Assessing van der Waals packing as a driving force in membrane protein association and folding

Gilbert Loiseau
Advisor: Alessandro Senes

Committee Meeting
August 31st, 2023
2:30 PM

Background

Proper membrane protein folding is necessary for essential biological functions such as cell signaling and gene regulation. Misfolding of membrane proteins often leads to disease phenotypes including growth defects and cancer. A variety of forces contribute to proper membrane protein folding including hydrogen bonding, weak polar interactions, and van der Waals packing. In order to fully understand how membrane proteins fold for proper function, it is necessary to elucidate the energetic contribution of each of these interactions to the folded state. The ability to drive the transition from the unfolded to the folded state has been characterized and quantified for hydrogen bonding and weak polar interactions, but research is lacking on the contribution of van der Waals packing. Previous research has demonstrated that disruption of packing within the core of bacteriorhodopsin destabilizes the protein structure (Faham et al., 2004; Joh et al., 2009) while membrane protein design has shown that optimized packing stabilizes a redesigned phospholamban structure (Mravic et al., 2019). However, outside of individual systems, the contribution of van der Waals packing to the folded state of membrane proteins has not yet been determined. With my research, I aim to characterize and quantify the extent at which van der Waals packing contributes to membrane protein association.

Results: Design, Sort-Seq, and fluorescence reconstruction of designed homodimers

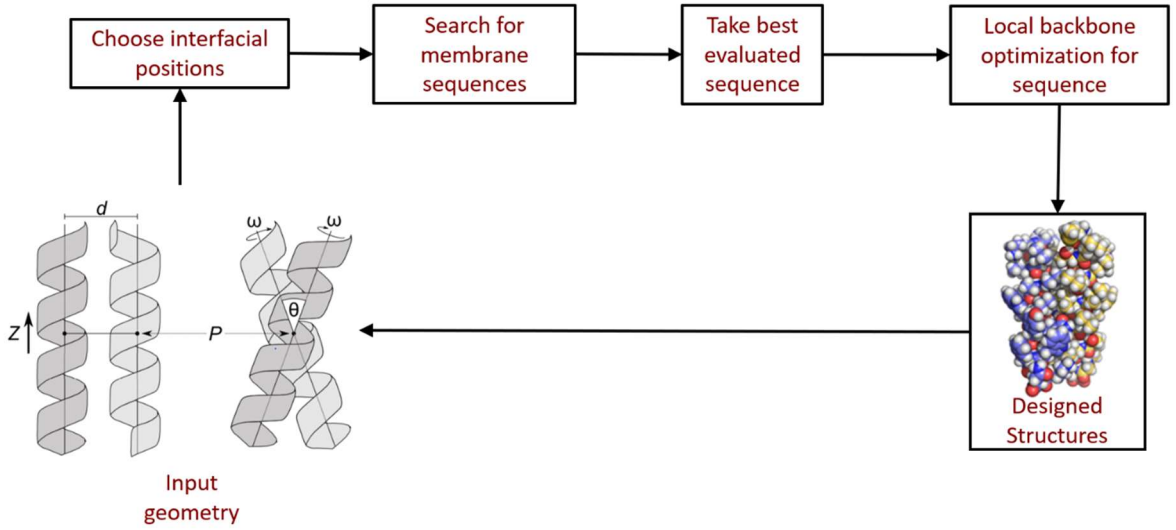

Figure 1. Flowchart of the computational design algorithm. A point from the density estimate obtained from geometries from the PDB (**Fig. S2B**) is input as a structural template. For each region of design, I have consulted the literature to determine interfacial positions (Anderson et al., 2017; Walters and Degrado, 2006) and common computational methods are used to search sequence space for well-packed homodimers. Through a combination of van der Waals packing, hydrogen bonding, and an implicit membrane energetic term, a sequence is chosen to undergo a local geometric repacks to stabilize the structure to the sequence, inspired by a recent backbone design paper (Huang et al., 2022; Kuhlman et al., 2003). This energy is compared to the monomeric state and the design is added to a pool of successfully design sequences if it is more stable. The algorithm then repeats this process until there are no more sequences with good packing at the interface. This process can be repeated will other common geometries from the PDB, allowing me to design hundreds of sequences with an array of expected dimerization based on sidechain packing energies.

