Computationally Designing Proteins as Recognition Elements that will Bind to Insulin Using RFDiffusion

Latherial Calbert, Jazmine Torres, A. Clay Richard, and Robert J. Pantazes
Department of Chemical Engineering, Auburn University

Motivation

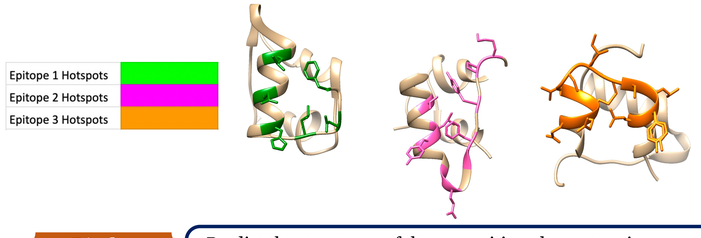
The BIO-SENS team is working together on a 4-stage project geared towards designing an in-line biosensor to aid in biomanufacturing. Proteins can be used as recognition elements. This stage's goal is to design 15 protein recognition elements to bind to the target analyte, insulin, spearheading the 1st stage of the project.

Background

Continuous in-line protein evaluation is essential for biotechnological progress. One significant target analyte identified by the Advanced Regenerative Manufacturing Institute is insulin. Insulin is a peptide hormone produced by beta cells of the pancreatic islets encoded in humans by the INS gene,[1] and machine learning enables the design of recognition elements for insulin binding by predicting structure and residues of the binder.

Method

Epitope Selection
Identify three regions located on insulin to target for designing protein binders.



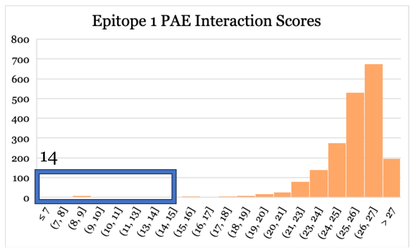
Binder Backbone Design
Predict the structures of the recognition elements using RFDiffusion to generate backbones. 2000 designs were generated per epitope.

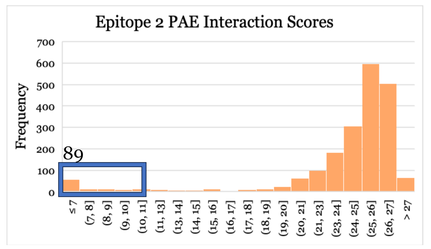
Binder Sequence Design
Predict the amino acid sequences for the previously generated backbone designs using ProteinMPNN.

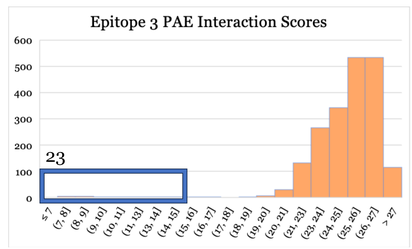
Property Calculation
Calculate the binding energy, buried surface area, and shape complementarity of the protein complexes using CHARMM fixed-backbone minimization, Rosetta all-atom minimization, and Rosetta Interface Analyzer

Final Design Selection
Down-select designs using the calculated properties to determine which are best to apply to the biosensor experimentally.

