St. Jude Children's Research Hospital

St. Jude BioHackathon

Title
Toolbox for convenient manipulation of AlphaFold output

Category
Processing Pipelines And Methods

Challenge
AlphaFold produces up to 5 predicted structures that contain slightly different conformations of the same modeled sequence. It then names files by appending an incremental value (from 0 to 5) according to the average pLDDT confidence score. However, in some cases, there is a specific region or domain of interest for which we will want to maximize this confidence score. AlphaFold also provides aligned error values for all residue-residue distances. Working with these values, however, is not straightforward.

Our objective with this challenge is to develop a library containing “convenience functions” for working with AlphaFold-generated models. The goal is to write several independent scripts that can output specific functions using AlphaFold models as input. Depending on the interest and the number of desired functionalities, it’s also tempting to build an object-oriented API to store and interact with AlphaFold output in a programmatic way.

Benefit
This would help us make the most of AlphaFold’s structure predictions by refining them further and extending the use-cases by allowing us to make better decsions.

Helpful Tools, Packages, or Software
AlphaFold 2, AlphaFold DB, Standard Python packages + numpy.

Test Data
PDB structures and AlphaFold models.