Chapter 1: Introduction

Table of Contents

[1.1 Introduction to membrane proteins 3](#_Toc171501167)

[1.2 The two-stage model of membrane protein folding 5](#_Toc171501168)

[1.3 Methods to study transmembrane helix oligomerization 7](#_Toc171501169)

[1.3.1 *In vitro* techniques 7](#_Toc171501170)

[1.3.2 *In vivo* assays 8](#_Toc171501171)

[1.4 Computational methods to study membrane protein structure 11](#_Toc171501172)

[1.4.1 Rosetta 11](#_Toc171501173)

[1.4.2 Molecular Software Library 12](#_Toc171501174)

[1.4.3 Topology Prediction and Docking Algorithms 13](#_Toc171501175)

[1.4.4 AlphaFold and RoseTTaFold 14](#_Toc171501176)

[1.5 Driving forces in membrane protein folding 16](#_Toc171501177)

[1.5.1 Hydrogen bonding and polar interactions 16](#_Toc171501178)

[1.5.2 Electrostatics and weak hydrogen bonding 18](#_Toc171501179)

[1.6 Understanding van der Waals as a driving force 22](#_Toc171501180)

[1.7 Thesis overview 27](#_Toc171501181)

[1.8 References 29](#_Toc171501182)

1.1 ****Introduction to membrane proteins****

A close-up of a string

Description automatically generated

**Figure 1.1 Types of Membrane Proteins.** Membrane proteins associate with a bilayer composed of hydrophilic heads and hydrophobic tails. Peripheral proteins are found at the interface of the bilayer and solute, while integral proteins are embedded into the membrane.

The cell membrane is a bilayer that separates internal cellular components from the outside environment. The membrane bilayer is composed of phospholipids, amphipathic molecules made of two distinct components: hydrophilic (water-loving) heads and hydrophobic (water-fearing) tails. To form the bilayer, hydrophobic tails are sandwiched between hydrophilic heads exposed to the soluble cell cytoplasm and the outer environment. Despite this separation, communication outside of the cell is critical to sense external stimuli and maintain cell survival. This process is regulated by a class of proteins tethered to the membrane: membrane proteins (MPs).

MPs can be broken down into two groups: peripheral MPs and integral MPs. Peripheral MPs are composed of both hydrophilic and hydrophobic components, allowing them to localize to the edges of the cell membrane while still exposed to the soluble environment. Unlike peripheral MPs, integral MPs are primarily hydrophobic; they are embedded within the membrane with minimal outside exposure. Integral MPs are made up of multiple structural subunits, such as β-sheets and α-helices. β-sheets often form open pores through the membrane, functioning as channels and transporters that allow ions and molecules to enter or exit the cell through the bilayer. Conversely, transmembrane (TM) α-helices are crammed into the membrane between lipids and assembled into complex multi-domain structures. Multiple TM α-helices (TMH) can associate in response to environmental stimuli, signaling activation and deactivation of the appropriate genes.

The sequestering of hydrophobic tails into the center of the membrane yields a hydrophobic environment; the core of the bilayer is devoid of polar interactions, which are responsible for the hydrophobic effect that drives soluble protein folding (Tanford, 1980; Yang et al., 1992). For MPs to fold within this hydrophobic environment, they must strike a delicate balance of interactions while surrounded by lipids. Mutations within MPs can drastically affect these interactions, preventing them from folding properly. Misfolding of TMHs has been implicated 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 that govern 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 and unfolding (Carpenter et al., 2008). Alternatively, researchers have focused on using model systems of MPs to better understand folding.

My research focused on using a model TMH system to investigate the extent that van der Waals packing can act as a driving force for MP folding and association. I developed an *in silico* protein design algorithm to study the association of single-pass TMH homodimers complemented with high-throughput experiments to validate my computational models. Before detailing my research and experiments, I review the contributions made to understanding driving forces in MP folding and association. I detail the forces and interactions involved in MP folding, while also highlighting the sequence and structural motif GASright, an important control for my research used to juxtapose differences between association by forces other than van der Waals packing. I then review the tools that have been used to study TMH association and folding before emphasizing the deficiency of research on the contribution of van der Waals packing.

**1.2 The two-stage model of membrane protein folding**

A diagram of a string and a string

Description automatically generated with medium confidenceEarly MP research focused on identifying membrane embedded regions within proteins using hydrophobicity analysis: navigating through the protein 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). This method was further developed to determine a charge bias known as the positive-inside rule, where charged AAs are likely to be found outside of the membrane (von Heijne, 1992). Multiple tools are now available, allowing researchers to easily identify TM regions from protein sequences (Wilkins et al., 1999).

**Figure 1.2 The two-stage model.** In stage 1, TM helices begin to form while the protein is inserted into the membrane. In stage 2, helices oligomerize and assemble into a fully folded protein as a result of thermodynamic interactions which include hydrogen bonding, electrostatics, and van der Waals packing.

