

Automated in situ CryoET Structure Determination with a Self-Configuring Workflow

Dari Kimanius

`dari.kimanius@biohub.org`

Biohub

Jonathan Schwartz

Chan Zuckerberg Biohub <https://orcid.org/0000-0002-8063-6951>

Daniel Ji

Biohub

Utz Ermel

Biohub <https://orcid.org/0000-0003-4685-037X>

Joshua Hutchings

Biohub

Rahel Woldeyes

Biohub

Zhuowen Zhao

Chan Zuckerberg Biohub <https://orcid.org/0000-0002-2355-2284>

Christof Henkel

NVIDIA

Mallak Ali

Biohub

Yue Yu

Biohub

David Agard

Biohub

Mohammadreza Paraan

Biohub

Bridget Carragher

Biohub

Brief Communication

Keywords: cryoET, Segmentation, Deep Learning, Sub-tomogram Averaging

Posted Date: August 28th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10374846/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: There is **NO** Competing Interest.

Automated in situ CryoET Structure Determination with a Self-Configuring Workflow

Jonathan Schwartz¹, Daniel Ji¹, Utz Ermel¹, Joshua Hutchings¹,
Rahel Woldeyes¹, Zhuowen Zhao¹, Christof Henkel², Mallak Ali¹,
Yue Yu¹, David Agard^{1,3}, Mohammadreza Paraan¹,
Bridget Carragher¹, Dari Kimanius^{1*}

^{1*}Biohub, Redwood City, 94063, CA, USA.

²Nvidia, Santa Clara, 95051, CA, USA.

³Department of Biochemistry & Biophysics, University of California,
San Francisco, USA.

*Corresponding author(s). E-mail(s): dari.kimanius@biohub.org;

Abstract

Cryo-electron tomography can reveal macromolecular structures in native cellular environments, but structure determination remains limited by expert-guided particle picking and dataset-specific processing. We present an automated framework combining Octopi, a self-configuring deep-learning particle detector, with py2rely, a standardized subtomogram averaging workflow. Starting from sparse target annotations, we applied the framework to 17 structure-determination tasks, yielding reconstructions at 3–20 Å resolution across purified macromolecules, isolated organelles, and intact cells.

Keywords: cryoET, Segmentation, Deep Learning, Sub-tomogram Averaging

Cryo-electron tomography (cryoET) reveals macromolecular complexes within native cellular environments in three dimensions [1]. From these volumes, subtomogram averaging (STA) determines the structure of complexes at high resolution [2]. Although cryoET data acquisition and tomogram reconstruction have become increasingly routine, structure determination by STA remains limited in scale, hindering high-throughput analysis in crowded, heterogeneous volumes typical of cellular tomography.

To determine structures from cryoET data at scale, two challenges must be addressed simultaneously: particles must be localized in complex cellular volumes, and the resulting coordinates must be reconstructed into structures through STA pipelines. Current approaches depend on expert-configured particle picking and extensive manual intervention during downstream processing, with workflows assembled separately for each dataset [3–5]. Unlike single-particle analysis, where first-pass automated workflows are available in tools such as CryoSPARC Live [6] and RELION [7], STA lacks an equivalent end-to-end route from candidate particles to reconstructed structures.

This missing end-to-end route also complicates the evaluation of particle-picking methods on experimental cryoET data. Benchmark datasets with available annotations have enabled systematic evaluation using coordinate-overlap metrics [8]. However, most experimental datasets lack corresponding ground-truth coordinates, limiting the applicability of annotation-based benchmarks. In such cases, STA can provide a structure-based readout of coordinate quality, since coordinate sets that recover more true particles are expected to yield higher-resolution structures [9]. For this comparison to be meaningful however, coordinate sets must be processed through the same automated STA workflow to minimize user bias. Thus, a generalizable, minimally supervised route from particle picking to structure determination is needed both for scalable structure discovery and for unbiased method evaluation.

To address this need, we developed a self-configuring particle-picking framework and linked it to an automated STA workflow for structure determination (Fig. 1a). The particle-picking component, Octopi, is a supervised deep learning framework that uses Bayesian optimization to automatically adapt model architecture and training choices to each new cryoET dataset (Fig. 1b). The STA component, py2rely, takes particle coordinates and tomographic tilt series as input and determines structures through a workflow built on RELION-5 [10] and the CCP-EM Pipeliner [11] (Fig. 1c). Both components operate through copick, a shared cryoET data interface that standardizes access to tomograms, annotations, and coordinates across tools [12]. We evaluated this framework on 17 structure-determination tasks spanning purified macromolecules, isolated organelles, and intact cells (Extended Data Table 1).

First, we applied Octopi to the Phantom dataset from the *CryoET Object Identification Kaggle Challenge*, where ground-truth coordinates are available for six macromolecular species in crowded pseudo-cellular tomograms [13]. To directly compare Octopi’s self-configuring strategy with an expert-configured baseline, we developed TopCup (**T**op **C**ryoET **U**-Net **P**icker), an open-source implementation of the top-ranked challenge entry, which was optimized specifically for the Phantom dataset. This annotated benchmark allowed us to evaluate particle-picking performance using conventional coordinate-overlap metrics, while also establishing the automated STA-based evaluation workflow in py2rely, which is used throughout this study to compare coordinate sets across datasets without ground-truth annotations.

Using the same evaluation metrics from the challenge, we compared Octopi to TopCup with conventional coordinate-overlap metrics (see Methods). Octopi achieved an F_β score of 0.763, matching TopCup, which scored 0.761, and approaching the top-ranked challenge entry, of 0.788 (Fig. 2e,f; Extended Data Fig. 2). Thus, without

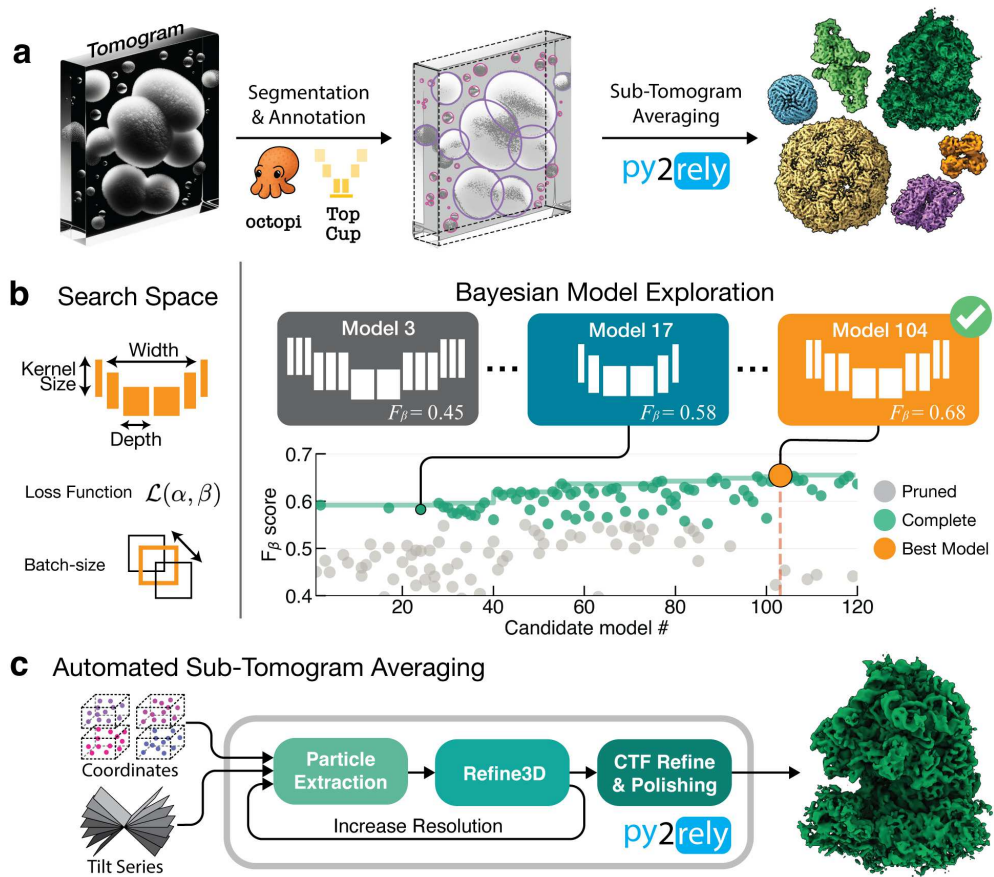


Fig. 1 Automated particle-picking and structure determination pipeline for cryoET. **a**, Overview of the cryoET pipeline. **b**, Octopi self-configures to new datasets through Bayesian hyperparameter search. The search space spans architectural choices and training choices (see Methods). Data shown are from an Optuna run on the CZCDP-10310 Phantom benchmark dataset; F_β improved from 0.45 (Model 3) to 0.68 (Model 104) over 120 trials, with representative network architectures shown above. **c**, Particle coordinates and tilt series feed into iterative subtomogram extraction, 3D refinement, and CTF refinement with Bayesian polishing in RELION-5. The full workflow is detailed in Extended Data Fig. 9.

dataset-specific model design, Octopi matched the coordinate-localization performance of a specialized, expert-configured model.

