Van der Waals packing facilitates membrane protein association

Abstract

Visual Abstract

Design algorithm, experiment, fluorescence correlation

Introduction

Membrane proteins comprise 25-30% of the proteins found within protein-coding genes of various organisms (Fagerberg et al., 2010). Proper membrane protein folding is critical for essential biological functions, including cell signaling, ion balance, and gene regulation. Misfolding of membrane proteins 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 understand how protein misfolding plays a role in disease states and progression, it is necessary to investigate how these proteins fold. However, studying membrane protein folding is inherently a difficult challenge because of their hydrophobic nature. Membrane proteins are difficult to express in yields high enough for biophysical experiments, and purification and solubilization of these proteins often lead to aggregation or unfolding (Carpenter et al., 2008). To combat these challenges, much of the research studying membrane protein folding is focused on understanding the biophysical forces that govern the folding process. Investigation of the biophysical forces that govern protein folding will allow us to determine why specific mutations result in diseases caused by misfolding. In addition, this knowledge can be applied to design new therapeutics that specifically target proteins in these misfolded states. Understanding the forces that enable membrane proteins to fold will contribute to our knowledge of how these fundamental forces are involved in human health and disease.

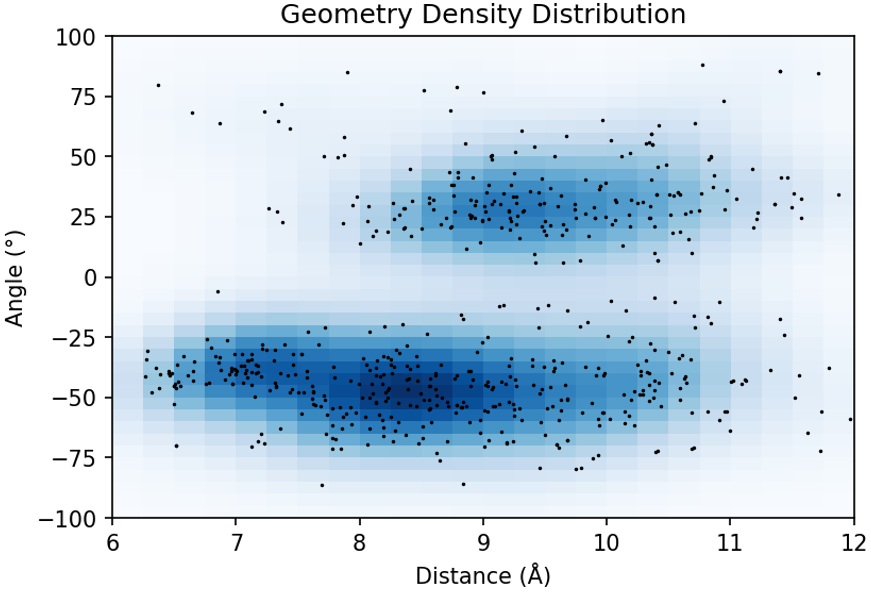
Proper membrane protein folding is regulated by a distribution of stabilizing hydrogen bonds, weak polar interactions, and van der Waals forces between the unfolded and folded states. Previous research has measured the contributions of both hydrogen bonding and weak polar interactions in the membrane and determined that these forces can drive membrane protein folding (Zhou et al., 2001; Yano et al., 2002; Johnson et al., 2007), but research is lacking on the contribution of van der Waals packing. This force is particularly important due to the nature of van der Waals interactions: Even if hydrogen bonding or polar interactions play a significant stabilizing role, because van der Waals occurs between any nonbonded atoms in close contact, it is a necessary force that is always present within the folded state. This means that van der Waals packing is essential for folding, but the extent at which packing can be a driving force for membrane protein folding is unclear. Without understanding the extent at which van der Waals packing contributes to folding, we cannot complete our understanding of how these other forces contribute to membrane protein association and folding.