In 1990, Popot and Engelman proposed the two stage model for MP folding: As the protein is threaded into the membrane, TMHs begin to form (stage 1) prior to stabilizing into a fully folded protein (stage 2). TMHs first reach a thermodynamic equilibrium with the lipid environment before undergoing stage 2, where individual TMHs oligomerize and assemble into the folded protein (Popot & Engelman, 1990, 2000). While stage 1 is driven by the hydrophobic effect to coordinate insertion of hydrophobic protein sequences into the membrane, stage 2 is governed by interactions between individual TM domains. Research on bacteriorhodopsin gives 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); lastly, extraction and reconstitution of individual helices of the protein were found to yield activity (Marti, 1998). Additional research on large protein complexes pushed the field forward, showing 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 to determine how minute changes in sequence and structure influence stability.

In the two-stage model, there are unique forces involved in each stage of folding. In the first stage, insertion of proteins into the membrane is driven by the hydrophobic effect, as MPs are more stable 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 between individual TM domains contribute to guiding the MP to the folded state.

Near the turn of the century, MP studies advanced our understanding of MP folding beyond the simplicity of the two-stage model. A third stage was considered, which accounted for 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). Rather than focusing on how bulk changes in forces impacts folding, researchers began to characterize the impact of individual AAs and the respective forces that drive MP folding.

1.3 Methods to study transmembrane helix oligomerization

To investigate MP folding, researchers have developed tools to study the oligomerization of TMHs. 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. The oligomerization process is essential for influencing cell gene expression, including epidermal growth factor receptors and proteins involved in tyrosine kinase signaling cascades (Kumari & Yadav, 2019). Furthermore, this thermodynamic information can be used to assess and validate computational models for designing and engineering novel proteins. In this section, I detail the tools and techniques that have been implemented to further understand the driving forces in MPs.

1.3.1 *In vitro* techniques

Early tools used to study MP folding monitored the reversible folding of MPs. *In vitro* techniques focus on expressing and solubilizing proteins into suitable membrane mimetics, such as detergents. An initial method studied the thermodynamics of TMH association by observing differences in mobility within SDS-PAGE gels. SDS-PAGE was used to tease the effect of point mutations in a variety of TM sequences, including GpA (Choma et al., 2000; Lemmon, Flanagan, Hunt, et al., 1992; Lemmon, Flanagan, Treutlein, et al., 1992; Zhou et al., 2000). Sedimentation equilibrium analytical ultracentrifugation (SE-AUC) is another technique that explores different folding states by varying the concentrations of detergents used to solubilize the protein. These samples are centrifuged at high speeds, resulting in a concentration gradient that is then analyzed to quantitatively ascertain the transition of the protein at different folding states. SE-AUC has been used to investigate mutations of GpA to better understand its thermodynamics of association (Doura & Fleming, 2004; Fleming et al., 1997; Fleming & Engelman, 2001). These techniques developed our understanding of the thermodynamics of TMH association. However, they were low throughput and limited to studying MPs solubilized in detergents.

Multiple *in vitro* techniques were developed to observe MP dynamics within membrane like environments. Disulfide cross-linking were used to measure TMH oligomerization in both micelles and lipid vesicles to investigate the interfaces of a variety of proteins (Cristian et al., 2003; Hastrup et al., 2001; Kovalenko et al., 2005; Lu et al., 2010). Pulse proteolysis quantitatively measures the thermodynamic stability of MPs by selectively denaturing and subsequently digesting the unfolded MP (Park & Marqusee, 2005). Using this technique to study bacteriorhodopsin folding uncovered that folding was 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 folding energy landscapes of GpA and mutants affecting its association (Blois et al., 2009; Hong & Bowie, 2011; Hong et al., 2013; Howarth et al., 2006; Huang et al., 2022). Compared to previous research, these methods allowed for studying MP folding thermodynamics in lipids and mixed micelles, closer to the environments of the cell membrane. While these techniques approach understanding proteins in native environments, other techniques were developed to further study MPs within cells at higher throughput.

1.3.2 *In vivo* assays

*In vivo* assays have been utilized to investigate the folding and association of MPs in their natural environment. Double mutant cycle analysis, a method that quantitatively measures interaction in protein structures, mutates two non-interacting residues within a protein to assess the impact of coupled residues on thermodynamic stability (Carter et al., 1984). Double mutant cycles in bacterial two-hybrid and protein complementation assays allow researchers to determine the strength of protein-protein interactions by changes in cell growth due to mutation, which can be monitored in high-throughput (Horovitz et al., 2019; Salinas & Ranganathan, 2018; Tarassov et al., 2008). Genetic reporter assays allow cells to 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 inhibits the β-galactosidase gene (Schneider & Engelman, 2003). Other reporter assay systems have utilized a chimera of the MPs of interest fused to ToxR, a dimeric transcription factor, to **A diagram of a cat and a cat

Description automatically generated**promote the expression of genes (Gurezka & Langosch, 2001; Russ & Engelman, 1999).