Next, we evaluate if these coordinate sets produced comparable reconstructions when processed through the same automated STA workflow. Coordinates from Octopi, TopCup, and the top-ranked challenge entry were processed using identical py2rely workflows. All three coordinate sets produced structures for each of the six Phantom targets, with final resolutions ranging from 3.5 to 11 Å and comparable reconstruction quality across methods (Extended Data Fig. 3). These results indicate that, given

sufficient true positives, py2rely can automatically perform structure determination across coordinate sets generated by different particle-picking methods.

The automated py2rely workflow also improved previously reported structural outcomes through iterative CTF refinement and per-particle frame alignment. For example, the ribosome reconstructed using picks from the top-ranked Kaggle entry reached 3.7 Å with the updated py2rely workflow, compared with the previously reported 6.8 Å reconstruction from the same challenge-derived coordinates [9]. Together, these results show that Octopi achieves competitive coordinate-localization performance on annotated benchmark data and that py2rely provides a consistent structure-based endpoint for comparing particle-picking methods.

We then applied this automated STA-based evaluation framework to native, membrane-rich cryoET datasets lacking ground-truth coordinates, where particle localization depends not only on local particle signal but also on cellular context. In unroofed HEK293 and HSC3 cells [14], Octopi was trained on two manually annotated tomograms containing 203 membrane-associated and 612 cytosolic ribosome coordinates. From these sparse labels and membrane annotations, Octopi distinguished the two ribosomal populations while simultaneously segmenting the surrounding membranes (Fig. 2a,b). The resulting subtomogram averages resolved both cytosolic and membrane-associated ribosomes at nanometer resolution. In the membrane-associated ribosome reconstruction, additional density was consistent with the Sec61-translocon-associated protein complex reported previously [15, 16] (Extended Data Fig. 4).

Membrane-bound targets present a greater challenge, as particle signal is dominated by the surrounding membrane. Using 157 manually annotated particles from isolated synaptic vesicles [17], Octopi identified V-ATPase candidates for downstream averaging. This target required additional curation, but nevertheless yielded a 19.5 Å reconstruction in which the V_0 and V_1 densities are clearly resolved (see Methods; Extended Data Fig. 5), approaching the 16.7 Å resolution reported in the original study. This was achieved despite only 95 of the 719 tomograms from the published dataset being available on the CryoET Data Portal. These results show that effective pickers can be trained from sparse labels even for challenging membrane-associated targets, provided that membrane context is learned jointly with particle picking.

More broadly, Octopi can incorporate larger-scale cellular features, such as vesicles, membranes, and organelles, when these contextual cues are informative for particle localization. This capacity for context-aware segmentation is further explored in the POPSICLE benchmark [8] (Extended Data Fig. 6).

In intact *E. coli* cells, Octopi recovered VLPs and GroEL in crowded bacterial cytoplasm without target-specific retuning (Fig. 2c,d). For GroEL, which is difficult to annotate manually because only a restricted set of high-contrast views is readily recognizable in tomographic slices, we bootstrapped training from template-matched coordinates. As few as 50 template-derived picks were sufficient to train an Octopi model that produced a sub-nanometer GroEL reconstruction (see Methods; Extended Data Fig. 7). For VLPs, a single Octopi model trained on three tomograms from one plunge-frozen preparation generalized to two independent data collections and a cryo-FIB lamella dataset with different signal-to-noise properties, yielding reconstructions between 3.0 and 3.9 Å across the three held-out datasets (Extended Data Fig. 8).

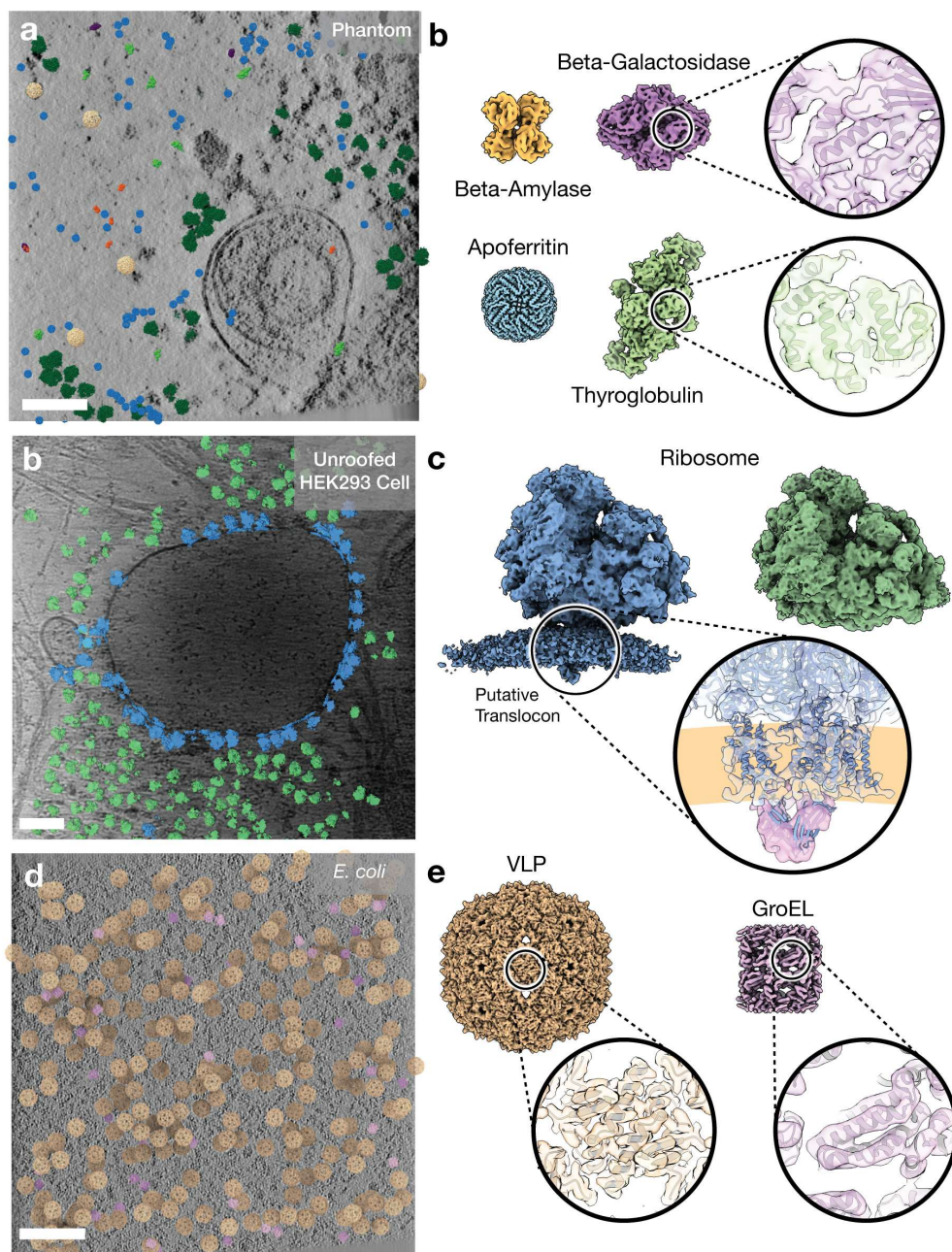


Fig. 2 Octopi detects diverse macromolecular targets. **a,b** Tomogram from the Phantom sample, used in the Kaggle macromolecular-localization challenge, with Octopi predictions overlaid for benchmark particle classes. **c,d** Representative unroofed-cell tomogram and reconstructions of cytosolic ribosomes and membrane-associated ribosomes in a crowded cellular environment. **e,f** An *E. coli* tomogram showing Octopi predictions for multiple intracellular targets in dense bacterial cytoplasm. Scale bars, 100 nm.

Octopi transfers across annotation sources and acquisition regimes, supporting its use as a general-purpose picker adaptable to new applications.

By interfacing directly with the CryoET Data Portal [18], Octopi and py2rely provide a community resource for in situ cryoET structure determination. This integration enables standardized method development and reproducible processing of tomograms and coordinates hosted in the portal, supporting structure determination across the growing body of publicly available cryoET data.

Overall, self-configuring particle picking coupled to standardized STA produced structures at 3 to 20 Å resolution. After a small set of manual annotations specifying the target molecule was provided, the workflow proceeded automatically for all targets tested except the synaptic-vesicle V-ATPase, which required manual class selection and mask adjustment. This exception identifies a clear direction for future development: improved automated handling of challenging membrane-associated targets. More broadly, the workflow provides a foundation for extending automation across additional processing decisions and applying the same high-throughput framework to a wider range of in situ targets.

