Chapter 1: Introduction

Table of Contents

[1.1 Introduction to membrane proteins 3](#_Toc163915844)

[1.2 Reviewing membrane protein folding 5](#_Toc163915845)

[1.2.1 The Two-stage model 6](#_Toc163915846)

[1.2.2 Determining driving forces in membrane protein folding 7](#_Toc163915847)

[1.2.3 Using computational tools to study membrane protein association 10](#_Toc163915848)

[1.3 Understanding van der Waals as a driving force 12](#_Toc163915849)

[1.4 GASright 17](#_Toc163915850)

[1.5 Thesis overview 19](#_Toc163915851)

[1.6 References 21](#_Toc163915852)

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

The cellular membrane is a bilayer composed of phospholipid molecules, separating DNA and internal features of the cell from the outside environment. Communication with the outside of the cell and sensing of external factors that impact the cell’s survival is primarily done via a class of proteins tethered to the cell membrane called membrane proteins (MPs). MPs comprise 25-30% of the proteins found within protein-coding genes (Fagerberg et al., 2010). MPs can be broken down into two groups: peripheral MPs are bound and localized to the edges of the cell membrane while integral MPs are embedded into the lipid bilayer. Integral MPs properly regulate cell homeostasis by folding into unique structures that span the length of the cell membrane. β-barrels are comprised of β-sheets that form open pores in the membrane, functioning as channels and transporters, allowing molecules to pass through the bilayer to enter or exit the cell. Multi-pass and single-pass MPs are made of TM α-helices that can associate into alternative structures in response to environmental stimuli, signaling the cell to respond using genetic signaling cascades. My research focuses on this model system of TMHs to better understand how changes in the system affect association.

Phospholipids are amphipathic molecules, composed of two distinct components: hydrophilic heads and hydrophobic tails. To form the bilayer, hydrophobic tails are sandwiched between the hydrophilic heads exposed to the soluble cell cytoplasm and outer environment. This sequestering of the hydrophobic portion of phospholipids into the center of the membrane yields an environment devoid of the strong polar interactions that drive soluble protein folding. For MPs to fold, they must strike a delicate balance of interactions between proteins and lipids within this hydrophobic environment. Mutations within MPs can drastically affect these interactions, preventing MPs from folding properly. Misfolding of transmembrane helices 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 to combat the progression of these diseases, it is necessary to understand the impact of the individual forces governing the folding process. However, studying MPs is inherently a difficult challenge. 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). Researchers alternatively have focused on using model systems of MPs to better understand folding.

My graduate work focused on developing an *in silico* protein design algorithm to study the association of single-pass transmembrane α-helix homodimers, complemented with high-throughput experiments to validate my computational models. In this introduction, I review the contributions made to understanding MP folding up until my research. I detail the forces and interactions involved in the membrane that contribute to MP folding, emphasizing the deficiency of research on the contribution of van der Waals packing. I then highlight the sequence and structural motif GASright, an important control for my research used to juxtapose differences between association by alternate forces outside of van der Waals packing.

* 1. Reviewing membrane protein folding

The elaborate nature of the lipid bilayer makes it difficult to directly study the forces involved in MP folding. As an initial approach, researchers aim to solve the structures of MPs to identify any structural features necessary for the folded state. However, solving MP structures is an inherently difficult task due to the need to express and solubilize MPs for experiments. Alternative approaches to studying folding utilize a combination of *in vitro* and *in vivo* experimental tools aiming to determine the rules that govern transmembrane (TM) folding. Folding of integral TM proteins involves a variety of energetic constraints resulting from the hydrophobic nature of the phospholipid bilayer. First, TMs must be translated and inserted into the membrane. The translocon complex assists the ribosome in the translation of MPs into the proper orientation in the bilayer. Once the protein is embedded into the membrane, the hydrophobic core of the bilayer yields an energetic penalty for any unpaired hydrogen bond donors and acceptors (Marinko et al., 2019; Popot & Engelman, 1990, 2000). TMs are inherently made up of amide nitrogens and carbonyl oxygens within the protein backbone. To satisfy the lack of hydrogen bonding within the hydrophobic membrane, TMs adopt standard α-helical and β-sheet structures where hydrogen bonds form along the protein backbone. Second, individual TM domains interact to stabilize the folded state. Many experiments and tools have been developed to tease out the interactions in this second stage. These tools strive to identify changes in stability between the unfolded and folded states, allowing researchers to uncover MP folding thermodynamics within a variety of systems. Furthermore, this thermodynamic information can be used to assess and validate computational models, which can be further implemented to design and engineer novel proteins. My research focuses on understanding the forces involved in the folding and association of TM α-helices (TMH). In this section, I will summarize the different experimental approaches that have been used to study the association and folding of MPs, with a particular focus on single pass TMHs.

1.2.1 The Two-stage model

Early MP research focused on identifying membrane embedded regions within proteins using hydrophobicity analysis: navigating through the amino acid (AA) sequence and scanning for stretches of hydrophobic AAs (Kyte & Doolittle, 1982). Hydrophobicity analysis was successful in predicting the helices in both bacterial photosynthetic reaction centers and bacteriorhodopsin (Engelman et al., 1982; Michel et al., 1986). It was further developed to determine a charge bias known as the positive-inside rule where charged AAs are likely to be found at outside of the membrane (von Heijne, 1992). These tools have been further developed (Wilkins et al., 1999) and were initially important to characterize TM regions of protein sequences.