**Figure 1.3 TOXCAT.** TOXCAT is an experimental assay that has been used to study the self-association of helices. The TM of interest is expressed bound to maltose binding protein (MBP) and ToxR, a dimeric transcription factor originally found in *V. cholerae*. When the TM associates, ToxR does as well, resulting in the expression of a gene (CAT) that can be measured to determine the strength of the association.

TOXCAT has been used to study TM helix-helix interactions, where the TM of interest is fused to dimeric transcription factor ToxR. When the TMs associate, ToxR dimerizes and promotes the expression of chloramphenicol acetyltransferase (CAT) which is measured to determine the strength of association. TOXCAT demonstrated that mutations of polar residues on GpA in the native membrane environment yield 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 aromatic AAs facilitates oligomerization (Johnson et al., 2007). TMH association as measured by TOXCAT correlates 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 the high-throughput assay TOXGREEN. The reporter gene CAT was 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). TOXGREEN can be used in high-throughput applications such as fluorescence activated cell sorting (FACS), where a library of TMs is 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 (cite sort-seq).

1.4 Computational methods to study membrane protein structure

In conjunction with experimental methods to study membrane protein folding, computational methods have been designed by evaluating previously solved MP structures. These methods look to further understand MP folding by establishing energetic terms that estimate the thermodynamics of MP folding. In this section, I review computational methods used to predict MP structures, highlighting unique features of each tool.

1.4.1 Rosetta

Rosetta houses a variety of energy functions and prediction tools for soluble environments, including the ability to dock or design proteins *de* novo (Chaudhury et al., 2011; Koehler Leman et al., 2017; Leman et al., 2020; Weitzner et al., 2017; Yarov-Yarovoy, Schonbrun, et al., 2006). These energetics include but are not limited to terms parameterized by CHARMM (van der Waals and electrostatics), a hydrogen bond and disulfide function curated from polar contacts found in ~8000 high-resolution crystal structures, and a side chain conformation energy based on the probability of occurrence from the Dunbrack rotamer database (Alford et al., 2017; RICHARDSON et al.). These Rosetta energy functions have been adapted to accommodate predicting helical TMs within the membrane environment. The updated functions include an energy term that separates the membrane into layers designating atoms as water-exposed, polar, interface, or hydrophobic (White & Wimley, 1999; Yarov-Yarovoy, Schonbrun, et al., 2006). Successful predictions helped discover structural details in MPs associated with voltage sensing and gating mechanisms (Vargas et al., 2012; Yarov-Yarovoy, Baker, et al., 2006).

Recently, RosettaMP was devised to enhance the functionality of MP prediction. With RosettaMP, TM helices are modeled *de novo* from sequence, the membrane bilayer is represented, and MP assembly is simulated (Koehler Leman et al., 2017). These tools increase accessibility to MP structures, improving the ability to visualize and predict structures of MPs that have not yet been solved, as well as enhancing membrane protein design (Duran & Meiler, 2018). Simultaneously, energetic predictions permit researchers to analyze structural mutations *in silico* prior to testing with *in vitro* or *in vivo* experiments (Thieker et al., 2022).

1.4.2 Molecular Software Library

A diagram of a structure

Description automatically generatedAnother computational tool for modeling protein structures is the Molecular Software Library, or MSL (Kulp et al., 2012). Similar to Rosetta, MSL contains 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, MSL was used to develop the CATM algorithm.

**Figure 1.4 CATM.** Using a refined set of dimer geometries for GASright proteins, the CATM algorithm predicts the stability and structure of GASright dimers from an input sequence.

Briefly, the CATM algorithm predicts structures of known and unknown TM dimers that associate by the GASright motif, and experimental studies have shown that it accurately estimates the energetics of association (Anderson et al., 2017; Díaz Vázquez et al., 2023; Mueller et al., 2014). CATM uses the Energy-Based conformer library applied at the 95% level for side chain mobility (Subramaniam & Senes, 2012). Energetics of predicted proteins are determined in CATM using the CHARMM 22 van der Waals function (MacKerell et al., 1998), the IMM1 membrane implicit solvation model (Lazaridis, 2003), and the hydrogen bonding function SCWRL4 (Krivov et al., 2009). Each of these energy terms is applied to optimize the dimer geometry by Monte Carlo (MC) backbone perturbation cycles where all parameters (xShift, zShift, axialRotation, and crossingAngle) are locally varied. The association energy is calculated as the energy of the dimer minus the energy of two monomers:

My research adapts the CATM algorithm to design structures with strong van der Waals packing in the absence of hydrogen bonding, allowing me to assess the extent at which packing can drive MP association. I further detail how I used MSL to design TM homodimers in chapter 3.