Utilizing an updated version of my design algorithm (**Fig. 1**), I designed 1039 sequences to be ordered for a CHIP. A majority of these designs are primarily driven to associate by an increase in calculated van der Waals packing, allowing me to assess the range at which van der Waals packing affects dimerization. In addition to my designed sequences, I made mutations at the interfacial residues of my sequences and predicted their energy scores. I have categorized these mutants in two ways: clashing mutants and void mutants. Clashing mutants were made by making mutations at positions of the designed interface with small amino acids, expecting that the large amino acid would lead to the interface no longer being amenable to form due to the increase in steric bulk. Void mutants were made by making mutations at positions with large amino acids, expecting the loss of atoms to decrease the stability of packing at the interface. The aim of this experiment is to result in a large pool of designs with mutants that can be used to trim our data down to only the structures with mutations that support our designed interface. Overall this resulted in 5303 sequences that were ordered and tested in Sort-seq (**Fig. S1**).

After cloning my sequences into the TOXGREEN plasmid, I ran three replicates of sort-seq (methods, Fig. S1A-B) to appropriately quantify the amount of association for my designed sequences and their mutants. These sequences were then sent for next generation sequencing to determine the counts present within the

appropriate bins. Using the counts from NGS (methods), I reconstructed the fluorescence of each of these sequences (Kosuri et al., 2013). In order to simplify our understanding of this fluorescence as dimerization, we compare the fluorescence of each of our sequences to that of a known strong dimer glycophorin A (Walters and DeGrado, 2006). In addition, we can assess whether our sequences are associating by comparing to the monomerizing mutant of GpA, known as G83I which has a glycine to an isoleucine mutation (Anderson thesis, 2019).