References

- [1] Beck, M., Baumeister, W.: Cryo-electron tomography: can it reveal the molecular sociology of cells in atomic detail? *Trends in Cell Biology* **26**(11), 825–837 (2016) <https://doi.org/10.1016/j.tcb.2016.08.006>
- [2] Bartesaghi, A., Lecumberry, F., Sapiro, G., Subramaniam, S.: Protein secondary structure determination by constrained single-particle cryo-electron tomography. *Structure*, 2003–2013 (2012) <https://doi.org/10.1016/j.str.2012.10.016>
- [3] Moebel, E., *et al.*: Deep learning improves macromolecule identification in 3D cellular cryo-electron tomograms. *Nature Methods* **18**(11), 1386–1394 (2021) <https://doi.org/10.1038/s41592-021-01275-4>
- [4] Teresa-Trueba, I., Goetz, S.K., Mattausch, A., *et al.*: Convolutional networks for supervised mining of molecular patterns within cellular context. *Nature Methods* **20**(2), 284–294 (2023) <https://doi.org/10.1038/s41592-022-01746-2>
- [5] Rice, G., *et al.*: TomoTwin: generalized 3D localization of macromolecules in cryo-electron tomograms with structural data mining. *Nature Methods* **20**(6), 871–880 (2023) <https://doi.org/10.1038/s41592-023-01878-z>
- [6] Punjani, A., Rubinstein, J.L., Fleet, D.J., Brubaker, M.A.: cryosparc: algorithms for rapid unsupervised cryo-em structure determination. *Nature methods* **14**(3), 290–296 (2017) <https://doi.org/10.1038/nmeth.4169>
- [7] Kimanius, D., Dong, L., Sharov, G., Nakane, T., Scheres, S.H.: New tools for automated cryo-em single-particle analysis in relion-4.0. *Biochemical Journal* **478**(24), 4169–4185 (2021) <https://doi.org/10.1042/BCJ20210708>

- [8] Schwartz, J., Ermel, U.H., Owens, C.B., Zhao, Z., Peck, A., Hart, G.L.W., Jensen, G.J., Carragher, B., Kimanius, D.: Popsicle: Benchmark datasets for segmentation and localization in cryoet. arXiv (2026) <https://doi.org/arXiv:2606.10255v1>
- [9] Peck, A., Hutchings, J., Schwartz, J., Yu, Y., Ermel, U.H., Kandel, S., Kimanius, D., Zhao, Z., Zheng, S., Artley, B., List, D., Silva Jr., S.A., Reade, W., Asuncion, J., Evans, K., Gadling, J., Khandwal, K., McCanny, S., McCarthy, D.G., Ni, J.X., Pourroy, J., Venkatakrishnan, M., Wang, Z.S., Agard, D.A., Potter, C.S., Carragher, B., Harrington, K., Paraan, M.: Lessons learned from a Kaggle challenge for particle picking in cryo-electron tomography. bioRxiv, 2025–1103686153 (2025) <https://doi.org/10.1101/2025.11.03.686153>
- [10] Burt, A., Toader, B., Warshamanage, R., Kügelgen, A., Pyle, E., Zivanov, J., Kimanius, D., Bharat, T.A., Scheres, S.H.: An image processing pipeline for electron cryo-tomography in relion-5. FEBS open bio **14**(11), 1788–1804 (2024)
- [11] Burnley, T., Palmer, C.M., Winn, M.: Recent developments in the ccp-em software suite. Biological Crystallography **73**(6), 469–477 (2017)
- [12] Ermel, U.H., Schwartz, J., Zhao, Z., Ji, D., Peck, A., Yu, Y., Paraan, M., Carragher, B., Frangakis, A.S., Harrington, K.I.S.: copick: An open dataset interface and toolkit for collaborative annotation and analysis of cryo-electron tomography data. Protein Science **35**(5), 70578 (2026) <https://doi.org/10.1002/pro.70578>
- [13] Peck, A., Yu, Y., Schwartz, J., Cheng, A., Ermel, U.H., Hutchings, J., Kandel, S., Kimanius, D., Montabana, E.A., Serwas, D., Siems, H., Wang, F., Zhao, Z., Zheng, S., Haury, M., Agard, D.A., Potter, C.S., Carragher, B., Harrington, K., Paraan, M.: A realistic phantom dataset for benchmarking cryo-ET data annotation. Nature Methods **22**, 1819–1823 (2025) <https://doi.org/10.1038/s41592-025-02800-5>
- [14] Sun, W.W., Michalak, D.J., Sochacki, K.A., Kunamaneni, P., Alfonzo-Méndez, M.A., Arnold, A.M., Strub, M.-P., Hinshaw, J.E., Taraska, J.W.: Cryo-electron tomography pipeline for plasma membranes. Nature Communications **16**(1), 10–10384146702556045 (2025) <https://doi.org/10.1038/s41467-025-56045-z>
- [15] Gemmer, M., Chaillet, M.L., Loenhout, J., Cuevas Arenas, R., Vismpas, D., Gröllers-Mulderij, M., Koh, F.A., Albanese, P., Scheltema, R.A., Howes, S.C., Kotecha, A., Fedry, J., Förster, F.: Visualization of translation and protein biogenesis at the er membrane. Nature **614**, 160–167 (2023) <https://doi.org/10.1038/s41586-022-05638-5>
- [16] Jaskolowski, M., Jomaa, A., Gamerdinger, M., Shrestha, S., Leibundgut, M., Deuerling, E., Ban, N.: Molecular basis of the trap complex function in er protein biogenesis. Nature Structural & Molecular Biology **30**, 770–777 (2023) <https://doi.org/10.1038/s41594-023-00990-0>

- [17] Kravčenko, U., Ruwolt, M., Kroll, J., Yushkevich, A., Zenkner, M., Ruta, J., Lotfy, R., Wanker, E.E., Rosenmund, C., Liu, F., Kudryashev, M.: Molecular architecture of synaptic vesicles. *Proceedings of the National Academy of Sciences* **121**(50), 2407375121 (2024) <https://doi.org/10.1073/pnas.2407375121>
- [18] Ermel, U., Cheng, A., Ni, J.X., Gadling, J., Venkatakrishnan, M., Evans, K., Asuncion, J., Sweet, A., Pourroy, J., Wang, Z.S., Khandwala, K., Nelson, B., McCarthy, D., Wang, E.M., Agarwal, R., Carragher, B.: A data portal for providing standardized annotations for cryo-electron tomography. *Nature Methods* **21**, 2200–2202 (2024)
- [19] Cardoso, M.J., Li, W., Brown, R., Ma, N., Kerfoot, E., Wang, Y., Murrey, B., Myronenko, A., Zhao, C., Yang, D., Nath, V., He, Y., Xu, Z., Hatamizadeh, A., Myronenko, A., Zhu, W., Liu, Y., Zheng, M., Tang, Y., Yang, I., Zephyr, M., Hashemian, B., Alle, S., Darestani, M.Z., Budd, C., Modat, M., Vercauteren, T., Wang, G., Li, Y., Hu, Y., Fu, Y., Gorman, B., Johnson, H., Genereaux, B., Erdal, B.S., Gupta, V., Diaz-Pinto, A., Dourson, A., Maier-Hein, L., Jaeger, P.F., Baumgartner, M., Kalpathy-Cramer, J., Flores, M., Kirby, J., Cooper, L.A.D., Roth, H.R., Xu, D., Bericat, D., Floca, R., Zhou, S.K., Shuaib, H., Farahani, K., Maier-Hein, K.H., Aylward, S., Dogra, P., Ourselin, S., Feng, A.: MONAI: An open-source framework for deep learning in healthcare (2022). <https://doi.org/10.48550/arXiv.2211.02701>
- [20] Akiba, T., Sano, S., Yanase, T., Ohta, T., Koyama, M.: Optuna: A next-generation hyperparameter optimization framework. In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 2623–2631 (2019). <https://doi.org/10.1145/3292500.3330701>
- [21] Lamm, L., Zufferey, T., Lamm, L., Righetto, R.D., Wietrzynski, W., Pöge, M., Martinez-Sanchez, A., Peng, T., Engel, B.D.: Membrain v2: an end-to-end tool for the analysis of membranes in cryo-electron tomography. *bioRxiv* (2024) <https://doi.org/10.1101/2024.01.05.574336>
- [22] Peck, A., Yu, Y., Paraan, M., Kimanius, D., Ermel, U.H., Hutchings, J., Schwartz, J., Serwas, D., Siems, H., Hill, N.S., Ali, M., Peukes, J., Greenan, G.A., Sheu, S.-H., Montabana, E.A., Carragher, B., Potter, C.S., Agard, D.A., Zheng, S.: Aretomolive: automated reconstruction of comprehensively corrected and denoised cryo-electron tomograms in real time and at high throughput. *Nature Methods* **23**, 1121–1125 (2026) <https://doi.org/10.1038/s41592-026-03093-y>
- [23] Liu, Y.-T., Fan, H., Jih, J., Tran, L., Zhang, X., Zhou, Z.H.: Isonet2 determines cellular structures at submolecular resolution without averaging. *bioRxiv* (2025) <https://doi.org/10.64898/2025.12.09.693325>
- [24] Chaillet, M.L., Roet, S., Veltkamp, R.C., Förster, F.: pytom-match-pick: A tophat-transform constraint for automated classification in template matching. *Journal of Structural Biology: X* **11**, 100125 (2025) <https://doi.org/10.1016/j.j>