A black background with a black square

Description automatically generated with medium confidence1.4.3 Topology Prediction and Docking Algorithms

**Table 1.** Docking tools, the websites they can be accessed at, and the key features for each tool.

Other methods used to determine 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 employs evolutionary-based, structural restraints to refine their docked structures (Braun T et al., 2015). TOPCONS can identify signal peptides separate from TM regions, and displays homology to known structures as well as a predicted ΔG of insertion (Tsirigos et al., 2015). TMDOCK applies 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 is available online, where users can input the sequence and additional information to guide the process.

1.4.4 AlphaFold and RoseTTaFold

Most notably, at the 2020 Critical Assessment of Structure Prediction (CASP) conference, Google’s DeepMind introduced the machine learning model AlphaFold. Unlike previously mentioned prediction algorithms, AlphaFold predicts structures without energetics. AlphaFold utilizes a combination of neural networks, 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 inclusion 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 online tools have been established to enhance access to these advanced protein prediction algorithms (Mirdita et al., 2022; Roberts et al., 2024). However, these machine learning algorithms are limited by the amount of information available. 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). These limitations are amplified in MPs due to the lack of MP structures, making small or complex TM proteins difficult to predict. To better understand the dynamics of association and folding in MPs, it is necessary to advance our knowledge of the forces involved in folding.

1.5 Driving forces in 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 aimed to solve the structures of MPs by identifying 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 (Carpenter et al., 2008). Alternative approaches to study folding utilized a combination of *in vitro* and *in vivo* experimental tools to determine the rules that govern TM folding. Folding of integral TM proteins involves a variety of energetic constraints resulting from the hydrophobic nature of the phospholipid bilayer. The translocon complex assists during translation, inserting TM domains into the membrane (White & von Heijne, 2004). TMs are composed of amide nitrogens and carbonyl oxygens within the protein backbone, atoms prone to forming hydrogen bonds. However, inserting hydrogen bonds into the hydrophobic core of the bilayer carries an energetic penalty (Marinko et al., 2019; Popot & Engelman, 1990, 2000). To satisfy the lack of hydrogen bonding within the membrane, TMs adopt standard α-helical and β-sheet structures where hydrogen bonds form along the protein backbone. Experimental tools have been developed to tease out folding interactions after insertion by using model MP systems. In this section, I will summarize advances in understanding driving forces in MP folding, with a particular focus on using single-pass TMHs.

1.5.1 Hydrogen bonding and polar interactions

Hydrogen bonding plays a key role in regulating MP structure and function, and many mutations on polar residues have been found to promote disease states (Choi et al., 2004; Partridge et al., 2002, 2004; Therien et al., 2001; Wehbi et al., 2008). Research characterizing the impact of polar residues on TMH association suggests that hydrogen bonding and polar interactions can drive TMH association. Using a wild-type like sequence of the GCN4 leucine zipper, a mutation from Asn to Val was found to decrease association on SDS-PAGE (Choma et al., 2000). Synthetic model poly-leucine peptides based on GCN4 were made with three different compositions of AAs at the interface, two being completely hydrophobic and the other hydrophobic with a single Asn. When tested for their ability to associate on SDS-PAGE, only the sequence with Asn was found to have equal amounts of monomers and dimers (Zhou et al., 2000). These results suggest that Asn plays a role in driving TMH association.

Further research began to characterize the impact of other polar AAs in TMH association systems. Poly-leucine based peptides were made with single AA mutations to a variety of polar residues and tested using the *in vivo* experimental assay TOXCAT. Their results showed that larger polar residues (Asn, Asp, Gln, and Glu) capable of being both hydrogen bond donors and acceptors drive association more than poly-leucine alone (Zhou et al., 2001). A similar study showed that replacing hydrophobic AAs with large polar AAs on the GCN4 peptide resulted in association with higher stabilities (Gratkowski et al., 2001). These studies suggest that large polar AAs drive association. However, large polar AAs are not often found in MP sequences, whereas small polar AAs Thr and Ser are more common due to their ability to more readily form hydrogen bonds with backbone carbonyls on the same helix (Gray & Matthews, 1984; Liu et al., 2002).

The peptides used in the previous studies were made of bulky hydrophobic AAs, possibly preventing Ser and Thr from playing the roles in association that they do in naturally occurring sequences. Using TOXCAT, a library of TM sequences that mutated the interface of known dimer glycophorin A (GpA) were screened for their ability to associate. A majority of the proteins found to associate were composed of Thr and Ser at the interface, suggesting that these AAs are important for association (Russ & Engelman, 2000). Additional investigation into the structure of GpA suggests that Thr 87 forms interhelical hydrogen bonds at the interface, supporting previous research that mutations at this residue disrupt dimerization (Lemmon, Flanagan, Hunt, et al., 1992; Lemmon, Flanagan, Treutlein, et al., 1992; MacKenzie et al., 1997; Smith et al., 2002). Alongside earlier research, this data suggests that hydrogen bonding is a driving force that strongly stabilizes TMH association.

