Specifics for assignment due 22 October 2018

You have already obtained access to phenix and Coot. The point of this assignment is to actually get experience using these toolkits (remember that "Coot" stands for "crystallographic object-oriented toolkit") to manipulate a crystal structure.

Begin by loading the three files related to bovine trypsin structures that were given in the assignment itself, namely the coordinate file for structure 2XTT, the sequence for structure 4I8H, and the structure factors for 4I8H. Put all three of those files in a folder (or "directory" if you prefer traditional Unix nomenclature). I encourage you to *look* those three files: they're all text files, so you can read them with any text editor or word processor. The exercise, then, is to rotate and translate the coordinates for structure 2XTT into the data for structure 4I8H, and then refine that structure and manually fit the coordinates so they more precisely align with the electron density while retaining their chemical integrity. While you're looking at things, I suggest you go to the Protein Data Bank (rcsb.org) and look at both of those structures' descriptions on the PDB site. Spend a few minutes looking at the structures' descriptions.

Having loaded the files into your directory, you can now invoke phenix. You should see a top-level menu that looks like this, but without the list of already-completed projects on the left side:

Ouit Preferences	?	Coot PVM		A home Solution her tools Ask for help					
Actions Job histor	Actions Job history								
Projects				Favorites					
Show group: All	groups	Manag	je	phenix.refine Automated refinement using diffraction data (X-ray, neutron,)					
Select 🖉 🧭 De	elete 🚽 New project	🐼 Settings		Data analysis					
ID	Last modified	# of jobs	R-free	Experimental phasing					
3hr2	Jan 27 2017 01:55	4	0.3129	Molecular replacement					
4CO3r	Jun 09 2014 10:45	10	0.3198	Model building					
5d9z	Jun 17 2016 10:58	4	0.3523	Beferement					
4g4m	Apr 14 2015 01:07 PM	8	0.3610	kennement					
RickOrtho	Sep 09 2014 11:23	4	0.3691	Phenix.refine					
1zff	Feb 02 2015 05:22	11	0.3928	Automated rennement using diffraction data (X-ray, neutron,)					
EAOAug	Aug 20 2013 10:12	9	0.4547	Cryo-EM					
1awq	Feb 15 2015 04:30	2	0.4658	Validation					
apbe12	May 24 2018 03:41	8	0.4698	1 too a de					
Leigh11	Nov 22 2015 11:56	16	0.4903	Ligands					
d5hsp	Jul 28 2016 10:58 AM	3	0.4909	Reflection tools					
d5o3uun	Feb 19 2014 02:57	10	0.5313	Mans					
5yce	Sep 23 2018 06:39	7	0.1801	inap3					
1to2	Sep 24 2018 06:52	3		Model tools					
4i8h	Sep 25 2018 03:14	5	0.1547	Other utilities					
4i8hr	Sep 25 2018 05:45	3	0.1622						
bgj55b	Sep 29 2018 07:34	14	0.1484						
1yph	Oct 03 2018 05:57	5	0.1733						
1qnj	Oct 05 2018 06:38	5	0.1406						
1iom	Oct 08 2018 11:57	7	0.1358						
Current directory:	/Users/ahoward/project	s/PDB/2B3Y		Browse					
PHENIX version 1.14-3	260-000			Project: 2b3y					

Click the button that says "New project" near the upper left-hand corner. You should see a menu that looks like this:

	Create project	
Please choose a only) and projec previously defin specific to this p	a simple project ID (alphanumeric and underscore characters at directory. The directory should not overlap with a ed project directory. PHENIX will store settings and files project in a subdirectory named '.phenix'.	
Project ID :		
Project directory :		Browse
Sequence file :		Browse
Add to group :	None	
	 ✓ Switch to this project Set up tutorial data 	
	Cancel	ОК

....

Type in a reasonable project ID, e.g. 4i8hs, and click on "Project directory". Use that browser to find the folder where you put your data: in my example, that would be /Users/myname/trypsin. Click on the folder name in the browser menu and click "Open". You'll return to the "Create project" menu, where you can specify a sequence file by clicking the browser to the right of "Sequence file:". Find the sequence file (4i8h.fasta.txt), click it, and click "Open". Leave the rest of the "Create project" menu alone and click OK. That will return you to the PHENIX home menu.

