**Thesis thoughts**

I think I’ve started writing this thesis stuff in another document somewhere, but since I haven’t yet found it, I’m going to just kind of start over with an outline, and then likely an introduction (taken from snippets of things I’ve written before, and then likely another version of an outline for the chapter for the public)

I was looking for a paper for a journal club and this was one I did on March 1, 2022:

<https://www.nature.com/articles/s41586-021-04383-5.pdf>

I think it had a pretty good intro on reasoning for designing backbones, and I liked the way they put it. So might be a good idea to go through it a bit more and take some notes on how they work it.

**Previous theses**

I also took a look at Samantha’s thesis and saw her intro involved a bunch of introduction to concepts like single pass membrane proteins, the importance of sequence motifs, methods used to study TM association, new technology that she used, and an overview. It’s 29 pages. I think the below is an okay look so far, maybe for the parts that I’m going to write on, but I think I’ll need a bit more detail. I need to relook at Samson’s thesis too to see how he put the intro together, as his modeling logic may be a good fit for my design logic? Chapter 2 should likely be my paper, Chapter 3 can be any other papers I contribute on (if any), Chapter 4 would be some next directions, and Chapter 5 will be the extra chapter for non-scientists.

I also have other theses that I found online for other people with primarily computational projects that end up being helpful.

**Some brain dump logics**

Since a lot of my research is focused on single pass, it might be a thought to talk about them? Although my work doesn’t exactly focus on those…it’s more on membrane proteins and the depth of knowledge we can gain using this methodology to determine how vdW impacts many of the protein structures. And if we can predict it using simple methods. By keeping all other forces constant and only looking at the dependency of protein association on vdW, we’re gaining knowledge of how strong vdW can be in having two helices interacting. Although larger protein complexes may have more vdW overall, if we can show that it’s an important force for two protein chains in the absence of other forces and additional chains, it could be…what? Maybe look at how the Degrado person’s thesis and paper is written? Just to see what he was thinking and how this pushes it forward a bit more. Where does my research fit? What niche? Why is teasing out the additional strength of a force, and studying it in so many additional conformations, important? Testing the overall geometric and energetic extent of vdW as a driving force for membrane proteins. If we can show that essentially randomly designed sequences are able to come together just based on geometry and putting it together with our energies (or if our energies are ass, then just using a geometric/fitting type of parameter) is able to show some sort of correlation. I think the important pieces of data we’re going to need to look at are the sequences with similar geometries (I need to start finalizing the pieces of code that do this so that I can just cut this part off completely and have confidence about it)

**Musings**

Is a thesis just a complete explanation of your research that is written in such a way that many of the questions someone might have are answered as you’re reading? Like thinking about what the next question would be, answering it, then again, and again and again. I guess I’ve never thought about writing in this style, but since it has to be so thorough, this might be the way to think about it.

Thesis Outline

**Introduction**

* The importance of understanding protein structure
  + What is protein structure? (basics of a static model to complex things such as MD; is there a good way that someone has explained this that I can put in my own words? Find a paper/thesis that does this)
  + Why is it important to understand a protein’s structure (therapeutics, understanding misfolding and why it happens, etc.; another one to find a paper or two for)?
  + What can understanding protein structure give us (I feel like this is kind distinct from the previous atm, but may change my mind later)?
  + How have people studied protein structure in the past and what are some things that have been used from protein structural information? Talk about the pdb and other databases?
  + Novel approach is protein design (when was the first design paper?); and how it has been used in understanding a complex set of proteins known as membrane proteins
    - It might be an interesting idea to look at how CASP pitched their meetings: what is the biological relevance (and then use that kind of wording for my public chapter and extract out what it means for vdW in this intro)

The above needs work: I’m not sure if I should use the above in it’s present order; it’s kind hard to integrate with the below parts, so maybe I should split it and add it to the below?

* Lack of understanding in membrane proteins:
  + Why are membrane proteins so important?
  + Define membrane protein (is this too much detail here? Should I move it to after comparing with soluble? Or just kind of state the below generally, then get into the specifics afterwards):
    - Folding and association
    - What is a membrane?
    - Why do proteins associate/fold in the membrane?
    - Needs to understanding membrane proteins
  + What problems have been solved in soluble proteins but haven’t yet been in membrane proteins? And why is this the case?
  + Why is it difficult to study membrane interactions vs soluble?
  + How has studying membrane protein interactions been done previously?
  + How has our lab pushed the field in understanding membrane protein folding and association?
  + Where do we see the membrane protein field going in the next 10 years?
* Our lack of understanding of vdW interactions (a lot of references here, but most specifically that Degrado paper that uses them, but doesn’t solve a tractable model for all proteins)
  + Define vdW and other forces of association
  + How have other forces been studied?
  + Why don’t we understand vdW/why is it a difficult force to study?
  + What has been done to study vdW so far? Why is my approach different (design of homodimers? Or using our knowledge of structure to predict and test interactions)?
* Previous methods to study…vdW? Membrane packing?
* Current and improving design methods? How can I saw that this is pushing the field if it isn’t really using those novel design algorithms? How can I push it to that point (other than just using those design algorithms? Implementing a similar workflow? I guess this question is more of how would I do this now since I know the power of machine learning?)
  + Would it be possible to read in the different geometries that these pack well at, the AAs that pack at the interface for each of these proteins, and essentially redesign these on standardized backbones using the method I have already (except maybe without energy, but with geometry as the main concern?)
    - So the machine learning method would kind of pick AAs based on how the local geometry around each of those interacting points in the geometry works. I think I would have a very limited space though because of the need to have a bunch of homodimers…but this may be more feasible with heterodesign?