In 1990, Popot and Engelman proposed a model for membrane protein folding called the two-stage model: As the protein is threaded into the membrane, TMHs begin to form (stage 1) separate from forming a completely folded protein (stage 2). TMHs first reach a thermodynamic equilibrium with the lipid environment before undergoing stage 2, where individual TMHs oligomerize, assembling into the folded protein (Popot & Engelman, 1990, 2000). While stage 1 is driven by the hydrophobic effect working to coordinate the ribosome and translocon to insert hydrophobic stretches of protein sequence into the membrane, stage 2 of MP folding is governed by an interplay. Research on bacteriorhodopsin gave credence to stage 2: denaturing two separate fragments of the protein and resuspending them in lipid vesicles results in an active, folded protein (Popot et al., 1987); two chemically synthesized TMHs of the protein were reconstituted in lipid vesicles with a larger fragment, resulting in the reformation of the bacteriorhodopsin shown by X-ray crystallography (Kahn & Engelman, 1992); and extraction and reconstitution of individual helices of the protein were found to yield activity (Marti, 1998). Further research on larger protein complexes continued to push the field forward, demonstrating that mutating the hydrophobic core of four-helix-bundle protein Rop and five-helix-bundle protein phospholamban decreases the stability of both proteins (Arkin et al., 1994; Munson et al., 1996). With reassembly of MPs being an effective model for studying MP folding, other groups continued to build on this research by exploring model systems of TMHs, aiming to determine how minute changes in sequence and structure influence stability.

To investigate MP folding in model systems, researchers have developed tools to study oligomerization of TMHs. This process contains biological relevance, as many MP receptors are activated through ligand binding, including epidermal growth factor receptors and other proteins involved in tyrosine kinase signaling cascades that influence cell gene expression (Kumari & Yadav, 2019). Much of the early research investigating early TMH oligomerization focused on a protein found in red blood cells, glycophorin A (GpA). GpA is a single-pass TM that self-associates forming a dimer, the simplest type of oligomer between TMHs. Multiple *in vitro* studies worked to define the interface of GpA, making point mutations along the protein and measuring the changes in dimerization on SDS-PAGE (Lemmon, Flanagan, Hunt, et al., 1992; Lemmon, Flanagan, Treutlein, et al., 1992). Using dimerization as a model system, researchers aimed to further characterize the thermodynamics of dimerization by monitoring changes in stability between the monomer and dimer state. Additional studies using *in vitro* techniques sedimentation equilibrium analytical ultracentrifugation (SE-AUC) and Förster resonance energy transfer (FRET) were able to determine differences in stability, effectively quantifying the thermodynamics of association for GpA (Fisher et al., 1999; Fleming et al., 1997).

Near the turn of the century, MP studies began to push the field forward to increase understanding of MP folding beyond the simplicity of the two-stage model. As an early advance passed the two-stage model, a third stage was considered taking into account the thermodynamic impact of ligand binding domains, folding of loops outside of the membrane, and inserting other domains into the bilayer (Engelman et al., 2003). Further studies began to characterize the impact of polar residues on TMH association: poly-leucine helices in SDS-PAGE exhibit association with polar mutations (Zhou et al., 2000; Zhou et al., 2001), synthetic model TMH based on the GCN4 leucine zipper in detergent micelles and biological membranes associate (Zhou et al., 2000), and the AA Asn was shown to mediate association in multiple TMHs (Choma et al., 2000; Gratkowski et al., 2001). Rather than simply focusing on how the bulk changes in forces, researchers began to attempt to characterize the impact of the individual AAs and the respective forces that drive folding.

1.2.2 Determining driving forces in membrane protein folding

A benefit of the simplicity of the two-stage model is that the forces involved in each stage of folding are unique. In the first state, insertion of the protein into the membrane is driven by the hydrophobic effect, where MPs are more stable found in the membrane than in the soluble environment. When MPs are being translated by the ribosome, a signal sequence on the protein directs translation to the translocon (Dalbey et al., 2011). Together, the translocon and ribosome individually thread hydrophobic segments of the protein into the membrane (Hessa et al., 2005; Rapoport, 2007).

In the subsequent stage of folding, van der Waals packing, electrostatics, hydrogen bonding, and weak polar interactions contribute to guide the MP to the folded state. After the structure of GpA was solved by solution nuclear magnetic resonance (NMR), groups further characterized their previous thermodynamic data (MacKenzie et al., 1997). Analysis of GpA dimerization in MD simulations and mutations on the NMR structure determined that mutations affect association resulting from changes in van der Waals interactions (Fleming et al., 1997; MacKenzie & Engelman, 1998; MacKenzie et al., 1997; Petrache et al., 2000). Additional research on polar residues suggested that hydrogen bonding and polar interactions plays a stabilizing role in MP folded structures (Choma et al., 2000; Gratkowski et al., 2001; Zhou et al., 2000; Zhou et al., 2001). However, much of the previous research up to this point was done by monitoring the reversible folding of MPs using SDS-PAGE, FRET, and SE-AUC, *in vitro* techniques using detergents to mimic the environment of the membrane while simultaneously denaturing the protein. Additionally, these techniques are relatively low throughput, only allowing for characterization of a few proteins at a time. Newer techniques were developed to combat these issues while teasing out the impact of driving forces in MPs.

Disulfide cross-linking has been used to measure TMH oligomerization in both micelles and lipid vesicles (Cristian et al., 2003). Pulse proteolysis quantitatively measures the thermodynamic stability of MPs by selectively denaturing unfolded MPs using urea, GdnHCl, pH, SDS, or heat and subsequently digesting the unfolded MP (Park & Marqusee, 2005). Bacteriorhodopsin folding was studied using this technique and found to be dependent on changing concentrations of mixed micelles, which was not determined previously (Schlebach et al., 2012; Schlebach et al., 2011). Steric trapping utilizes the streptavidin-biotin binding system to measure the binding affinity of associating TMs in lipid bilayers, and it has been used to further determine the kinetics of GpA folding and mutants affecting its association (Blois et al., 2009; Hong & Bowie, 2011; Hong et al., 2013; Howarth et al., 2006). These methods allow for studying MP folding thermodynamics in lipids and mixed micelles, closer to the environments of the cell membrane compared to previous research. While these techniques approach understanding proteins in native environments, other techniques were developed to further study MPs within cells in higher throughput.