Click on "Molecular replacement" on the right to show the following:

PHENIX home

() 🔀	? 🧾 🛫	<i> 🇱</i>	<i>s</i> \$
Quit Preferences	Help Citations Coot	PyMOL KiNG	Other tools Ask for help
Actions Job histo	ory		
Projects			Favorites
Show group:	Il groups 🗘 Ma	inage	phenix.refine Automated refinement using diffraction data (X-ray, neutron,)
Select 🖉	Delete 🚽 New project 🐼 Setti	ngs	Data analysis
ID	Last modified # of jobs	R-free	Experimental phasing
5h3v	Jan 05 2018 02:10 7	0.2823	Molecular replacement
rnaseh	Feb 21 2015 03:38 1	0.2833	
BG5Pn5	Apr 20 2017 01:42 1	0.2885	Phaser-MR (full-featured)
3psh	Jan 12 2015 10:03 4	0.2895	Maximum-likelihood molecular replacement with Phaser – use for all but
2gyp	Apr 10 2015 05:08 7	0.2906	
4u7q	Jun 08 2015 10:27 12	0.2981	Phaser-MR (simple one-component interface)
5a1j	Aug 03 2015 04:20 7	0.3043	Automated molecular replacement with Phaser – use only for structures
3psa31	Jan 19 2015 01:25 7	0.3105	with a single type of component
4r4d	Jan 21 2015 06:35 16	0.3106	생설 MRage - automated pipeline
3hr2	Jan 27 2017 01:55 4	0.3129	실실 실 Integrated model identification, preparation, and parallel MR search
1fx1	Oct 22 2017 07:18 16	0.3198	MR-Rosetta
4CO3r	Jun 09 2014 10:45 11	0.3267	AutoMR combined with Rosetta model improvement and AutoBuild for
5d9z	Jun 17 2016 10:58 4	0.3523	/ 🔨 difficult structures
4g4m	Apr 14 2015 01:07 PM 8	0.3610	Sculptor
RickOrtho	Sep 09 2014 11:23 4	0.3691	Modify a molecular replacement search model
1zff	Feb 02 2015 05:22 11	0.3928	Sculptor - Cost interface
EAOAug	Aug 20 2013 10:12 9	0.4547	Extension to Coot GIII for running Sculptor interactively
1awq	Feb 15 2015 04:30 2	0.4658	
apbe12	May 24 2018 03:41 8	0.4698	Ensembler
Leigh11	Nov 22 2015 11:56 16	0.4903	Create ensemble of models for molecular replacement
d5hsp	Jul 28 2016 10:58 AM 3	0.4909	Model building
d5o3uun	Feb 19 2014 02:57 10	0.5313	Refinement
5yce	Sep 23 2018 06:39 7	0.1801	
1to2	Sep 24 2018 06:52 3		phenix.refine
4i8h	Sep 25 2018 03:14 5	0.1547	Automated refinement using diffraction data (X-ray, neutron,)
4i8hr	Sep 25 2018 05:45 3	0.1622	Cryo-EM
bgj55b	Sep 29 2018 07:34 14	0.1484	Validation
1yph	Oct 03 2018 05:57 5	0.1733	
1ani	Oct 05 2018 06:38 5	0.1406	
Current directory	/Users/ahoward/projects/PDB/2E	3Y	Browse
DUENIX version 114	2260,000		Droject: 2b2v

Click on "Phaser-MR (simple one-component interface)", to bring up a menu that looks like this:

					Phaser-M	R (simp	le one-co	omponent i	nterface) (Pr	oject: 2b3y	()		
	\mathbf{X}	2	63	🕥 li	<u> </u>	1	>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>	3	20				
Pr	eferences	Help	Run	Abort S	ave	Xtriage	Sculptor	Ensembler	Advanced GU	I As	k for help		
6	Configure												4 0 >
G	nput files	Search options											4
	Input files a	ind options											
	compositi ensemble Title :	ion either by a and sequence	dding e; for n	a sequen nultiple e	ce file, or e nsembles, i	entering use the	the mole advanced	ecular weigh d Phaser-M	it below. This R GUI.	s interface	only suppor	ts a single search	
	Add fil	le Remov	ve file										
	Unit cell	:							Space g	group :		~	
	Data labe	als :				¢ م	hain type	:	protein	٥			
	Sequence	identities :				м	odel RMS	Ds :					
	Molecular	weight :				C	omponen	t copies :	1	0			
	Either the	RMSDs or seq quired if you d	uence lo not i	identitie	s are requir	red (one	per mod	lel, separat	d by spaces o	or commas), but not bo	oth. Molecular weight	
	is only rea			input a st		с.							

