Deep learning-based protein structure modeling and design

Minkyung Baek

School of Biological Sciences, Seoul National University, Republic of Korea

There has been considerable progress in protein structure prediction by leveraging genetic information through deep learning-based methods. In this talk, I'll present a three-track attention-based neural network named RoseTTAFold which transforms and integrates information at the sequence level, the distance map level, and the 3D coordinate level to generate accurate protein structures. It enabled high accuracy protein modeling and protein-protein interaction screening, and even de novo protein design. RoseTTAFold has been successfully applied to design a new protein scaffolding a given functional motif by further training the model on "inpainting" task that fills in additional sequence and structure based on given partial motifs to create a viable protein scaffold in a single forward pass. More recently, we extended our approach to develop a general deep learning framework for protein design including de novo binder deign, higher-order assembly design in addition of scaffolding functional motifs. We transformed RoseTTAFold to a diffusion model that have had considerable success in generative modeling of images and languages. By finetuning the RoseTTAFold on protein structure denoising tasks, we obtain a generative model of protein backbones (RFdiffusion) that achieves outstanding performance on unconditional design, symmetric oligomer design, scaffolding a functional motif design, and protein binder design.