MCC 2024 Protein Computational Learning Group

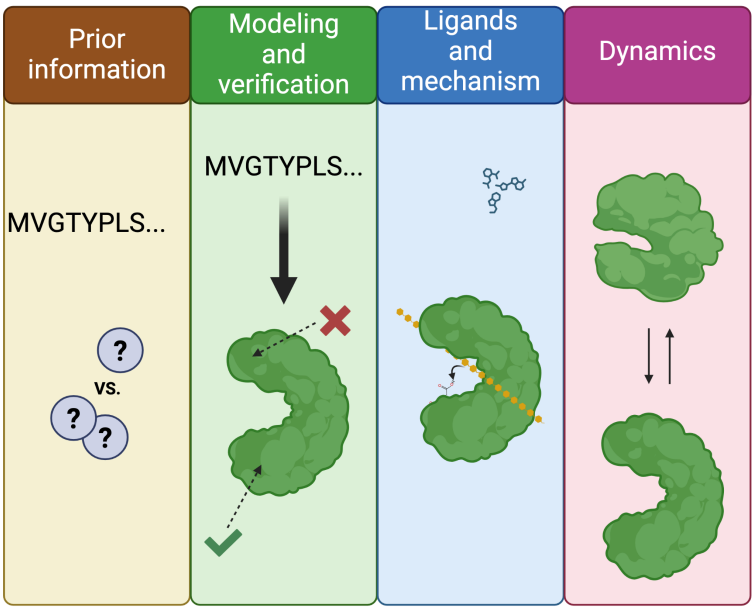
Table of Contents

Coming Soon!: DynaMut, CABS-Flex

Introduction and Motivation

The application of and access to computational techniques for predicting and understanding protein/biomolecule structure and function have increased significantly over the last ten years. Many of the softwares are free and some are “cloud-based” which has reduced access barriers significantly. However, there many hurdles to incorporating these techniques into the classroom, including knowing how to use them effectively, remain. Access has increased, but not accessibility.

This group aims to work as a community to increase accessibility to computational biochemistry by sharing knowledge and resources. Structured as a series of asynchronous, semi-guided assignments, participants can move at their own pace. Videos will be provided to give some guidance on using the software and there will be open-ended questions to provide direction. ***Asynchronous does not mean alone!*** There will be judgment-free office hours weekly (or so) but questions can always be asked in the Slack channel or via email (berndsce@jmu.edu).



Objectives

- Reduce using the phrase “I’m not a computer person”
- Use research-grade modeling and prediction software to describe protein structure and function
- Develop a working knowledge of how these software can be used to generate hypotheses or theoretical results to address questions of protein structure and function
- Outline some classroom or CURE activities for teaching students how to use the software

Links, dates, and times in 2024

General topics are provided for each meeting, however we are all busy! It is all about learning and developing skills which happen on an individual basis. Come, chat, and learn!

Zoom meeting link for office hours:

Hours on Tasks 1 and 2: July 9th, 12 pm to 1:30 pm ET (9 am to 10:30 am PT)

Hours on Tasks 3 and 4: July 23rd,

Hours on Tasks 5 and 6: July 30th(?)

Hours on Tasks 7 and 8: August 6th (?)

General documents will be added below on this page as needed.

Target protein: Arabidopsis AMY3

Task 1: Using Uniprot to determine what we "know" about AMY3

A common phrase in computational work is "garbage in, garbage out." A model is only is good as what was provided during assembly, therefore some prior information is needed. Before making a good model, it is important to know **a)** what is known already about the structure, such as partial structures or ones of closely related homologs to help verify the potential accuracy of the structure, and **b)** information on the primary and quaternary structure, which may influence the input parameters for the program.

In this task, you will explore two databases, Uniprot and the AlphaFold database, to learn what is known about AMY3 before modeling its structure in Task 2.