Enter a title, e.g. "phaser for 4I8H using 2xtt's coordinates", and then click on "Add file". You'll see a file browser containing your three input files. Select all three of them and click "Open": you can do this one at a time or all at once. One way or the other, when you're returned to the Phaser-MR menu, you should be able to see all three files, and their "Data type" entries should match what you expect them to be:

2xtt.pdb should be listed as "Search model", 4i8h-sf.cif should be listed as "X-ray data", and 4i8h.fasta.txt should be listed as "Sequence".

You should be able to see the unit cell for 4i8h in the line below the file menus. You can leave the defaults as they are for all but one of the items on the Phaser-MR menu. The one thing you'll need to specify is the "Sequence identities", which you can set to 90, meaning 90% accurate. Then click on the other tab within that menu, which is labeled "Search options". You should a menu that looks like this, apart from the directory names:



You could leave everything on this menu alone, but to save a bit of time I'd recommend pulling down the pulldown labeled as "Also try alternative space group(s):" and click "No alternatives". Then close that tab. The result should look something like this:

				Phaser-M	R (simp	le one-co	omponent i	nterface) (Pr	roject: 4i8	Bhs)		
	0	5	3 👩			1	3	30				
Prefere	ences Help	R	un Abort	Save	Xtriage	Sculptor	Ensembler	Advanced GL	ш	Ask for hel	D	
Confi	iqure											d d x
Innut	files Search	ontions										d b
Input	t files and optio	ns										
At I con ens Titl	least a reflection nposition eithe emble and see e : Phaser	ons file a er by ado quence; f for 418	nd one o ling a seq or multip H using	r more search i uence file, or e le ensembles, 2XTT as mode	models t entering use the el	forming a the mole advanced	i single ens cular weigh I Phaser-M	emble are re 1t below. Thi R GUI.	quired. Y is interfac	'ou must s e only sup	pecify the crystal ports a single search	
File	path								Format		Data type	
Q	/Users/ahowa	rd/proje	cts/PDB/	18HS/2xttr.pd	b				PDB	1	Search model	
Q	/Users/ahowa	rd/proje	cts/PDB/+	418HS/4i8h-sf.	cif				cif		X-ray data	
Q	/Users/ahowa	rd/proje	cts/PDB/	418HS/4i8h.fas	ta.txt				Sequen	ce	Sequence	
	Add file	Remove	file									
Uni	it cell: 54.3	842 58.4	4 66.47	6 90 90 90				Space	group :	P 21 21 2	~	
Dat	ta labels :	0	r4i8hsf,wav	elength_id=1,	¢ C	hain type	:	protein	٥			
Seq	juence identiti	es :	90		м	odel RMS	Ds :					
Mo	lecular weight	:			C	omponen	t copies :	1		0		
Eith is c	ner the RMSDs only required it	or seque f you do	ence iden not input	tities are requi a sequence fil	red (one e.	per mod	el, separate	ed by spaces	or comm	as), but no	ot both. Molecular weig	ht
O Idle									Pr	oject: 4i8hs	1	

Go up to the options at the top of phenix's "Phaser-MR" menu and click "Run". It will give you two choices: "Run now" and "Edit parameters and run". Click "Run now" and let it start. You should see some output in the Phaser-MR box, and after about fiteen minutes (on my laptop, anyway) you should get an output that looks something like this:

Phaser-MR (simple one-component interface) (Project: 4i8hs)	
Preferences Help Run Abort Save Xtriage Sculptor Ensembler Advanced GUI Ask for help	
Configure phaser_1	4 ▷ ×
Run status Graphs Results	4 Þ
Summary	
Phaser has found 1 MR solution(s). Image: Constraint of the solution of the solu	
Output files Title: Phaser for 4l8H using 2XTT as model Directory: /Users/ahoward/projects/PD8/4l8H5/phaser_1	
File name Contents	
A 418hs_phaser.1.mtz Phases and maps	
4ions_pnaser.1.pop Output model	_
The second secon	
ldle Project: 418hs	

Do look at the numbers near the upper left-hand corner. The log-likelihood gain (LLG) should be above 5000, and the TFZ should be above 40. If it is, you're ready to move on.

Close the Phaser window and look at your main Phenix menu. Partway down you'll see a category called "Refinement". Click that and click on "phenix.refine". You'll see a menu that includes an Input files menu. Populate that file menu with three files: your sequence file, which should already be there; your structure-amplitude file, which is the file 4i8h-sf.cif, and the output coordinate file from your phaser run, which you can find by clicking on the folder icon for the phaser run (probably phaser_1), and finding the output PDB file, which will have a name ending in something like 4i8hs_phaser_1.pdb. Leave the other items in their default form, except the Wavelength, which for this experiment is 0.5904. After you've put those names and that value in, your Input data tab should look like this:

	phenix.refine (F	Project: 4i8hs)								
💥 🦻 👘 🖾	🎶 🐮 🤰		Ch							
Preferences Help Run Abort Save	Graphics ReadySet TLS Res	traints Xtriage	Ask for help							
Configure				4 Þ ×						
Input data Refinement settings Output				4 Þ						
lob title : zeroth refinement										
Input files										
File path		Format	Data type							
Users/ahoward/projects/PDB/4I8HS/4i8h.fasta	txt	Sequence	Sequences							
/Users/ahoward/projects/PDB/418HS/418h-sf.c	f	cif	X-ray data, X-ray R-free							
V /Users/ahoward/projects/PDB/418HS/phaser_1	/4i8hs_phaser.1.pdb	PDB	Input model							
Add file Remove file Modify file data typ	Use symmetry from sele	ected file								
Space group : P 21 21 21	all 54.342 58.44 66.47	6 90 90 90								
	541542 50144 00147									
X-ray data and experimental phases										
Data lakala	D free label	fum ninn stir id 1 0 7	Test Regulation .							
Data labels : r#ionsi,wavelength_id=1, v	K-Tree label :	,wavelength_id=1, V	rest hag value : 0							
High resolution : (0.752)	Low resolution :	(28.892) F	Phase labels :	٥						
Wavelength 0,5904	Ontions									
harcienger .	optionality									
Neutron data										
Data labels :	D free label :		Test flag value :							
Data labels :	K-free label :	~	Test hag value :							
High resolution :	Low resolution :		Options							
Idle			Project: 4i8hs							
- IIII			,							

Now click on the second tab, labeled "Refinement settings". Turn on "XYZ (reciprocal-space)" and "XYZ (real-space)", "Occupancies", and "Individual B-factors" if they're not already turned on. Then click "Select Atoms", bringing up a new menu. You should then click on the "Refinement Strategy" tab. At the bottom of that menu is a question that says "Common configurations:" Pull that pulldown down and click on "All atoms except water anisotropic":



Click "Apply", and the screen will look like



Then click "Save and Close" in the upper lefthand corner, and you'll be returned to the Options menu. You'll make a few changes on that Options menu, and then you'll be ready to run phenix.refine. Specifically, under "Targets and weighting", shift the "Target function" to ML, turn on "Optimize X-ray/stereochemistry weight" and "Optimize X-ray/ADP weight" if they're not already on. Click "Refinement target weights". You'll see a menu with some choices. Take the defaults, except change the "Stereochemistry weight scale (wc)" to 1.2. Click OK on that menu so it can disappear. In the "Other options" section, turn on "Automatically add hydrogens to model", "Automatically correct N/Q/H errors", and "Update waters" if they're not already on. Click the "Number of processors" according to the number of separate CPUs in your computer; for many of you that number will be 4. The final result should look something like

