Chapter 1: Introduction

Table of Contents

[1.1 Introduction to membrane proteins 3](#_Toc163402392)

[1.2 Forces involved in membrane protein folding and association 5](#_Toc163402393)

[1.3 The van der Waals force 7](#_Toc163402394)

[1.4 GASright 11](#_Toc163402395)

[1.5 Studying membrane protein folding and structure 13](#_Toc163402396)

[1.5.1 Structural methods 13](#_Toc163402397)

[1.5.2 Quantitative methods 14](#_Toc163402398)

[1.5.3 Genetic reporter assays 15](#_Toc163402399)

[1.5.4 Computational methods 16](#_Toc163402400)

[1.5.5 Working at the experimental and computational interface 18](#_Toc163402401)

[1.6 Thesis overview 20](#_Toc163402402)

[1.7 References 22](#_Toc163402403)

* 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 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 using single-pass MPs as a model to understand the forces involved in the association and folding of α-helical MPs.

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 α-helical 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). To fully understand how to combat the progression of these diseases, it is necessary to understand the impact of individual forces involved in folding. 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). Additionally, MPs can have alternate folded states in different compositions of lipids, which is difficult to appropriately replicate *in vitro* or *in silico*. To combat these challenges, much of the research studying MPs folding is focused on combining *in vitro* or *in silico* studies with live cell research, aiming to more deeply understand the biophysical forces that govern folding in the MPs natural environment.

My graduate work focused on developing an *in silico* protein design algorithm to study the association of single-pass MP homodimers, complemented with high-throughput experiments to validate my computational models. In this introduction, I expand on 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 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. I then review current experimental and computational methods and rationalize the growing research that combines these methods to study the biochemistry of MPs.

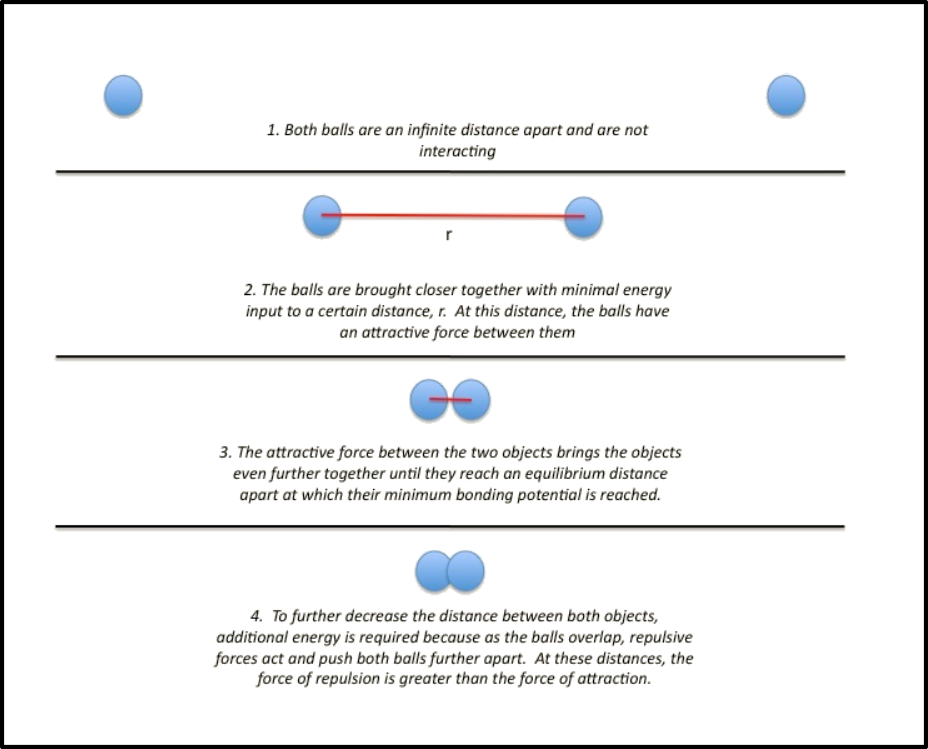
* 1. ****Forces involved in membrane protein folding and association****

Folding of integral transmembrane (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. My research focuses on understanding the forces involved in the folding and association of TM α-helices.