WHAT YOU GET: Information on your protein and identify potential structures

1. Go to Uniprot and search for AMY3 from *Arabidopsis thaliana*. Alternatively, use the Uniprot ID Q94A41
2. Go into the entry by clicking on the entry code Q94A41
3. Explore the page sections, guided by the questions below
    a. What is/are the function(s) of this protein?
    b. What cofactors are needed by this protein?
    c. Are the active site and substrate binding sites in the N or C-terminus of the protein?
    d. What is the subcellular location of this protein?
    e. What citations support the subcellular location? (Hint: look in the upper right corner of the subcellular location box)
    f. How is this protein processed in the cell? Specifically, are any protein regions removed in the cell?
    g. Are there any experimental structures of the protein?
    h. Is there any information on whether AMY3 forms dimers or higher order structures with itself?
4. Under Structure>3D structure databases, select the AlphaFoldDB link for Q94A41.
5. Explore AFDB page for AMY3 guided by the questions below
    a. Using the structure viewer, how many separate, folded domains are formed by the AMY3 sequence? (You can make the window within the page larger by pressing the wrench icon.)
    b. Is the AMY3 structure modeled with high confidence or mixed confidence?
    c. Compare the sequence in the structure to the processed sequence from 3f above. Is the AFDB structure showing the cell relevant form? Does it matter to the AMY3 structure?
    d. Using the PAE plot, left-click and drag on the dark green square regions to highlight those regions on the structure. The dark green regions are areas of the protein with apparently highest confidence in the folding. Do those regions correspond to folded regions or unfolded regions? (There is an explanation of the PAE plots at the bottom of the page for further information.)

What to have to move on: The cellularly relevant sequence and any information quaternary structure

Task 2: Model AMY3 using the AlphaFold3 server

Protein modeling used to be a dark art based on overlaying sequences on known protein structures and hoping that the sequence differences did not matter. However, the last 5 years have seen an explosive growth in the speed and accuracy of modeling biomolecules. One of the most exciting approaches is that of AlphaFold, which uses deep sequence alignment, coevolution of proximal side chains, and diffusion to predict protein structure and protein complex structure. In 2024, the AlphaFold3 server was launched allowing academic researchers access to one of the more accurate methods for structure prediction. As of June 2024, the AF3 server can predict protein, DNA, and RNA structures, complexes of those molecules and proteins bound to some ligands. The server takes only a few minutes to model large complexes and produces 5 models per run, allowing for comparison of the prediction consistency.

WHAT YOU GET: A model of your protein, complex, protein+ligand, or post-translationally modified protein

General AF3 use video

1. Navigate to the AlphaFold3 server
2. Make an account (if needed)
3. Clear the entry options
4. Paste in the relevant sequence of AMY3 based on 3f from Task 1. (Hint: AMY3 gets processed to remove an N-terminal sequence, so remove the necessary amino acids from the sequence)
5. Set the molecule type to Protein and copies should equal the oligomeric state identified from Uniprot in Task 1.
6. If there were any cofactors identified, add an entity and select ligand for any small molecules (ex. NAD) or ion (ex. Ca2+ or Na+).
    a. For small molecules, set copies to match the number of copies of the protein
    b. For ions, set copies to the number needed if known, other pick a reasonable number. Extra copies of the ion will be placed outside of the protein, so set the number of copies on the high side of possibilities.
7. Press Continue and preview job
8. Set the job name
9. Turn the seed option on (button on right), then select a random number.
10. Run the job.
11. The input sequence and settings will be on the page, repeat steps 7-10 with a new seed number.
12. The jobs list is shown at the bottom. Wait until both jobs finish (15 minutes or less).
13. View both jobs taking note of the pTM and ipTM scores. We will discuss these more in Task 4.

What to have to move on: Sequence of protein, downloaded data from both AF3 jobs, and

model quality statistics.

Task 3: Identify conserved areas of AMY3

Sequence conservation can identify structurally and functionally important regions. Overlaying conservation on the structure can further show important regions that are distant in primary structure but close in tertiary structure or 3-D space. ConSurf does this exact search. Given a structure or model of a protein, the program searches for similar sequences and maps conservation onto the PDB file. This information can then be used to help refine binding pockets or identify binding interfaces.

