Emap2sec+: Detecting Protein and DNA/RNA Structures in Cryo-EM Maps of Intermediate Resolution Using Deep Learning

Xiao Wang[1], Eman Alnabati[2], Tunde W. Aderinwale[3], Sai Raghavendra Maddhuri Venkata Subramaniya[4], Genki Terashi[5], Daisuke Kihara[6]

[1]No affiliation given [2]Purdue University, [3]Purdue University, [4]Purdue University, [5]Purdue University, [6]Purdue University

wang3702@purdue.edu

An increasing number of density maps of macromolecular structures, including proteins and protein and DNA/RNA complexes, have been determined by cryo-electron microscopy (cryo-EM). Although maps at a near-atomic resolution are routinely reported, there are still substantial fractions of maps determined at intermediate (~ 4 Å) or lower resolutions, where extracting structure information is difficult. Considering limited approaches developed for maps with protein-nucleic acid complexes under intermediate resolution, we report a new computational method, Emap2sec+, which identifies DNA or RNA as well as the secondary structures of proteins in cryo-EM maps of 5 to 10 Å resolution (Nat. Commn, 2021). Emap2sec+ uses the 3D deep residual convolutional neural network (3D-ResNet) as its core of the architecture. Emap2sec+ assigns structural labels with associated probabilities at each voxel in a cryo-EM map, which can help structure modeling in an EM map. It adopts a two-phase stacked neural network architecture, where predictions in the first phase are further smoothed in the subsequent second phase by incorporating the context of neighboring voxels. Emap2sec+ showed stable and high assignment accuracy for nucleotides in low-resolution maps. Detection accuracy is remarkably stable for binary classification of protein and nucleotides even in the maps of a low resolution of 8-10 Å. The code is available at https://github.com/kiharalab/Emap2secPlus together with other cryo-EM software https://kiharalab.org/emsuites.

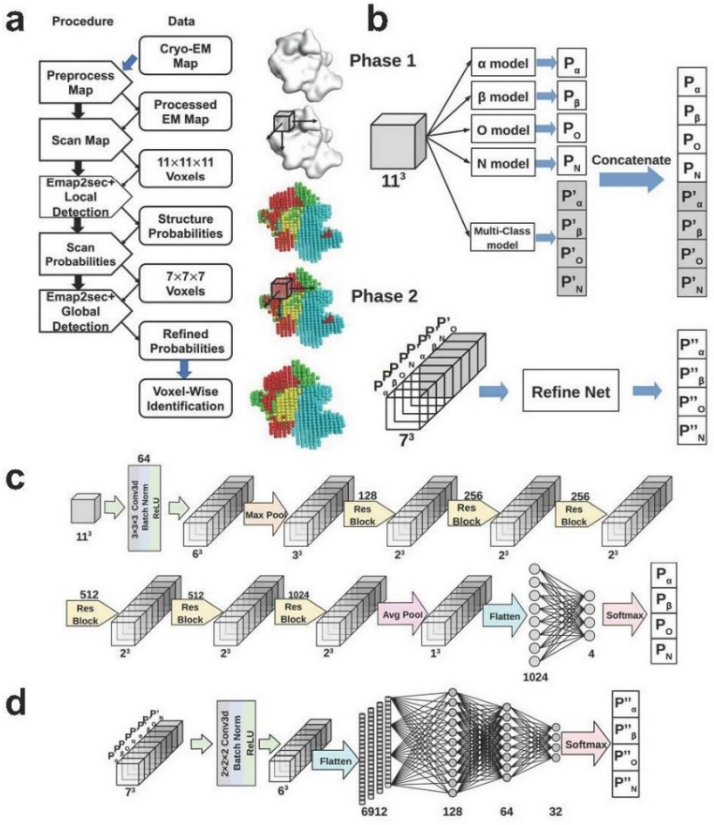

Figure 1