Chapter 1: Introduction

Table of Contents

[1.1 Introduction to membrane proteins 3](#_Toc162616848)

[1.2 Forces involved in membrane protein folding and association 5](#_Toc162616849)

[1.3 The importance of van der Waals packing 5](#_Toc162616850)

[1.4 Studying membrane protein folding and structure 7](#_Toc162616851)

[1.4.1 Structural methods 7](#_Toc162616852)

[1.4.2 Quantitative methods 8](#_Toc162616853)

[1.4.3 ToxR genetic reporter assays 8](#_Toc162616854)

[1.4.4 Computational methods 8](#_Toc162616855)

[1.4.5 Works at the experimental and computational interface 8](#_Toc162616856)

[1.5 GASright 9](#_Toc162616857)

[1.6 Thesis overview 10](#_Toc162616858)

[1.7 References 10](#_Toc162616859)

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

Membrane proteins (MP) are a vast category of proteins found tethered to the cellular membrane. MPs comprise 25-30% of the proteins found within protein-coding genes of various organisms (Fagerberg et al., 2010). Some are found bound to the edges of the membrane (peripheral) while others are embedded into the membrane (integral). This membrane is a bilayer composed of hydrophobic small molecules known as lipids, separating the cell from the outside environment. Integral MPs receive and orchestrate interactions between cells and the outside environment, relaying sensory information (G-protein coupled receptors), regulating pH and ion balance (Na+-K+ pumps), and promoting signaling cascades to regulate expression of important genes (receptor tyrosine kinases). However, the hydrophobic nature of this lipid bilayer environment results in a complex network of interactions responsible for MP folding and stability. Misfolding of MPs affects not only the aforementioned interactions but 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). To fully understand how MPs are involved in essential biological functions and to combat the progression of diseases, it is necessary to learn how and why MPs reach their folded state.

However, studying MP folding is inherently a difficult challenge because of their hydrophobic nature. MPs are difficult to express in high yields 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, a daunting task to replicate these biological conditions either *in vitro* or *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 cell research, aiming to explore and understand the impact of biophysical forces governing the folding process in a natural environment.

My graduate work focuses on developing an *in silico* algorithm to design novel single-pass membrane proteins known as homodimers, then using high-throughput experiments to test and validate my computational models. In this introduction, I expand on the forces found to be involved in membrane protein association, with a particular focus on the lack of knowledge on the impact of van der Waals packing. I then introduce current experimental and computational methods that inspired my research. I then highlight an important sequence and structural motif GASright, an important control for my research in developing information on van der Waals packing.

* 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. First, MPs must be translated and inserted properly into the membrane. This is accomplished through the aid of the translocon, a protein that in conjunction with the ribosome properly translates MPs into the proper orientation 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, composed of 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 separated into two groups: beta barrels, large proteins composed of repeating layers of beta sheets, and alpha helices, proteins twisted in a coil like helix structure. My research focuses on the latter group of proteins.

As alpha helices are threaded by the translocon into the membrane, the process to bring the MP to its folded state is regulated by a distribution of hydrogen bonds, weak polar and electrostatic interactions, and van der Waals forces between the unfolded and folded states. Hydrogen bonding regulates secondary structure within TM insertion, but hydrogen bonding between polar amino acids on opposing helices have been shown to drive membrane protein folding (Johnson et al., 2007; Yano et al., 2002; Zhou et al., 2001). Additionally, electrostatic interactions between positively charged residues (Lys) and electronegative ring amino acids (Tyr, Trp, Phe) results in association between helices (Johnson et al., 2007). However, not many studies have solely accounted for van der Waals forces between atoms coming in close contact.

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

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, maybe bring up that image that Alessandro has).

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. Lipid-lipid packing involves the interactions between individual lipid molecules that are necessary to keep the bilayer assembled. Lipid-protein packing occurs between these lipid molecules and parts of proteins exposed to the lipid environment. Protein-protein (or sidechain) packing focuses on the stability gained between individual or portions of proteins in close contact. Each of these forces plays a role in the ability for a protein to be stabilized in the bilayer. When an individual protein subunit is inserted into the membrane, it must destabilize the lipid-lipid packing with more favorable lipid-protein packing interactions. For protein-protein packing to occur, these newly formed lipid-protein interactions must be destabilized for a more favorable combination of protein-protein packing and lipid-lipid packing. 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 sidechain packing within the core of bacteriorhodopsin destabilizes protein structure (Faham et al., 2004; Joh et al., 2009). In addition, a recent study using MP design has shown that optimized sidechain packing can stabilize the folded state of the 5-helix bundle protein phospholamban (Mravic et al., 2019). Although these studies suggest that sidechain packing plays a significant role in stabilizing MP structure, there has not been much investigation of the energetic contribution of packing outside of individual MP systems. My research aims to characterize and quantify the extent to which sidechain packing is a driving force for MP folding for the general population of MP structures.

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

The complex nature of the lipid bilayer makes it difficult to directly study the forces involved in MP folding. Many groups focus on utilizing a combination of experimental and computational tools to estimate the contribution that these biophysical forces have on MP stability. In particular, thermodynamic experiments return results that can be relayed into information regarding the stability of proteins in specific conditions. This thermodynamic information can be used to assess and validate computational models, which can be further implemented to design and engineer novel proteins.

