RMSD for the alignment between 3hnf and 3hnd in the shared doc.

Looking at the alignment here, we can see that there is flexibility in the orientations of the two chains already. We can toggle the different models on and off in the model panel to make pairwise comparisons easier.

If we want to align a different chain, we can use selections in our mm command to specify that. **mm #1-2/A to #3/A**

How to make selections: https://www.rbvi.ucsf.edu/chimerax/docs/user/commands/atomspec.html

The basics:

Hierarchical Specifiers

Symbol	Reference Level	Definition	Examples
#	model	model number assigned to the data in ChimeraX (hierarchical, with positive integers separated by dots: N, N.N, N.N.N, <i>etc.</i>)	#1 #1.3
/	chain	chain identifier (case-insensitive unless both upper- and lowercase chain IDs are present)	/A
:	residue	residue number OR residue name (case-insensitive)	:51 :glu
@	atom	atom name (case-insensitive)	@ca

Chain B of 3HNF looks the most complete at the N-terminus so let's use the ATP cone domain (residues 1-92) from 3HNF for our model. 3HND has a substrate bound, so it may be a better model for the active site. Since there isn't a smaller active-site containing domain, we'll use residues 93-end from 3HND and build a "chimera" as our starting model.

2.3 There are many ways to do this "operation". Here we'll use a semi-manual approach to get comfortable manipulating .pdb files as text.

We'll select each piece we want to eventually stitch together and save to a new file in ChimeraX. Using a text editor we'll combine them into one file.

Here is how to save selections:

(https://www.cgl.ucsf.edu/chimerax/docs/user/commands/save.html)

On the ChimeraX command line you can navigate to the directory you want to save them in then run (assuming 3hnf is object #2 and 3hnd is object #3):

select #2/B:1-92, 805 save nterm_3hnf.pdb #2 selectedOnly true select #3/B:93-792, 801-803 save cterm_3hnd.pdb #3 selectedOnly true **2.4** Now we can open these files with your favorite text editor. Let's review the anatomy of a .pdb file.

I		Atom	#									
		1	Ator	n nar	ne							
			1	Doo	idu	o nome				000	DODOV	
				Res	lau	ename				Uccu	pancy	
					Cha	ain ID					B-factor	r (ADP)
						Residue	#			1		Atom type
												1
				1			Real sp	bace x, y,	z location			
		•	•	•	•	•				+	•	
	ATOM	5148	ĊB	ARG	B	80	87.540	-1.079	-8,861	1.00	45.85	ć
	ATOM	5149	CG	ARG	В	80	87.600	-0.568	-10.282	1.00	49.26	č
	ATOM	5150	CD	ARG	В	80	88.810	-1.144	-11.048	1.00	53.05	č
	ATOM	5151	NE	ARG	В	80	88.685	-0.950	-12.507	1.00	56.98	Ň
	ATOM	5152	CZ	ARG	В	80	89.472	-0.174	-13.274	1.00	60.14	C
	ATOM	5153	NH1	ARG	В	80	90.500	0.520	-12.756	1.00	61.48	N
	ATOM	5154	NH2	ARG	В	80	89.233	-0.095	-14.586	1.00	59.86	N
	ATOM	5155	N	ILE	В	81	86.332	-3.017	-6.780	1.00	43.39	N
	ATOM	5156	CA	ILE	В	81	86.039	-4.359	-6.293	1.00	42.86	С
	ATOM	5157	С	ILE	В	81	84.579	-4.463	-5.912	1.00	41.67	C
	ATOM	5158	0	ILE	В	81	83.866	-5.375	-6.355	1.00	41.38	0
	ATOM	5159	СВ	ILE	В	81	86.934	-4.714	-5.074	1.00	42.98	С
	ATOM	5160	CG1	ILE	в	81	88.398	-4.871	-5.539	1.00	43.78	С
	ATOM	5161	CG2	ILE	В	81	86.443	-5.994	-4.403	1.00	42.61	C
	ATOM	5162	CD1	ILE	В	81	89.451	-4.365	-4.548	1.00	43.89	C
	ATOM	5163	Ν	ALA	В	82	84.140	-3.502	-5.108	1.00	40.38	N
	ATOM	5164	CA	ALA	В	82	82.746	-3.422	-4.681	1.00	39.42	С
	ATOM	5165	С	ALA	В	82	81.763	-3.449	-5.869	1.00	38.58	С
	ATOM	5166	0	ALA	В	82	80.710	-4.084	-5.795	1.00	38.15	0

In your favorite text editor, combine the coordinates from nterm_3hnf.pdb and cterm_3hnd.pdb. Let's call that file **combine_3hnf_3hnd.pdb**. [Tips: you can delete everything that isn't atomic coordinates (ie REMARK, SEQRES, CRYST1 lines). Keep all the ATOM, ANISOU, and HETATM lines. The Residue numbers should be in proper sequence and all HETATM lines should go to the bottom. Remove any TER lines from the middle of the sequence. It doesn't matter if the atom # sequence is sensible].