- [25] Isensee, F., Jaeger, P.F., Kohl, S.A.A., Petersen, J., Maier-Hein, K.H.: nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods* **18**, 203–211 (2021) <https://doi.org/10.1038/s41592-020-01008-z>
- [26] Hatamizadeh, A., Nath, V., Tang, Y., Yang, D., Roth, H.R., Xu, D.: Swin UNETR: Swin transformers for semantic segmentation of brain tumors in MRI images. In: *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, pp. 272–284. Springer, Cham (2022)
- [27] Goddard, T.D., Huang, C.C., Meng, E.C., Pettersen, E.F., Couch, G.S., Morris, J.H., Ferrin, T.E.: Ucsf chimeraX: Meeting modern challenges in visualization and analysis. *Tools for Protein Science* **27**, 14–25 (2017) <https://doi.org/10.1002/pro.3235>
- [28] Ermel, U.: Copick/chimerax-copick: ChimeraX-copick: V1.6.0. <https://doi.org/10.5281/zenodo.18765347>
- [29] Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S.: Pytorch: An imperative style, high-performance deep learning library. In: *Advances in Neural Information Processing Systems*, vol. 32, pp. 8024–8035. Curran Associates, Inc., Red Hook, NY, USA (2019)
- [30] Harris, C.R., Millman, K.J., Walt, S.J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N.J., Kern, R., Picus, M., Hoyer, S., Kerkwijk, M.H., Brett, M., Haldane, A., Río, J.F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., Oliphant, T.E.: Array programming with numpy. *Nature* **585**, 357–362 (2020) <https://doi.org/10.1038/s41586-020-2649-2>
- [31] Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., Walt, S.J., Brett, M., Wilson, J., Millman, K.J., Mayorov, N., Nelson, A.R.J., Jones, E., Kern, R., Larson, E., Carey, C.J., Polat, Í., Feng, Y., Moore, E.W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E.A., Harris, C.R., Archibald, A.M., Ribeiro, A.H., Pedregosa, F., Mulbregt, P., Vijaykumar, A., Bardelli, A.P., Rothberg, A., Hilboll, A., Kloeckner, A., Scopatz, A., Lee, A., Rokem, A., Woods, C.N., Fulton, C., Masson, C., Häggström, C., Fitzgerald, C., Nicholson, D.A., Hagen, D.R., Pasechnik, D.V., Olivetti, E., Martin, E., Wieser, E., Silva, F., Lenders, F., Wilhelm, F., Young, G., Price, G.A., Ingold, G.-L., Allen, G.E., Lee, G.R., Audren, H., Probst, I., Dietrich, J.P., Silterra, J., Webber, J.T., Slavič, J., Nothman, J., Buchner, J., Kulick, J., Schönberger,

J.L., Miranda Cardoso, J., Reimer, J., Harrington, J., Rodríguez, J.L.C., Nunez-Iglesias, J., Kuczynski, J., Tritz, K., Thoma, M., Newville, M., Kümmerer, M., Bolingbroke, M., Tartre, M., Pak, M., Smith, N.J., Nowaczyk, N., Shebanov, N., Pavlyk, O., Brodtkorb, P.A., Lee, P., McGibbon, R.T., Feldbauer, R., Lewis, S., Tygier, S., Sievert, S., Vigna, S., Peterson, S., More, S., Pudlik, T., Oshima, T., Pingel, T.J., Robitaille, T.P., Spura, T., Jones, T.R., Cera, T., Leslie, T., Zito, T., Krauss, T., Upadhyay, U., Halchenko, Y.O., Vázquez-Baeza, Y., Contributors, S...: Scipy 1.0: fundamental algorithms for scientific computing in python. *Nature Methods* **17**, 261–272 (2020) <https://doi.org/10.1038/s41592-019-0686-2>

- [32] Hutchings, J., Ji, D., Zheng, S., Montabana, E.A., Ermel, U.H., Peck, A., Schwartz, J., Woldeyes, R., Paraan, M., Ali, M., Hill, N.S., Siems, H., Serwas, D., Cheng, A., Kimanius, D., Agard, D.A., Potter, C.S., Carragher, B., Yu, Y.: Evaluating the Volta phase plate for improved tomogram alignment in cryo-electron tomography. *IUCrJ* **13**(3), 260–272 (2026) <https://doi.org/10.1107/S2052252526002575>

Methods

Octopi

Target Generation. Training labels are represented as multi-class volumes. For each annotated particle, voxels within a sphere centered on the annotation are assigned to the corresponding particle class channel, with the sphere radius set to 80% of the configured particle radius. This reduced radius concentrates supervision on the particle interior and limits label overlap between neighboring particles in crowded tomograms. Continuous annotations, such as membranes or vesicles, can be encoded as additional channels and concatenated with the particle-localization channels.

Network and Training. Octopi uses a 3D U-Net implemented with MONAI primitives [19]. The architecture is configurable in encoder depth, feature-map width, downsampling stride, and the number of residual units at each level. Input tomograms are standardized to zero mean and unit variance, then sampled as 3D patches using class-balanced cropping that enriches patches for labeled particles, mitigating the strong foreground–background imbalance typical of cryoET data. Online augmentation includes random 3D rotations, axis flips, intensity jitter, and additive Gaussian noise. Networks are trained with AdamW for up to 1,000 epochs using Tversky, focal, or combined focal–Tversky loss.

Validation is performed voxel-wise: the candidate model runs inference over the entire validation volume, and the per-class predictions are compared against the ground-truth target masks across all voxels in the held-out set to compute precision, recall, and the F_β score. The model checkpoint achieving the highest validation score is saved as the final trained model.

Bayesian architecture search. Octopi self-configures to new datasets with Optuna [20], which performs Bayesian optimization over architectural and training hyper-parameters using a Tree structured Parzen estimator sampler. The search

space spans feature-map widths, strides, residual units, learning rate, batch size, and loss function weightings. Each trial trains a candidate model under the validation procedure described above, and trials whose intermediate score falls below the running median across completed trials are pruned, concentrating compute on promising regions of the search space. The model from the trial achieving the highest validation score is retained as the final model for downstream inference.

Inference. Trained models are applied to full tomograms to generate per-class 3D segmentation masks. Predictions are averaged over up to nine test-time augmentations (TTA), consisting of 3D rotations and axis flips. When multiple cross-validation models are available, their predictions are further averaged to form an ensemble, improving robustness. Particle coordinates are then extracted from the per-class masks using a watershed-based peak detector. Candidate particles with an effective radius between 50% and 100% of the specified per-class particle radius are retained, whereas candidates outside this range are discarded as noise.

Context-aware extraction of membrane-bound picks. Where a membrane or organelle segmentation is available, each pick is assigned to a membrane-proximal set if its Euclidean distance to the nearest segmented membrane voxel falls below a user-defined threshold, and to a distal set otherwise. For membrane-proximal picks, an initial orientation is assigned from the local membrane normal, estimated as the vector from the pick to the nearest organelle centroid.

TopCup

TopCup is a reimplement of the top-ranked segmentation solution from the 2024–2025 Kaggle CryoET Object Identification Challenge. The method employs an ensemble of 3D segmentation models based on the MONAI FlexibleUNet architecture, using ResNet34 and EfficientNet-B3 encoders. Training is performed using a multi-scale Dense Cross-Entropy loss with class-specific weighting to address class imbalance.

The network predicts per-class heatmaps from the decoder output, which are subsequently converted into particle coordinates using a non-maximum suppression (NMS)-based localization procedure. The model is trained primarily on annotated particle regions and relies on heavy data augmentation to improve generalization. These augmentations include random flips along all three spatial axes, random cropping, rotations, and mixup augmentation.

During inference, predictions from multiple models are ensembled and combined with TTA based on flip transformations. The resulting per-class heatmaps are averaged and postprocessed using NMS to produce the final particle coordinates.

py2rely

py2rely couples particle coordinates and aligned tilt series to create a final reconstruction through a fixed sequence of RELION-5 jobs (Extended Data Fig. 9). The pipeline can import coordinates saved in Copick, or accept previously formatted STAR-file inputs for RELION. py2rely proceeds through a binning cascade within a user-defined

range to unbinned, the full tilt series resolution with subtomogram extraction, 3D refinement, and post-processing performed at each level.