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. 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; Yano et al., 2002; 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). 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. 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.

* 1. ****The van der Waals force****

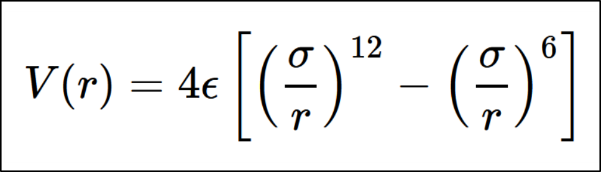
Figure 1: Van der Waals is an attractive force between two atoms at short distances.



*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*.

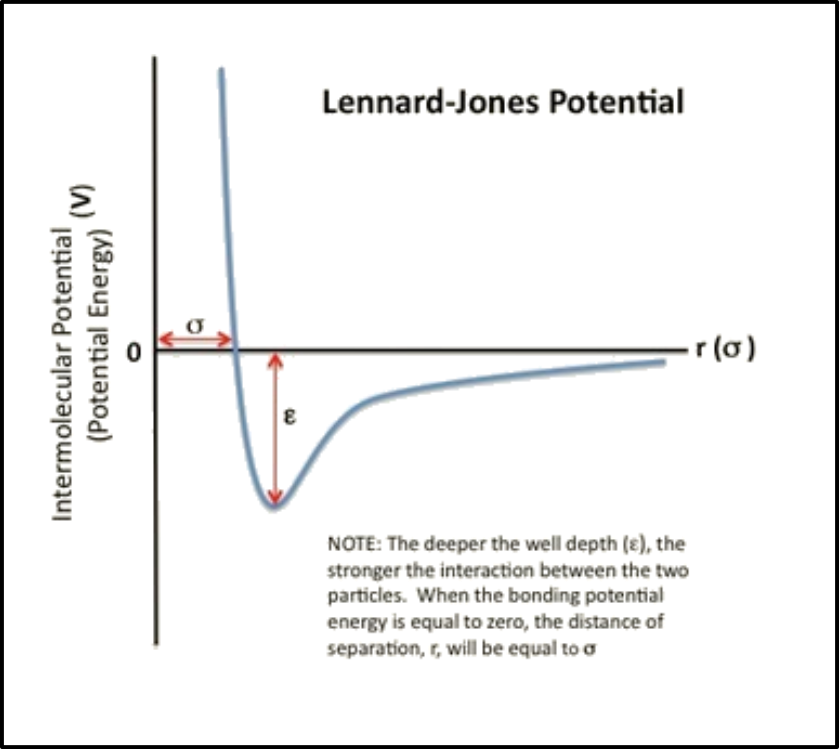


Figure 2: The Lennard-Jones potential, visualized.

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

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 determine the structures of MPs to identify structural features important for folding. However, solving MP structures is an inherently difficult task due to the need to express and solubilize MPs for experiments. Furthermore, many of these approaches cannot determine more than a single protein structure, making it difficult to use this information to understand the dynamic structural changes involved in folding. Many other groups focus on utilizing a combination of experimental and computational tools to estimate the contribution these biophysical forces have on MP stability. Thermodynamic experiments return results that can be relayed into information regarding the stability of proteins in specific conditions. 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 the different experimental and computational approaches that have been used to study the association and folding of single pass TMs, including the ability to solve and predict MP structures using both experimental and computational methods.

****1.5.1 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 protein structures after solving them to for others to access their findings. Initially, proteins structures were solved primarily using x-ray crystallography. X-ray crystallography has contributed to solving ~80% of MP structures (Kermani, 2021). However, producing crystals at high resolution remains a difficulty in crystallography. To bypass the need to crystallize proteins, many other MP structures have been solved by nuclear magnetic resonance (NMR). Solid-state NMR has bypassed the need for detergents in crystallography, obtaining structures of MPs less than 50 residues within lipid bilayers or nanodiscs (Liang & Tamm, 2016).