Chapters

* Exploring the geometric landscape of dimerization interactions (really just helices in close contact; is there a better way to word this?)
  + It might be a good idea to also outline the types of figures I’ll want to show here (although one of these will likely just be the paper and supplements, so may just give me ideas for those)
  + Have a figure that shows in detail how my design algorithm works, and explains some of the random detail (sequence entropy term, baselines, etc.; also, these should be supplementary figures that I should prepare soon)
* Protein design: assessing vdW packing as a driving force in membrane protein folding and association
  + What is protein design? Should I be going back to the idea of anfinsen’s hypothesis here? And talk about modeling?
  + Or should I instead be thinking about how with the tools we have
* What’s another chapter that would fit…machine learning? Or FRET as future directions? Maybe both
* PhD Research for the public: <http://scifun.org/Thesis_Awards/chapter_guidelines.html>
  + The way I see this chapter is that it is a novel for my time in graduate school:
    - What does it take to do scientific research?
      * Contrast time and expertise with energy and frustration
        + Think of the most frustrating thing that has ever happened to you; why was it frustrating? Why is this pursuit of finding/creating knowledge frustrating?
        + Applying knowledge to find more knowledge: but how do you decide that you’ve had enough? That you’ve done enough reading, application, etc. to essentially make new knowledge?
      * How are you supposed to expect something that you’ve never experienced (not knowing that graduate school would be this hard)?
      * The importance of a love for learning
        + Think of your favorite activity/interest: something I’ve loved for decades is basketball

Understanding the minutiae of different types of dribbles, how to create space on the court, when to pass and when to fake a pass, percentage of makes from different parts of the floor, etc. yadda yadda

* + - * + How about something that most people love:

Music: talk about listening to music and all of the tiny things that I love about it

When an artist does something new (reference some songs that it’s clear have changed the genre; likely Frank Ocean and mxmtoon)

* + - My love for membrane proteins and vdW
      * For years, I’ve been dealing with a back injury that hasn’t been explained by x-rays/MRIs/doctor’s/ etc. There are temporary ways to fix it (working out, running, etc.). This prompted my interest in science: why might my nerves be misfiring to tell my brain that there is pain in an area that doesn’t need to have pain? (explain that in leyman’s terms too)
      * Never did I expect to fall in love with membrane proteins: I think here I just need to kind of rant out what I love about my project: why is it so interesting to me? How did it become interesting to me? Why is it so difficult to accurately process and use all of the things I’ve learned in classes and apply it to research (how gels work simplified, what tags are, and how all of these can be combined for successful research)
      * This should basically be the leyman’s version of my intro
        + What is a protein? What is a membrane protein? Why are these proteins important?
        + How have people studied these proteins? What we know about them? What don’t we know energetically? Why?
        + Where does my work fit? What’s an analogy for the information we have about soluble proteins vs membrane proteins?
        + How will that push our understanding of membrane proteins? Why will it do so? Analogies to another thing that people understand
        + Don’t drag on here: make sure it’s clear and not bogged down in the details. Say just enough for people to understand and think of further questions if they want, but that doesn’t make someone think of further questions.
    - Data
      * How do I communicate the data? I think when I have it and can look at it, I’ll have a better understanding of what to do with this chapter (as of 2/18/2022, I have the computational data, but that is currently a bit complex; how can I simplify this…?
        + Studying protein association/interactions/coming together using fluorescence (don’t mention the transcription factor; just that they interact and fluoresce). Amount of fluorescence correlates to strength (is there an analogy here? Glow in the dark stars or something? Battery powered things (when batteries are dying/less strong, flickering lights/things don’t work as well and the proteins/vdW are the battery strength?)
        + Definitely talk about the game-like nature of my project: putting together these building blocks/legos and seeing if they fit well
    - I think after this it’s going to be a nice conclusion:
      * Thanking friends and family and others who supported
      * Talking about the hostilities of graduate school in a place that isn’t the best for people of color
      * What my love of learning has allowed me to do that I never expected:
        + Podcast, coding projects, music, etc.

Chapters that could get added?

* What could we do with design next? Can we optimize it for trimers? What if some of the things we design interact better as trimers?
* Heterodimer design: outside of just Josh’s GASright designs, what else could they be used for?
  + Find some resources about the importance of these heterodimeric interactions (this may be more important for the conclusion: talking about a lot of the proteins that end up misfolding (mutation that causes increased association? Where does this come from? First part is understanding if that is even the case)
  + Using this program in combination with other things (Rosetta and Alphafold) for design of other molecules of interest
  + Not designing on a standardized backbone (this might be something I want to do anyways: try to design using a random backbone, random geometry, picking positions, and testing association like that (I’d have to fix the sequence entropy term for this, but I think I could set this up as I work on the code in the coming week)
* Could I use machine learning on the data I’ve currently found?
  + Train my algorithm
  + Search for hotspots of association
  + Determine letter frequencies at different positions at interfaces in membrane proteins
    - Or differences in letter frequency in membrane vs soluble since I have that data too?
    - What could a paper on membrane vs soluble proteins look like? A review hasn’t been written in awhile? Maybe something like what Alessandro did, updating his database search with new information? Could also be interesting just to look at how alphafold designs soluble and membrane proteins and how similar they are?