*In vivo* genetic reporter assays have been utilized to investigate the folding and association of MPs in their natural environment. Cells express MPs of interest fused to a DNA binding domain that can either inhibit or promote transcription of a reporter gene. GALLEX is a two-hybrid system where TMs are fused to DNA binding domain LexA. Association of the TMs results in repression of the β-galactosidase gene (Schneider & Engelman, 2003). Other reporter assay systems have utilized a chimera of the MPs of interest fused to ToxR, dimeric transcription factor, to promote the expression of genes (Gurezka & Langosch, 2001; Russ & Engelman, 1999).

The assay TOXCAT has been used to study TM helix-helix interactions, where the TM of interest is fused to the dimeric transcription factor ToxR. When the TMs associate, ToxR dimerizes and promotes the expression of chloramphenicol acetyltransferase (CAT) which can be measured to determine the strength of association. TOXCAT demonstrated that mutations of polar residues on GpA in the natural membrane environment yields different results than the previous *in vitro* studies (Russ & Engelman, 1999; Zhou et al., 2000; Zhou et al., 2001). Johnson et al. expanded on these findings, suggesting that electrostatic interactions between charged and ring AAs facilitates oligomerization (Johnson et al., 2007). TOXCAT has also been shown to correlate to changes in the free energy of association of GpA and point mutations (Duong et al., 2007). Anderson et al. used TOXCAT to study the association of GpA and similar TMHs, suggesting that these proteins associate via a combination of hydrogen bonding and van der Waals interactions (Anderson et al., 2017). TOXCAT is a well-studied system for probing TMH association, determining the impact of individual AAs and their respective forces on the thermodynamics of association.

Recently, TOXCAT has been adapted into an alternative high-throughput assay TOXGREEN. The reporter gene CAT has been replaced with green fluorescent protein (GFP), allowing fluorescent readings to be used to assess the association levels of the TMs of interest and their corresponding mutants (Armstrong & Senes, 2016). The development of this assay allows it to be utilized in high-throughput applications such as fluorescence activated cell sorting (FACS), where a library of TMs can be expressed, sorted, and sequenced through next generation sequencing (NGS). The sequencing data can then be quantified to determine the relative association propensities for each protein present in the library. My research utilizes this technique to quantify the association for proteins that I computationally designed along with interfacial mutants predicted to break association. Using sort-seq, I aimed to determine the impact of van der Waals packing on association (cite sort-seq).

1.2.3 Using computational tools to study membrane protein association

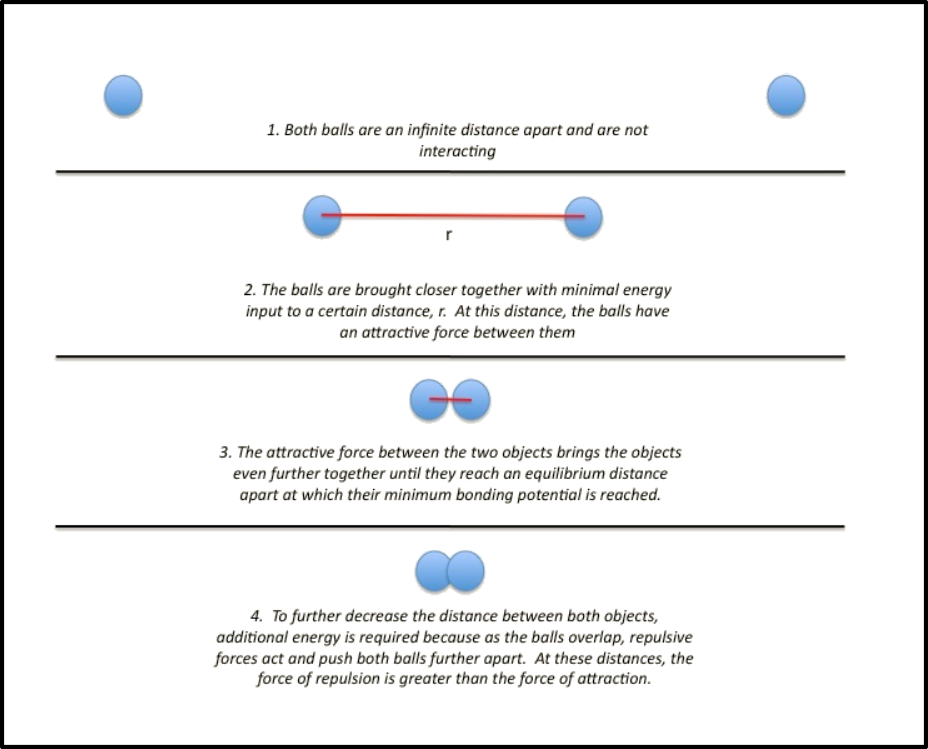
**Computational tools have been developed to help assess our understanding of the forces that drive MP association. These tools allow researchers to predict and design MPs *in silico* that can then be tested using the previously mentioned *in vitro* and *in vivo* experimental techniques. Utilizing protein design strategies is an impactful method for assessing MP folding using simple model systems (Ghirlanda, 2009). MP design to study TMH systems has been previously successful: peptides were engineered to associate with the TM helix of integrins and a cytokine receptor (Mravic et al., 2024; Shandler et al., 2011; Yin et al., 2007), an integral MP successfully transferred electrons across the lipid bilayer (Korendovych et al., 2010), a 4-helix bundle was designed to transport Zn2+ across the bilayer (Joh et al., 2014), and phospholamban was redesigned using packing interactions and shown to successfully fold (Mravic et al., 2019). By extracting backbone conformations from the PDB, sampling different AA conformations on the interface of a standardized backbone sequence, and assessing different conformations using known energetic functions, I designed thousands of TMHs to study in high throughput. I will further detail the computational tools used to study MP folding and detail my design algorithm in Chapter 3.**

