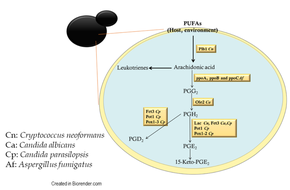
Using AlphaFold2 to Identify Novel Drug Targets Against *Cryptococcus*

Francisco Lugo Gonzales[1], Natalia Castro Lopez[2], Floyd Wormley Jr[2]
Department of Science and Engineering, Texas Christian University, Fort Worth Tx, USA
Department of Biology, Texas Christian University, Fort Worth Tx, USA

Abstract

Cryptococcus is an invasive fungus that causes cryptococcosis, an infection that highly affects immunocompromised people. There are currently a limited number of antifungals available to treat *Cryptococcus*, and with the increase in antimicrobial resistance, we need different alternatives to treat fungal infections. Our lab has identified proteins involved in the synthesis of eicosanoids, which are lipid signaling molecules involved in regulating the immune response. Moreover, fungi can produce eicosanoids using different enzymes that humans do, opening a line to identify new drug targets using these pathways. Previously, our lab had identified genes upregulated in the presence of the eicosanoid's precursor, arachidonic acid. Our goal is to use bioinformatics to predict and characterize the protein structure, using AlphaFold2, a machine learning application based on a deep neural network, and using this tool, identify small molecules that will bind to the proteins and help make drug design more efficient.

Introduction



Cn: *Cryptococcus neoformans*
Ca: *Candida albicans*
Cp: *Candida parasilopsis*
Af: *Aspergillus fumigatus*

Representation of the known fungal enzyme used for eicosanoid biosynthesis.

List of 20 genes upregulated in the presence of AA

Gene ID	Gene Name	Fold change
CNAG_03713	Efflux protein EncT	107.4742089
CNAG_00474	Hypothetical protein	65.8350302
CNAG_02146	Hypothetical protein	52.62822643
CNAG_00869	ATP-binding cassette transporter (AFR2)	47.22425607
CNAG_05741	Hypothetical protein	44.01660919
CNAG_02933	Quinone oxidoreductase	39.11930277
CNAG_05842	Cytochrome P450	38.88766928
CNAG_00904	Aflatoxin efflux pump AFLT	34.44989338
CNAG_02347	Hypothetical protein	30.15309438
CNAG_00605	Cytoplasmic protein	29.0276407
CNAG_03519	Cytoplasmic protein	28.61651601
CNAG_02087	Acyl-CoA-dependent ceramide synthase	28.41646
CNAG_04454	Hypothetical protein	24.90532224
CNAG_03754	Short-chain dehydrogenase/reductase SDR	21.8503731
CNAG_02839	Hypothetical protein	21.64550417
CNAG_07799	ABC transporter	19.19830023
CNAG_06653	Hypothetical protein	17.32678866
CNAG_03999	Hypothetical protein	16.83746159
CNAG_05544	Hypothetical protein	16.35662521
CNAG_03061	Multiple drug resistance protein	16.16406231

AlphaFold

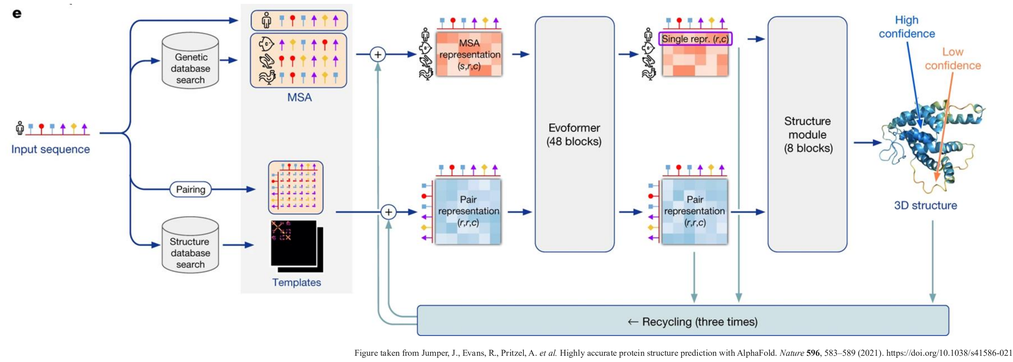
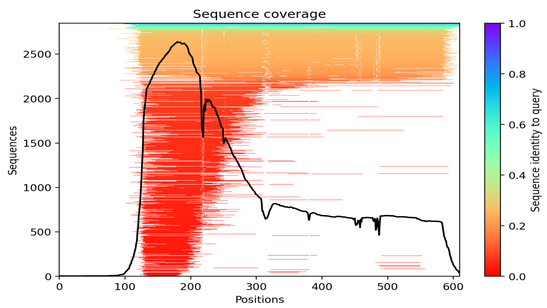


