Chapter 4: Future Directions

Table of Contents

Improving the design algorithm

* Energetics doesn’t seem to work outside of GASright. Design with other energetic terms/reparameterize the van der Waals term using the information that we have for these sequences that associate primarily through packing.
* Random ideas:
  + Alphafold is really good at predicting structures without energetics using multiple sequence alignments and …(shape fitting basically?).
    - Take the sequence space information for all predicted structures in alphafold and create an “energy” term
      * This term will make it more likely to get AAs that are typically found to associate in alphafold membrane protein models
    - See how well Alphafold is able to predict structures in different dimer geometric regions
      * Use this as a way to evaluate the designability of the space, effectively replacing our density plot of frequency with how well it’s likely that an area is to be computationally predicted (basically breaking down areas into pockets that we have a really good understanding of and ones that we don’t)
  + ML on the chosen geometric space?
  + Shape fitting
    - Energetics didn’t work very well, but many of our interfaces were suggested to be confirmed by our mutational data. Structurally, it seems like we’re able to design proteins at least semi-accurately
    - See if there are design papers that do shape fitting type stuff and see how they do it. We currently only have SASA implemented in MSL, but it doesn’t really work well.
    - If there is something, instead of using energetics during a monte carlo backbone optimization, use the fitting ability instead. Can then take that structure and try to calculate the energy post (maybe using something based on alphafold: Use energy as a way to say how designable the protein is, rather than how well it associates)

Improving association prediction

* Other energetics terms or reparameterizing our energetics for the other regions:
  + Is it possible that GASright is accurate with these energy terms because they are simply the most involved? What terms might we be missing? Does that Calpha-hbond somehow mitigate the error that we’re getting in our other designs? Could hbonding and packing actually be contributing less in other instances, as if the hbond is acting as a multiplier of sorts? Why don’t we break the hbonding component into regular and Calpha?

Heterodimer design

* Currently working with another graduate student in the lab to develop the current algorithm for heterodimers.
  + Talk about the complexity of heterodimers (increased parameters, needing to calculate the monomer energy for each helix, moving each helix individually rather than simultaneously with homodimers)
  + Finding the AAs that associate most often would be helpful here
  + Initially working with GASrights, since our energetics works best with GASright structures
  + An initial start:
    - Take the heterodimer space from geometries within the GASright region
      * For homodimers, I picked the most common xShifts and angles, then assessed the (…energies?) for a variety of zShifts and axial rotations to determine ones that would work
      * However, we also have that data from the MPs. It’s just currently a bit more difficult to interpret (I’m not sure how axial rotation and z are decided actually; maybe dependent on where the starting point of the helix is?)
        + What we might be able to do is just use the differences between axial rotations of the two helices and z shifts
        + OR we could actually make it standardized, by first choosing the closest AAs at the interface to basically define the interfacial parallelogram for these structures

Advancing forces research

* Packing: Rerun design without any AAs that have rings or potential hbonding
  + Try different backbones: polyAla, polyVal, etc.
    - With the same interface, determine differences in packing impacting association by the small changes in each AA
    - May give us some insight into how shapes/minute changes in packing affect association
      * Assumption is that leucine would pack the best and would pack less decreasing in size, but unsure if this is the case and how much of a decrease it would be.
      * If there’s a way to correlate the decrease with change in overall AA size, then we’ll gain a bit of insight into how different sized AAs affect the packing stability (can kind of try to correlate these energetics with the shape?)
* Hbonding and ring: Rerun design with only hbonding and ring AAs
* Sort-seq on any of these
* Machine learning on the terms to better predict, adding terms, etc.

Ideas for detecting protein concentration in high-throughput

* I think there was a paper I read for chem bio seminar that was pretty relevant
* If we could do this, then we’d be able to better normalize the fluorescence yield of each of our sequences rather than just by total fluorescence output, which assumes that our proteins are all expressing similarly (true for GAS and left, but for some reason not true for right)
  + Althought we currently only use total protein expression, so it’s possible that even though the right handed proteins express at a lower level than the others overall, they still insert the same