One caveat is if you have several closely related proteins within an organism with similar structure and reaction chemistry, but maybe different localization or expression (think about the 6 to 8 MDH enzymes or 7+ beta amylases found in plants). These proteins are likely to be included within the sequence alignment together so the results may reflect the family of enzymes, not a single member. This issue can be reduced by providing a sequence alignment specific to the ortholog or homolog of interest (e.g. only mitochondrial MDH enzymes). Finding homologs can be accomplished by using NCBI BLAST to find potential sequences and then aligning the results and trimming the sequences to include those of interest.

WHAT YOU GET: 3-D location of the conserved regions of your protein to support identification of ligand binding sites or protein-protein interfaces

1. Navigate to the ConSurf server
2. Upload your model of AMY3 to the server.
    a. You will likely need to convert the .cif file to a PDB by loading it into pyMol (or similar software) and saving the file as a PDB.
    b. This a good site: https://project-gemmi.github.io/wasm/convert/cif2pdb.html
3. Provide your email and a job name in the appropriate box
4. Choose to Submit with the Default Parameters
    a. Manual selection allows you to change the sequence alignment parameters
5. Wait for the email indicating that the job is complete. This can take anywhere from 20 minutes to 24 hours.
    a. Sometimes the email goes to the Junk folder
6. Explore the conservation of AMY3, guided by the questions below:
    a. Which end of the protein is more conserved?
    b. Is there more conservation within the folded domains or the unstructured regions?
    c. Are there pockets or surfaces which are highly conserved which may be interaction sites?
7. Download the results
8. The PDB files included in the results have the ConSurf score in the B-factor column of the PDB file. If you color the protein by b-factor or disorder in pyMol or mol* or whatever viewer you use, you can see the scores.
    a. You may need to rescale the color range as it runs from 0-10, not 0-100
    b. Cartoon or surface representations tend to work best
9. There is a CSV file which contains the numbers and conservation for plotting

Task 4: PAE viewer to analyze AlphaFold3 results

AlphaFold3 is a powerful tool to predict protein structure and complexes however the analysis of these results can be difficult. This is in part because the outputs of AF3 are presented with limited details and only one of the 5 data sets are shown. The PAE viewer is software that provides interactive ways to identify the high and low confidence areas of the structure. This is especially powerful for biomolecule complexes. We will use it here with just a single chain, but you may want to repeat the AF3 modeling and set the number of AMY3 copies to 2 to see how AF3 assesses the complex between the two subunits (experimental structure paper coming soon 🙂).

WHAT YOU GET: Information on which parts of the AF3 model are reliable and are appropriate for hypothesis generation

AF3 stats explanation page

Predicted aligned error (PAE) viewer explanation video

0. Download your AMY3 results from AlphaFold3 and unzip the file
1. Navigate to PAE viewer and select the Upload tab
2. Choose the model_0.cif file.
    a. This is the best model from AF3
3. In the Scores file part, upload the full_data_0 json file.
    a. There are three types of .json in the folder, a job_request, a summary_confidences, and a full_data. This latter file type typically is a much larger file than the other two
4. Press Upload
5. The analysis section has three main parts:
    a. The sequence bar where you can highlight specific sequences to locate them in the structure
    b. The structure window which is interactive
    c. The PAE window which you can use to highlight high or low confidence areas to locate them on the structure and sequence
6. Explore the data guided by the questions below:
    a. There are three areas of low error in the PAE window as indicated by the dark green color. What do these correspond to in the structure?
    b. Interdomain contacts that were proposed confidently would be depicted as dark green areas in the lower left or upper right quadrant away from the diagonal line. Is AF3 confident of any of the interdomain contacts?
    c. Are there low confidence regions within folded domains?

    d. How might you use the model confidence information going forward into docking or functional studies?
    e. Thinking back to the ConSurf task, how do the PAE plots compare to the highly conserved areas from ConSurf?
7. Optional: Remodel an AMY3 dimer in AF3 and analyse the results using PAE viewer. Answer the same questions and take note of any changes between the monomer and dimer models.
8. Open another instance of the PAE viewer in a separated tab and upload model_4 with the full_data_4 json file.
    a. Are the structure and PAE plots for model_0 and model_4 the same?
    b. What would be your interpretation if they were different?