****1.3 Understanding van der Waals as a driving force****

As individual TM helices are threaded into the membrane, an interplay of biophysical forces produces helix-helix association. This process is regulated by an intricate distribution of hydrogen bonding, electrostatic interactions, and van der Waals forces that govern the stabilities of the unfolded and folded states. Each of these interactions are driven by the types of amino acids (AAs) present within the TM. Removing hydrogen bonds within TMs decreases stability (Duong et al., 2007; Gratkowski et al., 2001; He & Hristova, 2008; Li et al., 2006; Stanley & Fleming, 2007). Hydrogen bonding not only regulates the secondary structure of TMs, but has been shown to drive TM helix association when polar amino acids (Ser, Thr) form interhelical hydrogen bonds between opposing helices (Johnson et al., 2007; Zhou et al., 2001). Additionally, electrostatic interactions between positively charged residues (Lys) and electronegative ring amino acids (Tyr, Trp, Phe) promotes association between helices (Johnson et al., 2007). Other charged and aromatic interactions have been shown to contribute similar stability as in water soluble proteins (Bañó-Polo et al., 2013; Burley & Petsko, 1985; Hong et al., 2009; Hong et al., 2007; Ulmschneider et al., 2017). However, hydrogen bonding and electrostatic interactions only highlight a subset of AAs typically present in MPs. The three AAs most frequently found in MPs (Leu, Ile, and Ala) are uncharged and lack the ability to hydrogen bond (Liu et al., 2002), thus TMs constituted of these AAs can only be stabilized by van der Waals forces.

Van der Waals occurs between atoms within close contact, including the interactions between MPs and the hydrophobic tails within the membrane. MP association motifs have been structurally characterized and studied, determining that tight van der Waals packing plays an important role in TMH association (Gurezka et al., 1999; Kim et al., 2004; MacKenzie et al., 1997; North et al., 2006; Russ & Engelman, 1999; Wu et al., 2005). Mutational studies on well packed residues in the core of MPs suggest that changes in packing can destabilize protein structure (Ash et al., 2004; Faham et al., 2004; Joh et al., 2009; Mravic et al., 2019; Yano et al., 2002). However, accounting for van der Waals between MPs and phospholipids is complex, and not many studies have successfully investigated the influence of van der Waals forces on MP stability.

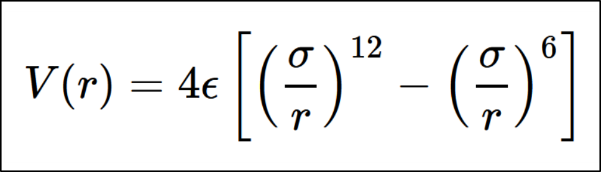
Figure 1: Van der Waals is an attractive force between two atoms at short distances.



*Adapted from chem.libretexts.org (Naeem)*

The physical properties of atoms are the foundation of intermolecular interactions. Atoms are composed of a nucleus of protons surrounded by an outer electron shell. The electron shell expands a finite distance away from the nucleus, constructing the space occupied by the atom, or the van der Waals radius. This is the closest distance another atom can approach without forming a chemical bond. When atoms are found at a distance smaller than their combined van der Waals radii, the opposing electron shells repulse, resulting in a strongly unfavorable interaction that pushes atoms away. However, as atoms approach this minimum distance, there is a weak attraction between protons in the nucleus of one atom and the electrons of another. This attraction is the core principle behind the van der Waals force: It is a favorable intermolecular interaction occurring between atoms in proximity (Figure 1).

The van der Waals force between two atoms can be calculated using the Lennard-Jones (LJ) Potential:



Eq. 1: The 12-6 Lennard-Jones Potential.

The LJ potential calculates the intermolecular potential (*V*) between two atoms at a specified distance (*r*), using the strength of attraction between the atoms (*ϵ*) and the distance where the potential is 0 (*σ*). This function expresses the repulsive force as (*σ*/r)12 while the attractive force is represented as (*σ*/*r*)6. As the atoms approach the minimum distance, there is a distance *r* that corresponds to the tightest attraction between the atoms *ϵ*.

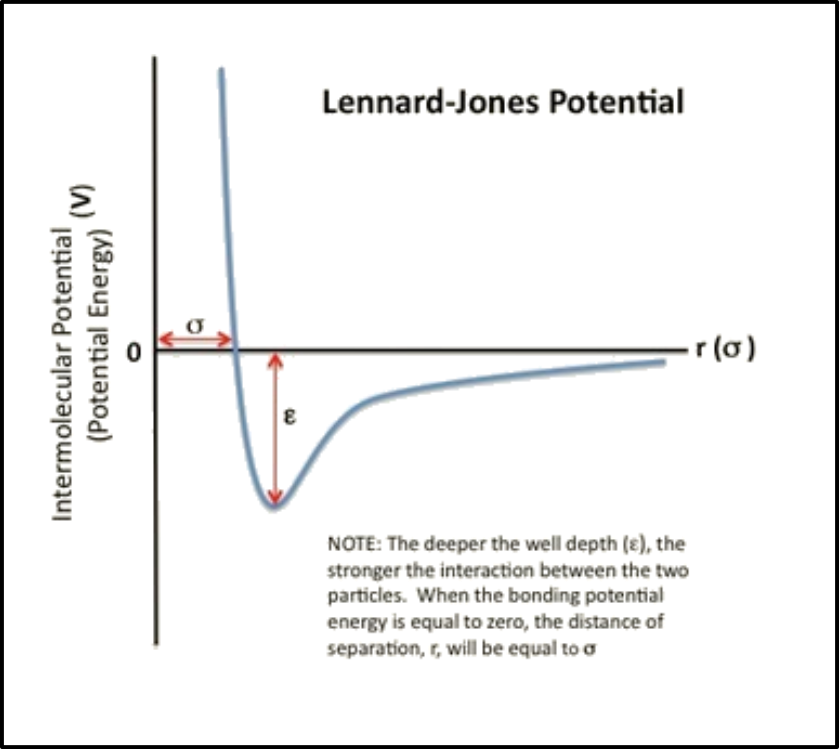


Figure 2: The Lennard-Jones potential, visualized.

