**Aim 1: High-throughput determination of how sidechain packing affects dimerization**

*Rationale:* The goal of this aim is to determine if solely changing the strength of interfacial sidechain packing influences membrane protein association.

*Experimental Design:* To accurately design membrane protein dimers, I first need to explore the geometric landscape at which dimerization is possible. However, there are no known structures of membrane protein dimers that associate due to solely van der Waals packing. Instead, by searching nonredundant protein structures in the PDB for geometries of any helices in close contact, I may find geometries that are favorable for association due strictly to sidechain packing. Each of these geometries will be standardized by a poly-Leu backbone, aiming to control for expression and insertion of our sequences. Then, only positions at the dimer interface will be mutated. With these geometries as a structural template, I will use dead end elimination, self-consistent mean field, and monte carlo to filter and search sequence space for amino acid combinations that pack at the dimerization interface. Using the CHARMM 22 van der Waals energy function, I will measure the stability of each sequence, determining the top 100 sequences at each unique geometry. Overall, computational design will result in a population of dimers whose stability can be compared by the amount of interfacial sidechain packing.

The designed population of sequences will be evaluated for their experimental stability using sort-seq, a high-throughput method which combines fluorescence activated cell sorting (FACS) with next-generation sequencing (NGS) to evaluate dimerization propensity of transmembrane domains. An oligo pool library consisting of my designed TM sequences will be cloned into TOXGREEN plasmids. These TOXGREEN plasmids use the reporter gene sfGFP as an output of dimerization, allowing for quantification (as described in methods). Based on the fluorescence output of sfGFP, cells are sorted into different bins with different fluorescence thresholds. These plasmids can then be purified out of the cells in each bin and enumerated via NGS. Based on the counts present within each bin, the fluorescence profile, and thus dimerization propensity, will be reconstructed for each dimer. Using this technique, I will be able to determine if there is a correlation between strength of sidechain packing and membrane protein association.

Image of design strategy here

*Expected Results/Interpretation:* Previous research has demonstrated that understanding the structures of natural proteins can contribute to both prediction and design (cites). In particular, if a structure is commonly found in nature but not dependent on a specific sequence, then designing sequences for this structure is likely to be more successful (Zhou and Grigoryan, 2014(?)). In my case, finding common geometries for helices in close contact in membrane protein with unique sequences is likely to yield successfully designed sequences for dimers. Much of the analysis for these geometries has been done, and I am currently in the process of designing sequences for these geometries. I expect that the designed sequence energies will correlate well to dimerization propensity found in sort-seq, resulting in a trend where sequences with weak sidechain packing results in a weak dimer and strong packing results in a strong dimer. However, if there is little to no correlation between our designed energies and dimerization propensity, we will have some evidence that although van der Waals packing is necessary for association, it is not a strong contributor to membrane protein association. Overall, these results will give insight into the extent at which sidechain packing contributes to stable helix-helix association in membrane protein folding.

*Potential Complications:* Despite the progress within the membrane protein field, only one membrane protein that relies solely on van der Waals packing has been designed (cite). In addition, none of the membrane proteins that have been designed are dimers (as far as I know? Need to confirm with another literature search). A likely reason for this is the difficulty that arises in structural confirmation, resulting in only 23 membrane protein dimers found in the PDB. Because my research is not focused on characterizing the structure of each design but rather on analyzing the trend that arises from changes in packing, structural characterization is not a limitation. However, if I decide that confirming the structure of some of my constructs is important to determine the accuracy of my designs, I will have a large population of proteins that I could test for expression, purification, and crystallization ability, making it more likely for me to get an acceptable crystal structure. Another caveat is that I am assuming that if a geometry is more prevalent in the PDB analysis, then it may be more amenable to design (cite Zhou and Grigoryan). To determine if this is the case, I will also design sequences for geometries that are not as prevalent within the analysis. If these designed sequences are overall less energetically stable than our other designs and have lower dimerization propensity, then it will support our assumption that common geometries are more favorable for design. Finally, because sort-seq does not measure free energy, it is not possible to confirm that all of my designed constructs form dimers instead of higher order oligomers. This instead will be addressed in aim 2.

**Aim 2: Quantification of the range at which sidechain packing influences dimerization**

*Rationale:* Although my previous aim will deduce the extent at which sidechain packing influences dimerization, it will not measure the biophysical strength of association. This aim will quantify the free energy range at which interfacial sidechain packing influences stability of dimerization.

*Experimental Design:* By solubilizing a constant protein concentration in increasing concentrations of detergent or lipid, I can make a gradient of mole fractions. These mole fractions influence the monomer to dimer ratio: increasing detergent will increase the amount of monomer, which can be detected by an increase in donor fluorescence and a decrease in acceptor fluorescence. By determining the change in fluorescence at different mole fractions using FRET, I can deduce an equilibrium curve for the transition between monomer to dimer, allowing me to determine the dissociation constant of each dimer. These dissociation constants, or Kds, will then be used to calculate the free energy of association (show equation?). In addition, my labeled proteins can be used to determine the oligomerization state of my designs to confirm that dimers are being formed rather than higher order oligomers. By measuring the amount of donor to acceptor excitation or quenching at increasing concentration of acceptor, the oligomerization state can be calculated for each design.