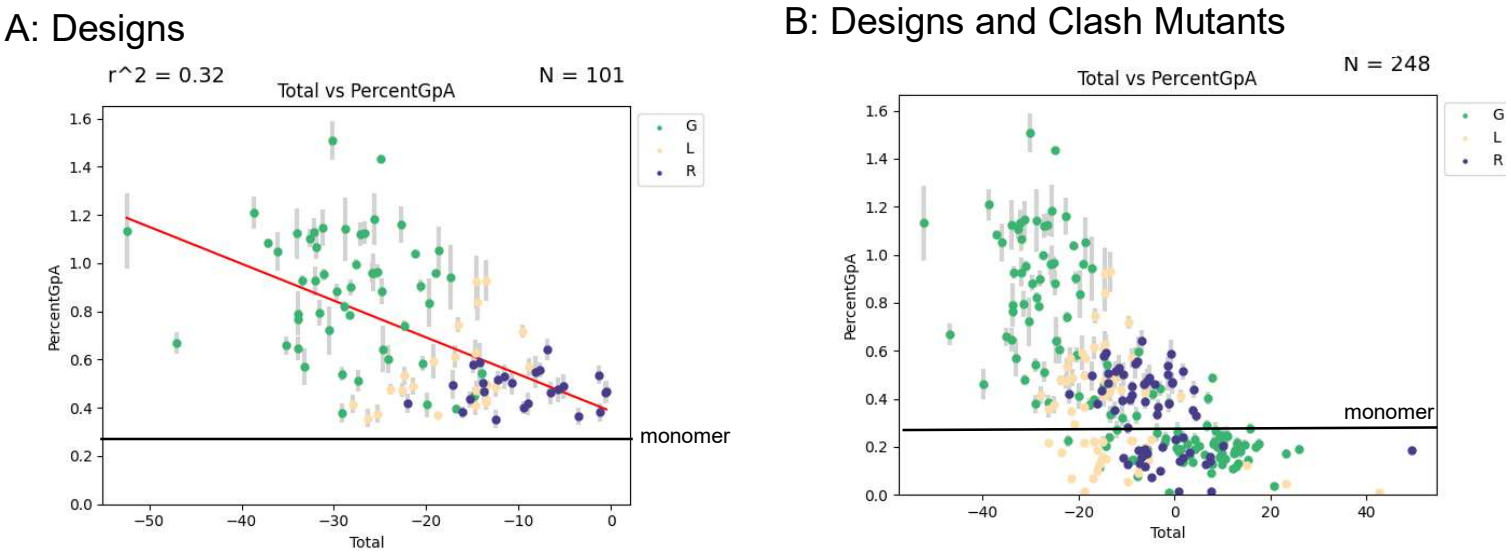

Figure 2. Predicted Energy vs Fluorescence (Percent GpA). The above graphs show the predicted energy score for a sequence on the x-axis vs the fluorescence in terms of a control strong dimer glycophorin A on the y-axis. The points on this graph are separated by the region that they were designed in: GASright (G) in green, Left (L) in yellow, and Right (R) in purple. **A)** Designed sequences on this graph were filtered using clash mutants that were found to break helix-helix association. This leaves 101 designed sequences (53 G, 24 L, and 25 R). **B)** Scatterplot of both designs and their respective clash mutants. A majority of sequences predicted to be clashing that have an energy score above 0 are found to not associate.

To better evaluate the extent at which van der Waals packing can help facilitate membrane protein association, I designed structures in three different geometric regions (**Fig. S2**). After filtering sequences by the clashing mutations that successfully break helix-helix association, I am left with a pool of 101 sequences where a majority (53) are GASright designs (**Fig. 2A**). These GASright designs are primarily stabilized by a combination of van der Waals packing and interhelical hydrogen bonding (Anderson et al., 2017). Within this data, GASrights are shown to have higher fluorescence than designs that are not stabilized by hydrogen bonding in the left and right handed regions. This data suggests that van der Waals packing can facilitate the association of weak dimers, but both the combination of hydrogen bonding and van der Waals packing may be necessary to facilitate strong dimerization. Additionally, when plotting energy scores of our designs and clash mutants, our energetics seem to be capturing the expected trend where mutants of these structurally supported designs have higher energy scores than their respective designs and do not associate (**Fig. 2B**).

Conclusion and Future Directions

I am currently in the process of analyzing sequences and their mutational data. The clash mutants resulted in a nice dataset of structures where helix-helix association was as expected with our design and mutant energy scores. However, the void mutants that were expected to break helix-helix association did not have as favorable a result. A majority of these void mutants resulted in dimer fluorescence (**Fig. S3**). Because these mutants only decreased the stability of van der Waals packing rather than making the interface completely unamenable to structurally associate, it is possible that many of these mutants have a small decrease in fluorescence compared to wild type sequences. However, small changes in fluorescence are difficult to determine the effect of each of these mutants with respect to their energy score. I am currently running structure prediction using the same energetic terms I used for design on these void mutant structures, which will allow me to better compare the expected fluorescence to our designed structures. Additionally, mutating the interface to a smaller amino acid may allow for tighter van der Waals packing at the interface, resulting in higher fluorescence. If this is the case, I may be able to utilize these mutant structures as additional designed sequences that suggest that small changes in van der Waals packing may or may not lead to fluctuation in fluorescence.

Currently, it appears that van der Waals packing can lead to a low level of association when compared to stronger dimers facilitated by hydrogen bonding in the GASright. One caveat may be that our current assay does not have a large enough dynamic range to determine the overall association strength by solely van der Waals packing. To better determine this, we may be able to assess the biophysical stability of these structures through the correlation seen between TOXGREEN and FRET (Díaz Vázquez et al., 2023). I will need to determine the expression of a variety of designed constructs to ensure that there is no expression variability between designs. I will perform western blots on a pool of designed constructs, and if there is no expression variability, I will be able to correlate our predicted stability in TOXGREEN to free energy, resulting in a biophysical assessment of the extent at which van der Waals packing can facilitate membrane protein association.