*Adapted from chem.libretexts.org (Naeem)*

Soluble proteins are driven to fold by the hydrophobic effect, where nonpolar amino acids are forced to the core of the folded state. These nonbonded atoms at the core of these proteins are found in tight contact with one another, compounding into a multitude of weak van der Waals interactions known as van der Waals packing. Although van der Waals packing is not a driving force for soluble protein folding, it is a necessary force that is always present in the folded state. For MPs situated in the core of the hydrophobic membrane, the hydrophobic effect does not drive MP folding. This means that MPs must rely on other forces to reach the folded state. Although hydrogen bonding and polar interactions have been found to drive MP folding, the extent at which packing contributes to the folded state is unclear. Like soluble protein folding, MP folding eventuates van der Waals packing. But because MPs are engulfed within the crowded lipid bilayer, it is difficult to tease out the influence that van der Waals packing has on MP folding.

The contribution of van der Waals packing to MP folding can be broken down into three distinct interactions: lipid-lipid packing, lipid-protein packing, and protein-protein packing. Lipid-lipid packing involves individual lipid molecules nudged tightly against each other to keep the bilayer assembled. Lipid-protein packing occurs between these lipid molecules and the lipid exposed protein shell. Protein-protein (or sidechain) packing focuses on the stability gained between fragments of proteins in close contact. Each of these interactions plays a role stabilizing an MP 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. This assortment of packing interactions takes place to keep the lipid bilayer intact while the MP reaches its folded state. But simultaneously accounting for all these interactions within the thermodynamics of MP folding is impractical using current technologies. Protein-protein (or sidechain) packing is a technically feasible starting point because of the ability to manipulate protein sequence and structure within a controlled environment.

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 role in stabilizing MP structure, there has not been much investigation of the thermodynamic 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.

****1.4 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 AAs 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. GASright proteins are frequently found to be involved in a variety of diseases: syndecan-2 overexpression has been found in colorectal cancer cell lines, neuropilin-1 has been shown to intensify symptoms of SARS-CoV-2, and Glycophorin A (GpA) misregulation is involved in sickle cell disease (Benedicto et al., 2021; Marshall et al., 2024; Vicente et al., 2013). Due to the prevalence of GASright proteins in medical applications as well as its well-defined sequence and structural features, many groups study these proteins to further understand the forces governing TM association.

The GASright motif’s unique sequence and defined structure has been shown to permit an uncommon structural feature. Small amino acids at the interface allows TM backbones to associate with a short interhelical distance, resulting in the formation of 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 they are less electronegative than typical nitrogen and oxygen donors. However, these carbons 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 the structures of GASright TMs found in natural sequences and testing their stability using *in vivo* TOXCAT, they showed that structures predicted to have more Cα–H bonds are more thermodynamically stability (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 *in vivo* experiments (Díaz Vázquez et al., 2023). These studies suggest that GASright proteins associate primarily through two forces: hydrogen bonding and van der Waals packing. By evaluating sequences designed to associate through solely van der Waals packing, I can differentiate the impact of packing (designs) versus both hydrogen bonding and packing (GASright) on association.

* 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 schoolwork 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 packing drives my designed proteins to associate mildly when compared to GASright proteins associating according to both hydrogen bonding and van der Waals packing.

**In Chapter 3**, I discuss the computational methods that have been used to study MPs in detail. My research aimed to utilize these growing technologies to discover the potential impact of biophysical forces on protein association. With improving experimental technologies, many other studies at the forefront of research utilize a combination of high-throughput experiments and computational analysis. I review similar studies and rationalize their impact on the field. Finally, I detail the computational inspirations for my project that involves data mining of the PDB and discuss the development of my protein design algorithm.

**In Chapter 4**, I describe a variety of future directions for studying van der Waals packing and other uses for my protein design algorithm. I discuss how to expand my protein design algorithm, detailing how to design heterodimers. Finally, I explore how my algorithm can be improved using 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 describe the premise of my research in leyman’s terms while simultaneously reflecting on lessons that I learned throughout my time in graduate school, giving 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>

Arkin, I. T., Adams, P. D., MacKenzie, K. R., Lemmon, M. A., Brünger, A. T., & Engelman, D. M. (1994). Structural organization of the pentameric transmembrane alpha-helices of phospholamban, a cardiac ion channel. *EMBO J*, *13*(20), 4757-4764. <https://doi.org/10.1002/j.1460-2075.1994.tb06801.x>

Armstrong, C. R., & Senes, A. (2016). Screening for transmembrane association in divisome proteins using TOXGREEN, a high-throughput variant of the TOXCAT assay. *Biochim Biophys Acta*, *1858*(11), 2573-2583. <https://doi.org/10.1016/j.bbamem.2016.07.008>

Ash, W. L., Stockner, T., MacCallum, J. L., & Tieleman, D. P. (2004). Computer modeling of polyleucine-based coiled coil dimers in a realistic membrane environment: insight into helix-helix interactions in membrane proteins. *Biochemistry*, *43*(28), 9050-9060. <https://doi.org/10.1021/bi0494572>

Bañó-Polo, M., Martínez-Gil, L., Wallner, B., Nieva, J. L., Elofsson, A., & Mingarro, I. (2013). Charge pair interactions in transmembrane helices and turn propensity of the connecting sequence promote helical hairpin insertion. *J Mol Biol*, *425*(4), 830-840. <https://doi.org/10.1016/j.jmb.2012.12.001>

