3.2 Introduction

To study MP folding, researchers aim to identify common structural patterns found among MP systems. The protein data bank (PDB) was established to collaborate and share discovered protein structures globally. This tool lets researchers deposit solved protein structures for others to access and evaluate their findings (Berman et al., 2000). Initially, protein structures were solved primarily using x-ray crystallography; X-ray crystallography has contributed to solving ~80% of MP structures (Kermani, 2021). MP structures have also been solved by nuclear magnetic resonance (NMR). Solid-state NMR has bypassed the need for detergents in crystallography, obtaining structures of MPs less than 50 residues within lipid bilayers or nanodiscs (Liang & Tamm, 2016). More recently cryo-EM has been used to solve MP structures. In addition to bilayers and nanodiscs, it is possible to solubilize and obtain the structures of MPs within detergents, saposin-lipoprotein nanoparticles, amphipols, and peptidiscs (Januliene & Moeller, 2021). Cryo-EM enables MP structures to be studied in a large variety of different environments, giving researchers the ability to study alternative structures of MPs by changing solubilization conditions.

Despite advancements in MP structural characterization, many efforts take years to ascertain conditions that successfully solve structures in high resolution. MPs make up ~30% of known protein coding genes and integral MPs make up 60% of all drug targets (Arinaminpathy et al., 2009; Overington et al., 2006). Yet only 4.6% of structures deposited in the PDB are MPs (April 2024; PDB). Solving the structures of MPs is difficult due to the need to mimic interactions found between the lipid bilayer and protein. Additionally, MPs are difficult to express in quantities necessary for structural experiments. Instead of focusing on structural determination, some groups aim to utilize information from known structures to advance MP research. Using previously solved protein structures as datasets, researchers have developed computational algorithms and tools that identify common motifs and patterns among MP structures. These tools utilize our current understanding of structures to deduce the impact of forces such as van der Waals packing or hydrogen bonding.

Computational tools have been developed to help assess our understanding of the forces that drive MP association. By deriving the contributions of these forces to protein stability, we can predict and/or design unknown proteins. Molecular dynamics simulations permit researchers to use established statistical and energetic potentials to simulate MP folding over time (Karplus & Petsko, 1990; MacKerell et al., 1998). Structure prediction tools use known information from previously solved structures to estimate the structure of MP folded states (Elofsson & von Heijne, 2007). Protein design strategies build on structure prediction, building unknown structures as simple model systems to assess the current understanding of MP folding (Ghirlanda, 2009). MP design to study TMH systems has been successful: peptides were engineered to associate with the TM helix of integrins and a cytokine receptor (Mravic et al., 2024; Shandler et al., 2011; Yin et al., 2007), an integral MP successfully transferred electrons across the lipid bilayer (Korendovych et al., 2010), a 4-helix bundle was designed to transport Zn2+ across the bilayer (Joh et al., 2014), and phospholamban was redesigned using packing interactions and shown to successfully fold (Mravic et al., 2019).

My research expands on previous prediction and design studies. I surveyed possible TMH dimer conformations by extracting backbone helix-helix conformations from MPs found in the PDB. I then sampled different AA combinations, designing the interface along a standardized backbone sequence. These designed proteins were predicted for their ability to associate using previously established energetic functions, and their stability was assessed using a complementary high-throughput assay. This combination of techniques allowed me to develop an algorithm to design thousands of TMHs to study in high-throughput. In this chapter, I detail the development of my computational algorithm and tools used to analyze my high-throughput data.