Initial Model Generation. Initial models are generated through one of three strategies depending on available prior information. When a reference structure is available, the reference is downsampled and low-pass filtered to match the first binning level; symmetric targets are additionally aligned to the symmetry axis to avoid biasing the initial orientation distribution. When no reference is available, RELION’s ab initio stochastic gradient descent is run on a subset of the input particles to produce a de novo initial model. When particle orientations are already known from a previous STA run, or from template matching estimates, the initial model is reconstructed directly from those particles without further searching.

Across this study, reference-based initialization was used for the Kaggle challenge and ab-initio initialize was used for the E. coli and V-ATPase sample structures.

AutoClass3D. py2rely optionally performs 3D classification without alignment after the first round of refinement to separate the dominant homogeneous population from these contaminants. When two classes are used, the class with the best estimated resolution is selected automatically and propagated to subsequent binning stages without user intervention; RELION’s estimated translation accuracy serves as a tiebreaker when estimated resolutions are identical. For higher class counts, the pipeline pauses after classification, and the user selects which classes to keep; the selected particles are combined into a single particle set and refinement resumes. AutoClass3D improved final resolution on the Phantom benchmark (Extended Data Fig. 10).

Automatic Mask Generation. Soft masks for refinement and FSC calculation are generated from the current reconstruction at each binning level without user intervention. At each stage, the merged half-map is low-pass filtered in Fourier space at a slightly lower resolution reported from the previous refinement job. The isocontour threshold passed to RELION’s MaskCreate is then set to the 98.5th percentile of the filtered volume’s voxel intensity distribution.

Iterative Polishing. After the final binning stage, py2rely enters an iterative polishing loop that alternates RELION’s per-particle, per-tilt CTF refinement and per-particle Bayesian polishing (beam-induced motion estimation). We measure the map resolution at the end of every iteration and terminate after three successive iterations without improvement. This loop typically converges in three to four iterations.

Minislab Curation. Coordinates produced by any of the three pickers can optionally be screened prior to subtomogram averaging using a slab-projection workflow. For each candidate coordinate, a thin tomographic slab centered on the particle is extracted from the reconstructed volume and projected along the z axis to yield a 2D particle image. The projected slabs are assembled into a stack and passed to RELION-5 reference-free 2D classification with a user-defined class count, mirroring the class-averaging strategy used in single-particle cryo-EM.

Classes displaying recognizable structural features are retained, and classes dominated by noise, membrane edges, ice, or carbon are discarded. The retained particle subset is forwarded to py2rely for 3D averaging at the binning level specified in the configuration. The step typically completes within minutes per dataset and is independent of the picking method, accepting Copick or STAR-formatted inputs. We used this workflow to clean the manual-annotation GroEL coordinates in Extended Data

Fig. 7; for the template-matched GroEL coordinates and all other targets in this study, py2rely was run directly without an intermediate slab-projection step.

Kaggle Benchmarking

We benchmarked Octopi, TopCup, and the top-ranking challenge entry coordinates on the Kaggle Phantom dataset [9, 13]. The dataset is partitioned into training, public test, and private test splits matching the original Kaggle challenge, and ground-truth annotations on the private split are accessed through the CryoET Data Portal under deposition CZCDP-10310.

Coordinate-overlap performance was scored using the F_β metric adopted by the challenge,

$$F_\beta = \frac{(1 + \beta^2) \text{precision} \cdot \text{recall}}{\beta^2 \cdot \text{precision} + \text{recall}},$$

with $\beta = 4$, which weights recall approximately sixteen times more heavily than precision and reflects the practical observation that subtomogram averaging tolerates moderate over-picking but is limited by the number of true positives recovered. A predicted coordinate counts as a true positive when its distance to a ground-truth annotation falls within the per-target tolerance defined by the challenge. Octopi and TopCup were trained on the seven training tomograms and evaluated on the private test set without challenge-specific tuning; the top-ranking challenge entry coordinates were taken from the CryoET Data Portal under deposition CZCDP-10319.

For the structural benchmark, all three coordinate sets were processed through identical py2rely workflows (Extended Data Fig. 3).

Ribosome Structure Determination in Unroofed Cells

Training annotations for cytosolic and membrane-bound ribosomes were drawn from two tomograms in dataset CZCDP-10426 (HEK293 cells), with one tomogram held out for validation. Plasma membrane segmentations were generated with MemBrain-seg [21] and provided as an additional dense target channel alongside the two ribosome classes during Octopi training.

The trained model simultaneously predicted the two ribosome states and plasma membranes on the remaining unroofed-cell datasets (CZCDP-10427 through CZCDP-10435) comprised of HEK293 and HSC3 preparations. Tilt series and predicted coordinates were imported from the CryoET Data Portal via the py2rely API and used for subtomogram averaging. Each target class was processed independently through the full py2rely pipeline. AutoClass3D was enabled for both runs, with two classes selected automatically at the first refinement stage.

Synaptic Vesicle V-ATPase Structure Determination

Tilt series from dataset CZCDP-10453/EMPIAR-12794 were aligned and reconstructed with AreTomo3 [22], and the resulting tomograms were denoised and missing-wedge-corrected with IsoNet2 [23]. The improved contrast of the denoised tomograms facilitated manual curation of a training set of 157 V-ATPase particles

across 11 tomograms. Synaptic-vesicle membranes were segmented with MemBrain-seg [21] and supplied as a dense membrane channel alongside the V-ATPase particle class during training. The picking model was trained in Octopi, using Optuna-based search to optimize the U-Net architecture and training hyperparameters. Applying the trained model to the denoised tomograms yielded 2,082 candidate V-ATPase coordinates from 95 tomograms.

These coordinates were passed to py2rely for subtomogram averaging. A *de novo* initial model was generated and used as the reference for 3D refinement at bin 4 (6.7 Å/pixel). 3D classification with a soft-edged mask covering the V₁ head identified a best-resolved class of 1,135 particles, refinement of which yielded a reconstruction at 22 Å. py2rely then applied per-particle CTF refinement and Bayesian motion polishing to improve the resolution to 19.5 Å. Soft solvent masks were generated computationally as raised-cosine envelopes centered on the complex: a sphere enclosing the V₁ head for classification and initial refinement, and an ellipsoid enclosing the V₁ and V₀ domains for final analysis.

GroEL in *E. coli*

GroEL coordinates in intact *E. coli* tomograms were obtained via two parallel workflows to compare manual and template-matched annotation strategies (Extended Data Fig. 7).

60 manually selected coordinates were used to train an Octopi model, which returned approximately 11,100 candidate picks. Because GroEL is difficult to annotate manually, only a restricted subset of high-contrast views is recognizable in tomographic slices. The initial model therefore over-picked substantially, and candidates were screened using Minislab curation [13]: thin slab projections were assembled into a stack and subjected to reference-free 2D classification in RELION-5, and 3 classes displaying recognizable GroEL features were retained before passing coordinates to py2rely.

In the template-matched workflow, the coordinates generated by template matching with pytom-match-pick [24] were used to train Octopi models across a range of seed sizes (5 to 600 picks), without any subsequent curation steps. For both workflows, py2rely used *de-novo* initial model generation without 3D classification.

POPSICLE Benchmark

POPSICLE (Particle/Object Picking & Segmentation In CryoET Learning & Evaluation) is a unified benchmark for cryoET segmentation and macromolecular localization [8]. It comprises four annotated datasets spanning prokaryotic and eukaryotic systems: bacterial and yeast tomogram sets for dense compartment segmentation, and the phantom dataset for sparse macromolecular localization. Octopi was trained and evaluated on each task using the benchmark’s standard data splits and evaluation protocols (Dice score for segmentation; recall-weighted F_β for localization) alongside nnU-Net [25] and SwinUNETR [26]. Segmentation results are shown in Extended Data Fig. 6. Across both tasks, Octopi achieved competitive performance relative to the benchmark architectures.

CryoET Data Portal Integration

Tomograms, tilt series, and ground-truth annotations were accessed through the CryoET Data Portal via the Copick API [12]. All Octopi training annotations, predicted coordinates, and reconstructed maps generated in this study are deposited under the deposition IDs listed in Data Availability. The shared Copick API allowed Octopi, TopCup, and py2rely to operate on the same metadata schema without conversion. All 3D renderings and volumetric visualizations were generated using UCSF ChimeraX with the Copick plugin [27, 28].

Software

The core Octopi, TopCup and py2rely functionality is implemented in Python. Deep learning components are implemented using PyTorch [29]. Numerical operators rely on NumPy and SciPy [30, 31]. Tomogram accessibility is enabled through Copick [12].