Figure taken from Jumper, J., Evans, R., Pritzel, A. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021). https://doi.org/10.1038/s41586-021-03819-2

Diagram of the model architecture of AlphaFold 2 system. Using input of amino acid sequence to query different protein databases, to create a Multiple Sequence Alignment (MSA), this allow the AF2 to identify similar protein to construct an initial representation of the structure. Then, the MSA and template are passed thought a transformer, which will look at the MSA and the pair interactions and create different and better iterations of the model. In the last part, of this model shows the structure model created using the MSA representation and the pair representation.

CNAG_03713

A)



B)

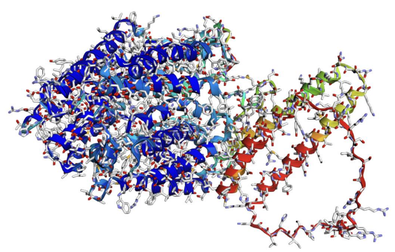
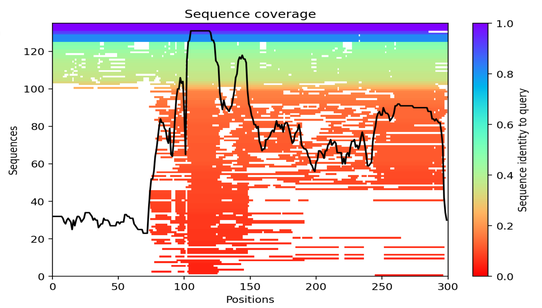


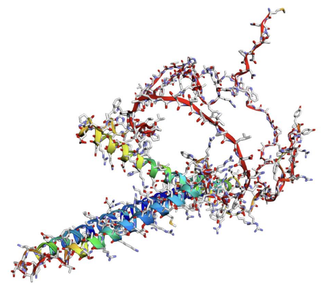
Figure 1. Predicted structure of CNAG_03713. Protein length 610 amino acids.
a The visual representation of the MSA is used to predict the protein structure based on existing sequences and structures closely related to the protein input. **b** is the predicted structure of CNAG_03713 found using AlphaFold2. pIDDT is predicted local distance difference test and shows how confident the prediction is. Anything >90 is highly accurate.

CNAG_00474

A)



B)



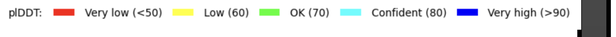


Figure 2. Predicted structure of CNAG_00474. Protein length 300 amino acids.
a The visual representation of the MSA is used to predict the protein structure based on existing sequences and structures closely related to the protein input. **b** is the predicted structure of CNAG_00474 found using AlphaFold2.

Findings and Future Directions

- The more sequences that a structure has, the more accurate the predictions were after putting the sequences through AlphaFold2. With the predictions, we were able to identify the molecules that could help to bind the proteins due to the accuracy of the predicted protein structures.
- We found that if a protein has more sequences, through AlphaFold2, it can identify more molecules compared to a protein that does not have as many structures as the program has more amino acids to work with.
- Next, using the structure we generated in AlphaFold2, we will use Prankweb to identify potential binding sites. Then we will use Autodock Vina or SwissDock tools to predict affinity to the binding site, first for known antifungal, then we will use small molecules, using this approach we can evaluate small molecules as therapy against *Cryptococcus in silico* before we perform studies *in vitro* and *in vivo*.

Acknowledgment

- I would like to thank Dr. Castro and Dr. Wormley for their assistance and support while conducting the research.