The contribution of van der Waals packing to membrane protein folding can be broken down into three distinct interactions: lipid-lipid packing, lipid-protein packing, and protein-protein packing. Protein-protein (or sidechain) packing, is a technically feasible starting point because of the ability to manipulate sequences and determine changes in stability due to mutation. Previous research has demonstrated that disruption of packing within the core of bacteriorhodopsin destabilizes protein structure (Faham et al., 2004; Joh et al., 2009). In addition, a recent study using membrane protein design has shown that optimized sidechain packing can stabilize the folded state of phospholamban (Mravic et al., 2019). Although it is known that sidechain packing plays a role in stabilizing membrane protein structure in these individual systems, the energetic contribution of sidechain packing to the folded state of membrane proteins more generally has not yet been determined. My research aims to characterize and quantify the extent at which sidechain packing is a driving force for membrane protein association. To do so, I am investigating the role that sidechain packing plays in the association of homodimers, a simple and tractable model for studying membrane protein folding (Popot and Engelman, 1990). Using large scale computational design on common dimeric backbone geometries found within the PDB and a complementary high throughput assay, we have demonstrated that sidechain packing can facilitate association for a variety of membrane protein structures. This research gives insight into the extent at which fundamental sidechain packing impacts membrane protein association and folding, something that has not yet been done despite the importance of van der Waals packing to membrane protein folding.

The contribution of van der Waals packing to stabilizing membrane protein structure remains a significant gap in our understanding of membrane protein folding (Hong et al., 2014). Determining how sidechain packing impacts the association of membrane protein subunits will give us a better understanding of how membrane proteins assemble to fold stabilized structures. By understanding how these forces contribute to stability, we will be closer to obtaining a holistic view of how all forces involved combine to stabilize membrane protein structure. My research will increase our knowledge of the fundamental rules of membrane protein folding, and add to our understanding of complex membrane protein mechanisms, such as oligomerization and conformational change necessary for essential biological processes including signal transduction and ion transport (Sanders and Myers, 2004; Gregersen et al., 2006). Eventually, this knowledge can be used to design new functional membrane protein structures, advancing the field of synthetic biology.

Results

**Design of membrane protein homodimers**



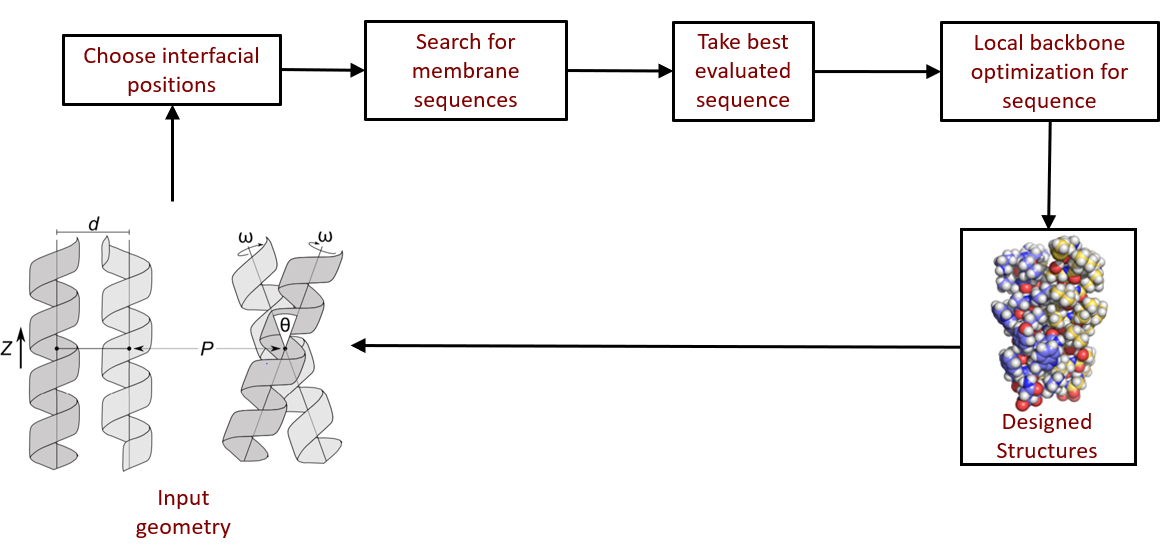


Figure 1. Membrane protein dimer design. (a) Helices within close contact (define how) were extracted from all membrane protein dimers submitted to the PDB in (date; 2019). Orientations of Proteins in membranes (OPM, cite) was used to only choose the proteins found in membranes. These geometric terms are plotted and overlaid over the density of these in space. (b) Design algorithm… (c)

