

APPLICATIONS OF THE EMERGING DEEP LEARNING TECHNIQUES FOR PROTEIN 3D STRUCTURE PREDICTION AND GENERATIVE DESIGN

Chugunov A.

During the last decade, deep learning algorithms have developed to the very impressive milestones in molecular and structural biology. Previously, computational techniques provided community with wide-spread genomic, proteomic and protein structure databases; and reliable methods for structure modeling, from comparative structure prediction to molecular dynamics simulations and docking. Although, one of the central problems of molecular biophysics – protein folding – remained largely unattainable.

Availability of genomic and (most importantly) metagenomic data from the DNA sequencing projects accelerated application of artificial neural networks to these immense datasets to uncover “deep” evolutionary information from these distantly related sequences – from general “appearance” of protein chains to hidden structural information suggesting protein 3D organization from multiple sequence alignments or even sole sequences. One of the most famous algorithms for neural prediction of protein structure – AlphaFold 2 – stands on the shoulders of previous metagenomic studies and advanced neural network research, introducing “transformer” architecture and “big language models” into bioinformatics.

Moreover, it appears that apparently different from structure prediction task – protein design, representing another biophysical enigma, – is just another type of prediction: prediction of protein sequence that would favorably fold into desired structure and would have wanted function. Variety of network architectures, among which is diffusion generative models, may assist in creation of protein sequences that never have existed and exhibit no structural or sequence similarity to any known protein. Experimental validation shows that at least in some cases these proteins in fact fold into a predicted structure, which seems to be a robust basis for future protein design efforts.

In this talk, I’ll give a brief overview of based on deep learning protein structure prediction methods and protein design efforts, and will try to answer a question, is the protein folding problem solved, finally?