Task 5: AlphaFill for identifying ligands and binding sites

Structure is often the first step to identifying the function and mechaism of action for a protein. One of the challenges in dsecribing function is identifying potential ligand binding sites. AlphaFill uses structures from the AlphaFold Database or a user provided model to identify all the potential ligands, model ligand binding, and if desired generate a docked, energy minimized model of the target protein bound to hte molecule. AlphaFill relies on data in the RCSB Protein Data Bank (PDB) and so contains relevent molecules and those which were necessary for crystallization. Thus these ligand bound models should be treated like a ligand bound crystal structure, good for a hypothesis or generating ideas while considered with some skepticism.

WHAT YOU GET: A ligand bound model of the target protein based on homology to experimental structures

AlphaFill explanation video

1. Navigate to the AlphaFill homepage
2. Enter either the Uniprot code for Amy3 (Q94A41) or upload the AF3 model.
    a. AlphaFill does not like headers in the structure files. If the file is rejected, open the file as a text file in NotePad or TextEdit, remove the header/remark lines up to the coordinates part, save and upload. This fixes 99% of issues with the file
3. Wait until the program finishes (typically less than 5 minutes).
4. The resulting page will show:
    a. A viewing window which is interactive and can shown potential non-covalent interactions
    b. Table of potential ligands. The codes are the ones used by the PDB and you can type these into the PDB website to get more information
        i. The % identity indicates how similar the binding site in your protein is to the one from the PDB. Higher % means a more identical binding site.
        ii. You can uncheck all the ligands at once by clicking on the check mark at the top of the table
        iii. Checking the box next to any ligand will show that ligand “docked” to the target protein.
        iv. Pressing the optimise link will energy minimize that ligand bound to the protein and produce a .cif file that can be used for MD and other studies
5. Based on the proposed function of AMY3 from Uniprot, work through the table and determine if any of the ligands are relevant. Optimize those structures
6. For multidomain proteins, sometimes the results are better or easier to process by modeling each domain individually and performing AlphaFill analysis on the individual domain. Additionally, the alignment search may be biased by domain with high sequence or structural homology to well-characterized proteins.
7. In AF3, model the carbohydrate binding module (CBM) in the AMY3 N-terminus and repeat AlphaFill analysis on just this domain. The CBM is approximately amino acids 96-387.

8. Model the AMY3 amylase domain in AF3 and repeat AlphaFill analysis on just this domain. The amylase domain is approximately amino acid 484 to the end of the sequence (837).
9. Compare the results of steps 5, 7, and 8.
    a. Were the results improved or different between the different modeling approaches?
10. Consider the binding site(s) identified by AlphaFill:
    a. Is the binding site in a conserved region of the protein? How does this affect your confidence?
    b. Is the binding site in a high model confidence region? How does this affect your interpretation of the results?

Task 6: Ligand docking with SeamDock and DiffDock

The ligand docked models from Task 5 are limited by the known structures in the PDB both in ligand diversity and position. Blind and targeted docking can fill in the gap by allowing the user to specify the ligand and the potential binding site. SeamDock is built on the "classical" AutoDock Vina approach of specifying a binding site within a cube. The cube is divided into a series of smaller cubes and the ligand is placed within these sub-cubes to determine the best position. DiffDock uses a diffusion model to "walk" the ligand around the protein until a reasonable position is found. This best repeated many times and the results clustered. A binding site can be supported if a majority of the runs reproduce a consistent site. A common workflow is to use DiffDock to identify potential sites, and then use SeamDock to compare potential sites and refine the positioning of the molecule.

