**Van der Waals packing facilitates membrane protein association**

**Abstract**

- TBD

- Start calling figures by a descriptive name instead of by numbers so that it’s easier to move them around later

- Start bulletting the story

**Graphical Abstract**

- Have a nice figure of how my design algorithm works, maybe tie it into the figure 1 (extracted geometries), finally output the protein and show fluorescence

- I also think visually for my presentation, having some sort of pymol driven video would be nice

**Premise/Intro**

- Membrane protein folding and association is important (look at Mravic et al. and other papers to understand how to write this)

* Structure is dictated by the interactions occurring at protein-protein interfaces
* Previous research has shown that hydrogen bonding and polar interactions act as forces that facilitate association
* Van der Waals packing is an necessary force, but the extent at which packing facilitates association is not well understood (why? Limitations of previous research like Mravic and Bowie?)
* This research uses a high throughput approach to study the impact of packing on membrane protein association
* Could talk about the fact that packing is a required feature,
  + not necessarily a driving force (bowie: cost of void is similar to voids in soluble; trading protein-protein interaction for lipid could mean that it’s necessary vs a driving force)
  + mention protein-protein packing vs protein-lipid packing
  + taking a thermodynamic approach
  + may also need to talk a bit about the balance with entropy

**Figure 1**

A diagram of a graph

Description automatically generated with medium confidence

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…

* Membrane protein design using the membrane protein geometric landscape from the pdb:
  + X proteins were extracted from orientations in membrane proteins
  + Any two helices in close contact (define) were extracted from each membrane protein structure (total number), and the helical geometric information was extracted using MSL
  + These helical geometries were used to determine a density plot of the potential designable space for membrane protein dimer geometries (figure)
  + Protein design using MSL (need to decide how much detail here; most should go into methods)
  + Designed x proteins with a range of expected stabilities
  + Talk about the design regions (GAS, right, left)
  + Have representations of each of the designs from the front and side view
* Maybe add an oval around each of the design sections corresponding to the colors that are later in the figures

**Figure 2**

**A diagram of a dna sequence

Description automatically generatedA diagram of a dna sequence

Description automatically generated**

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) 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.

* Describe experimental protocol and how TOXGREEN works with references to TOXCAT and TOXGREEN literature
  + Should we also add in detail for sort-seq here?
  + If sort-seq is added, include a figure that details how we convert the reconstructed fluorescence to percent GpA using TOXGREEN
  + Add in a sequence log for each class of design