More recently cryo-EM has been used to solve MP structures. In addition to bilayers and nanodiscs, it is possible to solubilize and obtain the structures of MPs within detergents, saposin-lipoprotein nanoparticles, amphipols, and peptidiscs (Januliene & Moeller, 2021). Cryo-EM allows MP structures to be studied in a large variety of different environments, giving researchers the ability to study alternative structures of these proteins by changing the solubilization conditions. Despite the advancements in MP structural characterization, many of these efforts take years to determine suitable solubilization conditions that return structures in high resolution. MPs make up only 4.6% of structures deposited in the PDB (April 2024; PDB). Lacking MP structures to assess, researchers focus on using other methods to study MP folding and association.

1.5.2 Quantitative methods

To determine changes in folding or association states of proteins, many groups take an *in vitro* approach using fluorescence. One popular approach is Förster resonance energy transfer (FRET). In FRET, fluorescent proteins, or fluorophores, are first attached to individual proteins. These fluorophores emit when hit with different wavelengths of light, and the emission spectrum of a donor fluorophore overlaps with the excitation spectrum of an acceptor fluorophore. If the tagged proteins associate, the fluorophores’ proximity results in energy transfer from the donor to the acceptor, producing light from the acceptor fluorophore. Conversely, if there is no association then the fluorophores are far apart, resulting in light emitting from the donor fluorophore. Using FRET, researchers have determined the thermodynamics of association for a variety of proteins, including GpA (Díaz Vázquez et al., 2023; Fisher et al., 1999; Maeda et al., 2018; You et al., 2005).

Mass spectrometry (MS) of MPs is a growing field, with more groups using it to study the stoichiometry and dynamics of protein association. MPs are ionized within MS compatible detergent solutions, and spectra analyzed to assess the homogeneity of the MP in solution (Laganowsky et al., 2013). However, MS is traditionally used to analyze large protein complexes. Native MS is a complementary approach that can be used to analyze a larger range of proteins, including small TMs. This method preserves noncovalent interactions in MPs by using non-denaturing ionization conditions, with aqueous buffers near physiological pH at lower temperatures and voltages. By determining the composition and abundance of distinct complexes, MS is able to return interaction data reflecting MP oligomeric states (Keener et al., 2021).

Sedimentation equilibrium analytical ultra-centrifugation (SE-AUC) is a technique that has been used to study the association of TM helices by solubilizing MPs in detergent and centrifuging the solution at high speeds. By analyzing the concentration gradients that form at different centrifugation speeds, researchers can calculate the equilibrium constant for various oligomeric states. Association energetics were calculated for multiple mutants of GpA, determining that differences in association were due to the loss of van der Waals packing (Doura & Fleming, 2004; Fleming et al., 1997; Fleming & Engelman, 2001). However, these *in vitro* techniques rely on producing large quantities of MPs and solubilizing them in appropriate solutions for each application, which can alter MP structure. Alternative experiments rely on *in vivo* approaches to study association and folding of MPs.

1.5.3 Genetic reporter assays

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). Another reporter assay known as 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). TOXCAT has been utilized to study the association of GpA and a variety of its mutants to better understand the impact of hydrogen bonding and van der Waals packing on association (Anderson et al., 2017; Russ & Engelman, 1999).

Recently, TOXCAT has been adapted into a 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 (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. I utilized this technique to quantify the association for the proteins that I computationally designed to determine the impact of van der Waals packing on association (cite sort-seq).

1.5.4 Computational methods

In conjunction with structural determination, computational methods have been invented to evaluate previously solved MP structures. Many methods look to further understand MP folding by establishing energetic terms that aim to estimate the thermodynamics of association. David Baker’s Lab at University of Washington are experts in this area, utilizing known energetic and structural information to develop the software suite Rosetta. Rosetta houses a variety of energy functions and prediction tools for soluble environments, including the ability to dock or design proteins *de novo*.