Data Availability

All tomograms, tilt series, and ground-truth annotations used in this study are publicly available on the CryoET Data Portal. The dataset IDs are 10426-10435 for Unroofed Cells, 10453 for the Synaptic Vesicles, 10359 and 10494 for the E. coli, and 10440, 10445, and 10446 for the Phantom Dataset. The octopi and TopCup predicted coordinates are deposited under the deposition ID 10358.

Code Availability

The open-source code for Octopi, TopCup, py2rely are openly available under the MIT License at <https://github.com/chanzuckerberg/py2rely>, https://github.com/czimginstitute/czii-cryoet_mlchallenge-winning_models, and <https://github.com/chanzuckerberg/octopi> respectively. The packages are distributed via the Python Package Index (PyPI) and can be installed using `pip install`.

Acknowledgments

We sincerely thank the HPC team at Biohub for computational support, and Liz Montabana and Mykhailo Kopylov for assistance with data collection. We also thank members from the CCPEM team including Matthew Iadanza, Colin Palmer, and Tom Burnley for early discussions with py2rely. This work was supported by Biohub and its donors, Priscilla Chan and Mark Zuckerberg.

Author Contributions

J.S. designed and implemented Octopi and py2rely. C.H. and Z.Z. designed and implemented TopCup. D.J. and U.E. assisted with the implementation of octopi. D.J., U.E., R.W., and J.H. assisted with the design and implementation of py2rely. M.P. and M.A. collected and processed the in situ E. coli datasets. R.W. processed the synaptic

vesicle dataset. Y.Y. collected and assisted in processing the VPP dataset. All authors contributed to writing and editing the manuscript.

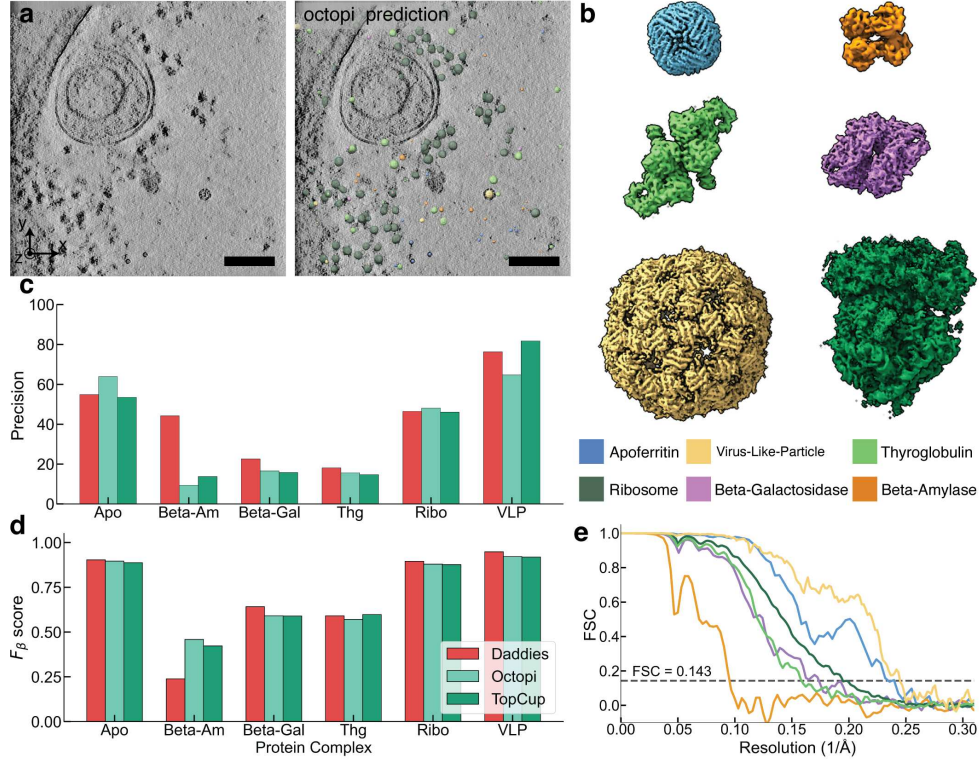
Competing Interests

The authors declare no competing interests.

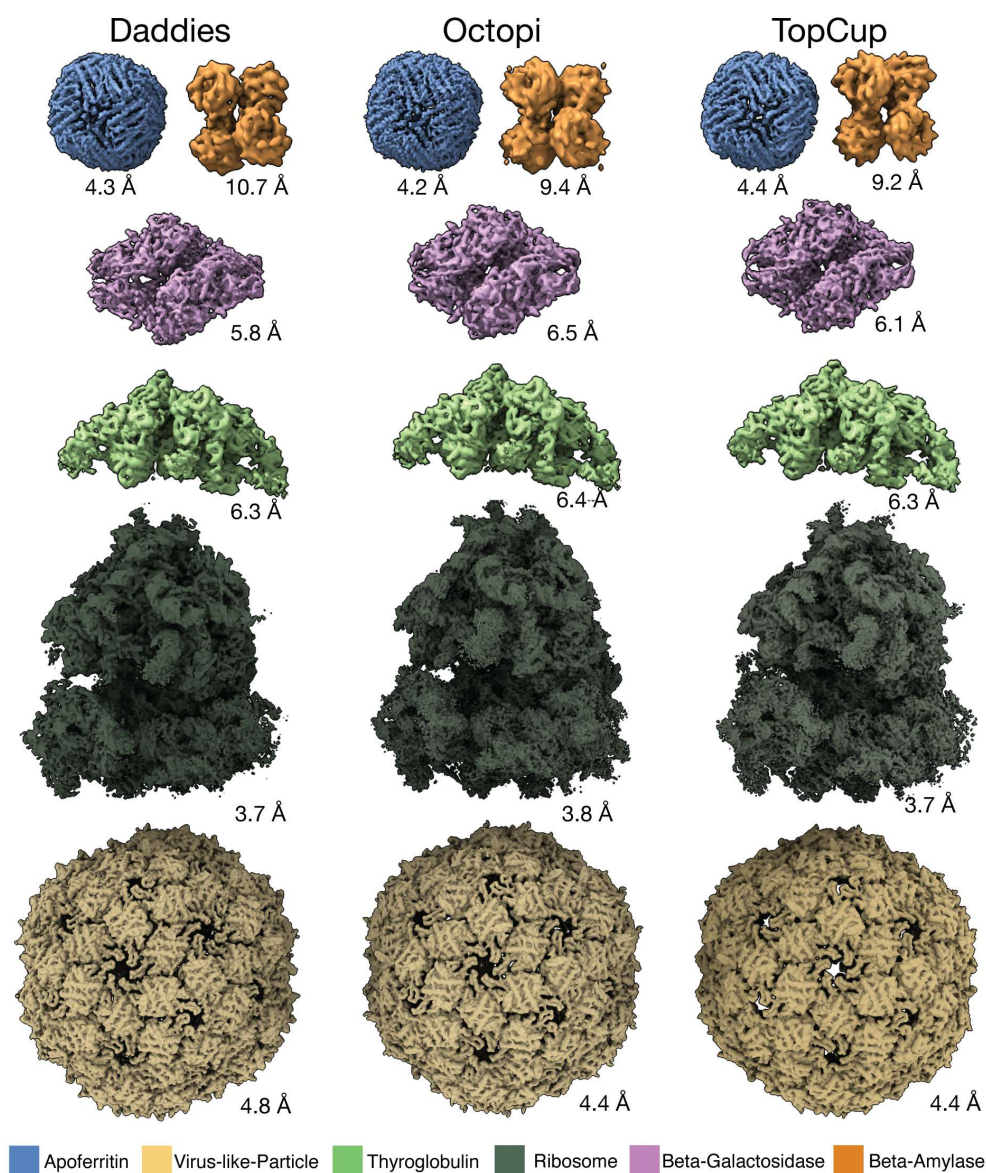
Extended Data

Extended Data Table 1 Summary of cryoET datasets and structures. For each biological target, the number of annotated coordinates and tomograms used to train the Octopi model is listed alongside the total picks returned by inference and the particle count contributing to the final subtomogram average. Rows marked Volta phase plate (VPP) and Non-VPP denote datasets acquired with and without a Volta phase plate, respectively [32].