2.5 After saving that file, open it twice in ChimeraX. If you want to make it easier to tell the two models apart in the model viewer, from the directory where you saved it, you can run **open combine_3hnf_3hnd.pdb name combine_copy1 open combine_3hnf_3hnd.pdb name combine_copy2**

We'll make a dimer by aligning one of those copies to chain A of a crystal structure dimer.

mm #5 to #3/A

2.6 Now that we have a dimer model we like, let's open the map.

(From a download from: <u>https://www.emdataresource.org/EMD-7006</u>, or from the provided files, or **open 7006 from emdb** or **open 7006 from emdb name a6map**). Two useful menus of note are the map menu and volume viewer. Play with the different representations to see how they look.



2.7 Next, we will fit our model. We'll start with a manual adjustment with the help of the "Right Mouse" features.

select #4-5 to select multiple models together. Then rotate and translate to fit approximately into the map. Making the map transparent will help.



We can use the fitmap command (or the Menu Tools>Volume Data>Fit in Map) to finish the fit once we're close. <u>https://www.rbvi.ucsf.edu/chimerax/docs/user/commands/fitmap.html</u>. [Note it can help to see where a model fits by making a smoother (or filtered) copy of the map. You can do this with the "Smooth" button in the Map menu or with the command **vop gaussian #6 sdev 2** (or whatever standard deviation value gives a map you like). This will make a new map object leaving your original map untouched].

fitmap #4 in #6 fitmap #5 in #6

If you're not close enough, the fitmap command won't give a good fit. Just keep trying until you get it right. Tip: if you use the Fit in Map box, in the options choose "Use map simulated from atoms". This will give you a correlation value that you can use to judge the fit. Put the correlation value you get from one of your fits into the shared doc.

2.8 Now let's save the two fitted chains relative to this map. We can't combine them into a single file in ChimeraX, but we'll do that in Coot soon.

save chimera_one.pdb #4 rel #6 save chimera_two.pdb #5 rel #6

We're done with ChimeraX for now! You can save your session or be bold and just close it.

Part 3 Now we'll use Coot to make a single file with both chains, we'll inspect the fit and make a few minor adjustments. On a Mac or Linux, I like to open coot from a terminal (running: /usr/local/bin/coot) so that files automatically save in the directory I want them to and I can see what coot is doing in that terminal.

3.1 Open coot and use File>Open Coordinates to open the two files you saved from ChimeraX. Use File>Open map to open the map file. Spend a few minutes orienting with the navigation.

Action	3-button mouse	Mac trackpad	Windows trackpad
Left click	Left button	click	click
Right click	Right button	Command-click	alt-click
Middle click	Middle button	Control-click	Control-alt-click
Rotate	Left mouse drag	Click drag	Click drag
Translate	Control-left mouse drag or middle mouse drag	Control-click drag	Control-click drag
Label	Shift-left mouse click (or double left mouse click)	Shift click or double click	Shift click or double click
Zoom	Right mouse drag	Command-two finger drag	Alt-drag
Clipping	Control-right mouse drag	Control-two finger drag	Control-alt drag
Rotate around Z	Control shift right mouse drag	Control-shift two finger drag	
Center	Middle mouse click		
Change map contour	Scroll wheel	Two finger scroll	

Disclaimer: This table may not be 100% accurate. Here's a good quick reference: https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/web/docs/crib-sheet.pdf

A couple of useful features to use right away are:

Edit>Map parameters to change how much of the map you can see.

Draw>Display manager gives you a window for managing everything you have open.

Try using different display modes in the Display Manager (see the purple arrow to the right) to get a feel for how things look.