Benedicto, A., García-Kamiruaga, I., & Arteta, B. (2021). Neuropilin-1: A feasible link between liver pathologies and COVID-19. *World J Gastroenterol*, *27*(24), 3516-3529. <https://doi.org/10.3748/wjg.v27.i24.3516>

Blois, T. M., Hong, H., Kim, T. H., & Bowie, J. U. (2009). Protein unfolding with a steric trap. *J Am Chem Soc*, *131*(39), 13914-13915. <https://doi.org/10.1021/ja905725n>

Burley, S. K., & Petsko, G. A. (1985). Aromatic-aromatic interaction: a mechanism of protein structure stabilization. *Science*, *229*(4708), 23-28. <https://doi.org/10.1126/science.3892686>

Choma, C., Gratkowski, H., Lear, J. D., & DeGrado, W. F. (2000). Asparagine-mediated self-association of a model transmembrane helix. *Nat Struct Biol*, *7*(2), 161-166. <https://doi.org/10.1038/72440>

Cristian, L., Lear, J. D., & DeGrado, W. F. (2003). Use of thiol-disulfide equilibria to measure the energetics of assembly of transmembrane helices in phospholipid bilayers. *Proc Natl Acad Sci U S A*, *100*(25), 14772-14777. <https://doi.org/10.1073/pnas.2536751100>

Dalbey, R. E., Wang, P., & Kuhn, A. (2011). Assembly of bacterial inner membrane proteins. *Annu Rev Biochem*, *80*, 161-187. <https://doi.org/10.1146/annurev-biochem-060409-092524>

Duong, M. T., Jaszewski, T. M., Fleming, K. G., & MacKenzie, K. R. (2007). Changes in apparent free energy of helix-helix dimerization in a biological membrane due to point mutations. *J Mol Biol*, *371*(2), 422-434. <https://doi.org/10.1016/j.jmb.2007.05.026>

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>

Engelman, D. M., Chen, Y., Chin, C. N., Curran, A. R., Dixon, A. M., Dupuy, A. D., . . . Popot, J. L. (2003). Membrane protein folding: beyond the two stage model. *FEBS Lett*, *555*(1), 122-125. <https://doi.org/10.1016/s0014-5793(03)01106-2>

Engelman, D. M., Goldman, A., & Steitz, T. A. (1982). [11] The identification of helical segments in the polypeptide chain of bacteriorhodopsin. In *Methods in Enzymology* (Vol. 88, pp. 81-88). Academic Press. <https://doi.org/https://doi.org/10.1016/0076-6879(82)88014-2>

Faham, S., Yang, D., Bare, E., Yohannan, S., Whitelegge, J. P., & Bowie, J. U. (2004). Side-chain contributions to membrane protein structure and stability. *J Mol Biol*, *335*(1), 297-305. <https://doi.org/10.1016/j.jmb.2003.10.041>

Fisher, L. E., Engelman, D. M., & Sturgis, J. N. (1999). Detergents modulate dimerization, but not helicity, of the glycophorin A transmembrane domain. *J Mol Biol*, *293*(3), 639-651. <https://doi.org/10.1006/jmbi.1999.3126>

Fleming, K. G., Ackerman, A. L., & Engelman, D. M. (1997). The effect of point mutations on the free energy of transmembrane alpha-helix dimerization. *J Mol Biol*, *272*(2), 266-275. <https://doi.org/10.1006/jmbi.1997.1236>

Ghirlanda, G. (2009). Design of membrane proteins: toward functional systems. *Current Opinion in Chemical Biology*, *13*(5), 643-651. <https://doi.org/https://doi.org/10.1016/j.cbpa.2009.09.017>

Gratkowski, H., Lear, J. D., & DeGrado, W. F. (2001). Polar side chains drive the association of model transmembrane peptides. *Proc Natl Acad Sci U S A*, *98*(3), 880-885. <https://doi.org/10.1073/pnas.98.3.880>

Gurezka, R., Laage, R., Brosig, B., & Langosch, D. (1999). A heptad motif of leucine residues found in membrane proteins can drive self-assembly of artificial transmembrane segments. *J Biol Chem*, *274*(14), 9265-9270. <https://doi.org/10.1074/jbc.274.14.9265>

Gurezka, R., & Langosch, D. (2001). In vitro selection of membrane-spanning leucine zipper protein-protein interaction motifs using POSSYCCAT. *J Biol Chem*, *276*(49), 45580-45587. <https://doi.org/10.1074/jbc.M105362200>

He, L., & Hristova, K. (2008). Pathogenic activation of receptor tyrosine kinases in mammalian membranes. *J Mol Biol*, *384*(5), 1130-1142. <https://doi.org/10.1016/j.jmb.2008.10.036>

Hessa, T., Kim, H., Bihlmaier, K., Lundin, C., Boekel, J., Andersson, H., . . . von Heijne, G. (2005). Recognition of transmembrane helices by the endoplasmic reticulum translocon. *Nature*, *433*(7024), 377-381. <https://doi.org/10.1038/nature03216>

Hong, H., & Bowie, J. U. (2011). Dramatic destabilization of transmembrane helix interactions by features of natural membrane environments. *J Am Chem Soc*, *133*(29), 11389-11398. <https://doi.org/10.1021/ja204524c>

Hong, H., Chang, Y.-C., & Bowie, J. U. (2013). Measuring Transmembrane Helix Interaction Strengths in Lipid Bilayers Using Steric Trapping. In *Membrane Proteins: Folding, Association, and Design* (pp. 37-56). Humana Press. <https://doi.org/10.1007/978-1-62703-583-5_3>

Hong, H., Joh, N. H., Bowie, J. U., & Tamm, L. K. (2009). Methods for measuring the thermodynamic stability of membrane proteins. *Methods Enzymol*, *455*, 213-236. <https://doi.org/10.1016/S0076-6879(08)04208-0>

