Abstract

Over four years have passed now since the groundbreaking debut of AlphaFold 2, “solving” the problem of protein structure determination. Since that time, the algorithm has ben improved and iterated upon, with future releases of the program improving predictive capabilities and expanding functionality. However, there are still numerous questions about human proteins that remain unanswered - many of which relate to the numerous protein-protein interactions ongoing within the human body. While AlphaFold works relatively well on single protein structure prediction, its capacity to predict multiple proteins in complex is lesser. Recent updates to AlphaFold - in particular the 2024 release of AlphaFold 3 - have improved this capacity, but it is currently unknown how well this program is able to predict the effects of point mutations on structures.

The upcoming study will seek to examine the capacity for various versions of AlphaFold to estimate the impact of single point mutations on the binding affinity of various protein-protein interactions. To this end, the SKEMPI 2.0 database will be used to provide empirical data to which the estimates of AlphaFold can be compared. Comparing the predicted ddG of AlphaFold to the empirically determined ddG of SKEMPI’s data will allow for an evaluation of the quality of AlphaFold’s predictions. The primary targets of this analysis are AlphaFold 2 and AlphaFold Multimer, with AlphaFold 3 also being evaluated if possible.

Protein-Protein Interactions

Protein-Protein interactions (or PPI) form the foundation of a significant number of biological processes. This is a blanket term used to cover a wide variety of interactions between a large number of proteins, and it is impractical to cover their entire breadth. Critically, though, PPIs are dependent on intermolecular forces, such as hydrophobic effects, hydrogen bonding, and dipole-dipole interactions. These forces can improve the proteins’ structure, causing the newly formed complex to be more stable than the two proteins in isolation. This improvement can be represented via thermodynamic principles as a change in Gibbs’ Free Energy, or ΔG. When ΔG is negative, this indicates that the resultant structure is thermodynamically favoured, and can occur spontaneously.

When a mutation occurs in a given protein, this can sometimes have an impact on the binding affinity of a PPI, typically through interfering with/creating new intermolecular forces. This changes the ΔG, and that measured change is referred to as ΔΔG (or ddG). This metric is one that is important to understanding the impact of a mutation on a given protein-protein interaction, and is one that many machine learning approaches fail to predict consistently or accurately. (Tsishyn et al., 2024)

AlphaFold

AlphaFold is a relatively new tool in the world of computational biology that has revolutionized the field of protein structure prediction using machine learning algorithms. Historically, determining the three-dimensional structure of a protein was reliant on direct visualisation techniques, such as x-ray diffraction. These techniques are able to empirically determine the structure of a protein, but are time-consuming and expensive. As such, using machine learning methods to approximate protein structure was a theoretically useful prospect for saving time, materials, and money.

The initial release of AlphaFold achieved success in 2018, but the program was not as well recognized until the debut of AlphaFold 2 at the CASP14 competition in 2020, wherein it achieved results that drastically outperformed its competitors. This catapulted the program to the forefront of protein structure prediction, where it remains today. (Jumper et al., 2021)

AlphaFold 2 is generally very accurate in its predictions, exhibiting a median RMSD error of just under one Angstrom for the protein backbone - less than the width of a Carbon atom. (Jumper et al.) This prediction quality, however, is only relevant for the prediction of a single protein. When predicting multiple proteins in a complex, accuracy drops significantly. Even when using an updated version tailored for use in multimeric structures, the success rate for antibody-antigen complexes has been found to be as low as 11 percent. (Yin et al., 2022) This thus leads to the study's hypothesis, an expectation that AlphaFold's performance in predicting the impact of PPI will be relatively poor.

The 2024 release of AlphaFold 3, however, promises significantly improved prediction quality for protein complexes, even specifically mentioning antibody-antigen complexes in its abstracts. The improvement in prediction accuracy for protein-protein complexes is relatively

small across all proteins, but immense for specifically antibody-antigen complexes, increasing from approximately 30% to approximately 65%. (Abramson et al., 2024)

