# Enhancement for MAINMAST, De Novo Main-Chain Tracing Method: Symmetric Multi-Chain Modeling, Local Refinement, and Graphical User Interface

Genki Terashi[1], Yuhong Zha[2], and Daisuke Kihara[1,3*]

[1.] Department of Biological Sciences, Purdue University, West Lafayette, IN, USA.
[2.] School of Computer Science, Carnegie Mellon University, Pittsburg, PA, USA.
[3.] Department of Computer Science, Purdue University, West Lafayette, IN, USA.
* Corresponding author: dkihara@purdue.edu

The significant progress of the cryo-EM poses a pressing need for software for structural interpretation of EM maps. Particularly, protein structure modeling tools are needed for EM maps determined around 4 Å resolution, where finding main-chain structure and assigning the amino acid sequence into EM map are still challenging problems. We have developed a de novo modeling tool named MAINMAST (MAINchain Model trAcing from Spanning Tree) for EM maps for this resolution range [1, 2]. MAINMAST builds main-chain traces of a protein in an EM map from a tree structure constructed by connecting points with a high density in the map without referring to known protein structures or fragments. The method has substantial advantages over the existing methods: i) MAINMAST directly constructs protein structure models from an EM density map without requiring reference structures; ii) The procedure is fully automated and no manual setting is required; iii) a pool of models are produced, from which a confidence score is computed that indicates accuracy of structure regions.

Here, we report substantial improvements of MAINMAST in three aspects. The largest improvement is that the method now can perform automatic map segmentation and structure modeling for symmetrical multi-chain complexes. The tree-graph structure that connects dense points are traced for multiple chains simultaneously in a symmetric fashion. Figure 1 and 2 are showing the examples of the multi-chain segmentation by MAINMAST and Segger v1.9.5 [4] (plugin in UCSF Chimera molecular visualization software [5]) on EMD-6551 (Magnesium channel CorA at 3.8Å resolution) and EMD-8118 (TRPV1 at 3.28Å resolution). In these two maps, MAINMAST successfully find the individual protein regions from the EM maps.

Moreover, the accuracy of a model is significantly improved by a new implementation of local sequence matching and structure refinement. The local matching protocol is also useful for identifying missing regions in a structure model, i.e. regions with a low density, in an EM map.

Finally, we developed a software plugin of MAINMAST for the UCSF Chimera [5], so that users can monitor structures at each step of a modeling procedure. The major functionalities include to generate and to display tree structures from local dense points in the map, main-chain traces, and reconstructed all-atom models. Through the interface, users can easily control parameters of MAINMAST and save and restore sessions [6].

References:
[1] G  Terashi and D Kihara, Nature Communications [Online] **9.1** 1618 (2018), https://www.nature.com/articles/s41467-018-04053-7 (accessed Feb 20th, 2019).
[2] G Terashi and D Kihara, Journal of Structural Biology **204.2** (2018) p.351-359.
[3] MAINMAST website, http://kiharalab.org/mainmast/index.html (accessed Feb 20th, 2019).
[4] GD Pintilie, et al., Journal of Structural Biology. **170.3** (2010): 427-438.
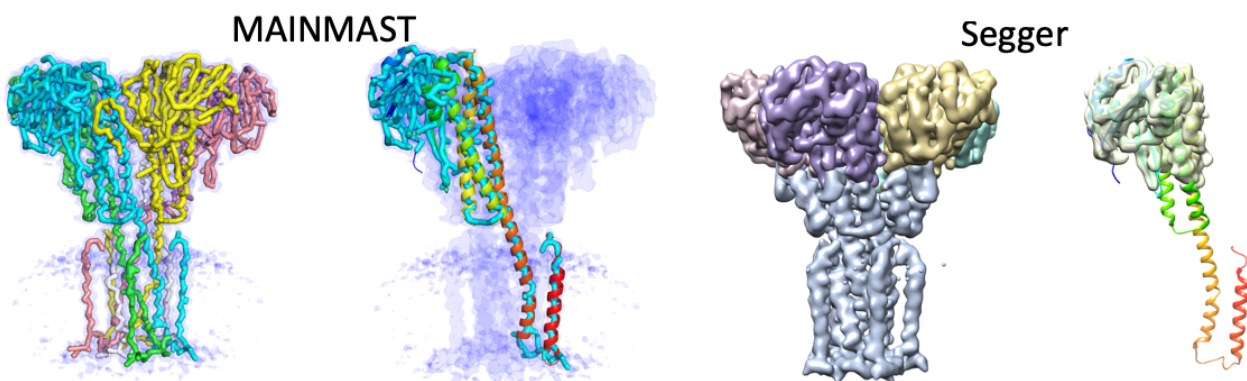[5] EF Pettersen, et al., Journal of Computational Chemistry. **25** (2004) p.1605–1612.