	X Coot	0.9 EL (ccp4)		
ile <u>E</u> dit <u>C</u> alculate <u>D</u> raw <u>M</u> easure	es Validate About Ligand			B/BC
🗄 🔍 Reset View 📄 Display Manag	er 🗝 🇞			Map
			Contraction of the	
	000	X Display	Manager	0
10 - 16 - 16 - 16 - 10 - 10 - 10 - 10 -	Maps All			¢.
M. M. Maria	3 emd_7006.map	🗹 Display	Scroll Properties Delete Ma	S - di
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1				5
				. I
12 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1				P2
14	Molecules All Last C	only		K
Cold State	1 chimera_one.pdb	☑ Display ☑ Active	Bonds (Colour by Atom)	elete Model
0 1278 AGIN	2 chimera two ndh		Bonds (Colour by Molecule)	alata Madal
The A GRAD ME	2 [chillera_cwo.pub	C Display C Active	Bonds (Colour by Chain) Bonds (Colour by Sec. Str.)	
	<u>.</u>		C-alphas/Backbone	
a de la companya de la		20	CAs + Ligands CAs+Ligs SecStr Col	
A A LONG	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1		Jones' Rainbow	
	CR wet as		Colour by Atom - No Waters Colour by B-factors - CAs	
10 10 State 14	Ling and a second se	7/6	Colour by B-factors - All	
			Colour by Occupancy	
		S.S.	12 4 13 4 5	
Carl Harry	All Standard Stand	States 1	Carl Carl States States	
CAR A LAND AND A	and the states	1 20 - 200 - 20	2 and the second	
Y Was Mode	A DECEMBER OF BRIDE		1.14	÷
ato New same		10	1000 - 1 1 1 1	Section 2
146 TO 12 (4)	A CAN PARA STATE			STATE OF STATE
A LONG TO A LONG TO A LONG TO A	/3!:2 CYS/B		63 - 10 to	
C. C. P. C. C. C. S. S.				1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
	a le l'and set		Stor Later	100
		2.00)		Co. Constant

3.2 First, we'll combine both chains into one file using:

Edit>Merge molecules (Coot automatically gives them unique chain IDs and reindexes the atoms) Pay attention to which file you are merging into. Save that file with a new name.

File>Save coordinates I called mine a2_chimera.pdb

You can close the extra copy of the merged chain from your Display Manager to keep the space clean.

3.3 We can use coots validation tools to do some analysis.

Validate>Density Fit analysis

We can find regions that don't fit the map well and inspect them. The region around residue 290 looks pretty bad. We can improve it using real space refinement in coot.



3.4 Different refinement modes are useful for different tasks. The two major modes are rigid body and dynamic refinement.

For dynamic refinement guided by map restraints (Real Space Refine Zone), we want to set the refinement weight using the Refine/Regularize control parameters. The weight value helps balance the relative contributions of the map fit vs. geometry during refinement. Click

"Estimate" to give a good starting refinement weight. The estimated value is a good place to start. If you need to adjust that weight, increasing it puts more weight on the map and decreasing it puts more weight on geometry.



Refine/Regularize Control.. Select Map...

🕙 Real Space Refine Zon

 \odot

t

22

55

Regularize Zone

Fixed Atoms..

Rigid Body Fit Zone

Rotate Translate

Dynamic

Rigid body

3.5 Once you get a good fit of that loop into the density, save your work. I like to write new files for each step of manual refinement. I'll call this one **a2_chimera_290loopflip.pdb**.

The full N-terminal domain also has a poor fit to the density, but we will try to fix this using phenix.refine.

4.0 Now we'll go back to chimera and build a full α_6 from our improved α_2 .

Open the map and your saved .pdb dimer. Open the dimer three times. I gave mine new names to keep them organized:

```
>open a2_chimera_290loopflip.pdb name a2_one
>open a2_chimera_290loopflip.pdb name a2_two
>open a2_chimera_290loopflip.pdb name a2_three
```

4.1 Fit the second and third dimer into the map as we did before (using manual rotate/translate and the fitmap tool).

4.2 Save the three files relative to the map (see the command above if you need a reminder). I named mine **a2_hexfit_one.pdb**, **a2_hexfit_one.pdb**, and **a2_hexfit_one.pdb**.

If we look at this model we notice the N-terminal domains of the neighboring dimers clash into each other. We already know the fit to the map is not good in this region, so all of this information is consistent. We will focus on this region at the next session.



5.0 The final step to get ready for phenix is to put all six chains into one pdb file. Coot is a fast way to do this. Open coot, open the three files you saved. Merge them into one file, and save that file (I called mine **a6_start.pdb**). You should now have a file with six chains, labeled A-F. Remember where you saved this, you'll need it for the next session!

