1.5.4 Computational methods

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

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

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

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

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

1.5.5 Working at the experimental and computational interface

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

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

For Chapter 3

Molecular dynamics (MD) simulations use statistical potentials and energetic functions to predict the structural and thermodynamic changes that MPs undergo within the membrane along a fixed timescale (Ash, Zlomislic, et al., 2004; Lindahl & Sansom, 2008). The development of the Monte Carlo algorithm alternatively can simulate similar thermodynamics utilizing decreasing temperatures to approach a stable folded state (Hansmann & Okamoto, 1999; Heimburg, 2000; Kokubo & Okamoto, 2004; Kuhlman et al., 2003; Müller et al., 2003). These approaches enhanced the ability for researchers to study MP thermodynamics.

MP computational tool reviews:

[**https://www.sciencedirect.com/science/article/pii/S000527361730233X**](https://www.sciencedirect.com/science/article/pii/S000527361730233X)

[**https://www.sciencedirect.com/science/article/pii/S0022283621003831**](https://www.sciencedirect.com/science/article/pii/S0022283621003831)

[**https://www.cambridge.org/core/journals/quarterly-reviews-of-biophysics/article/de-novo-protein-design-a-retrospective/FF37903868E1651D7E61A8495FB00B50**](https://www.cambridge.org/core/journals/quarterly-reviews-of-biophysics/article/de-novo-protein-design-a-retrospective/FF37903868E1651D7E61A8495FB00B50)

[**https://www.sciencedirect.com/science/article/pii/S0959440X11000972**](https://www.sciencedirect.com/science/article/pii/S0959440X11000972)

[**https://www.cell.com/fulltext/S0969-2126(11)00458-8**](https://www.cell.com/fulltext/S0969-2126(11)00458-8)

[**https://royalsocietypublishing.org/doi/10.1098/rstb.2016.0214**](https://royalsocietypublishing.org/doi/10.1098/rstb.2016.0214)

[**https://www.sciencedirect.com/science/article/pii/S0022283621003892**](https://www.sciencedirect.com/science/article/pii/S0022283621003892)