DiffDock reproduces AutoDock Vina binding sites with reasonable accuracy despite lacking a target site and may be less prone to bias than AutoDock Vina. Both methods are affected by poor models or, in the case of flexible proteins or proteins in complexes, a model which is not in the correct conformation to bind. When in doubt or if issues, try additional AF3 models.

WHAT YOU GET: The structure of a protein-ligand complex and, from SeamDock, relative deltaG values which can be used to compare relative ligand affinities.

SeamDock video

NOTE: Seam dock is a bit of a pain with large proteins, so you might use only the amylase domain or the CBM of AMY3.

0. Create a PDB of just the ligand OR use PubChem to find the SMILES of the target ligand. You can do this by importing the complex into pyMol or similar and deleting the protein, then saving the ligand alone.
1. Navigate to SeamDock and access the service through the portal
2. Upload the PDB of the ligand and make sure that it appears correct in the viewer window
    a. Use only the canonical SMILES not the isomeric SMILES.
    b. SDF inputs are also possible but seem to fail at a higher rate
3. Upload the receptor.
    a. You must do the ligand and receptor in the order that they appear on the screen
4. Adjust the box position and size until it encompasses the entire binding site.
    a. Recommend that you be generous with the box size initially
5. Under Docking parameters, leave the defaults unless you want to compare outcomes. They should be similar
    a. Mode number is the number of output positions
6. When ready, push launch docking and wait. Takes ~5 minutes for 4 docks
7. In the results window, only one dock is shown at a time. You can change the dock number by clicking in the table on the left.
    a. Weak interactions are listed below the window

    b. The affinity values are relative affinity and can be used to compare between runs and molecules, but are not absolute values which can be converted to K_D values.
8. All of the results combined in a single structure can be downloaded below the weak interactions table by saving the dock.pdb file.
    a. Depending on your viewing software, they may be listed as “States” or individual structures.
    b. In Pymol, they can be saved independently for further analysis by splitting the states. To split the states in the code line type split_states dock, prefix = conf
    c. This code splits the dock object into individual objects with the prefix of conf followed by a number. Adjust the dock name and prefix to suit personal needs.
9. Save the individual positions for MD

DiffDock

Diffdock guide video

0. As with SeamDock, using a domain structure is probably better.
1. Navigate to Diffdock web
2. Upload the target protein
3. Copy in the canonical SMILES for the ligand
4. Adjust the Samples per Complex as needed.
5. Scroll down and press Run DiffDock
    a. Wait 2-60 minutes to complete. There is sometimes an estimated time shown.
    b. A molecule will appear in the output window when done. If nothing appears, it typically means the docking failed.
6. When complete, download the .zip file and decompress the .zip
7. Within the results folder there is another folder which contains the structures.
    a. Open the PDB file in a viewer software
    b. Then drag or open the SDF files (all if possible) in the viewer.
    c. This will show the clustering the results and how consistent the dock is
    d. A cluster of many structures suggests a good binding site, scattered or diffuse positions suggests no binding sites were found OR the molecule cannot access the binding site.

Task 7: Cloud-based molecular dynamics

Molecular Dynamics or MD used to require significant infrastructure and technical knowledge to perform the most fundamental of tasks, including some that you have already completed in this series of exercises. However, the internet has led to cloud-based computing where scientists can access and use resources in distant locations. Sharing resources is great but not quite accessibility unless some effort is made to make these systems easier to use for non-specialists. Google Colab and the Making It Rain project now allow for use of Google's computing resources through a somewhat friendly interface. While there are time limitations and limits on the flexibility of what can be included in the simulation, for applications like understanding the effects of mutations or sequence changes the Colab is good enough.

This task is really half a task, Task 7 focuses on setting up the simulation while Task 8, takes the output and provides some analysis of the results. Task 9 will build on 7 and 8, to show how to prepare molecules for simulations with ligands bound or with authentic post-translational modifications.