phenix.terine (Project. 41015)	
X Preferences Help Dun Abort Sawe Graphics Predivide statistics Same statistics Same statistics Viriane Auf for help	
	4 Þ ×
Input data Refinement settings Output	4 1
Strategy	
Refinement strategy: ✓ XYZ (reciprocal-space) ✓ XYZ (real-space) Rigid body ✓ Individual B-factors Group B-factors TLS parameters ✓ Occupancies Anomalous groups Number of cycles: 3 O	
Select Atoms Note: selections can only be made for enabled options (e.g. NCS groups are available if "Use NCS" box is checked)	
Targets and weighting	
Target function : ML Image: Optimize X-ray/stereochemistry weight Image: Optimize X-ray/ADP weight	
Use NCS NCS type: torsion-angle	
Reference model restraints Use secondary structure restraints Use experimental phase restraints	
Refinement target weights Model interpretation NCS options ?	
Other options	
✓ Automatically add hydrogens to model ✓ Update waters Place elemental ions :	
Simulated annealing (Cartesian)	
Automatically correct N/Q/H errors Twin law : Number of processors : 4 2 ?	
Global refinement parameters Modify start model All parameters ?	
Idle Project: 418hs	

If that's roughly what you see, click "Run" and "Run now" as you did for the Phaser run.

Your phenix refine window will obtain some new tabs. Initially you'll see the "Run status" tab, but you can click back and forth between that tab and the "Log output" tab during the refinement run. Phaser will spawn a Coot window so you can see what your structure looks like while you're refining it. With an ultra-high-resolution data set like this it will take about an hour to complete the refinement, so find something productive to do while it's running. At a step partway through, the "Run status" tab will look like



There will be other tabs that appear as your run of phenix refine ends, including a MolProbity analysis, which will look like

						phenix	refine (Project: 4i	8hs)				
×	2 8	63 🖍	نا (2/2	*	s 🔊		**			
Preferences	Help	Run Abi	ort Save		raphics I	ReadySet 1	LS Restraints	Xtriage	Ask for help			
Configure	Refine 2										4 4	×
Log output	Run status R	esults N	lolProbity	Real-sp	ace correla	ation Aton	ic properties Hvg	irogens Sei	guence check		4	Þ
Summary	Basic geometry	Protein	Clashes								4	Þ
Validation s	ummary											
PRI Basic sta		These st represer resolution	atistics a its the ex on. 2.pdb :	re compu periment	ted using al resolut	the same t tion expecte	underlying distribu ed for a model of t	tions as the his quality; i	MolProbity web server. ideally the score should	The overall score be lower than the actual		
Rama	chandran out	iers:	0.00%	(Goal : -	c 0.2%)	Ramacha	indran favored:	99.10%	(Goal : > 98%)			
Rotan	ner outliers:		0.00%	(Goal :	196)	C-beta d	utliers:	0	(Goal : 0)			
Clash	score:		2.88			Overall	core:	1.08	(
Missing ato	ow validatio	on in Kil	NG									
Chain		Resid	ine		altic	oc.	Missing ator	18				
А		ILE 5	5				CD1	-				
For clarity, hydrogen atoms are not included.												

You're finished running phenix refine and can move on to examining your structure with Coot.

If your Coot window from within phenix.refine is up and running, you can use it; if not, you'll need to invoke Coot specifically and load up the PDB file (Refine_2/4i8hs_refine_2.pdb, or a similar name) from *File* \rightarrow *Open coordinates*, and the mtz file (Refine_2/4i8hs_refine_2.mtz or similar) from *File* \rightarrow *Auto Open MTZ*. You should see both the model and the map superimposed on parts of the model. I won't try to walk you through every detail of manual fitting, but I'll point out a few things. Before getting into the mechanics of fitting, start your Coot run by invoking symmetry: Draw \rightarrow *Cell&Symmetry*: say "Yes" to the *Master Switch* at the top, and click Apply and OK. Also, get the program to show you close contacts by clicking *Measures* \rightarrow *Environment Distances*: click on *Show Residue Environment* and click OK. Finally, click *Calculate* \rightarrow *Model/Fit/Refine* and, on the resulting menu, click *Select Map*, and click OK. Once you've done that, you may proceed.