In this section, I will summarize different experimental and computational approaches that have been used to study single pass MPs. After this I further explore recent studies at the interface of these methods that use both experiments and computation to further understand these forces that govern membrane protein folding and stability.

****1.4.1 Structural methods****

In 1995, the protein data bank (PDB) was established to collaborate and share protein structures globally online. This tool allows researchers to deposit structures alongside literature, allowing people from around the world to access and assess their findings. Initially, proteins structures were solved primarily using x-ray crystallography. However, …limits of x-ray for MPs. Thousands of other MP structures have been solved by nuclear magnetic resonance (NMR). … (how many of those are single pass?). More recently, techniques such as solid-state NMR and cryo-EM have been used to study MP structures within membrane mimetics such as liposomes and nanodisks. This allows the MP to be studied with the expectation that they will more closely resemble the structure in its native membrane environment. However, structural characterization of small MPs with a singular TM remains difficult despite these advancements in structural determination.

1.4.2 Quantitative methods

FRET

Mass spec?

In vitro?

1.4.3 ToxR genetic reporter assays

Method of choice for single pass TMs?

Multiple variations: GALLEX, TOXCAT, TOXGREEN, high throughput

1.4.4 Computational methods

In conjunction with structural determination, many groups focus on understanding MP structure and folding through the evaluation of previously solved MP structures.

Rosetta, MSL, etc.; docking …

Alphafold and history of how

New methods to predict: AI?

1.4.5 Works at the experimental and computational interface

With the rise of computational throughput over my time in graduate school, much research is now being done at the interface between experiments and computation. Groups utilize a combination of both experimental and computational tools to evaluate and assess their research and results. Namely, the rise of Alphafold as a confident prediction of protein structures has resulted in it being used to inform … for a variety of studies (cites). Additionally, with the ability to analyze data in high-throughput has led to technological developments in experimentation, such as … However, MP … limited. In Chapter 3, I will go into detail on …

* 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 found near 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 using *in vivo* TOXCAT, they showed that structures with more Cα–H bonds had a higher stability in the dimer state (Anderson et al., 2017). Additionally, Díaz-Vázquez et al. measured the free energy of association of GASright structures using *in vitro* FRET, concluding that the thermodynamic stability of GASright proteins is well correlated with the *in vivo* TOXCAT (Díaz Vázquez et al., 2023). Both studies posit that the primary forces involved in GASright association are a combination of hydrogen bonding and van der Waals packing. Because the goal of my research is to understand the extent at which packing can drive proteins to associate, this motif is an excellent control to evaluate my designed sequences against for their propensity to associate 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 the majority of my graduate school work to be published in the near future. In this paper, I determined that van der Waals packing is a weak driving force that leads to association of MP homodimers. I data mined the PDB for all solved MP structures to determine the best structures for computational design, developed a protein design algorithm and designed 1000s of proteins, and assessed the ability of proteins designed with solely van der Waals packing for their ability to associate using a high-throughput assay. I found that my designed proteins in regions outside of the GASright mildly associate in comparison to GASrights that associate according to both hydrogen bonding and van der Waals packing.

**In Chapter 3**, I discuss research at the interface of experiments and computation in detail, citing the inspirations for data mining, 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 my PhD journey through a chapter written for the public. I summarize the goals of my research in leyman’s terms while reflecting on the lessons that graduate school imparted onto me, giving my fully transparent thoughts on how my research affected my physical, emotional, and mental well-being. I showcase a bit of creativity in sharing these thoughts on how science and research has helped me grow during my time in graduate school.

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

Anderson, S. M., Mueller, B. K., Lange, E. J., & Senes, A. (2017). Combination of Cα-H Hydrogen Bonds and van der Waals Packing Modulates the Stability of GxxxG-Mediated Dimers in Membranes. *J Am Chem Soc*, *139*(44), 15774-15783. <https://doi.org/10.1021/jacs.7b07505>

Díaz Vázquez, G., Cui, Q., & Senes, A. (2023). Thermodynamic analysis of the GAS. *Biophys J*, *122*(1), 143-155. <https://doi.org/10.1016/j.bpj.2022.11.018>

Johnson, R. M., Hecht, K., & Deber, C. M. (2007). Aromatic and cation-pi interactions enhance helix-helix association in a membrane environment. *Biochemistry*, *46*(32), 9208-9214. <https://doi.org/10.1021/bi7008773>

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>

Yano, Y., Takemoto, T., Kobayashi, S., Yasui, H., Sakurai, H., Ohashi, W., . . . Matsuzaki, K. (2002). Topological stability and self-association of a completely hydrophobic model transmembrane helix in lipid bilayers. *Biochemistry*, *41*(9), 3073-3080. <https://doi.org/10.1021/bi011161y>

Zhou, F. X., Merianos, H. J., Brunger, A. T., & Engelman, D. M. (2001). Polar residues drive association of polyleucine transmembrane helices. *Proc Natl Acad Sci U S A*, *98*(5), 2250-2255. <https://doi.org/10.1073/pnas.041593698>