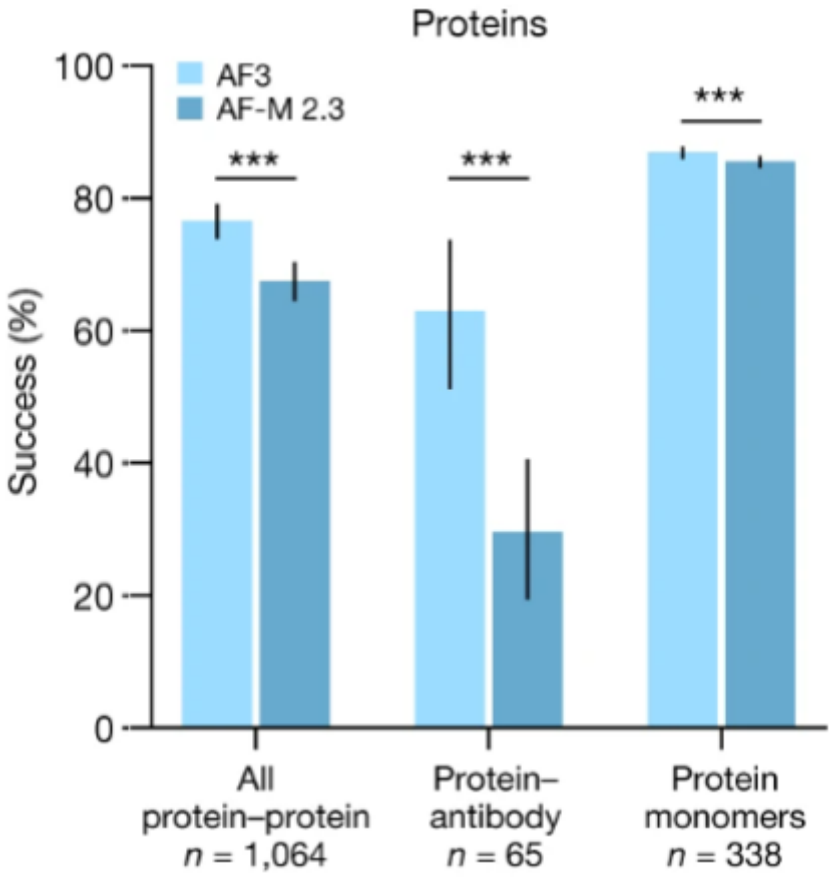

Figure 1: A bar graph displaying the improvement in prediction quality from AlphaFold Multimer to AlphaFold 3. The drastic improvement in antibody-antigen complexes can be observed adjacent to the impact across all protein-protein complexes. (Abramson et al.)

While this is undoubtedly a significant improvement, two essential questions remain unanswered: whether or not this new accuracy rate is sufficient to provide reliable results, and whether AlphaFold 3 is able to reliably predict binding affinities in mutant complexes.

AlphaFold uses two main metrics to estimate error: PAE and ipTM. PAE (predicted aligned error) is produced in a two-dimensional matrix that compares the position of a given residue (X axis position) assuming alignment on another residue (Y axis position). These regions are generally then colour-coded to illustrate regions of high and low confidence by positional

difference. ipTM (interface predicted template modeling) score is a measurement of the accuracy of the relative positions of the subunits in the complex. As a general rule, higher scores (approaching 1) are better, with 0.8 being a general threshold of high quality. (EMBL-EBI Training)

SKEMPI

SKEMPI v2.0 is a database that contains information on a wide variety of protein-protein interactions. For each interaction included, the database also contains information on a number of point mutations and their impact on the binding affinity of the interaction. The database was manually curated by its creators, featuring both data from its prior version (comprising approximately 40% of the data) in addition to new data gathered via manual review from a variety of literature.

SKEMPI divides its samples into four main categories: antibody-antigen interactions, protein-inhibitor interactions, pMHC/TCR interactions, and all others. The main three categories combined comprise just under half of the data set - the other ~50% being largely uncategorized. Of these samples, around 75% are single point mutations - the easiest to predict the outcome owing to their reduced complexity. These are the samples that are to be used within the study.

A major shortcoming of SKEMPI is that of the single point mutations recorded, more than 50% are mutations to Alanine. This massive overrepresentation is one of the biggest weaknesses of SKEMPI as a dataset, as this can lead to an inability to generalise results to wider arrays of mutants. Other overrepresented mutations include those from one aromatic residue to another, and charge swap mutations in which a positively charged residue is replaced with a negatively charged residue or vice versa. Comparatively, these biases are much less significant than the Alanine bias. These biases are to be taken into account in the course of the experimental work. (Jankauskaitė et al., 2019)

Workflow

The project workflow primarily involves the use of AlphaFold to predict the outcome of various mutations on protein-protein interactions. To begin, a set of PDB files were chosen from

the SKEMPI database. This subset was chosen through several requirements: first, a diverse array of protein complex structures were chosen, incorporating structures that varied in number of substrates, secondary structure, and interaction type. Second, interaction types were varied to ensure that all three major subcategories had some significant amount of representation. Third, a diverse array of crystal structure qualities were selected in order to determine the impact of crystal structure quality on AlphaFold performance. Finally, structures were prioritized that had relatively low Alanine bias - although this was largely unavoidable. These PDBs represent a total of 1043 mutations in SKEMPI, before removing multiple mutations. Each of these mutations has a recorded change in binding affinity, which will be compared to the predicted change by AlphaFold.