- 1. Go to Validate → Difference Map Peaks, accept the defaults, and click Find Peaks. You'll get a list of about 240 peaks, sorted by the absolute value of the difference between the observed density and the calculated density. Probably the topmost peak will be near the carbonyl oxygen of Asn 54, the carbonyl oxygen of Val 57, the OE2 of Glu 62, the OE1 of Glu 52, and two water molecules. This is actually the location of a calcium ion in this structure. So on the Model/Fit /Refine menu, click Place Atom At Pointer, pull down the pulldown to replace New Molecule with the name of your current PDB file, i.e. 25 ...HS/Refine_2/4i8hs_refine_2.pdb or something similar. Then click Ca (calcium) and click OK. That should place the calcium ion right where it needs to be.
- 2. The next highest peak will probably be on methionine 86, where the side-chain has two possible conformations, and since the sulfur scatters so effectively, it has a big difference peak. If you know how to add alternate conformations, do so; if you don't, skip this one.
- 3. The next highest peak will probably be an odd-shaped blob near the NE2 of His 40 and somewhat near the N of gly 175 and the Cα of Gln 174. This is actually a sulfate ion, so click *Place Atom At Pointer*, *SO4*, and *OK*. This will place the sulfate within the density but misorient it, so go up to the top of the Model/Fit/Refine menu, click Real Space Refine Zone, and click two atoms within your sulfate. Then click "Regularize Zone" and do the same thing. You may have to move it around a little using *Rotate / Translate Zone* to get it right.
- 4. The next several positive peaks will probably be water molecules. To add these, click *Place Atom At Pointer*, leave the identity as *Water*, and click OK: then move on to the next one.
- 5. Somewhere around peak 8 you will find a zone that looks like a benzene ring with three atoms

attached to one of the carbons. That's actually a benzamidine inhibitor that the experimenters inserted into the enzyme to keep it stable.

Click Extensions \rightarrow Modeling \rightarrow Add Other Solvent Molecules, and click Add a new Residue Type...

A dialog box will appear that says *Add new 3-letter-code/comp-id*; answer that with BEN which is the three-letter code for benzamidine. Then go back to the Solvent Ligands list, and you'll see benzamidine at the bottom of the list. Click on it, and it will magically appear in the structure. It will probably even be correctly oriented, but you may need to *Real Space Refine* it a little.

- 6. There are some amino acid residues within the protein whose side chains are seriously misaligned. Among them are Asn 56, Asn 61, Lys 139, Tyr 22, and Pro 74. Go to $Draw \rightarrow Go$ to *Atom* and click on the CB of each of these residues. You will see it in the center of your screen, and you can correct its orientation. In most cases you can do this with the *Rotamers*... button on the Model/Fit/Refine menu: click *Rotamers* and then step through the possible rotamers until you find one that matches the density. Click that one and then clean it up with *Real Space Refine Zone*.
- 7. If you look closely at the output of your MolProbity analysis, you'll see that it indicates that an atom is missing from Ile 55. Go to that residue and you'll see that, indeed, CD1 is missing. The easy way to put it in is to go to the *Model/Fit/Refine* menu and click *Mutate & Auto Fit*. A menu with the amino acids will appear; click on ILE, and the program will properly place the isoleucine.
- 8. Somewhere around peak 14 you will find another sulfate, in this case near the surface, not far from the amine nitrogen of gly 154. Do the same thing as you did with the first sulfate.

Once you've played around with improving the map one residue at a time like this, try a bulk approach: Click *Extensions* \rightarrow *All Molecule* \rightarrow *Stepped Refine*. A dialog box with a pdb filename will appear. Click OK. The program will step through the entire sequence, fitting each residue as it goes. In many places, nothing will happen; in a few, something will happen.

When you've done enough manual and automated improvements, save your work: $File \rightarrow Save$ Coordinates

and then respond to the dialog box when it asks for a filename.

What I will ask you to submit are:

- 1. The logfile from the phaser run.
- 2. The .txt output file from the phenix refine run.
- 3. The .pdb output file from the phenix.refine run.
- 4. The .pdb output file from the Coot effort.

As always, I'm happy to walk you through some of these steps.