Use of AlphaFold3 to predict single mutations in antibodies

Charles Bennington , **Principal Investigator:** William DeWitt, **Lab Mentor:** Maggie Russell

UW Medicine RESEARCH & GRADUATE EDUCATION

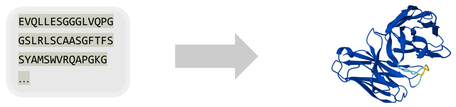
GONZAGA UNIVERSITY

Introduction

Alanine scanning mutagenesis: Within molecular biology, alanine sampling is a site-directed mutagenesis technique used to determine the contribution of a specific residue to the stability of the function.

Why do we do we use Alanine?
Alanine is used as a replacment for mutagenesis scans beacuse alanine is a neutral amino acid that is less likely to cause issues by just being alanine, meaning we get a more accurate map of the structral diffrences.

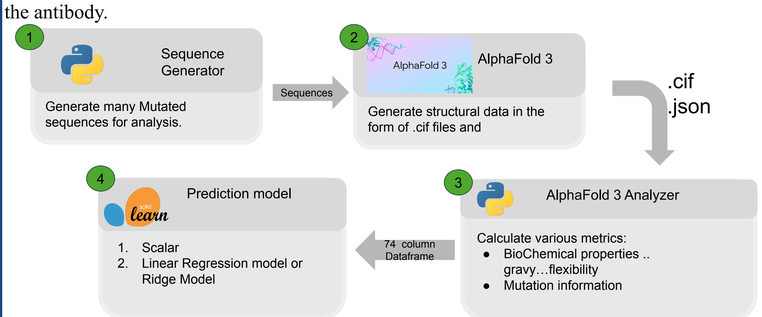
Alpha Fold3: A recent breakthrough in protein structural predictions. The third iteration of the alpha fold deep learning model, which predicts the 3D structure of proteins based on their Amino Acid sequence.



Cost Savings:
Currently, alanine sampling is a very time consuming and expensive process. If we could predict the sites that are more likely to lead to a decrease in the stability of the protein, we could cut costs.
This model is available for free, and and allows for the predictions on folded proteins. This AI model is much cheaper to run than a full alanine sample.

Methodologies

The goal of this research is to be able to predict the change in Protein Expression of a real lab observed value using a AlphaFold3. Delta Expr representing a change in overall stability of the antibody.



Why Alphafold3?
Alphafold3 is the most up to date model for predicting the structure of proteins, it was developed by google deepmind, and gives accurate predictions for what a protein looks like

Why python?
Python is considered the standard for the scientific computing world, some scientists have recently switched to Rust, but for the most part python is the standard.

AF3 jobs run : 243

Lines of code written: 2,100+

Pull Request comments : 291

Why sklearn?
Sklearn was choosen beacuse it was lightweight, and easy to understand.
Has this been tested?
This project has not been tested as a sub

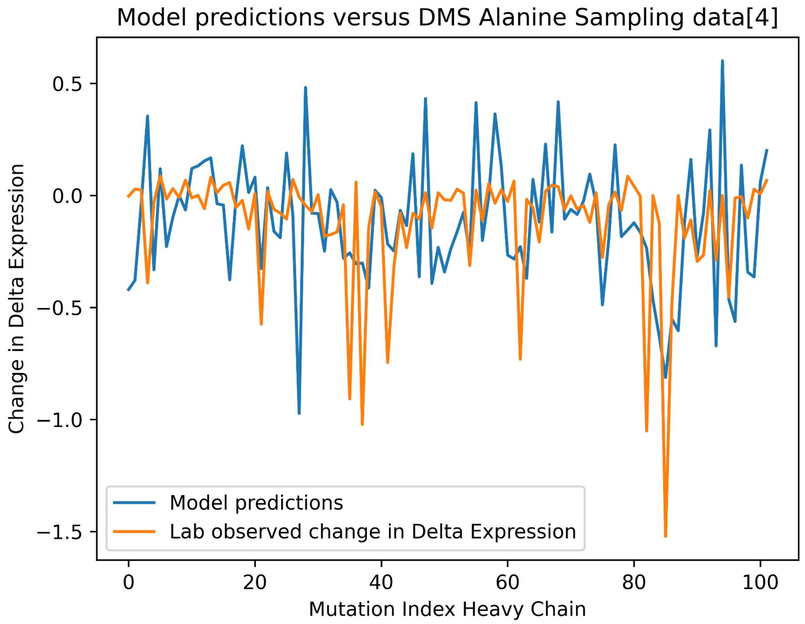


Fig 1: This displays the model written versus the real observed delta expr from a allaine sampling. This shows how effective the model is. As we can see the model is able to correctly predicts indexes that have a large difference in delta expr. This specific model was achieved by using multiple K-fold validators. The use of K-Fold validation ensures that we are getting the correct and real model performance.