The original Rosetta energy functions have been adapted to accommodate predicting helical TMs within the membrane environment. The updated functions include an energy term that models the membrane into layers classified defining atoms as water-exposed, polar, interface, and hydrophobic (White & Wimley, 1999; Yarov-Yarovoy, Schonbrun, et al., 2006). Successful predictions determined structural details in MPs associated with voltage sensing and gating mechanisms (Vargas et al., 2012; Yarov-Yarovoy, Baker, et al., 2006). Recently, RosettaMP was developed to enhance the functionality of MP prediction which includes modeling TM helices *de novo* from sequence, a representation of the membrane bilayer, and the ability to transform a protein into the membrane (Koehler Leman et al., 2017). These tools increase the accessibility to MP structure, improving the ability to visualize and predict structures of MPs that have not yet been solved. Simultaneously, energetic predictions enhance the knowledge in the field by allowing researchers to assess structural mutations *in silico* prior to testing with *in vitro* or *in vivo* experiments.

Another computational tool for membrane proteins is the Molecular Software Library, or MSL (Kulp et al., 2012). Similar to Rosetta, MSL houses a variety of tools to perform MP structure prediction. These tools include the ability to transform proteins in space, mutate specific residues, extract geometric information from known structures, and predict the structure and energetics of an input sequence. Additionally, it has been used to develop the CATM algorithm using two energy terms parameterized by CHARMM: hydrogen bonding and van der Waals packing. The CATM algorithm has been applied to predict the structures of both known and unknown TM dimers, and experimental studies have shown that it accurately predicts the energetics of association (Anderson et al., 2017; Díaz Vázquez et al., 2023; Mueller et al., 2014). My research adapts the CATM algorithm to design structures with strong van der Waals packing in the absence of hydrogen bonding to assess the extent at which packing can drive MP association.

Other methods for predicting interactions between MPs focus on predicting the topology or docking of individual TM helices. OCTOPUS predicts TM topology using a combination of Markov models and neural networks (Viklund & Elofsson, 2008). HADDOCK can apply experimental knowledge of the interface region between proteins to refine docking (de Vries et al., 2010; Dominguez et al., 2003). PREDDIMER utilizes a novel surface-based modeling approach to predict and screen TM dimers for conformation heterogeneity (Polyansky et al., 2012). EVFold uses evolutionary based structural restraints to refine their docked structures (Braun T et al., 2015). TMDOCK uses an all-atom model for helices, inserting them in the membrane and outputting a structure alongside a predicted ΔG of insertion and ΔG of association (Lomize & Pogozheva, 2017). Each of these methods was adapted into a webserver for online use, where users can input the sequence and additional information to guide the process.

Most notably, at the 2020 Critical Assessment of Structure Prediction (CASP) conference, Google’s Deepmind introduced the machine learning model AlphaFold. AlphaFold utilizes a combination of neural networks and training on multiple sequence alignments (MSAs) and solved protein structures to predict unknown structures to near atomic precision with a 95% confidence interval (Jumper et al., 2021). Shortly afterward, David Baker’s group introduced RoseTTAFold, improving on the Rosetta prediction by incorporating a similar architecture to AlphaFold, with the addition of a third track network that connects sequence, residue-residue distances, and atomic coordinates (Baek et al., 2021). Each of these methods drastically improved the ability to predict unknown protein structures using information from previously studied and solved proteins. With increasing interest in using these technologies, multiple free webservers have been established to enhance access to these advanced protein prediction algorithms (Mirdita et al., 2022; Roberts et al., 2024). However, there are limits to how well these prediction algorithms work. AlphaFold struggles to predict proteins with <30 homologs in their MSAs, and accuracy decreases for protein structures dependent on other protein-protein interactions, while RoseTTAFold has difficulty predicting higher-order oligomers (Agard et al., 2022). The disparity between soluble proteins and MPs in the PDB contributes to these limitations, making small TM proteins difficult to predict.