**Figure 1.** Segmentation of individual protein chains for EMD-6551 at 3.8Å resolution by MAINMAST (Left: whole and segmented tree-graph) and Segger (Right: whole and segmented regions). Stick models on the left two panels are tree-graph structures generated by MAINMAST. Cartoon models are native structures (PDB: 3jcf)
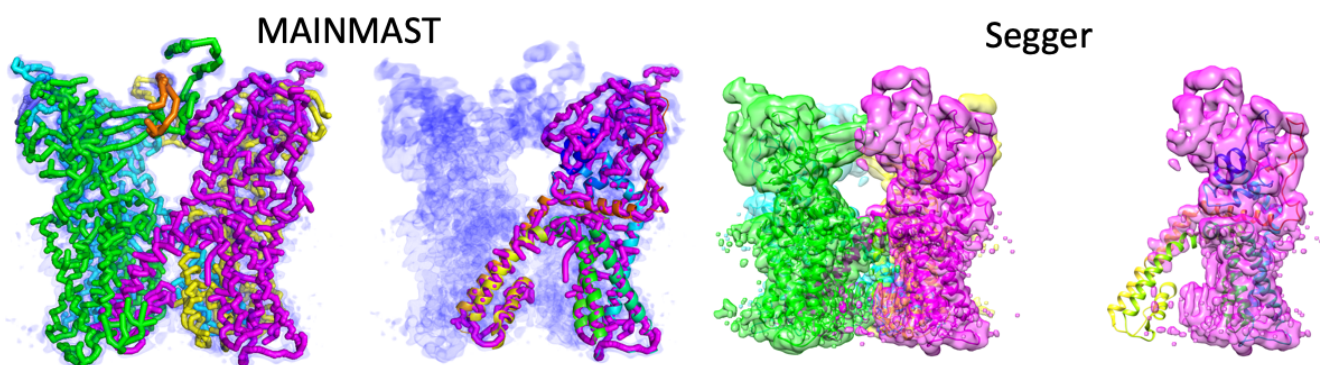


**Figure 2.** Segmentation of individual chains for EMD-8118 at 3.28Å resolution by MAINMAST and Segger. Cartoon models are native structures (PDB: 5irz)
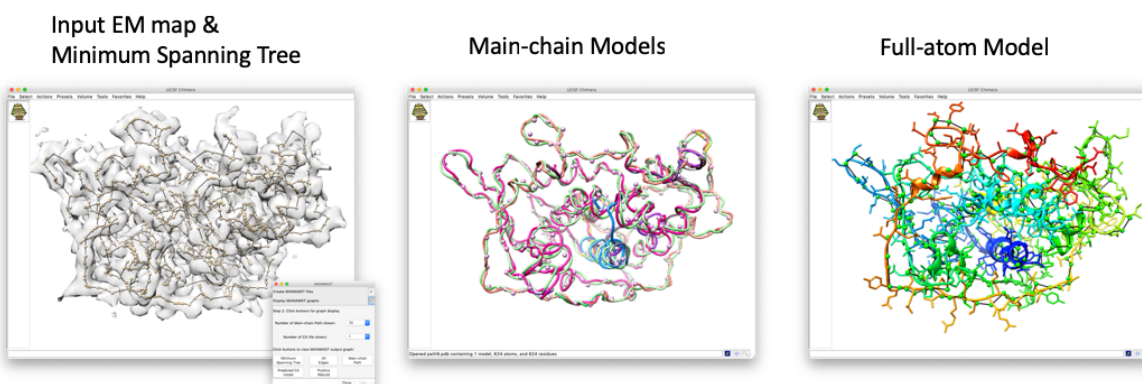


**Figure 3.** GUIs of the MAINMAST plugin for EMD-6374 at 2.9Å resolution. User can build protein models from an EM map.