[Insert design algorithm name here] designs homodimer membrane protein sequences by using an iteration of fixed backbone design and local backbone optimization. The sequence is initially optimized by evaluating van der Waals, hydrogen bonding, and an implicit membrane solvation model (cites). Simultaneously, we have developed a entropic term that determines the likelihood of a sequence to be found in membrane proteins by the comparison of amino acids found in membrane proteins (methods, cite). [design name] allows for the efficient design of hundreds to thousands of membrane proteins from a large variety of membrane protein geometries.

To design membrane protein dimers using [design name], we extracted the density of membrane protein helices within close contact found within the PDB (Figure 1A). This resulted in three distinct regions of high density: left handed, right handed, and GASright. ... More about the regions (maybe another figure to explain it, similar to SMA figure 3 to explain the controls)

A white square with black lines

Description automatically generatedA diagram of a dna sequence

Description automatically generatedA diagram of a dna sequence

Description automatically generatedUtilizing polyleu backbone for design

Figure 2. Controlling for expression and insertion variability. (a) [Design algorithm name] was run on a poly-leucine backbone, allowing for 8 variable amino acid positions. GASright and right handed positions are identical, while left handed positions are based on knobs into holes packing motifs found within coiled coils. (b) The variable amino acid positions were allowed to alternate between each of these amino acids, with GASrights designs being the only ones allowed to have glycine. (c) TOXGREEN fuses the designed protein to maltose binding protein (MBP) in the periplasm and ToxR, a dimeric transcription factor, in the cytoplasm. Dimerization of designs results in the dimerization of ToxR, and subsequently the transcription and translation of GFP as a readout of association strength.

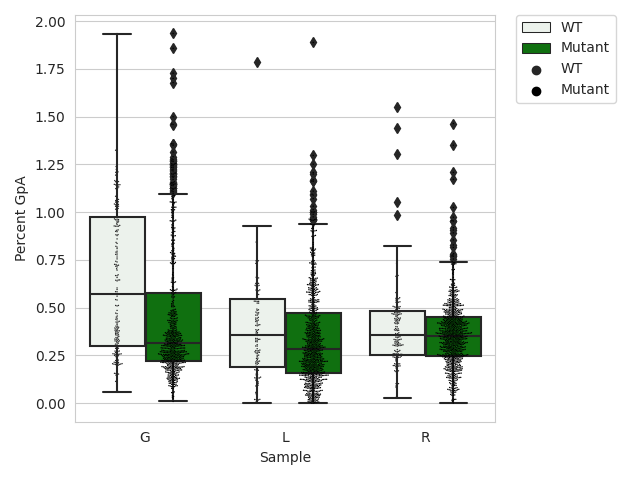
To reduce potential experimental heterogeneity of expression and insertion within the membrane, we chose 8 positions onto a standardized TM helix of 21 leucines, one of the most prevalent amino acids found in membrane proteins (cite). Poly-leucine has been used in previous studies and shown to have little self association, allowing us to posit that the association we see in our experiments is contributed primarily by our designed interface (cite 37, 39, 46, 56 from SMA). By using a poly-leucine backbone and the same number of positions for each of our designs, we expect to control for expression and insertion variability.