Day two - Refinement

You can find full documentation for Phenix here: <u>https://www.phenix-online.org/</u> Phenix tutorial videos are short and helpful: <u>https://www.phenix-online.org/documentation/reference/tutorial_channel.html</u>

Today we will do two rounds of refinement using phenix. First, a rigid body refinement followed by manual inspection and adjustment in Coot, and then a round of global minimization. The refinement runs will take ~10-20 minutes and during the downtime we can look at other useful phenix utilities and answer questions.

6.0 Open Phenix. In the home window make a new project for this tutorial.

Quit Preferences Help Citations	PH	HENIX home
Projects		Favorites
Show group: Unassigned projects	Manage	Data analysis
		Experimental phasing
Select V Delete	ct Settings	Molecular replacement
Betagal	Oct 07 2020 05:52 1	Model building
VLPs	Oct 07 2020 05:30 3	Refinement
2020_NCCAT_tutorial	Nov 24 2020 01:5 0	Cryo-EM
		Validation
		Ligands
		Reflection tools
		Маря
		Model tools
		PDB Deposition
		Other utilities and Program search
Current directory: /Users/czimanyi/Do	cuments/NCCAT_courses/Remote_Mo	odelBuild Browse
PHENIX version 1.18-3845-000		Project: 2020_NCCAT_tutorial

Give it a sensible name and put it in a directory you would like to do your work in. Phenix will automatically save outputs in that directory. I called my project **2020_NCCAT_tutorial**.

	Create project	
Please choose a underscore char only) and projec previously define	simple project ID (alphanumeric and acters t directory. The directory should not overlap with a ed project directory. PHENIX will store settings and files	
Project ID :	2020_NCCAT_tutorial	
Project directory :	/Users/czimanyi/Documents/NCCAT_courses/Remote_ModelBuild	Browse
Sequence file :		Browse
Add to group :	None	
1	Switch to this project	
	Set up tutorial data	
	Cancel	ОК

6.1 The first job we'll run is **PDB Tools** in the **Model Tools** section. We will convert the anisotropic B-factors to isotropic B-factors.

All programs in phenix have a similar workflow. When you choose the program it will give you a new window with two tabs. One tab ("Files") is for input/output information. The second tab ("Options") is for run parameters. Details on what files you need and what the parameter options are can be found in the documentation or you can use the "help" button.

		/	
PDB Tools (Project: 2020 NC	CAT tutorial)	DB Tools (Project: 202	20 NCCAT tutorial)
Reference Run Abort Ask for help		Preferences Help	
Configure	4 b ×	Configure	4 Þ 🗙
Files Options	4 Þ	Files Options	4 Þ
		Atom selections	
Job title : Isotropize model		Modify atom selection :	View/pick ?
- The set	Town I have been	Remove atom selection :	
Visers/czimanyi/Documents/NCCAT courses/Remote ModelBuilding 202	2 PDB Model file		View/pick ?
		Keep only atom selection :	View/pick ?
		Actions	
Add file Remove file		Truncate to poly-Ala Remove alternate conformers	iet segID to chain ID
		Clear segID field Convert SeMet residues to Met	
Output		Coordinates B-factors Occupancies Other modification	ons
File name : a6_isotropic.pdb		All settings	
o Idie	Project: 2020 NCCAT tutorial	O Idie	Project: 2020 NCCAT tutorial

Give your run a useful job title. Input your hexamer model from last session (mine was called a6_start.pdb). Give the output a good file name (I called mine **a6_start_isotropic.pdb**). The default selection is all atoms. That's what we want so you can leave all selection fields blank. Click the "B-factors..." button to find the editing options. Click "Convert to isotropic". Click ok to close that window.

Click on the Run cog to run. A log will appear and the job should finish in a few seconds.

