

#### Mapping the Zymomonas mobilis interactome

Sameer D'Costa Scientist, Computational Biology Group Throughput Computing 2025 (June 5<sup>th</sup>)





## Introduction

Who, What, Why

#### Introduction

- Scientist at Great Lakes Bioenergy Research Center (GLBRC at WEI)
  - Computational Biology Group (AI/ML)
- GLBRC mission
  - Creating biofuels and bioproducts
    - from non-food plant biomass
    - economically viable
- Zymomonas mobilis (Bacteria)
  - Really good at turning sugar into alcohol
  - Engineer for industrial use
  - What do the genes do?







#### Homo sapiens - Oxidative phosphorylation



#### https://www.kegg.jp/pathway/hsa00190





#### Zymomonas mobilis Electron Transport Chain







### What is the next complex in this chain?

- This is a really hard problem to solve computationally
  - ~2k genes encoding proteins
  - $\circ$  What do they all do?
  - Which proteins form complexes with other proteins?
- We have a few things in our favor
  - Predict which genes are localized to membranes or periplasm ~580
    - SignalP 6.0 and DeepTMHMM (CPU or GPU)
  - Predict individual protein structures and complexes
    - AlphaFold 3 (Google Deepmind) (data prep on CPU but inference on GPU)





#### AF3 predicts all Z. mobilis protein structures

- Predicting all ~2k single protein structures took a week at CHTC/GLBRC
  - Data pipeline (5 days on CPUs) searches 650GB databases for each protein sequence to create alignments
  - Inference pipeline (1.5 days on GPUs) select GPU minimum memory based on sequence size
  - Apptainer container ~3GB







#### AF3 can predict complexes, if precisely specified

GREAT LAKES BIOENERGY RESEARCH CENTER





### The problem is still too hard computationally

- No idea what the complexes are
- PPI (protein-protein interactions) subnetworks should allow us to brute-force predict complexes with AF3
- Expect ~2k actual PPIs among ~2M possible
- It will be hard to find all of them
- Hail Mary!
- CytC will form a complex with a new complex
- If not,
  - Run experiments guided by computational results
  - Save novel complexes for future work at GLBRC







# **Computational Plan**

## Deep learning to compute complexes in Z. mobilis





SCIENCE • 11 Nov 2021 • Vol 374, Issue 6573 • DOI: 10.1126/science.abm4805

#### Protein interaction networks revealed by proteome coevolution Science 365, 185-189 (2019)

Qian Cong<sup>1,2</sup>, Ivan Anishchenko<sup>1,2</sup>, Sergey Ovchinnikov<sup>3</sup>, David Baker<sup>1,2,4\*</sup>

- Bacteria have more genomic data available than eukaryotes
- Complexes are simpler
- Interacting partners are closer in the genome





#### Workflow







### Acknowledgements

- Jacob Eckmann
- Kevin Myers
- Mike Place

- CHTC facilitation team, HTCondor
  - Christina Koch
  - Andrew Owen
- AlphaFold 3 pipeline is at

https://github.com/sameerd/alphafold3/tree/hpc/hpc#running-alphaf old3-on-chtc

• Thanks to Patricia Tran for documentation and additional scripts