By observing MP structures, researchers have been able to identify and characterize hydrogen bonds between TM helices in multiple solved structures (Adamian & Liang, 2002; Freiberg et al., 2012; MacKenzie et al., 1997; White, 2005). Using double mutant cycle analysis, MPs with multiple TMHs were mutated to determine the contribution of interhelical hydrogen bonding to MP stability. The average contribution for hydrogen bonding in multiple proteins was found to be 0.5kcal/mol +/- 0.7 (Bowie, 2011). Using an SDS unfolding assay, the average contribution of eight hydrogen bonds was found to be 0.6 kcal/mol (Joh et al., 2008). Despite the relatively small contribution in larger protein, hydrogen bond energies estimated in vacuum have been calculated to be around ten times higher (Ben-Tal et al., 1997; Mitchell & Price, 1990; Rose & Wolfenden, 1993; Tsemekhman et al., 2007). Single mutants on a variety of MPs were also tested, determining on average that hydrogen bonding contributes similar stability to water soluble proteins (Bowie, 2011).

Hydrogen bonding and polar interactions are stabilizing forces in MP folding and association. In larger MP complexes, hydrogen bonding contributes similar stability as found in soluble proteins **(cite)**. However, the hydrophobic nature of the membrane suggests that hydrogen bonding contributes more to stability. Most polar AAs are able to form hydrogen bonds with the backbone of the TM, and this bond must be broken prior to forming a stabilizing interhelical hydrogen bond (Chamberlain & Bowie, 2004). Additionally, mutating polar AAs to nonpolar AAs disrupts association, suggesting that hydrogen bonding drives association of TMHs **(cite)**.

1.5.2 Electrostatics and weak hydrogen bonding

Electrostatics interactions, or cation-π, are a result of attractive forces between charged AAs (Lys and Arg) and the electron clouds of aromatic AAs (Phe, Tyr, His, and Trp). These interactions are found in a multitude of channels and G protein-coupled receptors, and are equally important for ligand binding of neurotransmitters, metal ions, and toxins (Infield et al., 2021). Charged AAs are not often found in MPs, but molecular dynamics simulations and potential of mean force calculations supports the thermodynamic stability of Arg in TMs (Ulmschneider et al., 2017). Electrostatic interactions have been studied between a variety of TMH interactions. Johnson et al. mutated a hydrophobic protein with a pair of charged and aromatic AAs. Using TOXCAT, they found that Lys coupled with Tyr, Trp, and Phe is able to drive these proteins to associate (Johnson et al., 2007). Another study looked at the role of aromatic AAs in the β-barrel outer MP OmpA, and using double mutant cycle analysis, found that each side chain contributes more than 1kcal/mol to stability (Hong et al., 2007). Additional SDS-PAGE analysis on helical hairpins demonstrated that TM-TM electrostatic interactions alongside helical turns promote folding (Bañó-Polo et al., 2013).

Similar to hydrogen bonding, electrostatics plays a strong stabilizing role in MP folding and association, able to drive association of TMHs with charged and aromatic interactions. Another class of forces that has been shown to strongly influence association of TMHs is facilitated by the GASright motif, 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: Gly, Ala, and Ser. These small residues define the interface of the motif (G/A/S)xxx(G/A/S), resulting in a short interhelical distance between two TM helices. The "right" subscript in GASright comes from an important structural feature in which TM helices associate at 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 have studied these proteins to further understand the forces governing TM association.

GpA is a well-studied protein that associates via the GASright motif. Multiple *in vitro* studies worked to define the interface of GpA, making point mutations along the protein and visualizing the changes in dimerization using 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 ascertain differences in stability, effectively quantifying the thermodynamics of association for GpA (Fisher et al., 1999; Fleming et al., 1997).

After the structure of GpA was solved by solution nuclear magnetic resonance (NMR), groups analyzed the structure to further characterize their thermodynamic data (MacKenzie et al., 1997). Analysis of GpA in molecular dynamics simulations coupled with mutations on the NMR structure showed that the alternative association resulted from changes in van der Waals interactions (Fleming et al., 1997; MacKenzie & Engelman, 1998; MacKenzie et al., 1997; Petrache et al., 2000). However, further investigation into the unique sequence and defined structure of the GASright motif has been shown to permit an uncommon structural feature. Small AAs at the interface allow 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 considered as 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 between one third and half of that of an N—H donor in vacuum (Scheiner et al., 2001; Vargas et al., 2000). Measurements of the stretching frequency of these bonds in GpA suggests that it could contribute 0.9 kcal/mol of stability to the dimer (Arbely & Arkin, 2004).

Further research on GASright TMs helped define the geometric structure for the network of hydrogen bonds. This research resulted in CATM, an algorithm that successfully predicted the structures of five known homodimer structures (Mueller et al., 2014). The CATM algorithm was used in conjunction with 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 stable (Anderson et al., 2017). Additionally, the free energy of association of GASright structures was measured 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: weak hydrogen bonding and van der Waals packing. Using a refined version of the CATM algorithm, I designed sequences to associate solely by van der Waals packing, leveraging GASright sequences as controls. By evaluating sequences designed that associate through van der Waals packing, I was able to differentiate the impact of packing (designed sequences) versus both hydrogen bonding and packing (GASright) on association.