**Figure 3**

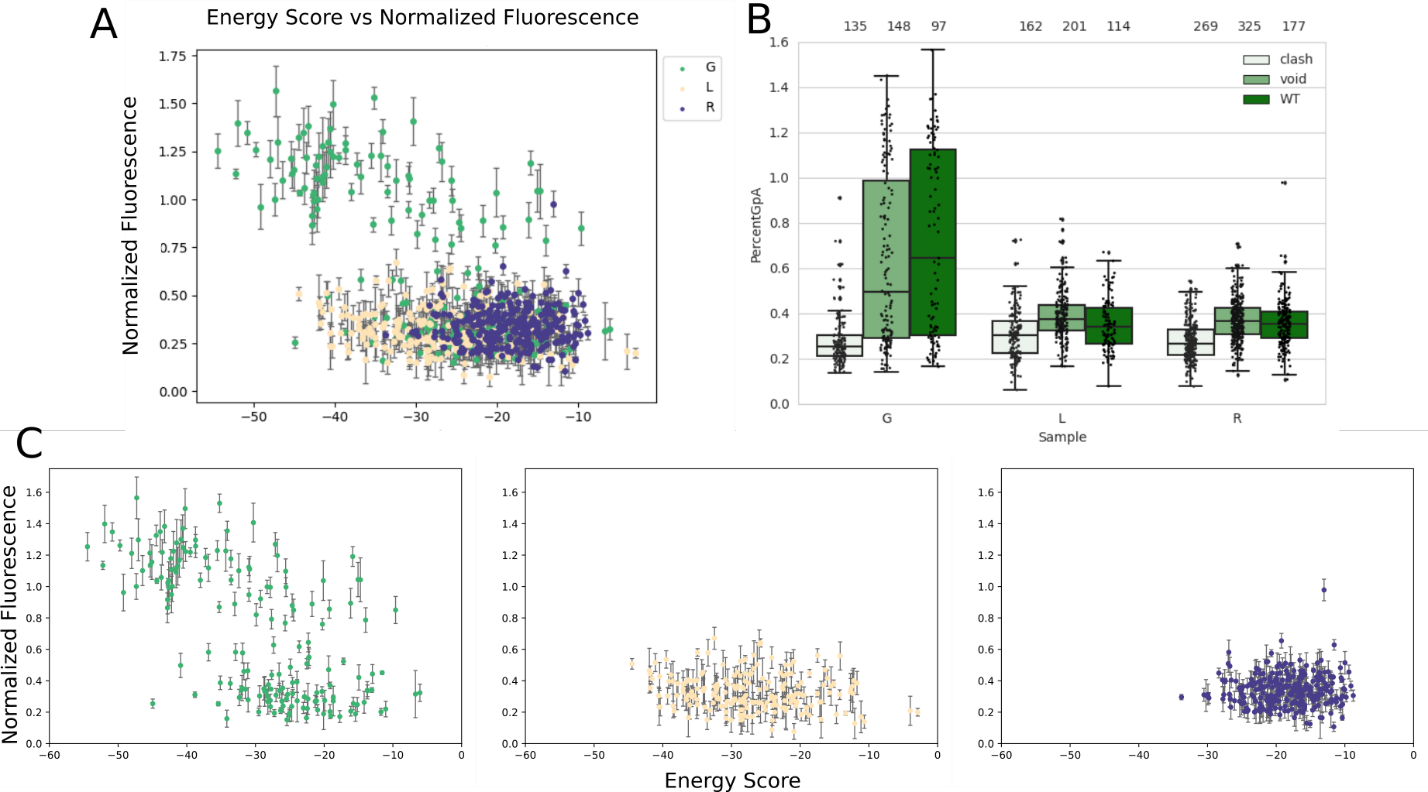
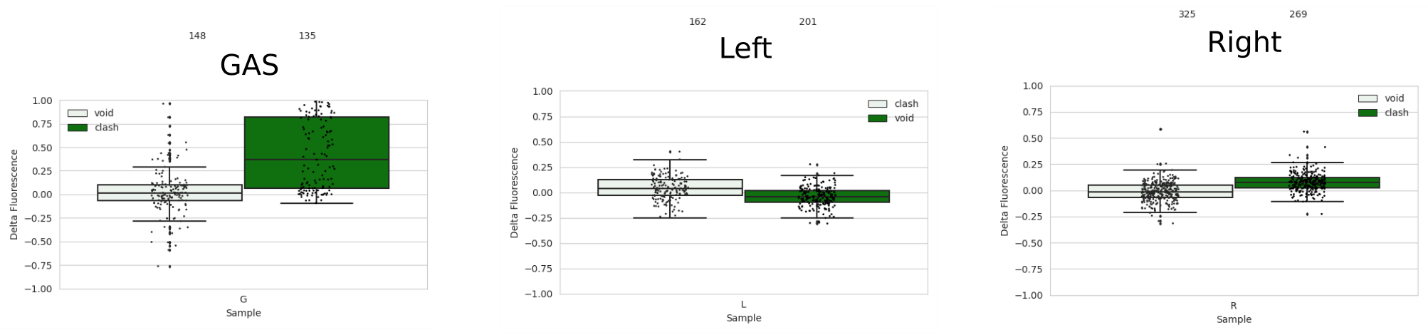


Figure 3. A) The normalized fluorescence plotted against the design energy score for all of the designed sequences in the dataset. The plot is separated by the geometric region of design with GASright in green, Left in yellow, and right in purple. B) The dataset is filtered for designs where at least 1 clash and void mutant is also found to fluoresce. The WT (dark green) fluorescence distribution is compared to the fluorescence distributions of their corresponding mutants. C) Separated graphs for normalized fluorescence against the design energy score in A.

* Fluorescence distribution vs design plots for all sequences
  + Haven’t checked significance between designs and mutants yet, but that info would go here (DO THIS)
  + Could show regression graphs for all of these sequences
  + Talk about GASright being the most stable, then left, then right
    - Importance of hydrogen bonding for membrane protein stability
  + Design energy score appears to better capture the GASright than the other structures
    - Mention potential for utilizing other energy terms (electrostatics)
* Add in a line around 0.4 (conservative limit for association

**Delta Fluorescence Figure**



Fluorescence of wild type designs subtracted by fluorescence of their corresponding mutants is plotted above.

* Plot this as an x vs y plot

Structures for the top 5 designs for each region

A screenshot of a computer

Description automatically generated

Figure: Frontal and side view of predicted structures for the top 5 fluorescent sequences for each region.

