Protein Bioinformatics Research 2021

Daisuke Kihara
dkihara@purdue.edu
http://kiharalab.org
Hockmeyer 229

MAINMAST: De novo Structure Modeling for medium (~4 Å) Resolution Maps

EM Map

Tracing

Cα model

EMD-6555 Porcine circovirus 2
Resolution: 2.9 Å

De-novo modeling: does not require known protein structures.
Fully automated: No need for visual inspection and human intervention.

(G. Terashi & D. Kihara, Nature Communications, 2018)
MAINMAST: (MAINchain Model trAcing from Spanning Tree)

- EM map
- Find local dense points (LDPs) by Mean Shift
- Connect All points by Minimum Spanning Tree
- Refine Tree Structure
- Thread sequence on the longest path
- Cα models ranked by threading score

EMD-6374 Cytoplasmic polyhedrosis virus resolution: 2.9Å

The longest path (red) does not always cover the whole protein structure.
Recognizing Secondary Structure Elements in EM maps

- An EM map is a 3D voxel data
- Use Convolutional Neural Network
- Scans an EM map with an input cube, and predicts the secondary structure of the protein in the middle of the cube (helix, strand, coil)

(Emap2sec: Detecting Secondary Structures in 6-10 Å EM maps)

(Subramaniya, Terashi, & Kihara, Nature Methods, 2019)
Examples of Structure Detection in real maps

Resolution
a: 6.0 Å  
b: 6.2 Å  
c: 6.8 Å  
d: 8.3 Å  
e: 9.1 Å  
f: 7.9 Å

Detecting DNA/RNA in 5-10 Å Maps

Emap2sec+ (Wang et al., Nature Commn., 2021)
DeepMAINMAST: Deep Learning Assisted De Novo Modeling

EMD-7770 beta-galactosidase 1.9 Å resolution

(Terashi et al., to be submitted 2021)

AttentiveDist: Protein Structure Prediction

Protein-Protein Docking

https://lzerd.kiharalab.org
**PL-PatchSurfer: Pocket-Ligand Comparison**

- Retinol
- Guanine
- Adenosine

Binding Pocket

Ligand Library

(Shin, Bures, Kihara, Methods, 2015)
(Hu, Zhu, Lyman, Bures, Kihara, I.J. Mol. Sci, 2014)

**Sequence-Based Function Prediction**

- Blue whale genome project
- Peptide toxicity prediction

(Jain & Kihara, in revision, 2018)
Genome Reannotation

Genome reannotation is a process where gene annotations are updated or revised to improve their accuracy and relevance. This can be done to correct errors, add new information, or improve the understanding of gene function. The diagram illustrates the reannotation status of various organisms, showing how much of the genome has been annotated in different categories: Annotated, High, Medium, Low, and None.

(Hawkins, Chitale, Luban, & Kihara, Proteins, 2009)

Structural & Functional Bioinformatics Group

The Structural & Functional Bioinformatics Group is a team dedicated to research and development in the field of bioinformatics. Their work involves the analysis and interpretation of biological data to understand the structure and function of biological systems.

@kiharalab  https://kiharalab.org