MAINMAST: De Novo Protein Structure Modeling for cryo-EM Maps Assisted by Structure Feature Detection by Deep Learning

Genki Terashi[1], Xiao Wang[2], Daisuke Kihara[3]

[1]Department of Biological Sciences, Purdue University [2]NA, [3]Purdue University

Genki Terashi

An increasing number of protein structures are determined by cryo-electron microscopy (cryo-EM) at ~4-5 Å resolution. However, tracing the main-chains and building full-atom models from EM maps of ~4-5 Å is still a challenging task. Here we present an update of our de novo protein structure modeling method for EM maps of the near atomic resolution, MAINMAST (MAINchain Model trAcing from Spanning Tree) (Terashi & Kihara, Nature Comm. 2018). MAINMAST constructs protein structure models from an EM density map without using known protein structures. MAINMAST builds main-chain traces of a protein from a tree graph structure constructed by connecting local maximum density points (LDPs) in the EM map. The latest version of MAINMAST can handle multiple protein chains with symmetry (Terashi, Kagaya, & Kihara, J. Chem. Inf. Model. 2020) and proteins with bound ligands in EM maps. For modeling of multiple protein chains, MAINMAST identifies pairs of equivalent LDPs that locate in symmetric positions. Then, multiple chain structures are modelled so that each of symmetric LDPs locates in each chains. For modeling of proteins with bound ligands, MAINMAST analyzes the tree-graph and first constructs main-chain model, and then detects the ligand-binding sites as unoccupied local graphs by the protein chain model. Once MAINMAST identifies the potential position of the bound ligand, MAINMAST segments the density around the binding site and models the ligand structure. Recently, we used deep learning, an extended version of Emap2sec+ (Wang, et al. Nature Commn. 2021), to improve the accuracy of the MAINMAST modeling. The new approach, DeepMainmast, perform modeling guided by detected residue and atom information from the map by deep neural network. We observed a drastic improvement of the modeling accuracy by the new approach. In the presentation, we discuss these methods with many modeling examples and other potential applications.