Methods

I aim to explore the effect of sidechain packing on homodimerization. I have identified common geometries from the PDB as structural templates for computational design of sequences of membrane protein sequences. I was able to identify areas of high density in the geometric space obtained from the PDB, allowing me to choose from a variety of geometries to explore for protein design (**Fig. S2B**). Each of these geometries is standardized with a poly-leucine backbone to control for expression and insertion of our sequences (Zhou et al., 2001; Anderson et al., 2017). To specifically vary the sidechain packing contributing to association and control for the possible interfaces, I chose positions that have been previously found to associate in these geometric regions in the literature (Anderson et al. 2017; Walters and Degrado, 2006). I then used well-known computational algorithms to filter and search sequence space for amino acid combinations that pack at the dimerization interface (Koehl and Delarue, 1994; Hansmann and Okamoto, 1999). Using a minimalistic set of energy functions that measure van der Waals packing, hydrogen bonding, and membrane implicit solvation (IMM1) (MacKerell et al., 1998; Lazaridis, 2003; Krivov et al., 2009), I have measured the stability of sequences at a variety of geometries, determining geometries where design of well-packed sequences is possible (Figure 1B). Using this subset of geometries, I designed a test subset of 1039 sequences that have been evaluated using a complementary high-throughput assay.

To evaluate successfully designed sequences, I used TOXGREEN, an *in vivo* dimerization assay that quantifies dimerization propensity through the output of sfGFP (**Fig. S1A**). This dimerization assay has been optimized for a high-throughput approach known as sort-seq, which combines fluorescence activated cell sorting (FACS) with next-generation sequencing (NGS) to evaluate dimerization propensity of TM domains (**Fig. S1B**). I expressed an oligo pool library consisting of my designed sequences and cloned them into TOXGREEN plasmids, allowing me to simultaneously study the association of hundreds to thousands of designed constructs. Based on the expression of sfGFP, cells were sorted into different bins with different fluorescence thresholds. These plasmids were then purified out of the cells in each bin and enumerated via NGS. Based on the counts present within each bin, the fluorescence profile, and thus dimerization propensity, was reconstructed for each dimer (**Fig. S1C**).