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

As individual TMHs 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 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 also drives TM helix association when polar AAs Ser and Thr form interhelical hydrogen bonds between opposing helices (Johnson et al., 2007; Zhou et al., 2001). Additionally, electrostatic interactions between positively charged Lys and electronegative aromatic AAs Tyr, Trp, and Phe promote 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 account for 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 form hydrogen bonds (Liu et al., 2002), thus TMs constituted of these AAs can only be stabilized by van der Waals forces.

Van der Waals forces occur 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.

A screenshot of a computer

Description automatically generated

**Figure 1.5 Van der Waals force.** Van der Waals forces between two atoms are non-existent at long range distances. However, van der Waals becomes an attractive force as the atoms approach up to a minimum distance. Once the minimum is reached, the force repulses these atoms away from each other.

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, pushing the 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.

A black and white square with a number and a symbol

Description automatically generated with medium confidenceThe van der Waals force between two atoms can be calculated using the Lennard-Jones (LJ) Potential:

A graph of a function

Description automatically generatedThe 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 1.6 Lennard-Jones Potential.** The potential used to estimate and quantify the van der Waals force between two atoms.

A close-up of a dna model

Description automatically generatedSoluble proteins are driven to fold by the hydrophobic effect, where nonpolar AAs 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 in van der Waals packing. However, 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.

**Figure 1.7 Different types of van der Waals packing.** Van der Waals packing can be separated into three interactions: lipid-lipid, lipid-protein, and protein-protein. Understanding the impact of each of these forces on folding is crucial to fully understand the impact that van der Waals has on MP folding and association.

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 (or sidechain) 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. Sidechain packing focuses on the stability gained between fragments of proteins in close contact. Each of these interactions plays a role in stabilizing a 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. 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 showed 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 on 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.

****1.7 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 MP association and folding. Prior research on the impact of packing to the folded state of MPs honed-in on singular systems, and I aimed to expand this knowledge to a larger variety of MP structures.

**In Chapter 2**, I present the majority of my graduate schoolwork, which will be published in the near future. In this paper, I found that van der Waals packing is a weak driving force that leads to association of MP homodimers. I data mined the Protein Databank (PDB) for all solved MP structures to determine the best structures for computational design, developed a protein design algorithm and assessed the ability for these proteins to association. I found that packing drives my designed proteins to associate mildly when compared to GASright proteins that rely on both hydrogen bonding and van der Waals packing.

**In Chapter 3**, I discuss my protein design method. With improving experimental technologies, many studies at the forefront of research utilize a combination of high-throughput experiments and computational analysis. My research paired high-throughput experiments with computational design explore the impact of biophysical forces on protein association in a large range of structures. I detail the computational tools used for my project that involves data mining of the PDB and 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 with simplified terminology while simultaneously reflecting on moments during 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.

****1.8 References****

Adamian, L., & Liang, J. (2002). Interhelical hydrogen bonds and spatial motifs in membrane proteins: polar clamps and serine zippers. *Proteins*, *47*(2), 209-218. <https://doi.org/10.1002/prot.10071>

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>

Alford, R. F., Leaver-Fay, A., Jeliazkov, J. R., O’Meara, M. J., DiMaio, F. P., Park, H.,…Gray, J. J. (2017). The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *Journal of Chemical Theory and Computation*, *13*(6), 3031-3048. <https://doi.org/10.1021/acs.jctc.7b00125>

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>

Arbely, E., & Arkin, I. T. (2004). Experimental measurement of the strength of a C alpha-H...O bond in a lipid bilayer. *J Am Chem Soc*, *126*(17), 5362-5363. <https://doi.org/10.1021/ja049826h>

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>

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>

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>

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>

Ben-Tal, N., Sitkoff, D., Topol, I. A., Yang, A.-S., Burt, S. K., & Honig, B. (1997). Free energy of amide hydrogen bond formation in vacuum, in water, and in liquid alkane solution. *The Journal of Physical Chemistry B*, *101*(3), 450-457.

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>

Bowie, J. U. (2011). Membrane protein folding: how important are hydrogen bonds? *Curr Opin Struct Biol*, *21*(1), 42-49. <https://doi.org/10.1016/j.sbi.2010.10.003>

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.

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>

Carpenter, E. P., Beis, K., Cameron, A. D., & Iwata, S. (2008). Overcoming the challenges of membrane protein crystallography. *Curr Opin Struct Biol*, *18*(5), 581-586. <https://doi.org/10.1016/j.sbi.2008.07.001>

Carter, P. J., Winter, G., Wilkinson, A. J., & Fersht, A. R. (1984). The use of double mutants to detect structural changes in the active site of the tyrosyl-tRNA synthetase (Bacillus stearothermophilus). *Cell*, *38*(3), 835-840. <https://doi.org/10.1016/0092-8674(84)90278-2>

Chamberlain, A. K., & Bowie, J. U. (2004). Analysis of side-chain rotamers in transmembrane proteins. *Biophys J*, *87*(5), 3460-3469. <https://doi.org/10.1529/biophysj.104.044024>

Chaudhury, S., Berrondo, M., Weitzner, B. D., Muthu, P., Bergman, H., & Gray, J. J. (2011). Benchmarking and analysis of protein docking performance in Rosetta v3. 2. *PloS one*, *6*(8), e22477.