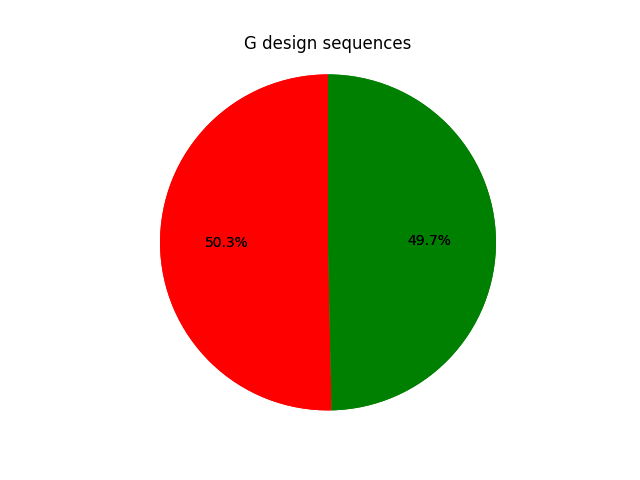
Figure 2A shows the chosen positions at the interface for each of the three regions. GASright sequences have been shown to be stabilized by a combination of van der Waals packing and interhelical hydrogen bonding (cite SMA). By utilizing the same interface for GASright within the right handed designs, we expect to be able to compare directly the potential contribution of van der Waals packing within our structures. Additionally, the left handed design interface is mirrored (better word) after the left handed coiled coil, which has been shown to allow for tight knobs into holes packing (cite). To specifically design membrane protein sequences, we developed a membrane protein entropic term based on the most prevalent hydrophobic amino acids found within membrane proteins (cite Alessandro paper and that other paper on AA abundance). GASrights were allowed to accommodate all of these hydrophobic amino acids including glycine at all positions, whereas right and left handed designs were limited to the rest of the hydrophobic amino acids to prevent the potential formation of interhelical hydrogen bonds. These designed helices were then expressed in plasmids that resulted in a chimeric protein with the dimeric transcription factor ToxR in the cytoplasm. As the helices dimerize, ToxR allows for the transcription and translation of green fluorescent protein as a reporter of association for our designs (cite TOXGREEN). This method has been further developed by our lab into a high-throughput assay that was utilized for this study.

**Sort-Seq: Assessing membrane protein association in high-throughput**

To experimentally assess the dimerization of our 1000 designed proteins and their corresponding mutants, we used sort-seq, a high-throughput method that utilizes TOXGREEN which has been used previously to assess TM homo-oligomerization *in vivo* (cite TOXCAT and green). This method utilizes the expression of a chimera of the designed proteins fused to the dimeric ToxR transcriptional activator. Dimerization of the TM helices results in the dimerization of ToxR, allowing it to bind to a promoter that activates the expression of the reporter gene green fluorescent protein (GFP). This fluorescence can be used as an output for the association of the designed proteins in a biological membrane. An oligo pool gene library of the designed proteins is cloned into *Escherichia coli* bacterial cells that are run (better word) through a fluorescence activated cell sorter (FACs), allowing us to measure the association strength of our designed proteins in high throughput (cite sort-seq/SMA thesis; methods for the code?).

**Fluorescence and association of designed structures and mutants**





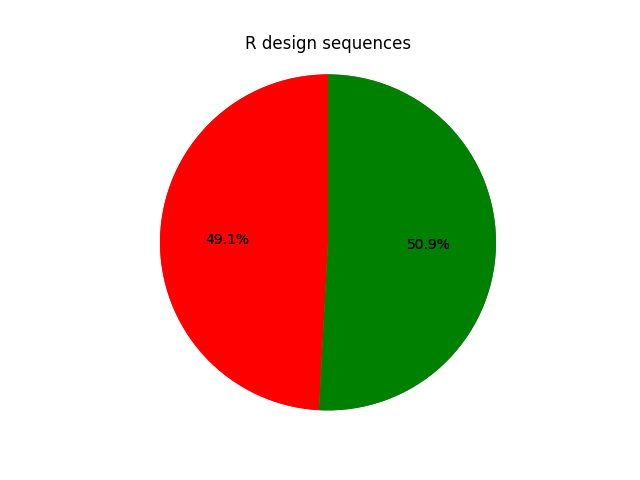
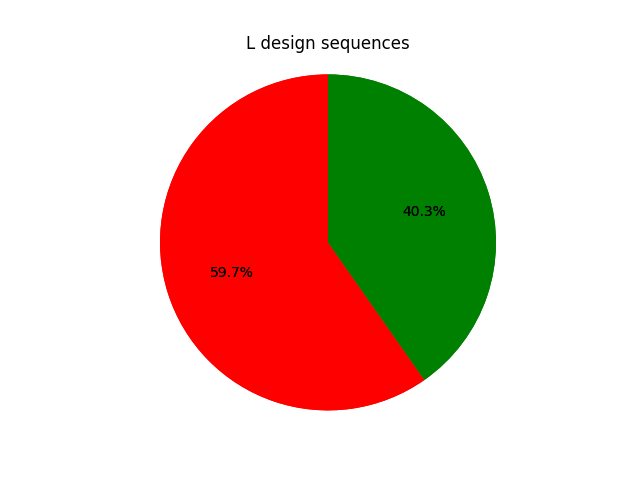