Supplementary Figures

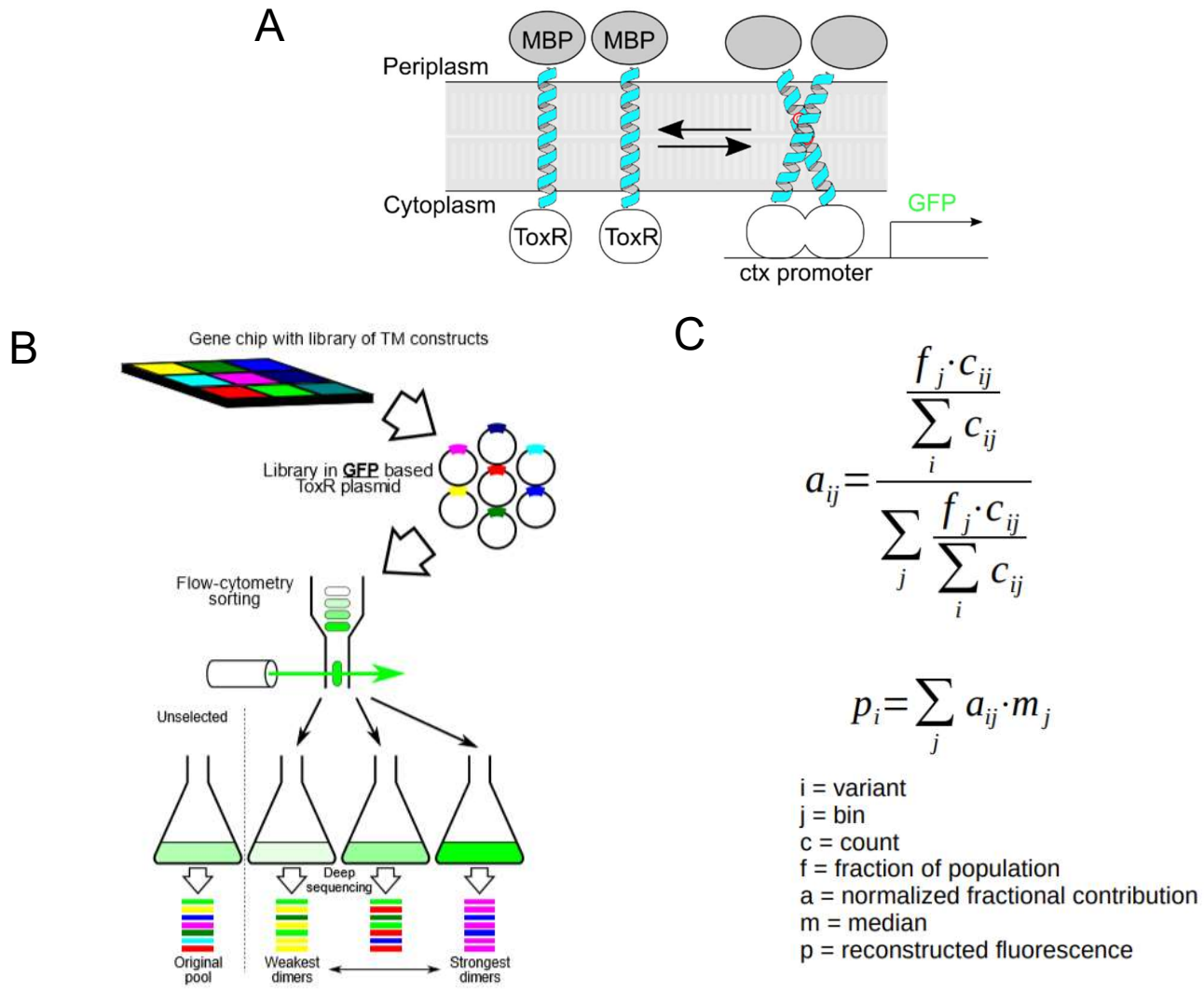

Fig. S1: TOXGREEN sort-seq schematic. A) TOXGREEN. TOXGREEN is an *in vivo* assay that reports TM helix-helix association through the expression of reporter gene sfGFP.
B) Sort-seq. Individual TM domains are synthesized on a chip using oligo pool technology and cloned into TOXGREEN plasmids. In the GFP-based assay, cells will fluoresce based on GFP expression correlating to the dimerization propensity of the expressed TM domain chimera. This GFP expression will be sorted by fluorescence-activated cell sorting (FACS) followed by next-generation sequencing of the bins.
C) Fluorescence Reconstruction. Fluorescence is reconstructed according to the above equation. Sequence counts (c) for each sequence (i) in each bin (j) is multiplied by the fraction of the population of the bin (f). This gives the total percent of a sequence within a particular bin. This is then divided by the total percent of sequence withing all bins, which results in the normalized fractional contribution of a sequence within a bin (a). After getting all of the contributions for a sequence within all bins, and contribution for a sequence per bin (a) is multiplied by the median of its respective bin (m). These are then summed together, resulting in the reconstructed fluorescence of a sequence across all bins (p).

