Chapter 1: Introduction

Table of Contents

[1.1 Introduction to single-pass membrane proteins 3](#_Toc161323742)

[1.2 Forces involved in membrane protein association 4](#_Toc161323743)

[1.3 GASright 5](#_Toc161323744)

[1.4 Studying membrane protein folding and structure 6](#_Toc161323745)

[1.4.1 Computational methods 6](#_Toc161323746)

[1.4.2 Experimental methods 6](#_Toc161323747)

[1.5 Overview of thesis 7](#_Toc161323748)

[1.6 References 8](#_Toc161323749)

* 1. ****Introduction to membrane proteins****

Membrane proteins (MP) are a vast category of proteins found tethered to the lipid bilayer. Some are found bound to the edges of lipid headgroups (peripheral) while others are embedded, spanning the length of the membrane (integral). The bilayer is a fluid system composed of many types of lipids, resulting in a complex network of interactions that are responsible for MP folding and stability. Proper MP association and folding is critical for a multitude of essential biological functions: … proteins are involved in signaling between … cells, (protein name) helps to regulate pH and ion balance, and (protein name) … genetic signaling cascades resulting in … such as (example).

MPs comprise 25-30% of the proteins found within protein-coding genes of various organisms (Fagerberg et al., 2010). Misfolding of MPs has been found to be involved in several human diseases such as Parkinson’s, cystic fibrosis, and cancer (Sanders and Myers, 2004; Gregersen et al., 2006). (expand on how it is involved in each of these diseases, a sentence on each). Many groups work to understand how protein misfolding plays a role in disease states and progression with a focus on understanding the mechanisms and forces involved in folding.

However, studying MP folding is inherently a difficult challenge because of their hydrophobic nature. MPs are difficult to express in yields high enough for biophysical experiments, and purification and solubilization of these proteins often lead to aggregation or unfolding (Carpenter et al., 2008). Additionally, proteins fold differently depending on the composition of lipids in the bilayer, creating a daunting task to replicate biological conditions *in vitro* and *in silico*. To combat these challenges, much of the research studying membrane protein folding is focused on combining *in vitro* or *in silico* studies with live cells, allowing for exploration of the biophysical forces governing the folding process in a natural environment.

* 1. ****Forces involved in membrane protein folding and association****

Folding of integral transmembrane (TM) MPs involves a variety of energetic constraints due to the hydrophobic nature of the lipid bilayer. Firstly, MPs must be translated and inserted properly into the membrane. This is accomplished through the aid of the translocon, a protein that helps to properly orient the protein in the bilayer. However, once the protein is embedded into the membrane, the hydrophobicity at the core of the bilayer results in an energetic penalty for any unpaired hydrogen bond donors and acceptors (Marinko et al., 2019; Popot & Engelman, 1990, 2000). This forces TMs, inherently composed of a backbone containing oxygen and nitrogen atoms, to conform to standard secondary structures with backbone amide protons and carbonyl oxygens satisfying the membranes lack of hydrogen bonding. These TMs can be broken down into two groups: beta barrels that … and alpha helices that ….

Proper membrane protein folding is regulated by a distribution of stabilizing hydrogen bonds, weak polar and electrostatic interactions, and van der Waals forces between the unfolded and folded states. Hydrogen bonding is typically penalized within TM insertion, but multiple groups have found that it is a … force involved in folding and association … . Additionally, electrostatic interactions … However, although van der Waals forces are an ever-present force with atoms coming into close contact, not many studies have solely accounted for the contribution of van der Waals packing interactions.

* 1. ****The importance of van der Waals packing****

This force is particularly important due to the nature of van der Waals interactions: Even if hydrogen bonding or polar interactions play a significant stabilizing role, because van der Waals occurs between any nonbonded atoms in close contact, it is a necessary force that is always present within the folded state. This means that van der Waals packing is essential for folding, but the extent at which packing can be a driving force for membrane protein folding is unclear (expand on this; could go into the other forces involved in the membrane?).

The contribution of van der Waals packing to membrane protein folding can be broken down into three distinct interactions: lipid-lipid packing, lipid-protein packing, and protein-protein packing (say something about the complexities of this force in the membrane and contrast with soluble). Protein-protein (or sidechain) packing, is a technically feasible starting point because of the ability to manipulate sequences and determine changes in stability due to mutation. Previous research has demonstrated that disruption of packing within the core of bacteriorhodopsin destabilizes protein structure (Faham et al., 2004; Joh et al., 2009). In addition, a recent study using membrane protein design has shown that optimized sidechain packing can stabilize the folded state of phospholamban (Mravic et al., 2019). Although it is known that sidechain packing plays a role in stabilizing membrane protein structure in these individual systems, the energetic contribution of sidechain packing to the folded state of membrane proteins more generally has not yet been determined. My research aims to characterize and quantify the extent at which sidechain packing is a driving force for membrane protein association.