Figure 3. Association of designs and mutants. Utilizing high-throughput assay sort-seq, the fluorescence corresponding to the association of 3026 sequences was reconstructed. (a) Designed sequences are compared to a monomeric control and split into those with a fluorescence below (red) and above (green) monomer. (b) Designs (white) are compared to their corresponding mutants (green) in a boxplot that suggests that wt designs associate more than their mutants on average.

The fluorescence of designed sequences and mutants corresponds to the association strength as determined by high-throughput sort-seq. Sequences designed with a negative crossing angle (right handed) resulted in a higher percentage of structures that can be considered dimers than those with a positive crossing angle (left handed) (Figure 3A). When compared to all of their respective mutants, designed sequences generally fluoresce more than mutants (check the p-values to see if significant). In particular, GASright designs, which were designed with hydrogen bonding, have significantly higher fluorescence than mutants. Both left handed and right handed designs are shown to have a higher mean fluorescence than their mutants (Figure 3B).

**Point mutations to support our designed interfaces**

To partially validate the designed structures, we opted to design mutations at interfacial positions with the expectation of decreased association (cite mutagenesis papers, SMA 15, 17, 57-59). Substitution of a small amino acid such as alanine or glycine with a large hydrophobic amino acid, such as isoleucine or phenylalanine, would prevent the helices from coming into close contact if our designed interface is correct. This should result in a significant decrease in van der Waals packing, supporting our designed structures and giving further insight into the contribution of packing on membrane protein association.

Figure 4. Fluorescence vs Total Energy

A graph of a number of dots and numbers

Description automatically generated with medium confidenceA graph of different types of amino acids

Description automatically generatedA graph of different colored squares

Description automatically generated

Figure 4. Evaluation of dimerization propensity as a function of van der Waals packing. (a) Graph of energy score vs the dimerization of designed sequences. (b) Boxplots comparing the dimerization propensity based on the number of potential hydrogen bonding amino acids. This data suggests that association is not impacted by increasing numbers of hydrogen bonding amino acids. (c) Boxplots comparing the dimerization propensity based on the number of ring amino acids. This data shows that sequences with three ring amino acids may have increased association, which could be due to an increase in stability by electrostatics and/or steric bulk.

To define our subset of structures, we filtered for only sequences where the mutant was monomeric. This resulted in 175 design sequences, where GASrights were increasingly stable following the expectation that we had from our energy score. Although the right and left handed designs do not match our design energy score as well as the GASrights, the overall distribution of our energy scores matches the resulting association: In terms of both energy score and association, GASrights are the strongest, right handed are the weakest, and left handed are in between (Figure 4). This data suggests that van der Waals packing can facilitate membrane protein association in the absence of hydrogen bonding.

Although our design energy scores suggest that van der Waals packing is driving membrane protein association, all of the sequences found in this final dataset contain at least 1 amino acid that can result in hydrogen bonding (serine or threonine). To determine the possibility of hydrogen bonding influencing the overall association strength of these designs, we compared the dimerization propensity between sequences with differing numbers of hydrogen bonding amino acids. Overall, the fluorescence expressed by both mutants and wild type sequences with varying numbers of hydrogen bonds are similar (Fig. 4b). Additionally, when looking at sequences with differing numbers of ring amino acids that could result in stability through electrostatics (cite), the fluorescence is similar between each group except for when there are 3 ring amino acids in a sequence (tryptophan, phenylalanine, or tyrosine) (Fig. 4c). However, there are a low number of sequences in this pool (x) compared to the others (y,z), and the median is similar between all three groups. It is possible that some of these structures are undergoing an increase in electrostatic stability, which is not a metric that we used for our design algorithm. However, the increase in number of atoms and steric bulk from these ring amino acids could also be playing a role in stabilizing through increased van der Waals interactions (cite).

