

Protein structure detection in intermediate-resolution cryo-EM maps using deep learning

Cryo-EM has established its position in structural biology as an indispensable method of choice for determining macromolecular structures due to recent technological breakthroughs. The recent years have observed a steep increase of biomolecular structures solved by cryo-EM, including those which were determined at a high, near-atomic resolution. On the other hand, there are still many structures solved at an intermediate resolution, 5 to 10 Å, or even at a lower resolution every year. Here, we developed a computational method, Emap2sec, which identifies the secondary structures of proteins, α helices, β sheets, and the other in an EM map of 5 to 10 Å resolution. Emap2sec uses convolutional deep neural network at its core of the algorithm, which has been recently very successful in 2D and 3D image recognition tasks. Emap2sec assigns the secondary structure to each grid points in an EM map. Thus, the local prediction is made in the context of a large region of the map.

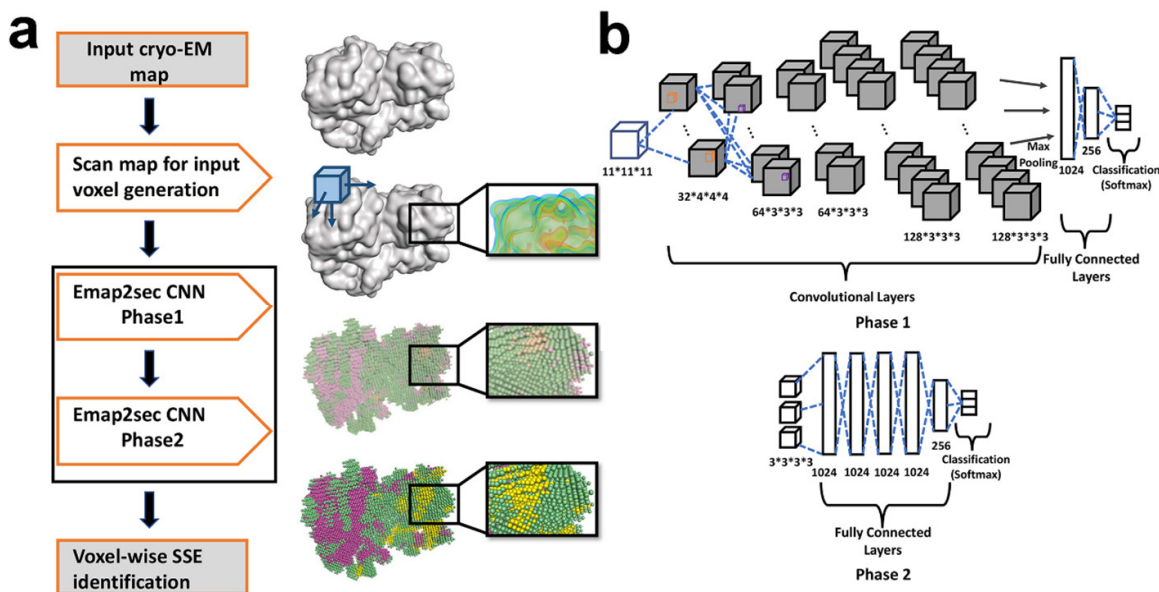


Fig. 1. The architecture of Emap2sec. a, the flowchart. Emap2sec takes a 3D EM density map as input, and scans it with a voxel of a size of $11\text{\AA} \times 11\text{\AA} \times 11\text{\AA}$. There are two phases in Emap2sec. The phase 1 network takes the normalized density values of a voxel and outputs the probability values of the three secondary structure classes. The phase 2 network takes the output from the phase 1 network and refines the assignment by considering assignments made to neighboring voxels. Finally, each voxel is assigned with a secondary structure class, which is the one with the largest probability among the three structure types. b, the architecture of the phase 1 and 2 deep neural networks. The phase 1 network has five convolutional neural network (CNN) layers followed by one max pooling layer. The last layers of the network are two fully connected (FC) layers. The phase 2 network consists of five fully connected neural nets followed by an output layer.