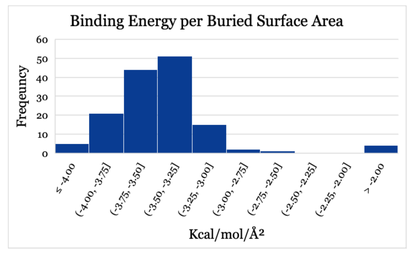
Results

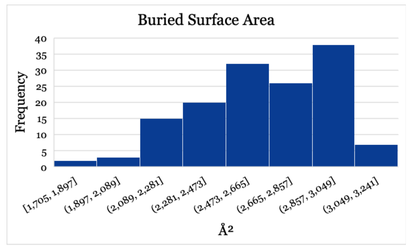


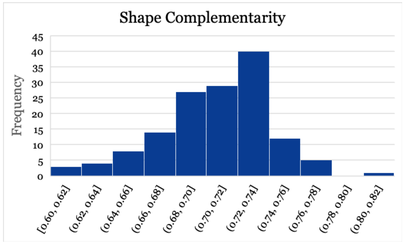




Predicted Aligned Error Interaction Scores (PAE) were calculated for all 2000 protein recognition element designs epitope. PAE indicates the expected positional error at residue x if the predicted and actual structures are aligned on residue y (using the Cα, N and C atoms). PAEs are measured in Ångströms and capped at 31.75 Å². The PAE score was the initial filtering metric for the designs. Epitope 1 and 3 was filtered by a PAE value less than or equal to 15, netting 14 and 23 designs respectively. Epitopes 2 was filtered by a PAE value of less than or equal to 10 netting 89 designs. Epitope 2 retained the most designs after the first filtering, and the subsequent filtering included all of the filtered designs from each epitope.



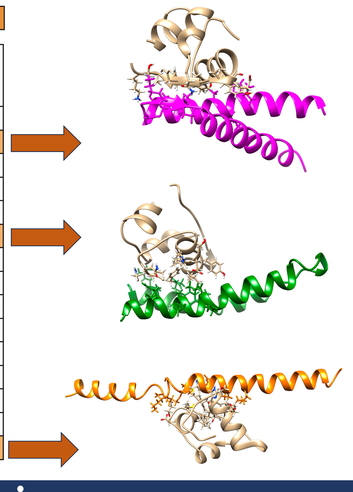




The last criteria used to filter the designs were Binding Energy per Buried Surface Area (BE/BSA), Buried Surface Area (BSA), Shape Complementarity (SC), and an in-house unpublished method that evaluates the stability of protein interfaces. Buried surface area is the total area of residues that are not on the surface of the protein and not exposed to the solvent. Binding energy per buried surface area is a criteria that quantifies the change in Gibbs free energy per buried surface area. Shape complementarity measures how well the protein-protein interface fits together in terms of stability. Higher values of the BE/BSA and SC factors indicate a higher likelihood of binding success, and a threshold value of 750 BSA is sufficient to increase the likelihood of experimental binding success. In order to find optimal filter values, gradient descent was used to help find the balance between each factor's maximum value in association with finding the largest amount of applicable protein designs.

Top 15 Protein Recognition Element Designs

Insulin Epitope	Design/ Version	Predicted Alignment Error	Shape Complementarity	Binding Energy per Buried Surface Area Kcal/mol/Å²
2	586_1	5	.73	-3.27
2	724_1	6	.74	-3.77
2	700_1	6	.73	-3.62
2	790_1	6	.73	-3.58
2	89_1	6	.73	-3.86
1	440_0	8	.74	-3.32
2	16_0	5	.72	-3.59
2	538_1	6	.71	-3.57
2	586_0	5	.79	-3.83
2	661_0	5	.72	-3.41
2	819_0	7	.72	-3.49
2	626_1	7	.78	-4.02
2	277_0	8	.76	-3.64
2	289_0	8	.73	-3.64
3	480_1	8	.77	-3.86



Conclusion

- Binder designs with PAE scores less than or equal to 10 tend to be promising for experimental success.
- High SC and BE/BSA increase success likelihood in experiments, while a threshold value BSA value of 750 is sufficient to increase the likelihood of experimental success.
- Epitope 2 was the most promising site, as most of the designs with a PAE less than or equal to 10 came from this site.
- The next step is to test all 15-recognition element designs experimentally, verifying that they will bind to the insulin analyte. Although this research focused on developing binders to insulin, the workflow may be repeatable for other proteins.
- This study's results can further the study of other recognition elements.

Future Work

The designed recognition elements obtained from this research will be further evaluated using experimental testing by the Balog Laboratory.

Acknowledgements

We would like to thank the Pantazes Lab and would like to express our gratitude to the entire BIO-SENS team for their support. This research is supported by the National Science Foundation EPSCoR Research Award (#2119237).

References:
(1) Voet D, Voet JG (2011). *Biochemistry* (4th ed.). New York: Wiley.
(2) Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., Yuan, D., Stroe, O., Wood, G., Laydon, A., Žídek, A., Green, T., Tunyasuvunakool, K., Petersen, S., Jumper, J., Clancy, E., Green, R., Vora, A., Lutfi, M., Figurnov, M., ... Velankar, S. (2022). AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic acids research*, *50*(D1), D439–D444. https://doi.org/10.1093/nar/gkab1061