6.2 When it's done, find the output file (it should be in a subdirectory of your project directory called "PDBTools_#) and open it in a text editor to see what happened. It should look like this:

Before	Before After				
🔴 🔿 🕑 a6.pdb		000		a6_isotropic.pdb	
ATOM S647 N MET A 1 222:106:153.0966 163.032 1.00108.61 ATOM S647 N MET A 1 1255 16633 9668 700 -5133 -1331 ATOM S647 N MET A 1 1255 16633 9668 700 -5133 -1331 ATOM S648 CA MET A 1 1255 161372 9513 G11 -0905 -1225 ANTSOU S648 CA MET A 1 1222:556 155.6644 162.032 1.00102.74 ANTSOU S680 MET A 1 1222:556 155.6666 161.0601 1.00102.74 ANTSOU 5850 MET A 1 1227.5737 1919 381 -4682 -1310 ANTSOU 5850 MET A 1 1287.979 155.118 161.2531 1.00104.53 ANTSOU 5851 CB MET A 1 13661 153.65 942 831 -5081<1337	N N C C C C C C C S S C C N N C C C C	ATOM 58 ATOM 5	845 N MET A 1 346 CA MET A 1 346 C MET A 1 348 C MET A 1 348 C MET A 1 348 C MET A 1 350 C MET A 1 351 SD MET A 1 353 N 115 A 2 355 C HIS A 2 355 C HIS A 2 355 C HIS A 2 356 G HIS A 2 357 BUTS A 2 358 360 D15 A 2 368 361 IE1 HIS A 2 366 366 C VAL A 3 365 C VAL A 3 365 C VAL A 3	Practic Practic 222.166 168.896 163.022 222.166 156.964 152.345 222.556 156.946 152.345 222.556 156.946 152.947 222.556 156.946 151.946 222.556 156.946 151.946 224.977 158.118 161.163 218.378 158.129 159.759 218.144 156.433 159.760 222.196 155.379 162.239 222.300 152.854 161.961 222.340 152.854 161.963 223.442 154.954 163.991 223.442 154.224 165.549 224.931 154.242 165.549 223.442 154.2424 165.949 224.951 153.316 166.434 224.621 155.316 166.434 224.621 166.2246 178.142 224.621 166.622 167.718 224.649 164.642 178.142	.00198.02 N .00195.23 C .00192.74 C .00192.74 C .00192.74 C .00194.54 C .00194.56 C .00194.56 S .00191.33 C .00195.02 N .00198.36 C .00195.78 C .00995.41 O .00995.41 C .00995.41 C .00995
ATOM 5858 0 HIS A 2 221.170 152.819 161.901 1.00 95.41 ATOM 5858 0 HIS A 2 21.176 152.819 161.901 1.00 95.41	0	ATOM 58	367 CB VAL A 3 368 CG1 VAL A 3	222.735 150.361 159.202 1	.00 89.81 C
ATOM 5859 CB HIS A 2 224.091 154.035 163.091 1.00 99.19 ANISOU 5859 CB HIS A 2 13985 15256 8447 115 -4068 -925	c	ATOM 58 ATOM 58	869 CG2 VAL A 3 870 N ILE A 4	222.486 151.619 158.372 1 222.845 148.700 162.146 1	L.00 91.11 C
ATOM 5860 CG HIS A 2 223.467 153.501 164.338 1.00 99.41 ANISOU 5860 CG HIS A 2 14315 15172 8284 220 -4071 -732	C	ATOM 58 ATOM 58	871 CA ILE A 4 872 C ILE A 4	223.543 147.606 162.845 1 224.024 146.585 161.833 1	.00 89.89 C
ANISOU 5861 NDI HIS A 2 223.412 154.224 165.509 1.00102.10 ANISOU 5861 NDI HIS A 2 14899 15521 8375 358 -4301 -690	N	ATOM 58	873 0 ILE A 4 874 CB ILE A 4	222.623 146.883 163.864 1	1.00 80.18 0

The output file is our input for real space refinement.

6.3 Open a Real Space Refinement window from the Cryo-EM menu.

Load the map, model, and ligand restraints (dtp.cif, ttp.cif, gdp.cif).

[To do refinement, we need ligand restraints for all ligands in our model file. If you don't already have them, you can generate them with phenix.elbow. We won't go into details on how to make good restraint files, but these are important and care should be taken in making them (my favorite source is asking a trusted friend for theirs). For today, we've provided .cif files with restraints for the three ligands in the model you've made.]

Give the resolution (3.3) and a sensible job title and output name (I called **mine a6_nterm_rb**.

In the parameters tab, for Run options, choose only rigid body (uncheck all others).