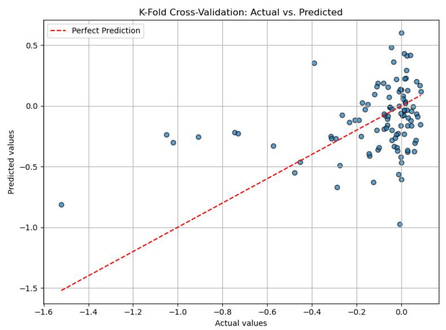
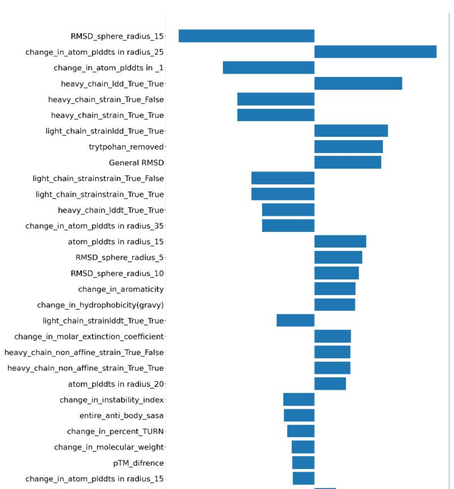


Fig 2: This displays the accuracy of the model, and how close it is in general to getting the correct prediction. As you can see the model tends to guess a little bit low, something we are currently working on. This dataset is compared to the replay dataset from a prior paper. [4]

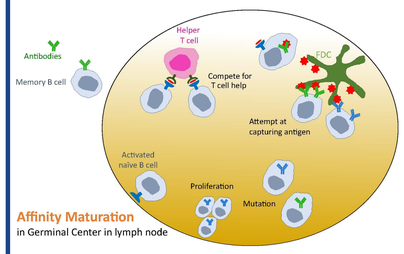
Fig 3: This shows the top coefficients for the model, and which metrics inferred from the alphafold3 model were most used.

Looking at this model we can see that
- Change in plddts
- General RMSD
- Strain
- Radius of gyration



Background

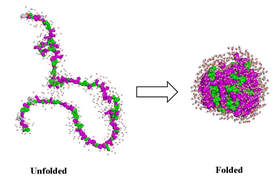
What are antibodies?
Antibodies are little Y-shaped proteins made by the body to Identify and neutralize antigens(bad guys). Your bodies B-cells helps mass product them using somatic hypermutation.



Somatic Hypermutation: B-cells make antibodies. For an antibody to bond(and destroy) an antigen(bad virus / thing) it needs to have the right shape. The B-cell brute forces the right shape, by creating many different possible combos of the anti-body until it gets the best one. Each Different antibody is measured by affinity maturation (Essentially, how well it bonds). The higher the affinity maturation, the more of that specific antibody it will make.

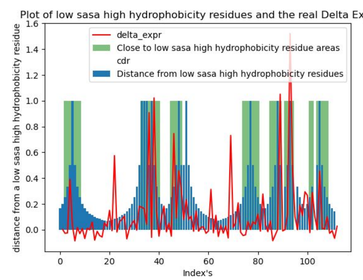
What is a deep mutational scan?
A deep mutational scan is where scientists try to simulate the process for somatic hypermutation within a lba setting, typically this process is for when they are looking for new proteins or new combinations. This process also helps us understand how and what contributes to the folding of the protein.

Protein Folding?
Upon their creation proteins will “fold”, this is where they will curl up into a final “energetically stable shape”. Most of the time this works, and the protein folds, however sometimes the protein fails to fold properly, causing it to unravel.



Next Steps

1. **Testing on more data:** Testing this method on more data, and looking to see if it works for more and different antibodies.
2. **Creation of Python library:** creation of a python library to make the pipeline for predicting antibody stability values much easier and more successful than they are now.
3. **Exploration of expr predictive residues**: residues that have a high correlation with protein misfolding.



References

1. Durham E, Dorr B, Woetzel N, Staritzbichler R, Meiler J. Solvent accessible surface area approximations for rapid and accurate protein structure prediction. J Mol Model. 2009 Sep;15(9):1093-108. doi: 10.1007/s00894-009-0454-9. Epub 2009 Feb 21. PMID: 19234730; PMCID: PMC2712621.
2. Trejos M, Aristizabal Y, Aragón-Muriel A, Oñate-Garzón J, Liscano Y. Characterization and Classification In Silico of Peptides with Dual Activity (Antimicrobial and Wound Healing). Int J Mol Sci. 2023 Aug 23;24(17):13091. doi: 10.3390/ijms241713091. PMID: 37685896; PMCID: PMC10487549.
3. McBride, John M., et al. "AlphaFold2 Can Predict Single-Mutation Effects." *Physical Review Letters*, vol. 131, no. 21, 2023, p. 218401. https://doi.org/10.1103/PhysRevLett.131.218401
4. DeWitt, W. S., Vora, A. A., Araki, T., Galloway, J. G., Alkutkar, T., Bortolatto, J., ... & Victora, G. D. (2025). Replaying germinal center evolution on a quantified affinity landscape. *bioRxiv*, 2025-06.