Hong, H., Park, S., Jiménez, R. H., Rinehart, D., & Tamm, L. K. (2007). Role of aromatic side chains in the folding and thermodynamic stability of integral membrane proteins. *J Am Chem Soc*, *129*(26), 8320-8327. <https://doi.org/10.1021/ja068849o>

Howarth, M., Chinnapen, D. J., Gerrow, K., Dorrestein, P. C., Grandy, M. R., Kelleher, N. L., . . . Ting, A. Y. (2006). A monovalent streptavidin with a single femtomolar biotin binding site. *Nat Methods*, *3*(4), 267-273. <https://doi.org/10.1038/nmeth861>

Joh, N. H., Oberai, A., Yang, D., Whitelegge, J. P., & Bowie, J. U. (2009). Similar energetic contributions of packing in the core of membrane and water-soluble proteins. *J Am Chem Soc*, *131*(31), 10846-10847. <https://doi.org/10.1021/ja904711k>

Joh, N. H., Wang, T., Bhate, M. P., Acharya, R., Wu, Y., Grabe, M., . . . DeGrado, W. F. (2014). De novo design of a transmembrane Zn²⁺-transporting four-helix bundle. *Science*, *346*(6216), 1520-1524. <https://doi.org/10.1126/science.1261172>

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>

Kahn, T. W., & Engelman, D. M. (1992). Bacteriorhodopsin can be refolded from two independently stable transmembrane helices and the complementary five-helix fragment. *Biochemistry*, *31*(26), 6144-6151. <https://doi.org/10.1021/bi00141a027>

Kim, S., Chamberlain, A. K., & Bowie, J. U. (2004). Membrane channel structure of Helicobacter pylori vacuolating toxin: role of multiple GXXXG motifs in cylindrical channels. *Proc Natl Acad Sci U S A*, *101*(16), 5988-5991. <https://doi.org/10.1073/pnas.0308694101>

Korendovych, I. V., Senes, A., Kim, Y. H., Lear, J. D., Fry, H. C., Therien, M. J., . . . Degrado, W. F. (2010). De novo design and molecular assembly of a transmembrane diporphyrin-binding protein complex. *J Am Chem Soc*, *132*(44), 15516-15518. <https://doi.org/10.1021/ja107487b>

Kumari, N., & Yadav, S. (2019). Modulation of protein oligomerization: An overview. *Prog Biophys Mol Biol*, *149*, 99-113. <https://doi.org/10.1016/j.pbiomolbio.2019.03.003>

Kyte, J., & Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *J Mol Biol*, *157*(1), 105-132. <https://doi.org/10.1016/0022-2836(82)90515-0>

Lemmon, M. A., Flanagan, J. M., Hunt, J. F., Adair, B. D., Bormann, B. J., Dempsey, C. E., & Engelman, D. M. (1992). Glycophorin A dimerization is driven by specific interactions between transmembrane alpha-helices. *J Biol Chem*, *267*(11), 7683-7689.

Lemmon, M. A., Flanagan, J. M., Treutlein, H. R., Zhang, J., & Engelman, D. M. (1992). Sequence specificity in the dimerization of transmembrane alpha-helices. *Biochemistry*, *31*(51), 12719-12725. <https://doi.org/10.1021/bi00166a002>

Li, E., You, M., & Hristova, K. (2006). FGFR3 dimer stabilization due to a single amino acid pathogenic mutation. *J Mol Biol*, *356*(3), 600-612. <https://doi.org/10.1016/j.jmb.2005.11.077>

Liu, Y., Engelman, D. M., & Gerstein, M. (2002). Genomic analysis of membrane protein families: abundance and conserved motifs. *Genome Biol*, *3*(10), research0054. <https://doi.org/10.1186/gb-2002-3-10-research0054>

MacKenzie, K. R., & Engelman, D. M. (1998). Structure-based prediction of the stability of transmembrane helix-helix interactions: the sequence dependence of glycophorin A dimerization. *Proc Natl Acad Sci U S A*, *95*(7), 3583-3590. <https://doi.org/10.1073/pnas.95.7.3583>

MacKenzie, K. R., Prestegard, J. H., & Engelman, D. M. (1997). A transmembrane helix dimer: structure and implications. *Science*, *276*(5309), 131-133. <https://doi.org/10.1126/science.276.5309.131>

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>

Marshall, J. N., Klein, M. N., Karki, P., Promnares, K., Setua, S., Fan, X., . . . Fontaine, M. J. (2024). Aberrant GPA expression and regulatory function of red blood cells in sickle cell disease. *Blood Adv*, *8*(7), 1687-1697. <https://doi.org/10.1182/bloodadvances.2023011611>

Marti, T. (1998). Refolding of bacteriorhodopsin from expressed polypeptide fragments. *J Biol Chem*, *273*(15), 9312-9322. <https://doi.org/10.1074/jbc.273.15.9312>

Michel, H., Epp, O., & Deisenhofer, J. (1986). Pigment-protein interactions in the photosynthetic reaction centre from Rhodopseudomonas viridis. *EMBO J*, *5*(10), 2445-2451. <https://doi.org/10.1002/j.1460-2075.1986.tb04520.x>

Mravic, M., He, L., Kratochvil, H. T., Hu, H., Nick, S. E., Bai, W., . . . DeGrado, W. F. (2024). De novo-designed transmembrane proteins bind and regulate a cytokine receptor. *Nat Chem Biol*. <https://doi.org/10.1038/s41589-024-01562-z>

Mravic, M., Thomaston, J. L., Tucker, M., Solomon, P. E., Liu, L., & DeGrado, W. F. (2019). Packing of apolar side chains enables accurate design of highly stable membrane proteins. *Science*, *363*(6434), 1418-1423. <https://doi.org/10.1126/science.aav7541>

Munson, M., Balasubramanian, S., Fleming, K. G., Nagi, A. D., O'Brien, R., Sturtevant, J. M., & Regan, L. (1996). What makes a protein a protein? Hydrophobic core designs that specify stability and structural properties. *Protein Sci*, *5*(8), 1584-1593. <https://doi.org/10.1002/pro.5560050813>