Strategy			
Run : minimization_global morphing	rigid_body local_grid_sear	ch	
Max iterations : 100	Macro cycles : 3	-	
Target bonds rmsd : 0.01	Target angles rmsd : 1.0		
Select Atoms 🛛 Use secondar	v structure restraints 🛛 🗹 Use NCS		
Strategy Options			
Morphing : once	3		
Simulated annealing : once	Options		
Reference model restraints	Options		
Other Options			
Scattering table : electron	🗘 Weight :	Resolution factor : 0.25	
Nproc : 1	Random seed : 0		
☑ Rotamer restraints	amachandran restraints		
🗹 Refine ncs operators 🛛 🗸 S	now per residue		
	tic linking All parameters		
Model interpretation Automa	, a parameter sin		

We can also define what rigid bodies to use. We're going to start by moving only the N-terminal ATP cone domain (residues 1-92 and the bound dATP, residue 805).

To do this, click the "select atoms" button (see purple arrow above). In the Rigid Body tab, click the Add Selection button. If you double click on the added cell you can type in selections. Our selection is:

"chain A and resid 1:92 or chain A and resid 805"



Phenix selection syntax is here:

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

You can add as many selections as you'd like and check that they make sense using the "Validate syntax" button. Add a selection for each of the six chains. Click "Save and Close" to finish.

Back in the Refinement Settings window, Set Macro cycles to 1, all else can stay as defaults.

Click run. This took ~10 minutes to run on my MacBook.

6.4 When it's finished you'll get new tabs showing you the location of the output files and validation results for the refined model:



Put the CC(main chain) value for chain D in the shared doc.

7.1 Let's inspect in Coot using the dynamic link (otherwise open a new coot session and load in the files). After coot opens with your refined model, also open the input model so we can see what happened. You should notice a substantial shift of the N-terminus that now has a better fit to the density (try using CA trace mode to simplify the view):

Take some time to explore in coot and see if you find areas that need work. Because the map has D3 symmetry, choose only one chain to work on (changes can be propagated to the other chains with symmetry later). Here are a few ideas to find places that need work:

 Use Validate>Unmodelled blobs to find places where there is map density but no model.



- Use the validation tables in the phenix window to navigate to side chains with outlier rotamer geometry.
- Lys158 and Lys162 aren't fully modeled. Add those side chain atoms in.

Once you've made a few manual adjustments, save your work (I called mine **a6_nterm_manual_coot.pdb**). We'll use this file as the input for a round of full minimization.

8.1 In the phenix real-space refinement window, add the file you saved from Coot as your input for the next round of refinement (and remove the old one if it's still there). Give your job a useful title (ie "first round of minimization - 3 cycles") and change the file name prefix for the output file if you want. In the refinement settings, check minimization_global. Make sure macro cycles is set to 3 (this is just to make it faster today).

Click run.

8.2 If you want to try other options, you can run multiple refinements simultaneously. Let's also try "morphing" and see if it helps. This will take longer to run (about ~30 min. longer on my MacBook Pro) but it won't slow down your other run. Check both minimization_global & morphing, give it a useful job tile and output file name prefix and run it.

8.3 When the refinement is done we will inspect the results together. Put the CC(main chain) for whatever chain you modified most in coot in the shared doc.

This is as far as we will work for this tutorial. There is certainly much more building and refinement one can do!

Things we didn't do: Building from scratch. There are lots of helpful tools in the Calculate menu for de novo building (some of my favorites are "mutate residue range" and the "Other modelling tools" menu with the "Place helix here" and "Place strand here" functions)

Advanced topics: Automated building

Buccaneer (part of CCP4): <u>http://www.ccp4.ac.uk/html/cbuccaneer.html</u> ArpWarp (part of CCP4 - for high resolution maps):

https://arpwarp.embl-hamburg.de/

Phenix map_to_model:

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

Advanced topics: Ligands

In our current model, we actually have the wrong ones!! dATP 805 is correct but TTP (802) and GDP (803) were not included in the sample used to generate this map, we should replace them with dATP and CDP respectively. You can use File>Get Monomer to get common molecules from a standard library. The default is to open them as a new model so don't forget to merge them into your working model before saving your pdb.

Sources of ligand coordinates and restraint files:

Cambridge Crystallographic Data Center: <u>https://www.ccdc.cam.ac.uk/</u> HIC-Up -- Hetero-compound Information Center - Uppsala (Hic-up): <u>xray.bmc.uu.se/hicup/</u> Crystallography Open Database: <u>http://www.crystallography.net/cod/</u> Acedrg program in CCP4:

https://www2.mrc-Imb.cam.ac.uk/groups/murshudov/content/acedrg/acedrg.html