*Expectations/Interpretation:* I expect that FRET will allow me to determine the strength of association for a subset of my designed dimers. FRET should recapitulate the correlation of packing to dimerization that I see in aim 1, confirming that strength of packing contributes to the free energy of association. In addition, FRET will allow me to quantify the strength of packing and the range at which packing contributes to association. By taking dimers with the weakest association, FRET can determine the strength of packing necessary for dimerization to occur. In addition, by taking dimers with strongest association, FRET will determine the maximum strength at which packing can contribute to association. Overall, I expect that as sidechain packing is increased, the free energy of association will also increase. However, if this is not the case and there is little correlation between strength of sidechain packing and association, FRET can still be used to determine the free energy of packing that contributes to association in my designed constructs. This result would suggest that sidechain packing contributes to association but cannot increase association without other forces such as hydrogen bonding and electrostatic interactions. Redesigns of my constructs with polar amino acids that lead to these forces can then be used to support this finding.

*Potential Complications:* FRET involves a variety of complications including protein expression and labeling. For efficient labeling to occur, it is important to have a high concentration of protein. However, membrane protein are difficult to express and solubilize while ensuring that the structure of the protein is maintained. …problems with expression… After expressing and purifying a set of weak, mild, and strongly association proteins, I will have to label half with a donor fluorophore and the other half with an acceptor fluorophore. …Problems with labeling… One final problem that is a bit more difficult to fix is dimerization of proteins with the same label, or donor-donor or acceptor-acceptor association. We account for this possibility while calculating the Kd from the FRET efficiency. …explain the concept of the math here…

Different geometries may have different numbers of interacting amino acids. This will influence the not only influences the van der Waals packing within the dimer, but also van der Waals interactions with the lipid bilayer. To limit these, we will design a set amount of positions on standardized helices, keeping much of the interactions with the backbone and the lipid bilayer consistent between our designs. (add in complications with the method; what are limitations of the method and why is it important that we use FRET; then do the same with FRET)

One thing that we might be assuming from our pdb analysis is that certain geometries being more prevalent than others may correlate to stability. However, this is not the case and we will have to do a multitude of designs for a variety of these geometries in order to see which will work…

Rationale 1: Optimized sidechain packing has been shown to stabilize oligomerization of a designed protein (cite). However, …other factors could have played a role… If it is possible to design stable structures using solely sidechain packing, then it should be possible to measure packing using dimerization, a simple and tractable model system. Dimerization has been used as a model system to study the energetics of folding due to its simplicity and … . In addition, our previous research predicting the structures of dimer using only van der Waals and hydrogen bonding demonstrates our expertise … using this model (more detail?). Utilizing computational techniques often used in prediction and design, I aim to design sequences that dimerize with form van der Waals packing. I will then test these sequences using high-throughput in vivo sort-seq, a high-throughput method that measures dimerization propensity. This aim will determine if the amount of sidechain packing correlates to dimerization propensity, demonstrating that optimized …

If it is possible to design a stable oligomer using solely sidechain packing, then it should be possible to determine how packing influences association using dimerization, a simple and tractable model system for studying …

Aim 1 stuff

Aim 1a: Design of dimers dependent on optimized sidechain packing

What will this aim tell us?

Using computational techniques to search sequence space , such as dead end elimination, self-consistent mean field, and monte carlo, we can design sequences that are reliant solely on van der Waals packing.

protein design, we can engineer sequences with a variety of van der Waals packingBy designing sequencesfor membrane protein dimers, we aim to study the fundamental association present due to specific changes in van der Waals packing. We can then design an array of dimers with variations in van der Waals packing. To match our high-throughput design, we will measure our results using an *in vivo* high-throughput assay known as sort-seq. Our approach is to use design and sort-seq to determine the extent at which van der Waals packing can influence dimerization.

(fix this rationale; talk more about why we are doing design and study in high throughput; I might just be able to get rid of first sentence, then edit through the next couple, emphasizing high-throughput research)

Experimental Design:

Even with recent advances in membrane protein design, no membrane protein dimers have yet to be designed. Our lab are experts in predicting how sequence affects structure and energy. We have previously shown that these predictions correlate well to experimental stability.

In addition, structural prediction and protein design utilize similar algorithms to search the protein energetic landscape for stable structures. Using these algorithms, I will be able to design thousands of dimers with weak to strong sidechain packing (as described in methods). I will then determine if the amount of sidechain packing correlates to dimerization propensity using a high-throughput sort-seq method which combines fluorescence activated cell sorting (FACS) with next-generation sequencing (as described in methods).

How much detail do I need and what do I need to mention here:

* Already talked about rationale for my experiments in the innovation; now I just need to mention the actual experiments and how they work in detail

PDB analysis:

* What is it? Why did we do it?
  + Haven’t mentioned at this point; mention after describing computation; Much of our prediction work settles on using specific geometric templates that maximize energies; However, lack of membrane proteins specifically for vdW results in no template; aim to find a template by using helices in close contact in membrane proteins; the more unique, the more likely it is designable (why?)

Computation

* Algorithms (mention them and what they will generally be used for)
* How does the design method work and what will it result in

Sort-seq

* What is it and how does it work
* Mention that it’s specifically for homodimers?

Explain how to do it; not the rationale behind it!!!