Naeem, R. *Lennard-Jones\_Potential*. <https://chem.libretexts.org/Bookshelves/Physical_and_Theoretical_Chemistry_Textbook_Maps/Supplemental_Modules_(Physical_and_Theoretical_Chemistry)/Physical_Properties_of_Matter/Atomic_and_Molecular_Properties/Intermolecular_Forces/Specific_Interactions/Lennard-Jones_Potential>

North, B., Cristian, L., Fu Stowell, X., Lear, J. D., Saven, J. G., & DeGrado, W. F. (2006). Characterization of a Membrane Protein Folding Motif, the Ser Zipper, Using Designed Peptides. *Journal of Molecular Biology*, *359*(4), 930-939. <https://doi.org/https://doi.org/10.1016/j.jmb.2006.04.001>

Park, C., & Marqusee, S. (2005). Pulse proteolysis: a simple method for quantitative determination of protein stability and ligand binding. *Nat Methods*, *2*(3), 207-212. <https://doi.org/10.1038/nmeth740>

Petrache, H. I., Grossfield, A., MacKenzie, K. R., Engelman, D. M., & Woolf, T. B. (2000). Modulation of glycophorin A transmembrane helix interactions by lipid bilayers: molecular dynamics calculations. *J Mol Biol*, *302*(3), 727-746. <https://doi.org/10.1006/jmbi.2000.4072>

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>

Popot, J. L., Gerchman, S. E., & Engelman, D. M. (1987). Refolding of bacteriorhodopsin in lipid bilayers. A thermodynamically controlled two-stage process. *J Mol Biol*, *198*(4), 655-676. <https://doi.org/10.1016/0022-2836(87)90208-7>

Rapoport, T. A. (2007). Protein translocation across the eukaryotic endoplasmic reticulum and bacterial plasma membranes. *Nature*, *450*(7170), 663-669. <https://doi.org/10.1038/nature06384>

Russ, W. P., & Engelman, D. M. (1999). TOXCAT: a measure of transmembrane helix association in a biological membrane. *Proc Natl Acad Sci U S A*, *96*(3), 863-868. <https://doi.org/10.1073/pnas.96.3.863>

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>

Schlebach, J. P., Cao, Z., Bowie, J. U., & Park, C. (2012). Revisiting the folding kinetics of bacteriorhodopsin. *Protein Sci*, *21*(1), 97-106. <https://doi.org/10.1002/pro.766>

Schlebach, J. P., Kim, M. S., Joh, N. H., Bowie, J. U., & Park, C. (2011). Probing membrane protein unfolding with pulse proteolysis. *J Mol Biol*, *406*(4), 545-551. <https://doi.org/10.1016/j.jmb.2010.12.018>

Schneider, D., & Engelman, D. M. (2003). GALLEX, a measurement of heterologous association of transmembrane helices in a biological membrane. *J Biol Chem*, *278*(5), 3105-3111. <https://doi.org/10.1074/jbc.M206287200>

Shandler, S. J., Korendovych, I. V., Moore, D. T., Smith-Dupont, K. B., Streu, C. N., Litvinov, R. I., . . . DeGrado, W. F. (2011). Computational design of a β-peptide that targets transmembrane helices. *J Am Chem Soc*, *133*(32), 12378-12381. <https://doi.org/10.1021/ja204215f>

Stanley, A. M., & Fleming, K. G. (2007). The role of a hydrogen bonding network in the transmembrane beta-barrel OMPLA. *J Mol Biol*, *370*(5), 912-924. <https://doi.org/10.1016/j.jmb.2007.05.009>

Ulmschneider, M. B., Ulmschneider, J. P., Freites, J. A., von Heijne, G., Tobias, D. J., & White, S. H. (2017). Transmembrane helices containing a charged arginine are thermodynamically stable. *Eur Biophys J*, *46*(7), 627-637. <https://doi.org/10.1007/s00249-017-1206-x>

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.

Vicente, C. M., Ricci, R., Nader, H. B., & Toma, L. (2013). Syndecan-2 is upregulated in colorectal cancer cells through interactions with extracellular matrix produced by stromal fibroblasts. *BMC Cell Biol*, *14*, 25. <https://doi.org/10.1186/1471-2121-14-25>

von Heijne, G. (1992). Membrane protein structure prediction. Hydrophobicity analysis and the positive-inside rule. *J Mol Biol*, *225*(2), 487-494. <https://doi.org/10.1016/0022-2836(92)90934-c>

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>

Wilkins, M. R., Gasteiger, E., Bairoch, A., Sanchez, J. C., Williams, K. L., Appel, R. D., & Hochstrasser, D. F. (1999). Protein identification and analysis tools in the ExPASy server. *Methods Mol Biol*, *112*, 531-552. <https://doi.org/10.1385/1-59259-584-7:531>

Wu, T., Malinverni, J., Ruiz, N., Kim, S., Silhavy, T. J., & Kahne, D. (2005). Identification of a multicomponent complex required for outer membrane biogenesis in Escherichia coli. *Cell*, *121*(2), 235-245. <https://doi.org/10.1016/j.cell.2005.02.015>

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>

Yin, H., Slusky, J. S., Berger, B. W., Walters, R. S., Vilaire, G., Litvinov, R. I., . . . DeGrado, W. F. (2007). Computational design of peptides that target transmembrane helices. *Science*, *315*(5820), 1817-1822. <https://doi.org/10.1126/science.1136782>

Zhou, F. X., Cocco, M. J., Russ, W. P., Brunger, A. T., & Engelman, D. M. (2000). Interhelical hydrogen bonding drives strong interactions in membrane proteins. *Nat Struct Biol*, *7*(2), 154-160. <https://doi.org/10.1038/72430>

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>