WHAT YOU GET: A sense of the dynamics or motion of the protein and, with the right controls, how changes in sequence can affect the structure and dynamics, which can affect the function.

Colab based simulation video guide

Things to remember when using a colab

- Unless you download the file/image to your computer OR store it in Google Drive, it does not save
- You must maintain an internet connection the entire time, you cannot close the window and come back
- With a free Colab account, there is a 12-hour connection limit. You can restart each day and if you have the strides in the simulation set right, you can pick up near to when the connection was lost.

1. Sign into your Gmail/Google account
2. Create a Google folder associated with this task
3. Upload your .PDB file to this folder and make sure the name is short and lacks punctuation or spaces
    a. model.pdb is a good name
    b. You can use the PDB of the full-length protein or one of the domains.
4. Navigate to the Step 1 Colab
    a. Make a copy of the Colab by going to File>Save a Copy in Drive
    b. There are instructions on the Colab to follow, which will be summarized here
    c. This colab is only a suggested guide, it can work for MD but can be changed to suit your instructional goals

    d. Punctuation and spaces are really important here so be careful
5. In Colab, code blocks are run by pushing the “play” button to the left of the text.
    a. A running block will show a moving swirl are the stop button
    b. In some instances you need to wait for a block to finish before running the next
    c. Once a block is run, you cannot change the settings without re-running the code block
6. Run the Install Conda Colab block
    a. YOU MUST WAIT FOR THIS BLOCK TO FINISH BEFORE MOVING ON!
7. Run Install Dependencies and Import Google Drive.
    a. While these blocks are running, edit the input information under loading your .pdb file
8. Follow the instructions in the Colab to complete the simulation.
    a. The suggested defaults should be fine.
    b. If you are running the full-length protein, it will be a slow simulation.

I want to study the effects of a point mutant on the protein, how would I do that?

1. Model the new sequence in AlphaFold3
2. Create a PDB from the AF3 model
3. Create a new folder in Google
4. Upload the mutant PDB file
5. Run the Colab with the mutant.
    a. Keep all of the settings the same as for the unchanged protein!

Task 8: Analysis of MD simulations

The amount of information provided by a simulation can be very overwhelming, and knowing not only where to start but also whether it means anything can be challenging. Like a laboratory experiment, having an analysis workflow is very helpful. The linked analysis Colab provides that initial workflow to describe the general behavior of the molecule in solution such as dynamics and pKa values which can be parsed to describe specific events such as distances between amino acids. The instructions in the Colab are pretty descriptive so here I will describe what different values can mean and where they can be extended on in further analysis or even laboratory experiments.

MD analysis video guide

MD Analysis colab

WHAT YOU GET: A description of the movement of the molecule and every atom in the molecule in a hypothetical universe. Comparison of these types of data between ligand bound and unbound or with mutations can provide an atomic/molecular-level description of the possible structural effects for a biochemical behavior.

1. **RMSD: Root Mean Square Deviation**
    a. A plot of RMSD vs. simulation time can indicate whether a molecule has reached "equilibrium" in the simulation → if not, the simulation needs to run longer.
    b. Most comparisons are of RMSD relative to the first structure in the simulation so the RMSD value increases and hopefully plateaus within the simulation time
    c. Analysis of data when the simulation is in pre-equilibrium are generally less reliable.
    d. Usually need at least 50 frames of simulation data where the RMSD is level with time for a reasonable analysis
    e. 2-D heatmap plots of RMSD vs. RMSD can show stable times when the protein was in stable conformations
    f. This plot tells me if I can "trust" my simulation
2. **RMSF: Root Mean Square Fluctuation**
    a. Reflects the movement of an amino acid during the simulation
    b. Plots usually show RMSF vs. amino acid position.
    c. The higher the RMSF value, the more motion or dynamics that amino acid had in the simulation
    d. This is the first piece of information to compare when looking for effects of ligands, PTMs, or mutations
3. **pK_A values**
    a. pK_A values from a simulation are determined by the side chain environment, thus they can be a proxy for localized change in conformation/structure.

b. An amino acid with a wide-range of pK_A values might be one in a dynamic area. Narrow range of values = consistent surroundings.
c. Should correlate with the RMSF data
d. This is the second piece of information I use to compare simulations BUT it is more sensitive to the structure being at equilibrium, so check that RMSD plot!