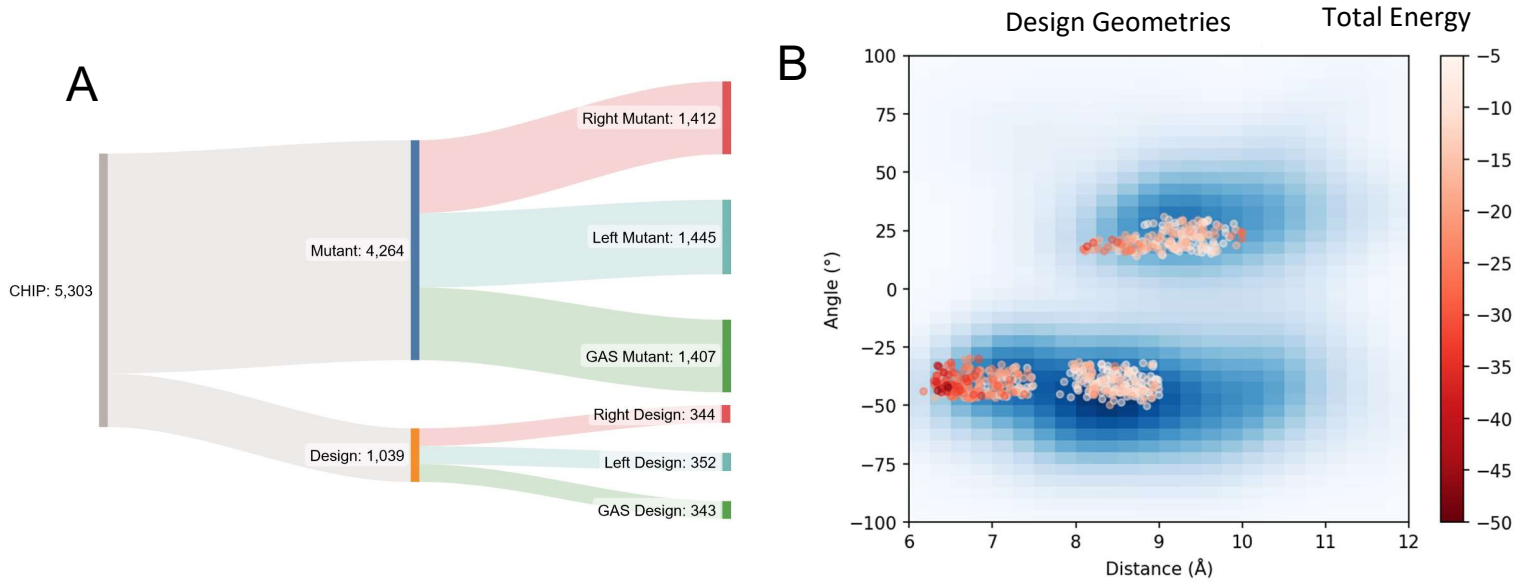

Fig. S2: CHIP Design Breakdown. A) Flow Chart of CHIP Order. I ordered a CHIP of 5303 sequences, 1039 of which were designed and 4264 mutants at positions designed to break dimer association.
B) Design Geometries by Energy Score. I designed sequences in the three most densely populated portions of of the membrane protein geometric space: GASright (6.5-7.5 Å distance), Right handed (8-9 Å distance), and Left handed (8-10 Å distance with a positive crossing angle). GASright designs are prominently stronger because this region is amenable to association via Cα hydrogen bonding.