* 1. ****Studying membrane protein folding and structure****

(what do we gain from understanding biophysical forces? The ability to design and engineer novel proteins, exploring and predicting their structures, and training to do both *or something like that*) In addition, this knowledge can be applied to design new therapeutics that specifically target proteins in these misfolded states. (better ending sentence about how learning in depth about these forces will improve human life)

****1.4.1 Experimental methods****

* **Talk about the experiments used in the past, what were they able to discover**
* **The establishment of experimental databanks like the PDB and memprotDB and others**
* **Go into the quantity increase in the last 15 years** 
  + **talk about the difference between number of structures in the PDB with number of structures predicted by Alphafold (and compare how long it took and maybe now how many papers are citing this powerful tool for their experiments)**
* **Combining computation expands the ability to think of more challenging and complex ways to tackle experiments**
* **The important methods for matching this high-throughput age of science:**
  + **FACS, gene chips, NGS, cloning strategies, expression strategies, machines**
  + **Other things that are able to match the throughput of computation**

****1.4.2 Computational methods****

* Talk about the importance of papers at the interfaces of all disciplines, needing to understand the computation
* Talk about the need to understand and utilize it as a tool to increase what we can learn and the rate at which we learn about these small systems, could even cite the lady who recently imaged the black hole because of her ability to store and process so much data
* Rosetta -> Alphafold -> all the new methods that are still coming up
* The use of AI in analyzing data and building new structures
* Thoughts about how to evaluate the field in the presence of AI?
* I think the above are just mentionings, because I would like to go more into detail for how my stuff worked later in the thesis and how it plays off of some of these
  1. ****GASright****

The GASright is one of the most prevalent sequence and structural motifs found in TM proteins (Walters & DeGrado, 2006). GAS is an acronym for the three amino acids typically found in the sequence: Glycine, Alanine, and Serine. These small residues define the interface of the motif (G/A/S)xxx(G/A/S), resulting in a short interhelical distance between TM helices. Right originates from the structural features in which TM helices associate with a right-handed crossing angle. Additionally, it is frequently found in a variety of biological systems involved with immunology, metabolism, and cancer. Due to its potential importance in medical applications as well as its well-defined sequence and structural features, GASright proteins have been used as a simple and tractable system to further understand forces governing TM association.

The GASright motif’s unique sequence and defined structure has been shown to permit an uncommon structural feature. The short interhelical distance allows TM backbones to come in close contact, forming a network of weak hydrogen bonds where donors are Cα carbons and acceptors are carbonyl oxygens on the opposite helix (Cα–H∙∙∙O=C, or Cα–H bonds). Carbon atoms are not commonly associated with hydrogen bond donors because carbon is a less electronegative atom than other donors nitrogen and oxygen. However, these carbon atoms are … by electronegative withdrawing groups on the peptide backbone, increasing their electronegativity. Estimates from quantum mechanics calculations suggest that the stabilizing energy of an Cα–H bond may contribute one third to one half of that of an N—H donor in vacuum (Scheiner et al., 2001; Vargas et al., 2000). Anderson et al. utilized a combination of computational structure prediction and the experimental assay TOXCAT to determine the influence of this network of Cα–H bonds. By predicting and analyzing GASright motifs found in natural sequences, they showed that structures with more Cα–H bonds had a higher stability in the dimer state. Additionally, Díaz-Vázquez et al. measured the free energy of association of GASright structures using *in vitro* FRET, concluding …. These studies posit that the primary forces involved in GASright association are a combination of hydrogen bonding and van der Waals packing. … makes this motif a … control to evaluate the propensity for TM association using solely van der Waals packing.

* 1. ****Thesis overview****

My graduate research focused on using computational protein design in combination with high throughput assays to determine the extent at which van der Waals packing contributes to membrane protein association and folding. Prior research on the impact of packing to the folded state of membrane proteins honed-in on singular systems, and I aimed to expand this knowledge to a larger variety of membrane protein structures.

**In Chapter 2**, I present most of my work to be published in the near future. In this paper, I data mined the PDB for all solved membrane protein structures, developed a computational protein design algorithm, and assessed the ability of proteins designed with solely van der Waals packing for their ability to associate using a high-throughput assay. … conclusion

**In Chapter 3**, I discuss my computational methods in detail, citing the inspirations for datamining, the design algorithm, and the analysis that led to the conclusions found in my paper.

**In Chapter 4**, I describe a variety of future directions for my protein design project, expanding on what can be improved upon and how I would design proteins with the tools available today.

**In Chapter 5**, I share a collaboration with the SciFun program at UW-Madison, detailing science through a chapter for the public. I reflect on the lessons that graduate school imparted onto me throughout my time here, giving my fully transparent thoughts on how my research affected my physical, emotional, and mental well-being. … on key events that helped me grow.

* 1. ****References****

Marinko, J. T., Huang, H., Penn, W. D., Capra, J. A., Schlebach, J. P., & Sanders, C. R. (2019). Folding and Misfolding of Human Membrane Proteins in Health and Disease: From Single Molecules to Cellular Proteostasis. *Chem Rev*, *119*(9), 5537-5606. <https://doi.org/10.1021/acs.chemrev.8b00532>

Popot, J. L., & Engelman, D. M. (1990). Membrane protein folding and oligomerization: the two-stage model. *Biochemistry*, *29*(17), 4031-4037. <https://doi.org/10.1021/bi00469a001>

Popot, J. L., & Engelman, D. M. (2000). Helical membrane protein folding, stability, and evolution. *Annu Rev Biochem*, *69*, 881-922. <https://doi.org/10.1146/annurev.biochem.69.1.881>

Scheiner, S., Kar, T., & Gu, Y. (2001). Strength of the Calpha H..O hydrogen bond of amino acid residues. *J Biol Chem*, *276*(13), 9832-9837. <https://doi.org/10.1074/jbc.M010770200>

Vargas, R., Garza, J., Dixon, a. D. A., & Hay, B. P. (2000). How Strong Is the Cα−H···OC Hydrogen Bond? *Journal of the American Chemical Society*, *122*, 4750-4755.

Walters, R. F., & DeGrado, W. F. (2006). Helix-packing motifs in membrane proteins. *Proc Natl Acad Sci U S A*, *103*(37), 13658-13663. <https://doi.org/10.1073/pnas.0605878103>