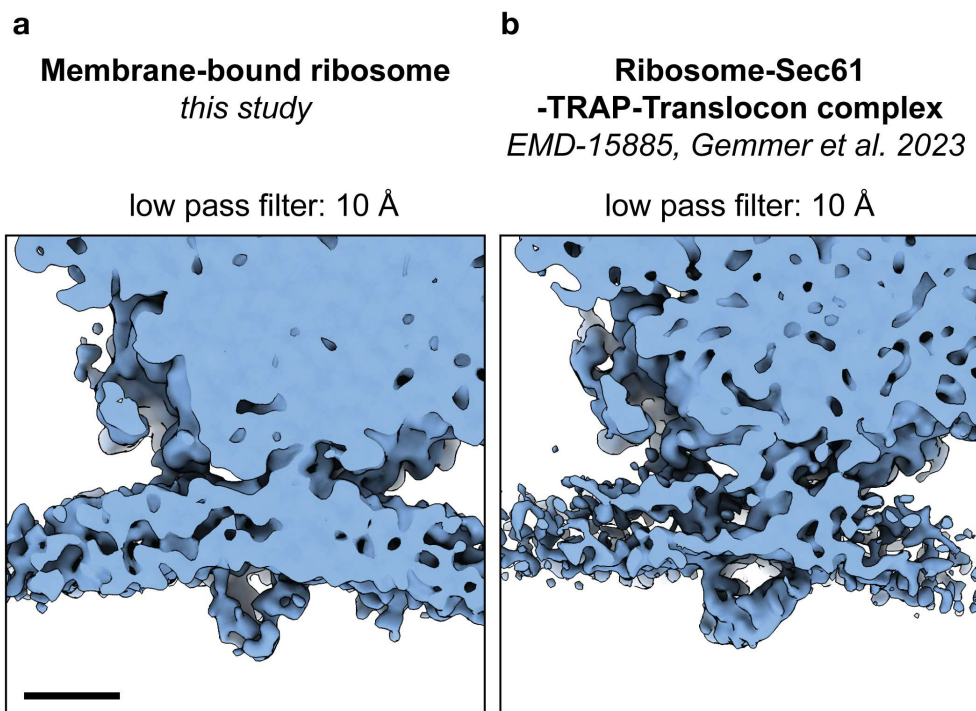
Structure	Dataset	# Train Picks / Tomograms	# Picks Found	# Picks in Final Structure	Resolution (Å)
Membrane-Ribosome	Unroofed Cells	203 / 2	53,369	20,179	10.9
Cytosolic Ribosome	Unroofed Cells	612 / 2	3,653	1,734	6.8
V-ATPase	Synaptic Vesicles	157 / 11	2,082	1,135	19.5
VLP	E. Coli	1,500 / 3	131,000	131,000	3.0
VLP	E. Coli	1,500 / 3	57,800	31,400	3.9
VLP	E. Coli	1,500 / 3	122,000	40,000	3.2
GroEL	E. Coli	50 / 12	845	845	7.4
Apo ferritin	Phantom	375 / 7	34,941	34,941	4.2
Beta-Amylase	Phantom	83 / 7	17,976	7,509	9.4
Beta-Galactosidase	Phantom	103 / 7	15,456	5,375	6.5
80S-Ribosome	Phantom	331 / 7	50,836	33,563	3.8
Thyroglobulin	Phantom	244 / 7	33,865	12,371	6.4
VLP	Phantom	113 / 7	4,768	4,768	4.4
80S-Ribosome	VPP	159 / 3	8996	3710	6.8
80S-Ribosome	Non-VPP	225 / 4	15643	11479	4.8
VLP	VPP	25 / 1	490	290	7.0
VLP	Non-VPP	22 / 14	514	323	4.8



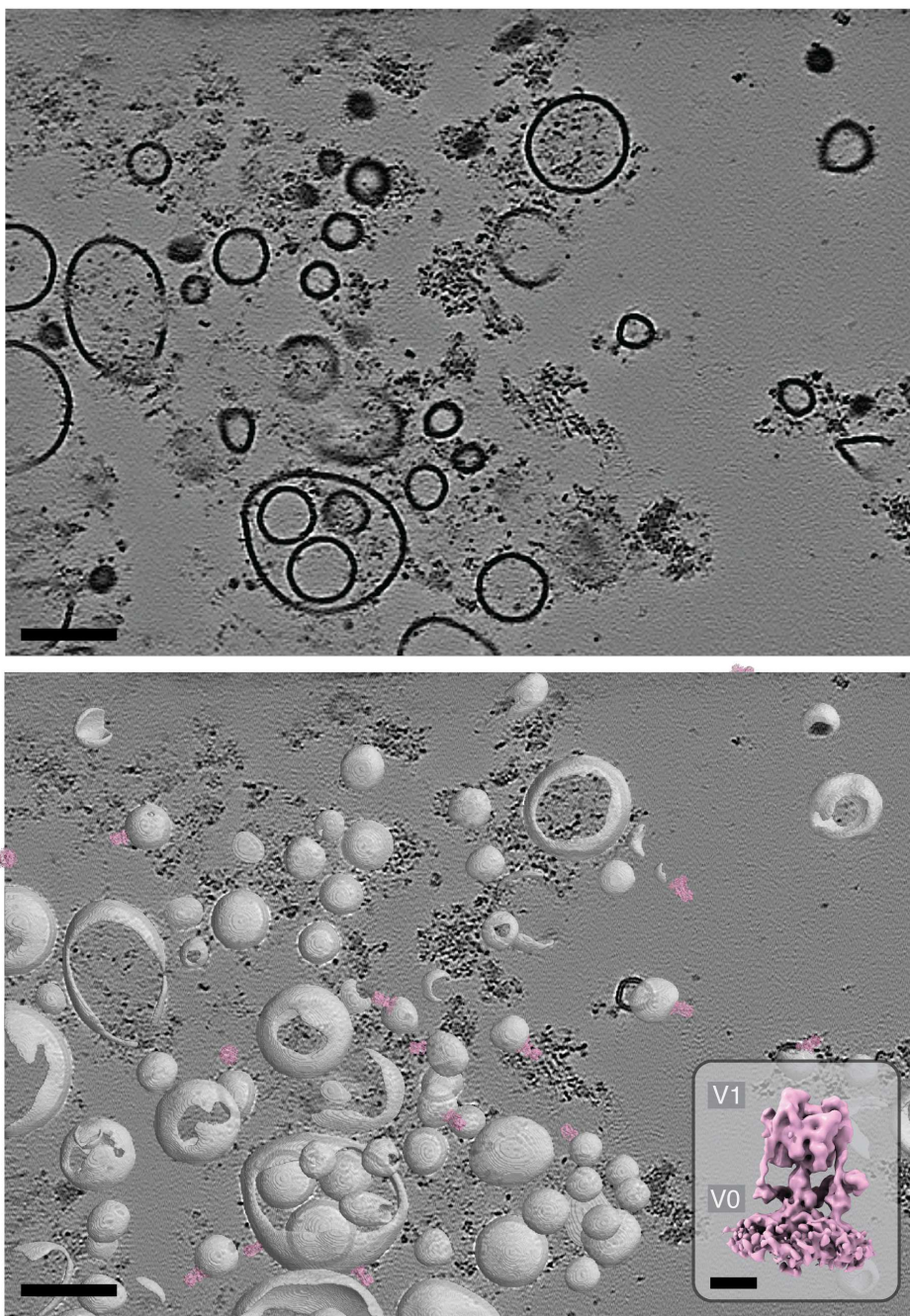
Extended Data Fig. 2 Benchmark evaluation on the Kaggle Phantom Dataset. **a**, Representative tomographic slice (left) and the corresponding Octopi prediction (right), with raw predicted coordinates shown as colored spheres corresponding to the six particle species. Scale bar, 100 nm. **b**, Reconstructions for the six benchmark targets: apoferritin, beta-amyase, beta-galactosidase, ribosome, thyroglobulin, and virus-like particle. **c**, Precision of particle localization for Daddies, Octopi, and TopCup, defined as the fraction of predicted coordinates overlapping ground-truth annotations. **d**, F_β score for the same three models, with β weighted toward recall. **e**, Fourier shell correlation (FSC) curves for each target's subtomogram average obtained by py2rely using octopi's particle coordinates.