Discussion

Figure 5. Contribution of van der Waals packing… how can I make these figures a bit more easy to understand? Bar graphs of average? It would be nice here if we had some way to say confidently that our vdW designs make sense. Could we have a bar graph of average deltaG for sequences with no ways to have hydrogen bonding vs not? And then electrostatics as well? How many are there of those even?

A graph with colored dots

Description automatically generated with medium confidenceA graph of different colored dots

Description automatically generated

Figure 5. Contribution of van der Waals packing …

Van der waals facilitates…x amount…approximately x% of GASright, known for Hbonding. Can we conclude anything about the contribution of van der Waals in GASrights and other regions of proteins? Are there any other studies that have left/right handed structures that do experiments that demonstrate packing/dimerization ability? Or portions of proteins found in some of these structures (maybe coiled-coil FtsLB data? What conclusions if any can be drawn from these? If they exist, could I take a look at the distribution of known proteins and their sequences/interfacial residues (any way to do this computationally?) and compare to them? I could at the very least compare to the residues found in the structures that I extracted from and see if this gives me any insight? Would it be possible to look at them positionally? Or at least look at the most similar structures by RMSD or something?

Could have a figure here for the confirmation of designed structures using a different method? TMDOCK could really be useful here because of the compared deltaG. Honestly, I likely could run TMDOCK in low throughput on just the proteins that match up with our clashing data, and maybe compare the deltaG values just to see how an experimental deltaG compares to an alternate energetic prediction than mine? Assuming all of the proteins work, I think it could be worth it

I think I’m going to likely say something about ring stacking and maybe cite some papers on it; unfortunately, I don’t think I have a single sequence with neither a hbonding AA (S, T, G) or electrostatic AA (F, Y, W). This will make claims a bit harder, but I think I can just have a caveat for that for the electrostatic ones. Oof might make it a little harder: every single sequence seems to have at least one hbonding aa. Is there any way I can show that although this is the case, they aren’t positionally likely to be able to hbond? Like looking for sequences with these at a farther distance?

Acknowledgements

Twist, Flow core, Raman lab, CBI, SciMed, TMDOCK maybe, should send an email

Methods

* Many supplementary figures
  + More details in the starter outline

**References**

1. Fagerberg, L., Jonasson, K., von Heijne, G., Uhlén, M. & Berglund, L. (2010). Prediction of the human membrane proteome. Proteomics 10, 1141–1149.
2. Sanders, C.R., and Myers, J.K. (2004). Disease-Related Misassembly of Membrane Proteins. Annu. Rev. Biophys. Biomol. Struct. 33, 25–51.
3. Gregersen, N., Bross, P., Vang, S., and Christensen, J.H. (2006). Protein Misfolding and Human Disease. Annu. Rev. Genom. Hum. Genet. 7, 103–124.
4. Carpenter, E.P., Beis, K., Cameron, A.D., and Iwata, S. (2008). Overcoming the challenges of membrane protein crystallography. Current Opinion in Structural Biology 18, 581–586.
5. 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.
6. Yano, Y., Takemoto, T., Kobayashi, S., Yasui, H., Sakurai, H., Ohashi, W., Niwa, M., Futaki, S., Sugiura, Y., and Matsuzaki, K. (2002). Topological Stability and Self-Association of a Completely Hydrophobic Model Transmembrane Helix in Lipid Bilayers. Biochemistry 41, 3073–3080.
7. Johnson, R.M., Hecht, K., and Deber, C.M. (2007). Aromatic and Cation−π Interactions Enhance Helix−Helix Association in a Membrane Environment †. Biochemistry 46, 9208–9214.
8. 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.
9. 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.
10. 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.
11. Popot, J.L., and Engelman, D.M. (1990). Membrane protein folding and oligomerization: the two-stage model. Biochemistry 29, 4031–4037.
12. Hong, H. (2014). Toward understanding driving forces in membrane protein folding. Archives of Biochemistry and Biophysics 564, 297–313.

Supplement

**Table Totals**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Total | WT | Mutant |
| G | 944 | 247 | 697 |
| L | 1024 | 149 | 875 |
| R | 1058 | 200 | 858 |