Chapter 1: Introduction

Table of Contents

[1.1 Introduction to membrane proteins 3](#_Toc162807252)

[1.2 Forces involved in membrane protein folding and association 4](#_Toc162807253)

[1.3 The importance of van der Waals packing 5](#_Toc162807254)

[1.4 GASright 7](#_Toc162807255)

[1.5 Studying membrane protein folding and structure 8](#_Toc162807256)

[1.5.1 Structural methods 9](#_Toc162807257)

[1.5.2 Quantitative methods 10](#_Toc162807258)

[1.5.3 Genetic reporter assays 11](#_Toc162807259)

[1.5.4 Computational methods 12](#_Toc162807260)

[1.5.5 Working at the experimental and computational interface 14](#_Toc162807261)

[1.6 Thesis overview 15](#_Toc162807262)

[1.7 References 16](#_Toc162807263)

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

The cellular membrane is a bilayer composed of phospholipid molecules that separate DNA and other internal features of the cell from the outside environment. Communication outside the cell and sensing of external factors impacting 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. Due to the hydrophobic nature of the membrane, integral MPs are composed of unique structures allowing them to regulate cell homeostasis. β-barrel MPs are composed of β-strands that form open pores in the membrane, acting as channels for molecules to pass in and out of the cell. Multi-pass and single-pass MPs are made of α-helices that can take on alternate structures, 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 α-helical MP folding.

The hydrophobic nature of the phospholipid bilayer results in a complex network of interactions responsible for folding and stability. Phospholipids are amphipathic molecules, composed of … hydrophilic heads and hydrophobic tails. The bilayer forms a sandwich with phospholipid heads on the outside exposed to the soluble environment, with hydrophobic tails sequestering the hydrophobic tails between membrane, tucked away from any contact with the soluble environment. **… (really describe how the bilayer looks here) …, to set the stage for talking about the internal membrane interactions, that feels lacking right now)** Misfolding of MPs affects not only the has been found to be involved in several human diseases such as Parkinson’s, cystic fibrosis, and cancer (Sanders and Myers, 2004; Gregersen et al., 2006). To fully understand how MPs are involved in essential biological functions and to combat the progression of diseases, it is necessary to understand the forces involved in folding. However, studying MP folding 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, proteins fold differently depending on the composition of lipids in the bilayer, a daunting task to replicate these biological conditions either *in vitro* or *in silico*. To combat these challenges, much of the research studying membrane protein folding is focused on combining *in vitro* or *in silico* studies with live cell research, aiming to explore and understand the impact of biophysical forces governing the folding process in a natural environment.

My graduate work focused on developing an *in silico* protein design algorithm to study the association of single-pass membrane proteins known as homodimers, complemented using high-throughput experiments to test and validate my computational models. In this introduction, I expand on the forces and interactions involved in membrane protein association, with a particular focus on the lack of knowledge on the impact 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 movement combining these methods and how it inspired my research.

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

Folding of integral transmembrane (TM) helices involves a variety of energetic constraints due to the hydrophobic nature of the phospholipid bilayer. First, TMs must be translated and inserted properly into the membrane. This is accomplished through the aid of the translocon, a protein that in conjunction with the ribosome translates MPs into the proper orientation in the bilayer. **(if the above part is expanded on about the bilayer properly, the following should make more sense)** Once the protein is embedded into the membrane, the hydrophobic core of the bilayer results in an energetic penalty for any unpaired hydrogen bond donors and acceptors (Marinko et al., 2019; Popot & Engelman, 1990, 2000). TM helices, inherently made up of **… amide nitrogens and carbonyls oxygens along the protein backbone, would generally not be favorable in this environment.** **However, this lack of favor in the bilayer …** forces TMs helices to adopt a standard α-helical structure where backbone amide protons and carbonyl oxygens undergo intrahelical hydrogen bonding to stabilize within the hydrophobic membrane.