PDB	Rfree	Mutation Count	Qualities
1JTG	0.205	275	Mostly Alanine, Unique Structure
5E9D	0.238	16	Diverse Mutations, TPC
3SZK	0.277	21	All Alanine
3BN9	0.264	35	Majority Alanine, AB/AG and Pr/PI
2J0T	0.267	23	Diverse Mutations, Unique Structure
1KTZ	0.216	27	Mostly Alanine, Unique Structure
3SGB	0.181	295	Diverse Mutations, Pr/PI, Estimated Rfree
3MZG	0.228	91	Mostly Alanine, Unique Structure
1DAN	0.218	130	Mostly Alanine, Unique Structure
1VFB	0.241	72	Majority Alanine, AB/AG
1OGA	0.221	58	Mostly Alanine, TPC

Table 1: A list of the 11 chosen protein-protein complexes to be used in the experiment. Includes the Rfree value of their crystal structure (or estimated Rfree, if none was provided) and the number of mutations for that complex in the SKEMPI database. Furthermore, notes are included on the Alanine bias found in the mutations, uniqueness of the protein complex structure, and whether the complex falls into one of the three major categories.

To calculate this predicted change, AlphaFold 2 will be used to predict the binding affinity between the two structures for both the mutant and wild-type sequences. The difference

between these two affinities, known as ddG, can be compared to that same difference as recorded in SKEMPI to determine the relative accuracy of AlphaFold. This ddG is the change in the ΔG outlined in the section on protein-protein interactions. This benchmarking process will then also be performed using AlphaFold Multimer and AlphaFold 3, assuming time allows.

Finally, as an additional quality assurance method, the error metrics of PAE and ipTM, produced by AlphaFold, will be used as a measure of evaluating structure quality. Below is an example PAE calculated by ColabFold to illustrate how it can be used to determine structure quality.

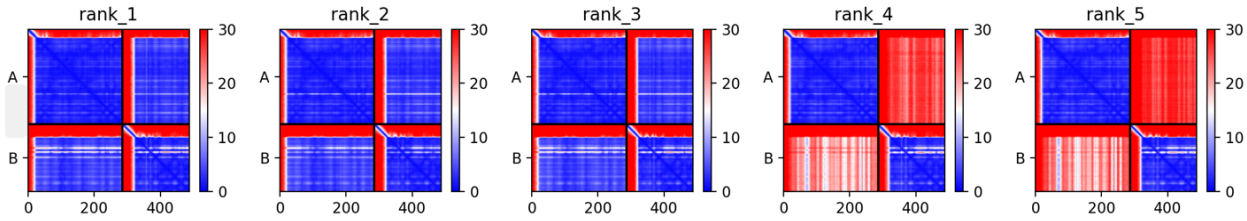

Figure 2: A plot of the PAE from the five predicted structures of PDB 1JTG by AlphaFold 2, ranked by quality. The red areas represent high distances depending on where the structure is aligned, indicating a risk of inaccuracy. It can be observed that fewer red regions are present at better-ranked models, but areas near the beginning of each chain remain uncertain.

In conclusion, the project aims to continue along this course, evaluating the AlphaFold programs and their capacity to predict ddG. It is hypothesized that this capacity will be relatively poor for AlphaFold 2, and improve somewhat as the models advance. However, given the inconsistent rate of protein-protein complex prediction, even for more recent versions of AlphaFold, it is not expected to see consistent accuracy. Regardless, this benchmarking will provide further insights into the limitations and strengths of AlphaFold, providing new avenues of innovation in protein structure prediction.

Citations

Abramson, J., Adler, J., Dunger, J. *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* 630, 493–500 (2024). https://doi.org/10.1038/s41586-024-07487-w

EMBL-EMI Training. AlphaFold: A practical guide. DOI:10.6019/TOL.AlphaFold-w.2024.00001.1

Jumper, J., Evans, R., Pritzel, A. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589 (2021). https://doi.org/10.1038/s41586-021-03819-2

Justina Jankauskaitė, Brian Jiménez-García, Justas Dapkūnas, Juan Fernández-Recio, Iain H Moal, SKEMPI 2.0: an updated benchmark of changes in protein–protein binding energy, kinetics and thermodynamics upon mutation, *Bioinformatics*, Volume 35, Issue 3, February 2019, Pages 462–469, https://doi.org/10.1093/bioinformatics/bty635

Matsvei Tsishyn, Fabrizio Pucci, Marianne Rooman, Quantification of biases in predictions of protein–protein binding affinity changes upon mutations, *Briefings in Bioinformatics*, Volume 25, Issue 1, January 2024, bbad491, https://doi.org/10.1093/bib/bbad491

Yin R, Feng BY, Varshney A, Pierce BG. Benchmarking AlphaFold for protein complex modeling reveals accuracy determinants. *Protein Science*. 2022; 31(8):e4379. https://doi.org/10.1002/pro.4379