

# Beyond Ribosomes: In Situ Structural Biology of a Challenging Target in *C. Reinhardtii*

**Dr Jessica Heebner**<sup>1</sup>, Sofie van Dorst<sup>2</sup>, Dr. Ron Kelley<sup>1</sup>, Dr. Martin Obr<sup>1</sup>, Dr. Sagar Khavnekar<sup>1</sup>, Dr. Xianjun Zhang<sup>1</sup>, Dr. Saikat Chakraborty<sup>1</sup>, Dr. Ricardo Righetto<sup>2</sup>, Dr. Florent Waltz<sup>2</sup>, Dr. Alicia Michael<sup>2</sup>, Dr. Wojciech Wietrzynski<sup>2</sup>, Dr. Grigory Tagiltsev<sup>3</sup>, Dr. John Briggs<sup>3</sup>, Dr. Juergen Plitzko<sup>3</sup>, Dr. Ben Engel<sup>2</sup>, Dr. Abhay Kotecha<sup>1</sup>

<sup>1</sup>Thermo Fisher Scientific, Hillsboro, United States, <sup>2</sup>Biozentrum, University of Basel, Basel, Switzerland, <sup>3</sup>Department of Molecular Structural Biology, Max Planck Institute of Biochemistry, Martinsreid, Germany

## Background

Understanding high-resolution protein structure in the context of the whole cellular environment is the vision of visual proteomics. With the advent of high-throughput cryo-FIB and cryo-electron tomography (CryoET), paired with cutting-edge computational techniques, achieving such an ambitious goal is no longer a far-reaching dream. The new generation of cryo-FIB from Thermo Fisher Scientific uses plasma ions, which reduce redeposition and ion beam damage, substantially improving throughput. In preparation for a large-scale effort towards visual proteomics of the model organism *Chlamydomonas reinhardtii*, we have prepared enough lamellae to allow for acquisition of more than 1800 tomograms of different cellular compartments. Initial results demonstrate that the data is of sufficient quality to achieve sub-nanometer resolution (6Å) for the 80S ribosome using a fraction of the dataset (six tomograms). Together with collaborators, multiple proteins were selected as targets for sub-tomogram averaging (STA). One of these proteins required the development of a novel workflow. Photosystem II (PSII), which is present in very high abundance in the chloroplast, proved to be a challenging target as it is a membrane embedded protein with only a small region protruding into the lumen of the thylakoid membrane. The denoising neural network cryoCARE was implemented on all tomograms in the dataset to increase template matching accuracy. Unfortunately, in the case of PSII, denoising tended to blur or erase the small density that projects from the membrane which led to significant under picking when template matching was performed.

## Methods

An alternative strategy that paired a regression denoising UNet with a semantic segmentation UNet was employed to generate candidate coordinates for further averaging and classification. Regression denoising: Synthetic training data for regression UNet was generated using CryoTomoSim. Using a box of 400x400x50 pixels, a mixture of medium and small proteins were modeled in four layers at 7.84 Å /pixel, exact protein identity is not important to training. 10 membrane vesicles were modeled. Iterations for each layer were 500, 500, 4000, 8000, and particle density was 0.8. Finally, the vitreous ice option was used. The output or ideal tomogram was simulated at -1 micron defocus, -89 to 89 tilt, 0.5 degree tilt increment and total dose = 0. The input or noisy tomogram was simulated at -3 micron defocus, -60 to 60 tilt, 3 degree tilt increment, total dose 80. Both datasets are loaded into Dragonfly 2022.2 (Object Research Systems) and used as training input and output for a regression 2.5D UNet with architecture: depth level 5, initial filter count 64, slice count 5, patch size 128, Loss function ORSMixedGradientLoss. Training proceeded for 46 epochs using a total of 15,840 patches. Segmentation: 2.5D Segmentation UNet for PSII was trained using Dragonfly 2022.2 according to protocol. In brief, Segmentation Wizard was used for manual annotation of 5-6 training slices and a generic UNet of the following architecture was trained: (2.5D: 3 slice, depth level: 5, initial filter count: 64, patch size: 128). Training data included slices from tomogram numbers: 24, 373, 473, and 900. Aside from patch size, all hyperparameters are left as default. Training labels were Membrane, ATP Synthase, Ribosomes, PSII and Background. Post segmentation, PSII class was extracted for each

tomogram, split into connected components, and center of mass X, Y, Z coordinates were calculated for each label and exported to CSV for subtomogram extraction. Once trained, both UNets are applied to 41 tomograms to generate 52,000 PSII candidates which can be extracted for the STA workflow.

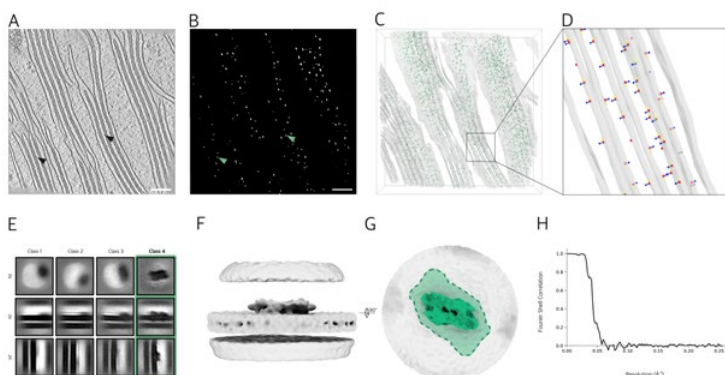
### Results

With the initial set of 52,000 candidate coordinates, one high-quality class emerged during classification and as of the writing of this abstract an 18Å structure has been achieved (Fig 1). Ongoing work is in progress to apply this workflow to an additional 200 tomograms. With more candidate coordinates, optimized alignment of sub-tomograms, and further post-processing, we are confident we can improve the resolution of the in situ structure of PSII.

### Conclusion

The scale of this dataset is exciting, but the huge number of molecular complexes within living cells makes it difficult to identify, confirm the identity of, and determine each structure by just one group. Achieving a full visual proteome of *C. reinhardtii* will necessitate a large collaborative effort. Challenging targets such as PSII demonstrate the need for new, creative methods or combinations of techniques to facilitate in situ structural determination. To that end, we would like to create an open access database for *C. reinhardtii* to accelerate annotation and curation, enable further cell biology research, and develop new computational tools for in situ cryo-ET. Along with sharing the raw data, reconstructed tomograms, denoised datasets, and structural determinations, we will provide high quality segmentations of selected datasets created using 2.5D (Dragonfly) and 3D (MemBrain Seg) UNets. This project has the potential to provide invaluable insights into cellular processes and will hopefully lay the foundation for future large-scale studies of other species.

### Graphic:



### Keywords:

CryoET, Deep Learning, Visual Proteomics

### Reference:

1. Purnell, C. et al. Rapid Synthesis of Cryo-ET Data for Training Deep Learning Models. <http://biorxiv.org/lookup/doi/10.1101/2023.04.28.538636> (2023) doi:10.1101/2023.04.28.538636.
2. Heebner, J. E. et al. Deep Learning-Based Segmentation of Cryo-Electron Tomograms. *JoVE J. Vis. Exp.* e64435 (2022) doi:10.3791/64435.
3. Lamm, L. et al. MemBrain v2: An End-to-End Tool for the Analysis of Membranes in Cryo-Electron Tomography. <http://biorxiv.org/lookup/doi/10.1101/2024.01.05.574336> (2024) doi:10.1101/2024.01.05.574336.