CDT-BAI project submission 2024

1. Primary Supervisor Name
Edward Wallace

2. E-mail Address *
Edward.Wallace@ed.ac.uk

3. University and/or Research Web Site Address *
https://ewallace.github.io

4. College Affiliation
CSE

5. School, Department, or Institute Affiliation
Institute for Cell Biology, School of Biological Sciences

6. Any Additional Relevant Affiliations
Centre for Engineering Biology

7. Primary Supervisor Name
Christopher W. Wood

8. E-mail Address *
Chris.Wood@ed.ac.uk

9. University and/or Research Web Site Address *
https://www.wellswoodresearchgroup.com

10. College Affiliation
CSE

11. School, Department, or Institute Affiliation
Institute of Quantitative Biology, Biochemistry and Biotechnology. School of Biological Sciences

12. Any Additional Relevant Affiliations
Centre for Engineering Biology

18. If an external partner is not part of the project now please identify possible partners if you have some in mind and/or any strategy you have for developing one

We have not yet engaged an external partner. EW will discuss with Gilead Sciences, who develop antifungal drugs for human health. Another possible external partner is Dr Tihana Bihacic at St. George's University of London, a clinical professor who works on fungal infectious disease. The role of an external partner would be to advise on how to make the protein analysis more relevant for drug development and clinical practice.

19. Primary Research Theme Alignment
AI for Biomedical & Health Informatics

20. Other Research Theme Alignment
AI for Biomedical Engineering

21. Project Title
Systematic identification of conserved protein functions in human fungal pathogens

22. Project Summary Description (max. 400 words)

Fungal pathogens of humans are a major global threat to human health, as highlighted by the 2022 World Health Organisation fungal priority pathogens list. The project aims to computationally identify potential drug targets in fungal pathogens. One major challenge in antifungal drug development is that fungi are relatively closely related to humans, making it difficult to target essential fungal proteins without also targeting their human homologs. New data and AI approaches make it possible to address this systematically, by combining fungal genome data and functional genetic screens of essential proteins with protein structural predictions and structural similarity analysis. The project aims first, to assemble a list of the proteins that are present in WHO critical priority fungal pathogens and absent in humans, using both sequence and structural data. This list will be compared to published data on the functional importance of proteins for fungal virulence, that would inform which conserved protein families to prioritise for follow-up. Second, the project aims to look in detail at AI-predicted molecular surfaces to find features such as pockets, active sites, and binding sites, that are conserved within these fungal pathogens and not in humans. This molecular surface analysis is more fine-grained and could identify candidate drug target surfaces that are distinct in fungi from in humans. Overall, the project aims to leverage the power of protein structural prediction and big data to prioritise candidate antifungal drug targets at a large scale.

23. Project Description (max c.1000 words, c.2-3 pages of A4 at 12pt) *

Abstract

Fungal pathogens of humans are a major global threat to human health, as highlighted by the 2022 World Health Organisation fungal priority pathogens list. The project aims to computationally identify potential drug targets in fungal pathogens, by combining fungal genome data and functional genetic screens of essential proteins with AI-powered protein structural predictions and structural similarity searches. The overall goal is to generate prioritised lists of proteins, and of functionally important protein surfaces, that are conserved in fungal pathogens but absent in humans. These lists can be then be used to aid future functional investigations and drug development efforts.

Introduction

Fungal pathogens cause millions of infections per year. The 2022 World Health Organisation fungal priority pathogens list identified 19 fungi that represent the greatest threat to public health, including 4 critical priority pathogens (World Health Organization 2022). The threat of pathogens to human health is compounded by the small number of antifungal drug classes and the increasing emergence of antifungal drug resistance (Fisher et al. 2020). One major challenge in antifungal drug development is that fungi are closely related to humans within the tree of eukaryotic life, making it difficult to target essential fungal proteins without also targeting their human homologs.

The project aims to computationally identify potential drug targets in fungal pathogens. New data and AI approaches make it possible to address this systematically. Fungal genome data and annotations give confident lists of proteins and allow calculation of sequence similarity (Li et al. 2021). Functional genetic screens of proteins in fungal pathogens indicates which proteins are important for fungal survival and virulence (Goranov and Madhani 2014). Structural prediction methods such as AlphaFold2 create detailed and testable hypotheses for protein structure and function (Varadi et al. 2022), and allow structural similarity searches that can be more accurate than sequence-based approaches (van Kempen et al. 2024).

First, the project aims to assemble a list of the proteins that are present in WHO critical priority fungal pathogens and absent in humans, using both sequence and structural methods. This list will be compared to published data on the functional importance of proteins for fungal virulence, that would inform which conserved protein families to prioritise for follow-up.