Choi, M. Y., Cardarelli, L., Therien, A. G., & Deber, C. M. (2004). Non-native interhelical hydrogen bonds in the cystic fibrosis transmembrane conductance regulator domain modulated by polar mutations. *Biochemistry*, *43*(25), 8077-8083. <https://doi.org/10.1021/bi0494525>

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>

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>

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>

Duran, A. M., & Meiler, J. (2018). Computational design of membrane proteins using RosettaMembrane. *Protein Science*, *27*(1), 341-355. <https://doi.org/https://doi.org/10.1002/pro.3335>

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>

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>

Freiberg, A., Kangur, L., Olsen, J. D., & Hunter, C. N. (2012). Structural implications of hydrogen-bond energetics in membrane proteins revealed by high-pressure spectroscopy. *Biophys J*, *103*(11), 2352-2360. <https://doi.org/10.1016/j.bpj.2012.10.030>

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>

Gray, T. M., & Matthews, B. W. (1984). Intrahelical hydrogen bonding of serine, threonine and cysteine residues within alpha-helices and its relevance to membrane-bound proteins. *J Mol Biol*, *175*(1), 75-81. <https://doi.org/10.1016/0022-2836(84)90446-7>

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>

Hastrup, H., Karlin, A., & Javitch, J. A. (2001). Symmetrical dimer of the human dopamine transporter revealed by cross-linking Cys-306 at the extracellular end of the sixth transmembrane segment. *Proc Natl Acad Sci U S A*, *98*(18), 10055-10060. <https://doi.org/10.1073/pnas.181344298>

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>

Horovitz, A., Fleisher, R. C., & Mondal, T. (2019). Double-mutant cycles: new directions and applications. *Curr Opin Struct Biol*, *58*, 10-17. <https://doi.org/10.1016/j.sbi.2019.03.025>

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>

Huang, B., Xu, Y., Hu, X., Liu, Y., Liao, S., Zhang, J.,…Liu, H. (2022). A backbone-centred energy function of neural networks for protein design. *Nature*, *602*(7897), 523-528. <https://doi.org/10.1038/s41586-021-04383-5>

Infield, D. T., Rasouli, A., Galles, G. D., Chipot, C., Tajkhorshid, E., & Ahern, C. A. (2021). Cation-π Interactions and their Functional Roles in Membrane Proteins. *J Mol Biol*, *433*(17), 167035. <https://doi.org/10.1016/j.jmb.2021.167035>

Joh, N. H., Min, A., Faham, S., Whitelegge, J. P., Yang, D., Woods, V. L., & Bowie, J. U. (2008). Modest stabilization by most hydrogen-bonded side-chain interactions in membrane proteins. *Nature*, *453*(7199), 1266-1270. <https://doi.org/10.1038/nature06977>

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>

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>

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>

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>

Kovalenko, O. V., Metcalf, D. G., DeGrado, W. F., & Hemler, M. E. (2005). Structural organization and interactions of transmembrane domains in tetraspanin proteins. *BMC Struct Biol*, *5*, 11. <https://doi.org/10.1186/1472-6807-5-11>

Krivov, G. G., Shapovalov, M. V., & Dunbrack, R. L. (2009). Improved prediction of protein side-chain conformations with SCWRL4. *Proteins*, *77*(4), 778-795. <https://doi.org/10.1002/prot.22488>

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>

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>

Lazaridis, T. (2003). Effective energy function for proteins in lipid membranes. *Proteins*, *52*(2), 176-192. <https://doi.org/10.1002/prot.10410>

Leman, J. K., Weitzner, B. D., Lewis, S. M., Adolf-Bryfogle, J., Alam, N., Alford, R. F.,…Barth, P. (2020). Macromolecular modeling and design in Rosetta: recent methods and frameworks. *Nature methods*, *17*(7), 665-680.

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>

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>

Lu, C., Mi, L. Z., Grey, M. J., Zhu, J., Graef, E., Yokoyama, S., & Springer, T. A. (2010). Structural evidence for loose linkage between ligand binding and kinase activation in the epidermal growth factor receptor. *Mol Cell Biol*, *30*(22), 5432-5443. <https://doi.org/10.1128/MCB.00742-10>

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>

MacKerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J.,…Karplus, M. (1998). All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B*, *102*(18), 3586-3616. <https://doi.org/10.1021/jp973084f>

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>

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>

Mitchell, J. B., & Price, S. L. (1990). The nature of the N H… ︁O C hydrogen bond: An intermolecular perturbation theory study of the formamide/formaldehyde complex. *Journal of computational chemistry*, *11*(10), 1217-1233.