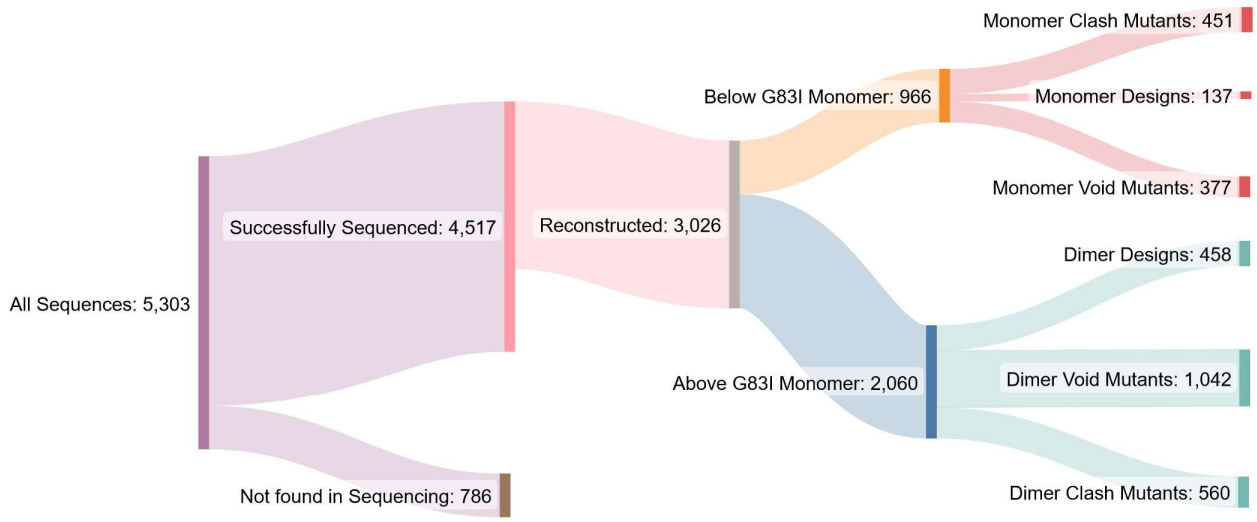

Figure S3. Flow Chart of CHIP Fluorescence Breakdown. Of the 5303 designed sequences ordered on the CHIP, 4517 were successfully found within the NGS of my sorted population. Within that 4517, I was able to successfully reconstruct 3026 sequences that were present in all replicates of my data. When compared to the fluorescence of the monomerizing mutant control, 2060 sequences were found to be more fluorescent. These were then broken down into designs and the different mutant sequences that were made prior to design. To get the dataset shown in Figure 2, I took the 451 monomer clash mutants and filtered out sequences that had at least 2 clashing mutants and at least a 50% decrease in fluorescent signal from the wild type design.

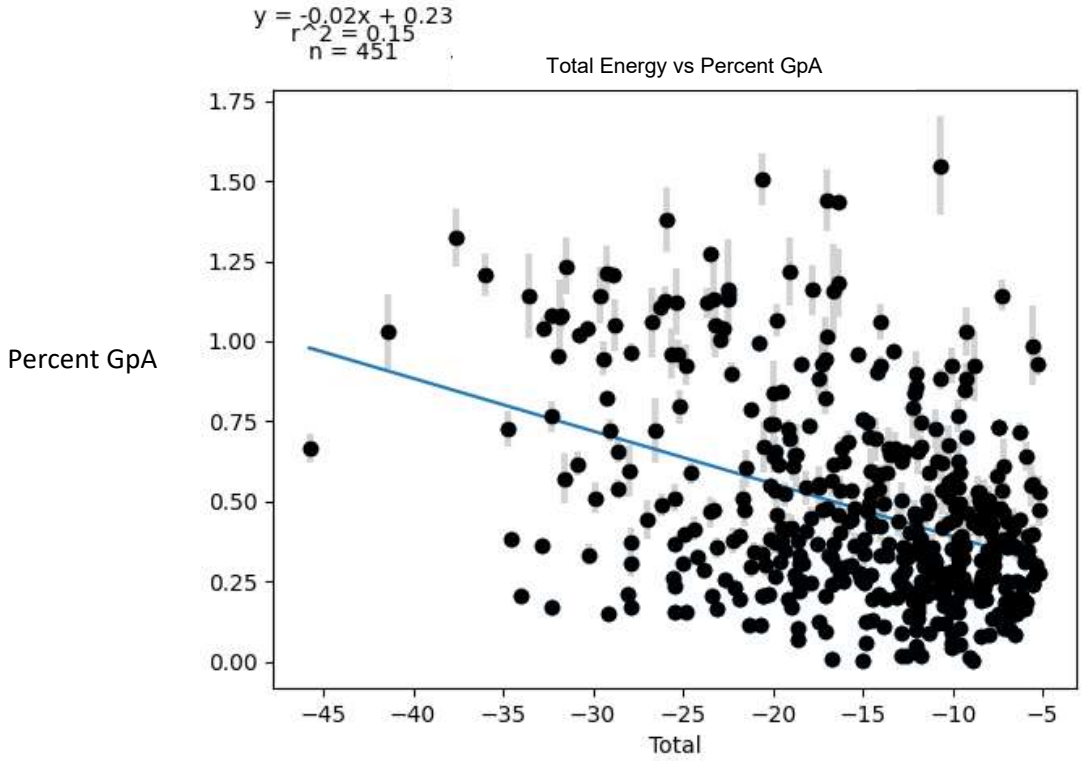

Figure S4. Unfiltered Clashing Dataset. The above graph shows all of the wild type designed sequences that had at least one clashing mutant below the G83I monomer threshold (25% GpA).