1.5.5 Working at the experimental and computational interface

With the rise of computational technologies and experimental throughput over my time in graduate school, a lot of research is now being done at the interface between computation and experiments. The development of AlphaFold and RoseTTAFold allows researchers to predict structures of their proteins, making informed decisions on potential residues to mutate to see an impact in the structure or function (cites). With the ability to analyze data in high throughput using computation, there is a rising boon for complementary high-throughput experiments. My research is at the interface of computational and experimental design, utilizing computational tools to predict and create homodimer structures with novel sequences to be tested with a complementary high-throughput assay.

In Chapter 3, I will further detail studies at the experimental and computational interface. I cite studies that inspired and enhanced both the experiments and computational analysis within my research and describe the methods for my computational design and analysis at length.

* 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 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 research at the interface of experiments and computation 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****

Agard, D. A., Bowman, G. R., DeGrado, W., Dokholyan, N. V., & Zhou, H. X. (2022). Solution of the protein structure prediction problem at last: crucial innovations and next frontiers. *Fac Rev*, *11*, 38. <https://doi.org/10.12703/r-01-0000020>

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>

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>

Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G. R., . . . Baker, D. (2021). Accurate prediction of protein structures and interactions using a three-track neural network. *Science*, *373*(6557), 871-876. <https://doi.org/10.1126/science.abj8754>

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>

Braun T, Koehler Leman J, & OF, L. (2015). **Combining Evolutionary Information and an Iterative Sampling Strategy for Accurate Protein Structure Prediction**. In: PLOS Computational Biology.

de Vries, S. J., van Dijk, M., & Bonvin, A. M. (2010). The HADDOCK web server for data-driven biomolecular docking. *Nat Protoc*, *5*(5), 883-897. <https://doi.org/10.1038/nprot.2010.32>

Dominguez, C., Boelens, R., & Bonvin, A. M. (2003). HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J Am Chem Soc*, *125*(7), 1731-1737. <https://doi.org/10.1021/ja026939x>

Doura, A. K., & Fleming, K. G. (2004). Complex interactions at the helix-helix interface stabilize the glycophorin A transmembrane dimer. *J Mol Biol*, *343*(5), 1487-1497. <https://doi.org/10.1016/j.jmb.2004.09.011>

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>

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>

Fleming, K. G., & Engelman, D. M. (2001). Specificity in transmembrane helix-helix interactions can define a hierarchy of stability for sequence variants. *Proc Natl Acad Sci U S A*, *98*(25), 14340-14344. <https://doi.org/10.1073/pnas.251367498>

Januliene, D., & Moeller, A. (2021). Single-Particle Cryo-EM of Membrane Proteins. *Methods Mol Biol*, *2302*, 153-178. <https://doi.org/10.1007/978-1-0716-1394-8_9>

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>

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., . . . Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*(7873), 583-589. <https://doi.org/10.1038/s41586-021-03819-2>

Keener, J. E., Zhang, G., & Marty, M. T. (2021). Native Mass Spectrometry of Membrane Proteins. *Anal Chem*, *93*(1), 583-597. <https://doi.org/10.1021/acs.analchem.0c04342>

Kermani, A. A. (2021). A guide to membrane protein X-ray crystallography. *FEBS J*, *288*(20), 5788-5804. <https://doi.org/10.1111/febs.15676>

Koehler Leman, J., Mueller, B. K., & Gray, J. J. (2017). Expanding the toolkit for membrane protein modeling in Rosetta. *Bioinformatics*, *33*(5), 754-756. <https://doi.org/10.1093/bioinformatics/btw716>

Kulp, D. W., Subramaniam, S., Donald, J. E., Hannigan, B. T., Mueller, B. K., Grigoryan, G., & Senes, A. (2012). Structural informatics, modeling, and design with an open-source Molecular Software Library (MSL). *J Comput Chem*, *33*(20), 1645-1661. <https://doi.org/10.1002/jcc.22968>

