Efficient and accurate prediction of protein structures and interactions using artificial intelligence

Minkyung Baek

Department of Biochemistry, University of Washington, Seattle, WA98195, USA

An accurate protein structure prediction from its amino acid sequence is a longstanding challenge in computational biology. Considerable progress has recently been made by leveraging genetic information through deep learning-based methods. In this talk, I'll present a three-track attention-based neural network named RoseTTAFold. In this model, information at the 1D sequence level, the 2D distance map level, and the 3D coordinate level are successively transformed and integrated to generate accurate protein structures. The three-track network produces structure predictions with accuracies approaching those of AlphaFold in CASP14, enables a rapid solution of challenging X-ray crystallography and cryo-EM structure modeling problems, and provides insights into the functions of proteins of currently unknown structure. The network also enables rapid generation of accurate protein-protein complex models from sequence information alone, reducing computational cost of traditional approaches which require modeling of individual subunits followed by docking. We further use a combination of RoseTTAFold and AlphaFold to screen through paired multiple sequence alignments for 4.3 million pairs of yeast proteins, identify 1,505 likely to interact, and build structure models for 106 previously unidentified assemblies and 806 that have not been structurally characterized. These complexes, which have as many as 5 subunits, play roles in almost all key processes in eukaryotic cells and provide broad insights into biological function.

Bio: Minkyung Baek is a postdoctoral scholar in the Department of Biochemistry at the University of Washington. She studied Chemistry at the Seoul National University where she received her B.A. and Ph.D. degrees. During her Ph.D., she developed computational methods to predict protein-small molecule and protein-protein complex structures. Since she joined the Baker Lab at the University of Washington as a postdoc in 2019, she has focused on developing an AI-based method for protein structure prediction and de novo protein design.