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>

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>

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>

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>

Partridge, A. W., Therien, A. G., & Deber, C. M. (2002). Polar mutations in membrane proteins as a biophysical basis for disease. *Biopolymers*, *66*(5), 350-358. <https://doi.org/10.1002/bip.10313>

Partridge, A. W., Therien, A. G., & Deber, C. M. (2004). Missense mutations in transmembrane domains of proteins: phenotypic propensity of polar residues for human disease. *Proteins*, *54*(4), 648-656. <https://doi.org/10.1002/prot.10611>

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>

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>

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>

RICHARDSON, J. S., KEEDY, D. A., & RICHARDSON, D. C. “THE PLOT” THICKENS: MORE DATA, MORE DIMENSIONS, MORE USES. In *Biomolecular Forms and Functions* (pp. 46-61). <https://doi.org/10.1142/9789814449144_0004>

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>

Rose, G. D., & Wolfenden, R. (1993). Hydrogen bonding, hydrophobicity, packing, and protein folding. *Annual review of biophysics and biomolecular structure*, *22*(1), 381-415.

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>

Russ, W. P., & Engelman, D. M. (2000). The GxxxG motif: a framework for transmembrane helix-helix association. *J Mol Biol*, *296*(3), 911-919. <https://doi.org/10.1006/jmbi.1999.3489>

Salinas, V. H., & Ranganathan, R. (2018). Coevolution-based inference of amino acid interactions underlying protein function. *Elife*, *7*. <https://doi.org/10.7554/eLife.34300>

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>

Smith, S. O., Eilers, M., Song, D., Crocker, E., Ying, W., Groesbeek, M.,…Aimoto, S. (2002). Implications of threonine hydrogen bonding in the glycophorin A transmembrane helix dimer. *Biophys J*, *82*(5), 2476-2486. <https://doi.org/10.1016/S0006-3495(02)75590-2>

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>

Subramaniam, S., & Senes, A. (2012). An energy-based conformer library for side chain optimization: improved prediction and adjustable sampling. *Proteins*, *80*(9), 2218-2234. <https://doi.org/10.1002/prot.24111>

Tanford, C. (1980). *The hydrophobic effect: formation of micelles and biological membranes 2d ed*. J. Wiley.

Tarassov, K., Messier, V., Landry, C. R., Radinovic, S., Molina, M. M. S., Shames, I.,…Michnick, S. W. (2008). An in vivo map of the yeast protein interactome. *Science*, *320*(5882), 1465-1470.

Therien, A. G., Grant, F. E., & Deber, C. M. (2001). Interhelical hydrogen bonds in the CFTR membrane domain. *Nat Struct Biol*, *8*(7), 597-601. <https://doi.org/10.1038/89631>

Thieker, D. F., Maguire, J. B., Kudlacek, S. T., Leaver-Fay, A., Lyskov, S., & Kuhlman, B. (2022). Stabilizing proteins, simplified: A Rosetta-based webtool for predicting favorable mutations. *Protein Science*, *31*(10), e4428. <https://doi.org/https://doi.org/10.1002/pro.4428>

Tsemekhman, K., Goldschmidt, L., Eisenberg, D., & Baker, D. (2007). Cooperative hydrogen bonding in amyloid formation. *Protein science*, *16*(4), 761-764.

Tsirigos, K. D., Peters, C., Shu, N., Käll, L., & Elofsson, A. (2015). The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides. *Nucleic Acids Research*, *43*(W1), W401-W407. <https://doi.org/10.1093/nar/gkv485>

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, 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>

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>

Wehbi, H., Gasmi-Seabrook, G., Choi, M. Y., & Deber, C. M. (2008). Positional dependence of non-native polar mutations on folding of CFTR helical hairpins. *Biochim Biophys Acta*, *1778*(1), 79-87. <https://doi.org/10.1016/j.bbamem.2007.08.036>

Weitzner, B. D., Jeliazkov, J. R., Lyskov, S., Marze, N., Kuroda, D., Frick, R.,…Gray, J. J. (2017). Modeling and docking of antibody structures with Rosetta. *Nature protocols*, *12*(2), 401-416.

White, S. H. (2005). How hydrogen bonds shape membrane protein structure. *Adv Protein Chem*, *72*, 157-172. <https://doi.org/10.1016/S0065-3233(05)72006-4>

White, S. H., & von Heijne, G. (2004). The machinery of membrane protein assembly. *Curr Opin Struct Biol*, *14*(4), 397-404. <https://doi.org/10.1016/j.sbi.2004.07.003>

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>

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>

Yang, A.-S., Sharp, K. A., & Honig, B. (1992). Analysis of the heat capacity dependence of protein folding. *Journal of Molecular Biology*, *227*(3), 889-900. <https://doi.org/https://doi.org/10.1016/0022-2836(92)90229-D>

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>

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>