Review of AlphaFold

The blog discusses the intricacies of AlphaFold3, aiming to provide a detailed, user-friendly guide of the complex model. AlphaFold is a revolutionary model that can predict 3D protein structures from a given sequence of amino acids. AlphaFold3 expands its capabilities to predict structures and interactions of a broader range of biological modules, including DNA, RNA and ligands. The impact of AlphaFold has been revolutionary in the field of biology and drug discovery and has dramatically accelerated protein structure prediction.

The core of AlphaFold 3 builds upon the successful architecture of AlphaFold 2, with significant enhancements to handle a broader range of molecular types and interactions. The blog discusses 3 main sections of AlphaFold3 – (i) Input Preparation, (ii) Representation Learning, and (iii) Structure Prediction. During input preparation, the user provides sequences of some molecules to predict structures and these are embedded into numerical tensors. Furthermore, the model retrieves a collection of other molecules from a database that is presumed to have a similar structures (multiple sequence alignment – MSA and templates) to the user-provided molecules. MSA and templates are included so that any protein with known structures can inform the model for predicting the structure of a new protein. Atom-level representations are created for the molecules which are then aggregated to form token-level representations. In the representation learning phase, improved representation of the token-level tensors are learned by simultaneously improving the MSA and pair representations. It does a series of operations independently on these two representations then also enables cross-talk between them.

Once the representations are learned, AlphaFold uses Diffusion model to predict the 3d structure of the protein. It introduces Invariant Point Attention, a transformer-based neural network architecture designed specifically for three-dimensional structures.

Strengths

1. High accuracy in predicting molecular structures and interactions in the protein
2. Significant improvement in drug-like interaction predictions and helps in accelerating drug discovery
3. It can model a wide range of biomolecules, including amino acids, DNA, RNA, ligands, etc.

Weakness

1. There are potentials for hallucinations in structure predictions. A detailed study of how this can be improved needs to be done
2. How to incorporate the medium in which the protein structure is being folded. Different mediums will have different ways in which the molecules combine.

Improvements:

1. Enhancing the model's ability to predict dynamic molecular processes
2. Detect and tackle hallucinations