WHERE DO I GO FROM HERE?
From the MD data, it should be possible to identify interesting areas to followup with lab experiments or more detailed analysis AND/OR explain biochemical data

- Different active site conformation ⇒ lower activity?
- The first substrate alters the protein conformation ⇒ ordered binding mechanism?
- Serine phosphorylation lowers the active site general base pK_A value ⇒ lower/higher activity?

UNDER DEVELOPMENT

Task 9: Simulations with ligands, mutants, or post-translational modifications

Proteins do not exist in isolation nor are they static molecules. Proteins interact with small molecules as substrates or regulators and they can be modified post-translationally in several ways. In all cases, the protein structure, dynamics, and function can be affected. Because these situations may be fleeting in the cell, building models of the proteins with ligands or modifications can show the effects. Models of modified or ligand-bound proteins also are useful for the development of therapeutics.
In this task, we will model a post-translationally protein and generate the files needed to simulate the protein, similar to what we did in Task 7. If we essentially repeat Task 7, why does it need a distinct task? The difference is the modification. Molecular dynamics requires a forcefield, which describes the chemistry in our system, some modifications have chemical parameters that are not found in protein forcefields, thus we must generate some parameters for them before the simulation, this is the purpose of CHARMM-GUI. Then we will use the CHARMM-GUI outputs in a Colab similar to the Colab in Task 7. From there, everything is the same as in Tasks 7 and 8.

Ligands present similar challenges to post-translational modifications, unusual chemistry. Thus a similar workflow must be followed for protein-ligand complexes.

CHARMM-GUI walkthrough for a post-translationally modified protein

MD simulation colab

WHAT YOU GET: Data on how a PTM, ligand, or sequence variation affects the dynamics of the protein.

****NOTE: if you want to compare the simulation of the prevalent variant or wild-type to the modified protein, you need to prep the wild-type through this method as well. Just don't make any changes to the protein sequence or modification state!*

1. Sign up for a CHARMM-GUI account and then sign into the server.
2. From the CHARMM-GUI site, go to the input generator then solution builder
    a. The walkthrough linked above starts here if you prefer to follow the video
3. Upload a PDB to the site
4. Go through the steps and add a PTM to your protein
    a. UniProt lists known sites of modification, if there are any, under PTM/Processing. This is a great place for students to explore the data a bit and find literature on known sites and their effects.

    b. AMY3, unfortunately, only has one modification, a disulfide.
5. Remember at step 4 to indicate **AMBER outputs**
6. Download and unzip the CHARMM-GUI output folder
7. Copy the .rst7, .parm7, and .pdb file from the AMBER folder to a Google Drive folder
8. Navigate to the MD simulation Colab for CHARMM-GUI outputs
9. Follow the instructions to run the simulation
10. Analyze the simulation as in Task 8.

I want to simulate a ligand-bound structure (this process is temperamental…)

1. Create a PDB or .cif file with the protein bound to the ligand
2. Upload it to CHARMM-GUI
3. In step 2, make sure the box next to the ligand is checked
4. In step 3, choose “Use CHARMM general force field….”
    a. This is where things usually go wrong, especially for exotic molecules
    b. Malate, citrate, etc. usually are ok; NAD/NADH is 50/50
    c. I’m working on a more reliable set of steps
5. Proceed through the rest of the steps as before
    a. Make sure to choose CHARMM36m for forcefield and AMBER outputs
6. If you get through CHARMM-GUI, simulations *should* work fine