Laganowsky, A., Reading, E., Hopper, J. T., & Robinson, C. V. (2013). Mass spectrometry of intact membrane protein complexes. *Nat Protoc*, *8*(4), 639-651. <https://doi.org/10.1038/nprot.2013.024>

Liang, B., & Tamm, L. K. (2016). NMR as a tool to investigate the structure, dynamics and function of membrane proteins. *Nat Struct Mol Biol*, *23*(6), 468-474. <https://doi.org/10.1038/nsmb.3226>

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>

Lomize, A. L., & Pogozheva, I. D. (2017). TMDOCK: An Energy-Based Method for Modeling α-Helical Dimers in Membranes. *J Mol Biol*, *429*(3), 390-398. <https://doi.org/10.1016/j.jmb.2016.09.005>

Maeda, R., Sato, T., Okamoto, K., Yanagawa, M., & Sako, Y. (2018). Lipid-Protein Interplay in Dimerization of Juxtamembrane Domains of Epidermal Growth Factor Receptor. *Biophys J*, *114*(4), 893-903. <https://doi.org/10.1016/j.bpj.2017.12.029>

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>

Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., & Steinegger, M. (2022). ColabFold: making protein folding accessible to all. *Nat Methods*, *19*(6), 679-682. <https://doi.org/10.1038/s41592-022-01488-1>

Mueller, B. K., Subramaniam, S., & Senes, A. (2014). A frequent, GxxxG-mediated, transmembrane association motif is optimized for the formation of interhelical Cα-H hydrogen bonds. *Proc Natl Acad Sci U S A*, *111*(10), E888-895. <https://doi.org/10.1073/pnas.1319944111>

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>

Polyansky, A. A., Volynsky, P. E., & Efremov, R. G. (2012). Multistate organization of transmembrane helical protein dimers governed by the host membrane. *J Am Chem Soc*, *134*(35), 14390-14400. <https://doi.org/10.1021/ja303483k>

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>

Roberts, J. B., Nava, A. A., Pearson, A. N., Incha, M. R., Valencia, L. E., Ma, M., . . . Keasling, J. D. (2024). Foldy: An open-source web application for interactive protein structure analysis. *PLoS Comput Biol*, *20*(2), e1011171. <https://doi.org/10.1371/journal.pcbi.1011171>

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>

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>

Vargas, E., Yarov-Yarovoy, V., Khalili-Araghi, F., Catterall, W. A., Klein, M. L., Tarek, M., . . . Roux, B. (2012). An emerging consensus on voltage-dependent gating from computational modeling and molecular dynamics simulations. *J Gen Physiol*, *140*(6), 587-594. <https://doi.org/10.1085/jgp.201210873>

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>

Viklund, H., & Elofsson, A. (2008). OCTOPUS: improving topology prediction by two-track ANN-based preference scores and an extended topological grammar. *Bioinformatics*, *24*(15), 1662-1668. <https://doi.org/10.1093/bioinformatics/btn221>

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>

White, S. H., & Wimley, W. C. (1999). Membrane protein folding and stability: physical principles. *Annu Rev Biophys Biomol Struct*, *28*, 319-365. <https://doi.org/10.1146/annurev.biophys.28.1.319>

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>

Yarov-Yarovoy, V., Baker, D., & Catterall, W. A. (2006). Voltage sensor conformations in the open and closed states in ROSETTA structural models of K(+) channels. *Proc Natl Acad Sci U S A*, *103*(19), 7292-7297. <https://doi.org/10.1073/pnas.0602350103>

Yarov-Yarovoy, V., Schonbrun, J., & Baker, D. (2006). Multipass membrane protein structure prediction using Rosetta. *Proteins*, *62*(4), 1010-1025. <https://doi.org/10.1002/prot.20817>

You, M., Li, E., Wimley, W. C., & Hristova, K. (2005). Forster resonance energy transfer in liposomes: measurements of transmembrane helix dimerization in the native bilayer environment. *Anal Biochem*, *340*(1), 154-164. <https://doi.org/10.1016/j.ab.2005.01.035>

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>