The architecture of Emap2sec is shown in Figure 1. Emap2sec scans an EM map with a cube of 11^3\AA^3 and makes the secondary structure detection for the center position of each cube. Emap2sec has a characteristic

two-phase architecture of the neural network, where an output of the first network is further refined by the second network reducing noise in the prediction. The performance of Emap2sec was tested on two datasets of EM maps, a dataset of 34 simulated EM maps at 6.0 Å and 10.0 Å and a dataset of 43 experimental EM maps whose resolution ranges from 5.0 to 9.5 Å. The overall accuracy at each amino acid level was 83.1 % and 79.8 % on average for the simulated maps at 6.0 Å and 10.0 Å, respectively. On the experimental map dataset, the accuracy was 64.4 % on average with the highest recording 91.6 %.

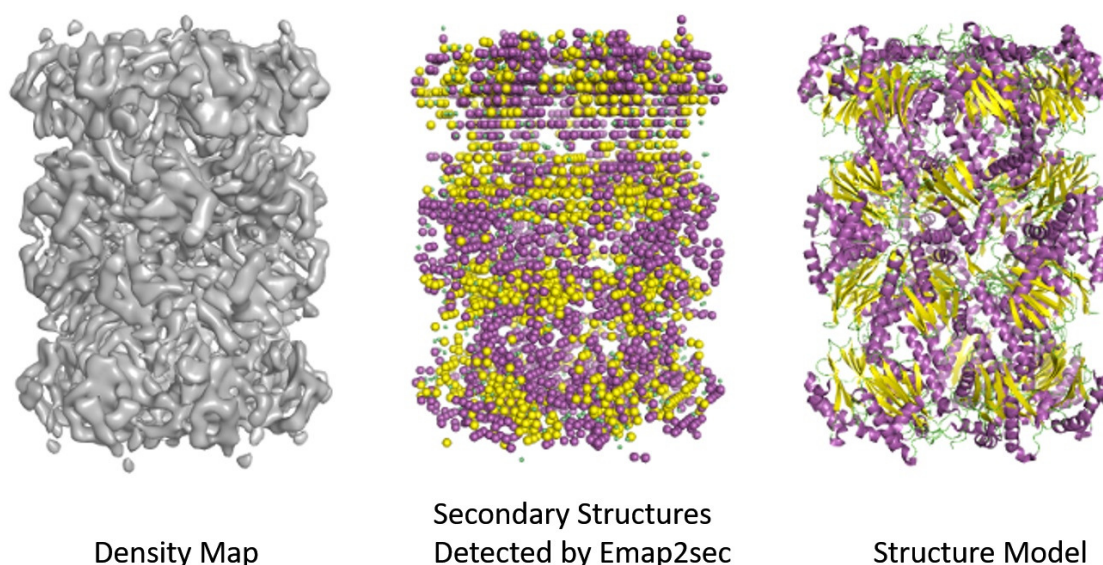


Fig. 2. Protein secondary structure detection by Emap2sec applied to Archaeal 20S proteasome (EMD-1733), a map resolution of 6.8 Å. Left, the cryo-EM density map; middle, the secondary structures detected by Emap2sec. magenta dots, alpha helices; yellow, beta strands; green, loops. Right, the atomic-detailed structure associated with this density map.

Emap2sec is available at kiharalab.org (<https://kiharalab.org/emsuites/>). An updated method, Emap2sec+, which also detects DNA/RNA structures in cryo-EM maps is now developed and made available at the same website.

The preprint of the Emap2sec+ paper is released at bioRxiv (<https://www.biorxiv.org/content/10.1101/2020.08.22.262675v1>).

Sai Raghavendra Maddhuri Venkata Subramaniya
Department of Computer Science, Purdue University, West Lafayette, IN, 47907, USA

Publication

[Protein secondary structure detection in intermediate-resolution cryo-EM maps using deep learning](#)

Maddhuri Venkata Subramaniya SR, Terashi G, Kihara D
Nat Methods. 2019 Sep