As individual domains of the MP are threaded into the membrane, the TMs begin to interact. This process is regulated by an intricate distribution of hydrogen bonding, weak polar and electrostatic interactions, and van der Waals forces that govern the stabilities of the unfolded and folded states. Hydrogen bonding not only regulates secondary structure within TMs, but has been shown to drive MP folding when interhelical hydrogen bonds form by polar amino acids (Ser, Thr) on opposing helices **(flip this sentence)** (Johnson et al., 2007; Yano et al., 2002; Zhou et al., 2001). Additionally, electrostatic interactions between positively charged residues (Lys) and electronegative ring amino acids (Tyr, Trp, Phe) results in association between helices (Johnson et al., 2007). However, these interactions only highlight a subset of amino acids (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). Van der Waals forces occur between any atoms within close contact, meaning that these interactions are ever-present during the MP folding process. However, not many studies have solely accounted for the influence of van der Waals forces on MP folding and association. **(maybe flip this entire paragraph? I need to emphasize AAs as building blocks for proteins earlier on)**

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

Johannes Diderik van der Waals won the 1910 Nobel Prize in physics for his contribution to thermodynamics. **(If you’re gonna mention him, build on him more geez)** Named after him, the van der Waals force is an intermolecular interaction that occurs between atoms in close proximity. Each atom has its own van der Waals radius, or an atom’s distance of closest approach without forming a covalent bond. **(build, maybe flip this and the next sentence, going to basics atoms being made of electrons, thus the shell and vdW radii, also could go into the experiment that he used to test this because that could be interesting too)** This distance is dependent on the space that an atom occupies in its outer electron shell. When atoms are found at a distance smaller than their combined van der Waals radii and unable to form a covalent bond, the result is a strong repulsive force pushing the atoms away from each other. However, as atoms approach this minimum distance, the result is the weak attractive force known as van der Waals. **(better end)**

Soluble proteins are driven to fold by the hydrophobic effect, where nonpolar amino acids are forced to the core of the folded state. The nonbonded atoms at the core of these proteins are found in close contact, resulting in a weak attraction by van der Waals forces also known as packing. **(rephrase to instead end with tight packing)** Although van der Waals is not a driving force for soluble protein folding, it is a necessary force that is always present in the folded state. However, in the hydrophobic environment of the membrane, the hydrophobic effect does not drive MP folding. This means that MP folding and association must rely on other forces to reach their 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, van der Waals packing is a necessary force involved in MP folding. But because MPs are engulfed within lipids, it is difficult to tease out the thermodynamic impact of van der Waals on the MP folded state. **(reword, but getting there)**

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 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 individual or portions of proteins in close contact. Each of these forces plays a role in the ability for a protein to be stabilized in the bilayer. When an individual protein subunit is inserted into the membrane, it must destabilize the lipid-lipid packing with more favorable lipid-protein packing interactions. For protein-protein packing to occur, these newly formed lipid-protein interactions must be destabilized for a more favorable combination of protein-protein packing and lipid-lipid packing. Protein-protein (or sidechain) packing, is a technically feasible starting point because of the ability to manipulate sequences.

Previous research has demonstrated that disruption of sidechain packing within the core of bacteriorhodopsin destabilizes protein structure (Faham et al., 2004; Joh et al., 2009). In addition, a recent study using MP design has shown that optimized sidechain packing can stabilize the folded state of the 5-helix bundle protein phospholamban (Mravic et al., 2019). Although these studies suggest that sidechain packing plays a significant role in stabilizing MP structure, there has not been much investigation of the energetic contribution of packing outside of individual MP systems. My research aims to characterize and quantify the extent to which sidechain packing is a driving force for MP folding for the general population of MP structures.