References

1. Faham, S., Yang, D., Bare, E., Yohannan, S., Whitelegge, J.P., and Bowie, J.U. (2004). Side-chain Contributions to Membrane Protein Structure and Stability. Journal of Molecular Biology *335*, 297–305.
2. Joh, N.H., Oberai, A., Yang, D., Whitelegge, J.P., and Bowie, J.U. (2009). Similar Energetic Contributions of Packing in the Core of Membrane and Water-Soluble Proteins. J. Am. Chem. Soc. *131*, 10846–10847.
3. Mravic, M., Thomaston, J.L., Tucker, M., Solomon, P.E., Liu, L., and DeGrado, W.F. (2019). Packing of apolar side chains enables accurate design of highly stable membrane proteins. Science 363, 1418–1423.
4. Zhou, F.X., Merianos, H.J., Brunger, A.T., and Engelman, D.M. (2001). Polar residues drive association of polyleucine transmembrane helices. Proc Natl Acad Sci U S A 98, 2250–2255.
5. Anderson, S.M., Mueller, B.K., Lange, E.J., and 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, 15774–15783.
6. Koehl, P., and Delarue, M. (1994). Application of a Self-consistent Mean Field Theory to Predict Protein Side-chains Conformation and Estimate Their Conformational Entropy. Journal of Molecular Biology 239, 249–275.
7. Hansmann, U.H.E., and Okamoto, Y. (1999). New Monte Carlo algorithms for protein folding. Current Opinion in Structural Biology 9, 177–183.
8. MacKerell, A.D., Bashford, D., Bellott, M., Dunbrack, R.L., Evanseck, J.D., Field, M.J., Fischer, S., Gao, J., Guo, H., Ha, S., et al. (1998). All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B 102, 3586–3616.
9. Lazaridis, T. (2003). Effective energy function for proteins in lipid membranes. Proteins 52, 176–192.
10. Krivov, G.G., Shapovalov, M.V., and Dunbrack, R.L. (2009). Improved prediction of protein side-chain conformations with SCWRL4. Proteins 77, 778–795.
11. Fleming, P.J.and F.M. Richards (2000) Protein Packing:Dependence on Protein Size, Secondary Structure and Amino Acid Composition. J.Mol.Biol. 299, 487-498.
12. Armstrong, C.R., and Senes, A. (2016). Screening for transmembrane association in divisome proteins using TOXGREEN, a high-throughput variant of the TOXCAT assay. Biochim Biophys Acta 1858, 2573–2583.
13. Anderson, S. M. (2019). Understanding the GAS_{right} motif: sequence, structure, and stability. Thesis. University of Wisconsin, Madison, WI.
14. Kosuri, S., Goodman, D.B., Cambray, G., Mutalik, V.K., Gao, Y., Arkin, A.P., Endy, D., and Church, G.M. (2013). Composability of regulatory sequences controlling transcription and translation in Escherichia coli. Proc. Natl. Acad. Sci. U. S. A. 110, 14024–14029.
15. Huang, B., Xu, Y., Hu, X., Liu, Y., Liao, S., Zhang, J., Huang, C., Hong, J., Chen, Q., Liu, H. (2022). A backbone-centred energy function of neural networks for protein design. Nature 602, 523–528.
16. Kuhlman, B., Dantas, G., Ireton, G.C., Varani, G., Stoddard, B.L., Baker, D. (2003). Design of a Novel Globular Protein Fold with Atomic-Level Accuracy. Science 302, 1364–1368.
17. Walters, R.F.S., DeGrado, W.F., 2006. Helix-packing motifs in membrane proteins. Proc Natl Acad Sci U S A 103, 13658–13663.
18. Díaz Vázquez, G., Cui, Q., Senes, A. (2023). Thermodynamic analysis of the GASright transmembrane motif supports energetic model of dimerization. Biophysical Journal 122, 143–155.