Second, the project aims to look in detail at AI-predicted molecular surfaces to find features such as pockets, active sites, and binding sites, that are conserved within these fungal pathogens and not in humans. This molecular surface analysis is more fine-grained and could identify candidate drug target surfaces that are distinct in fungi from in humans.

Overall, the project aims to leverage the power of protein structural prediction and big data to prioritise candidate antifungal drug targets at a large scale.

Research Challenge

The research challenge is to identify proteins, and protein surface and structural features, that are conserved in human fungal pathogens but absent in humans. Furthermore, to prioritize these based on some kind of confidence score as well as according to published functional data. This challenge will require critical engagement with a variety of data types (sequence data, structural predictions, functional genomics screens) and a variety of machine learning and AI methods (sequence homology, structural predictions, structural similarity), in order to evaluate the quality of predictions and obtain reliable results.

Data & Methodology

A range of structural and evolutionary bioinformatics approaches applied to public data, including:
- Fungal annotated genome and functional genomics data, from FungiDB (Basenko et al. 2018). Evolutionary analysis of sequences, gene order, and phylogeny.
- Protein structure prediction with AlphaFold2 (Jumper et al. 2021), AlphaFold database (Varadi et al. 2022), and OpenFold (Ahdritz et al. 2024).
- Detailed structural analysis using structural analysis software from the Wells Wood group.

RRI/Ethical Considerations

This project will use publicly available data on fungal genomes and proteomes, meaning that there will be no privacy or medical concerns. We don't envisage any ethical concerns arising from the work.

We do not expect this project to develop any protectable IP, as a prioritized list of potential drug targets without extensive functional validation would not be valuable.

Expected Outcome & Impact

The expected outcome is a prioritized list of proteins that are conserved and functionally important in human fungal pathogens. These computational hypotheses can be used to design future experiments for function (investigating the importance of conserved yet neglected proteins), conservation (by cross-complementation of proteins from one fungus to another), and druggability (by screening against compound libraries). This will certainly lead to insight into fungal biology and fungal evolution, and could lead to long-term impact in accelerating drug design against fungal pathogens.

References

Ahdritz, Gustaf, Nazim Bouatta, Christina Floristean, Sachin Kadyan, Qinghui Xia, William Gerecke, Timothy J. O'Donnell, et al. 2024. "OpenFold: Retraining AlphaFold2 Yields New Insights into Its Learning Mechanisms and Capacity for Generalization." *Nature Methods*, May. https://doi.org/10.1038/s41592-024-02272-z.

Basenko, Evelina Y., Jane A. Pulman, Achchuthan Shanmugasundram, Omar S. Harb, Kathryn Crouch, David Starns, Susanne Warrenfeltz, et al. 2018. "FungiDB: An Integrated Bioinformatic Resource for Fungi and Oomycetes." *Journal of Fungi (Basel, Switzerland)* 4 (1). https://doi.org/10.3390/jof4010039.

Fisher, Matthew C., Sarah J. Gurr, Christina A. Cuomo, David S. Blehert, Hailing Jin, Eva H. Stukenbrock, Jason E. Stajich, et al. 2020. "Threats Posed by the Fungal Kingdom to Humans, Wildlife, and Agriculture." *MBio* 11 (3). https://doi.org/10.1128/mBio.00449-20.

Goranov, Alexi I., and Hiten D. Madhani. 2014. "Functional Profiling of Human Fungal Pathogen Genomes." *Cold Spring Harbor Perspectives in Medicine* 5 (3): a019596.

Jumper, John, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, et al. 2021. "Highly Accurate Protein Structure Prediction with AlphaFold." *Nature* 596 (7873): 583–89.

Kempen, Michel van, Stephanie S. Kim, Charlotte Tumescheit, Milot Mirdita, Jeongjae Lee, Cameron L. M. Gilchrist, Johannes Söding, and Martin Steinegger. 2024. "Fast and Accurate Protein Structure Search with Foldseek." *Nature Biotechnology* 42 (2): 243–46.

Li, Yuanning, Jacob L. Steenwyk, Ying Chang, Yan Wang, Timothy Y. James, Jason E. Stajich, Joseph W. Spatafora, et al. 2021. "A Genome-Scale Phylogeny of the Kingdom Fungi." *Current Biology: CB* 31 (8): 1653-1665.e5.

Schrödinger, LLC. 2015. "The PyMOL Molecular Graphics System, Version 1.8."

Varadi, Mihaly, Stephen Anyango, Mandar Deshpande, Sreenath Nair, Cindy Natassia, Galabina Yordanova, David Yuan, et al. 2022. "AlphaFold Protein Structure Database: Massively Expanding the Structural Coverage of Protein-Sequence Space with High-Accuracy Models." *Nucleic Acids Research* 50 (D1): D439–44.

World Health Organization. 2022. *WHO Fungal Priority Pathogens List to Guide Research, Development and Public Health Action*. Geneva: World Health Organization.