* Overall they look well packed, with some slight voids in the left and right regions
* Add in the energies for each of these, and the SASA (compare that to the VDW per square angstrom (vdw/SASA))

**Amino acid distribution per Fluorescence**

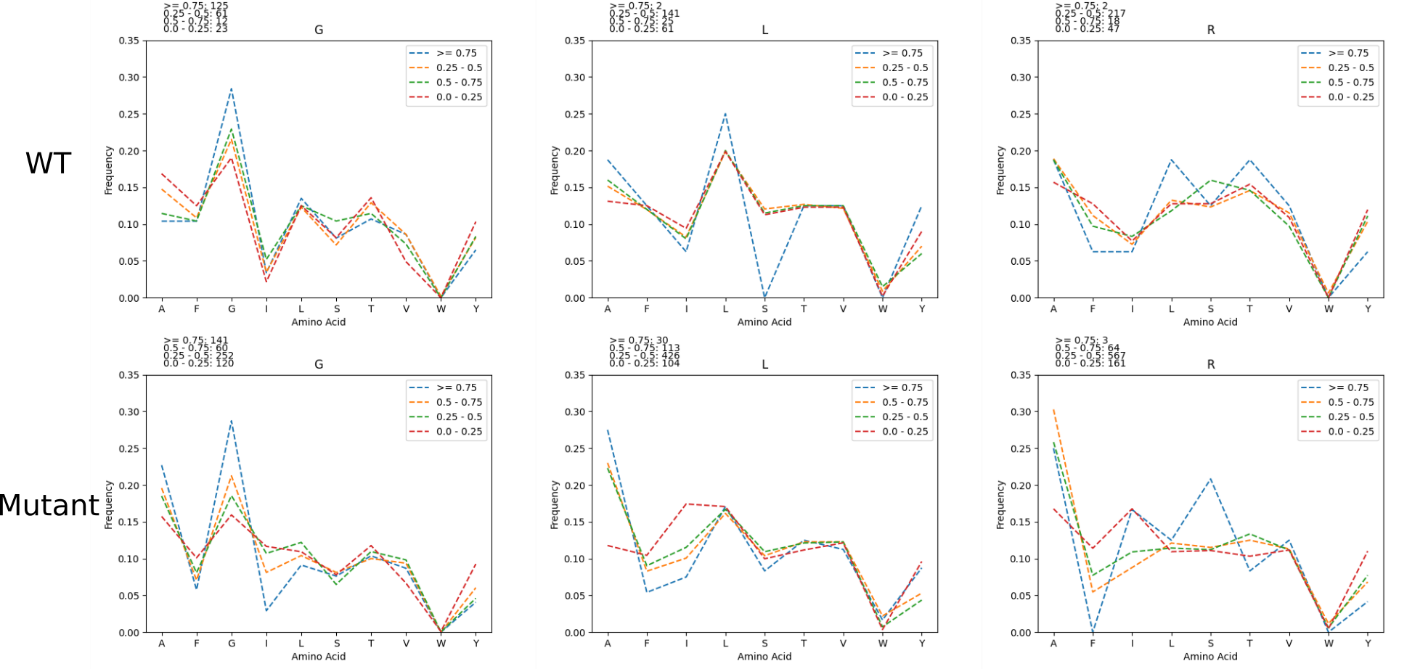


Figure: The amino acid frequency was calculated for sequences based on their fluorescence (>= 0.75, 0.5 – 0.75, 0.25-0.5, and < 0.25). This was done for designs and all mutant sequences.

* In all regions, isoleucine is more prevalently found in sequences with lower fluorescence
* Phenylalanine is also more prevalently found in mutant sequences with lower fluorescence
* Tyrosine is kind of variable, but more prevalently found in mutant sequences with lower fluorescence
  + Suggest that these AAs disrupt association
* Alanine is more prevalent in sequences with higher fluorescence
  + Suggests that it helps facilitate association
* Distributions of hbond AAs (Ser, Thr) are similar throughout, suggesting that potential hbonding by these AAs is not the main driving force for association

**Delta G figure**

A graph of a graph with numbers and dots

Description automatically generated with medium confidence

Figure 4. Converting fluorescence to free energy.

* Detail here for how we convert from fluorescence to delta G
  + Cite Gladys paper
* Talk about how we trimmed to this dataset:
  + This image is for:
    - At least 1 clashing mutant < 35% OR the difference between the WT and mutant is > 50%
  + Give more information about clashing mutations here (probably also earlier in the design section)
* Talk about the impact of van der Waals packing in left and right versus GASright