CONSIDER FOR TALK

2025 SEA Faculty Meeting Abstract

Baylor University

Waco TX

Corresponding Faculty Member: Jonathan Lawson (Jonathan\_Lawson@baylor.edu)



Jonathan N Lawson

(fold)Seeking to Make Sense of Genome-Wide AlphaFold Data

Jonathan N Lawson

Data from AlphaFold and Foldseek can enhance phage annotations, but a systematic rubric is needed to guide students through this process. We collected genome-scale AlphaFold results to explore how these tools can facilitate more accurate annotations and provide a baseline dataset to better define the role of structure prediction in bioinformatic analysis of phage genomes. All predicted ORFs of two Arthrobacter phages were input into Alphafold as monomers, dimers, trimers, and as monomers with various ligands, including ATP, NAD, ssRNA, dsDNA, and the adjacent gene. Students collected results in a common folder and reported PTM and iPTM in a shared spreadsheet. Over 80% of the predicted ORFs had folded monomer PTM scores > 0.5. Patterns in PTM/iPTM values varied even with structural proteins known to polymerize. Interestingly, folding with ATP universally increased PTM scores with 28% having an iPTM score >0.8. Folding with nucleic acids yielded likely interactions based on visual inspection, but PTM/iPTM scores were often below thresholds due to the influence of sequence and length of the nucleic acids on these scores. These results provided a foundation for independent projects that explored a variety of questions. These included exploring Foldseek differences using the different .cif output files, comparing AlphaFold predicted structures with phams, asking if Foldseek results varied with common ligands, exploring possible interactions with hypothetical phage proteins and host defense systems. The genome-wide application of AlphaFold and Foldseek can enhance annotations; however, independent projects have highlighted numerous considerations in evaluating these data.