* 1. ****GASright****

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

The GASright motif’s unique sequence and defined structure has been shown to permit an uncommon structural feature. The short interhelical distance allows TM backbones to come in close contact, forming a network of weak hydrogen bonds where donors are Cα carbons and acceptors are carbonyl oxygens on the opposite helix (Cα–H∙∙∙O=C, or Cα–H bonds). Carbon atoms are not commonly associated with hydrogen bond donors because itis less electronegative than typical nitrogen and oxygen donors. However, these carbons are found near **(could specify the groups here if you wanted)** electronegative withdrawing groups on the peptide backbone, increasing their electronegativity. Estimates from quantum mechanics calculations suggest that the stabilizing energy of an Cα–H bond may contribute one third to one half of that of an N—H donor in vacuum (Scheiner et al., 2001; Vargas et al., 2000). Anderson et al. utilized a combination of computational structure prediction and the experimental assay TOXCAT to determine the influence of this network of Cα–H bonds. By predicting and analyzing GASright motifs found in natural sequences using *in vivo* TOXCAT, they showed that structures with more Cα–H bonds had a higher stability in the dimer state (Anderson et al., 2017). Additionally, Díaz-Vázquez et al. measured the free energy of association of GASright structures using *in vitro* FRET, concluding that the thermodynamic stability of GASright proteins is well correlated with the *in vivo* TOXCAT (Díaz Vázquez et al., 2023). Both studies posit that the primary forces involved in GASright association are a combination of hydrogen bonding and van der Waals packing. Because the goal of my research is to understand the extent to which packing can drive proteins to associate, this motif is an excellent control to evaluate my designed sequences against for their propensity to associate using solely van der Waals packing. **(could probably work on wording in this last portion)**

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

The complex nature of the lipid bilayer makes it difficult to directly study the forces involved in MP folding. Many groups focus on utilizing a combination of experimental and computational tools to estimate the contribution 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. This thermodynamic information can be used to assess and validate computational models, which can be further implemented to design and engineer novel proteins. In this section, I will summarize different experimental and computational approaches that have been used to study the association and folding of single pass TMs, including the ability to solve and predict MP structures using both experimental and computational methods. **(needs to be better, talk about generally what structural method say, thermodynamic say, etc.)**

****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). **(check Gladys and Samantha’s again, but I think being this vague about things like nanodiscs is fine)**

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 conditions that yield publishable results in high resolution. MPs make up only 4.6% of structures deposited in the PDB (April 2024; PDB). Lacking structures for MPs to assess and study, researchers focus on using other methods to study MP folding and association. **(reword ending sentence)**

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 that emit different wavelengths of light, or fluorophores, are first attached to individual proteins. The emission spectrum of a donor fluorophore overlaps with the excitation spectrum of an acceptor fluorophore. If the proteins associate, then the donor fluorophore transfers its energy to the acceptor fluorophore, resulting in emission of the acceptor fluorophore. Conversely, if there is no association, there will only be donor fluorescence. 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). **(could go into more detail on how the spectra biophysically works and energy transfer happens if you want)**

Mass spectrometry (MS) of MPs is a growing field, with more groups using it to study 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 also been used to study TM helix-helix interactions. The TM of interest is fused to ToxR, a dimeric transcription factor, and maltose binding protein to determine proper insertion. 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 gene for CAT has been replaced with green fluorescent protein (GFP), allowing fluorescent readings to be used to assess the association levels of the MPs of interest (Armstrong & Senes, 2016). The development of this assay allows it to be utilized in high-throughput applications such as cell sorting, 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. **(probably can build on this more, but might be fine since it’ll be in the paper)**

1.5.4 Computational methods

In conjunction with structural determination, computational methods have been created to evaluate previously solved MP structures. Many methods look to further understand MP folding by developing energetic terms that aim to estimate the thermodynamics of association. The Baker Lab at University of Washington are experts in this area, utilizing known energetic and structural information for the development of 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. 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 used 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). 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. 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. (slightly better ending sentence? Haven’t mentioned that small TMs aren’t often found?)

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 my designed proteins in regions outside of the GASright mildly associate in comparison to GASrights that associate according to both hydrogen bonding and van der Waals packing.

**In Chapter 3**, I discuss research at the interface of experiments and computation in detail, citing the inspirations for data mining, the design algorithm, and the analysis that led to the conclusions found in my paper.

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

**In Chapter 5**, I share a collaboration with the SciFun program at UW-Madison, detailing my PhD journey through a chapter written for the public. I summarize the goals of my research in leyman’s terms while reflecting on the lessons that graduate school imparted onto me, giving my fully transparent thoughts on how my research affected my physical, emotional, and mental well-being. I showcase a bit of creativity in sharing these thoughts on how science and research has helped me grow during my time in graduate school.

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

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>

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.

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>

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>

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>

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>

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.

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>