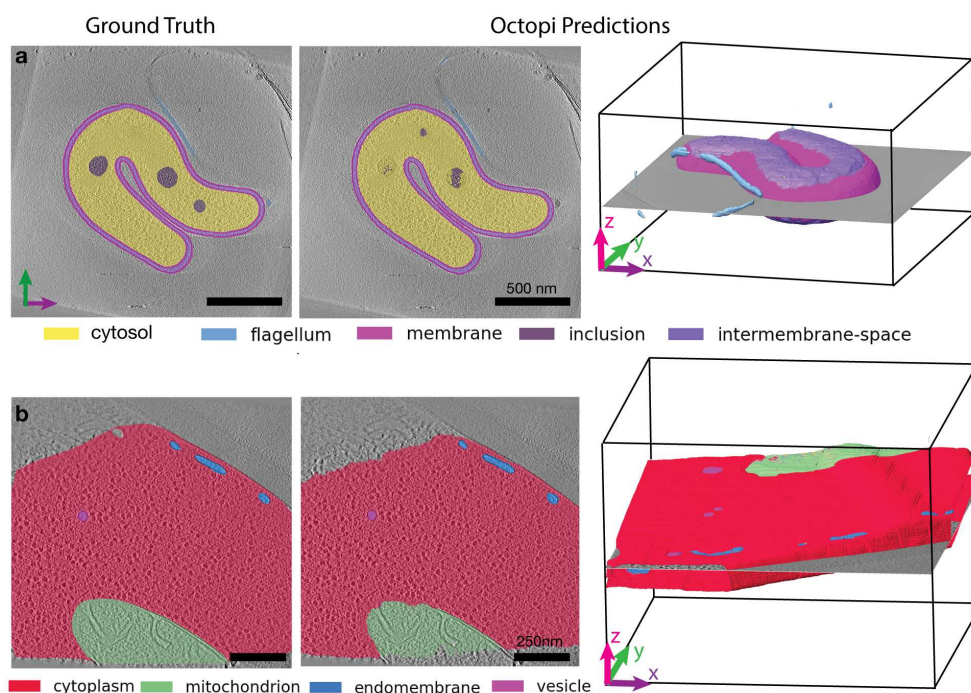
Extended Data Fig. 3 Subtomogram averages obtained by the top-ranking challenge entry (Daddies), TopCup, and Octopi in the Kaggle Phantom Challenge. Reconstructions for the Phantom challenge targets: apo-ferritin, beta-amylase, beta-galactosidase, thyroglobulin, ribosome, and virus-like-particle generated from coordinates produced by the Daddies model, Octopi and TopCup. All maps were refined using the same py2rely workflow. Reported resolutions are shown on the bottom right of each structure.



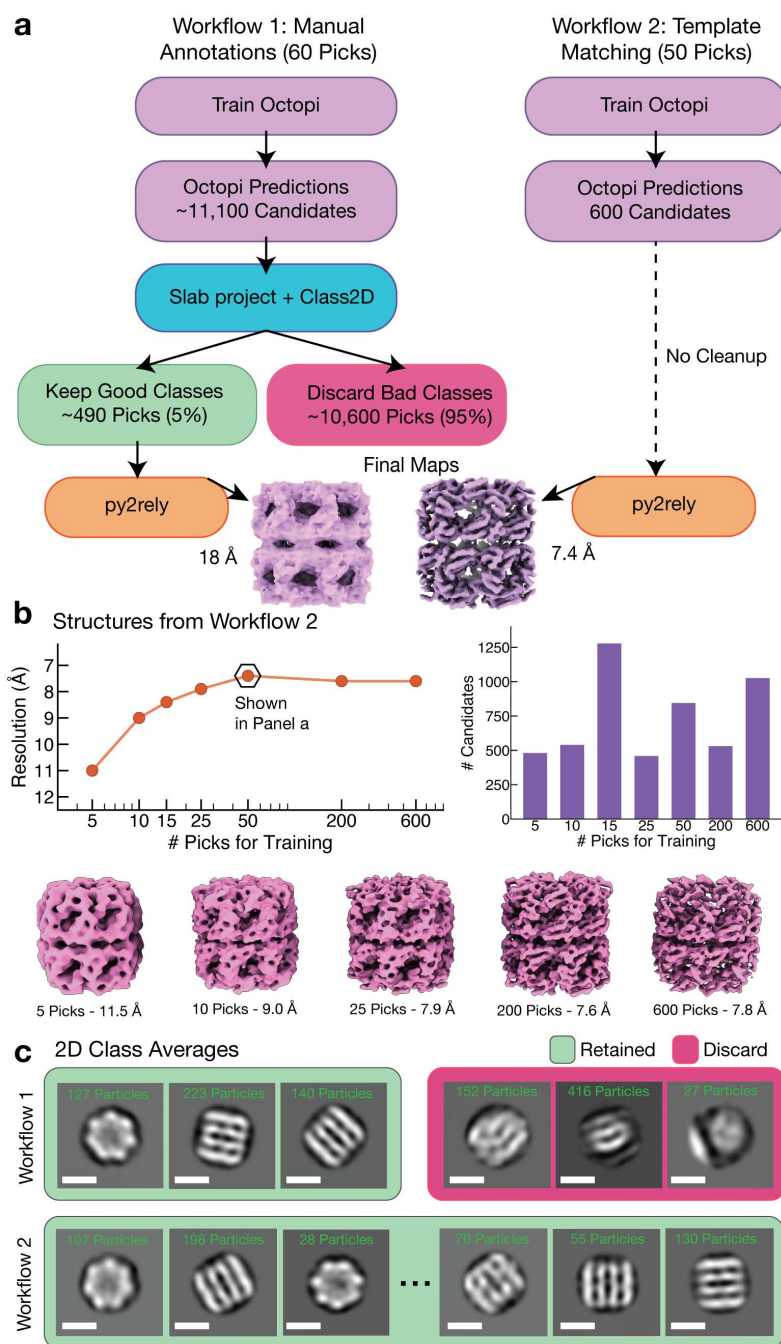
Extended Data Fig. 4 Comparison of membrane-associated density to a published translocon map. a, Close-up view of the membrane-associated density from the membrane-bound ribosome subtomogram average obtained in this study, low-pass filtered to 10 Å. Scale bar, 5 nm. **b** Equivalent view of the Ribosome-Sec61-TRAP translocon complex (EMD-15885) [15], low-pass filtered to 10 Å for comparison.



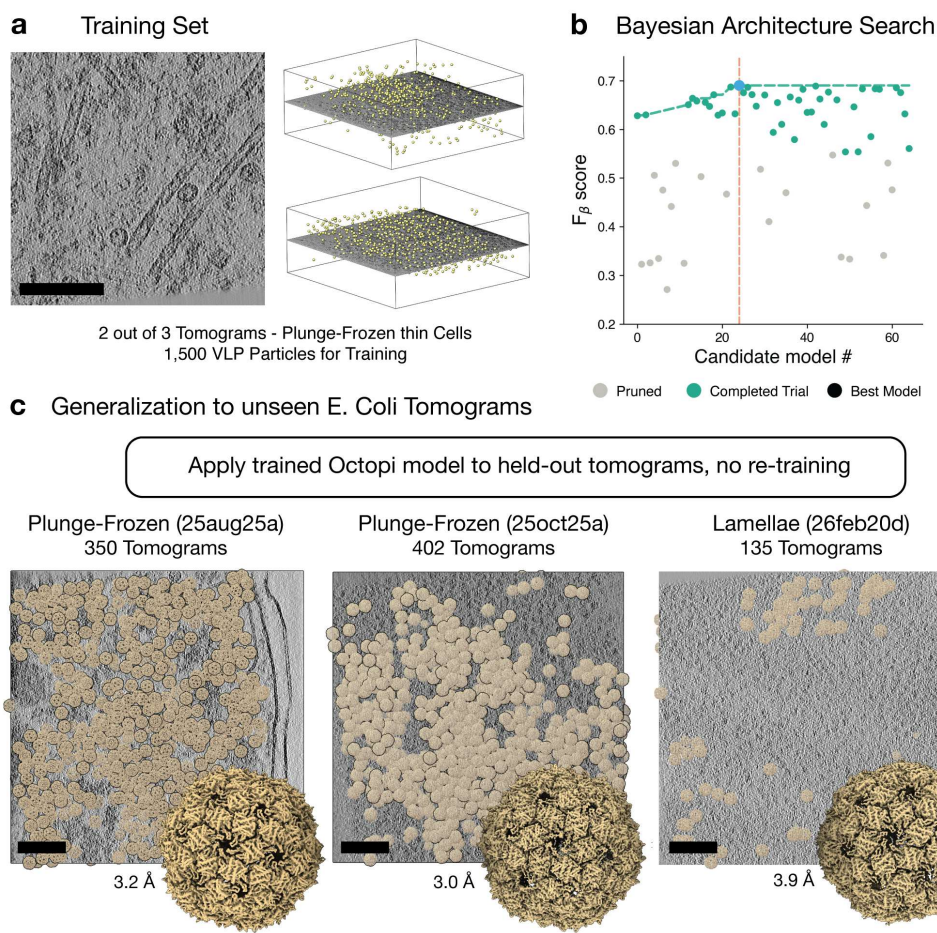
Extended Data Fig. 5 Octopi applied to V-ATPase in synaptic vesicles. **a**, Representative tomogram slice from the synaptic vesicle dataset showing Octopi-predicted V-ATPase coordinates (pink) overlaid on segmented vesicle membranes. Scale bars, 100 nm (main panel), 5 nm (inset).



Extended Data Fig. 6 Demonstration of Octopi for Organelle Segmentation. Octopi was trained and evaluated on the bacterial and yeast segmentation tasks of the POPSICLE benchmark [8] using expert annotations on the CryoET Data Portal (depositions CZCDP-10350 and CZCDP-10351). **a**, Ground truth (left) and Octopi predictions (middle) for a bacterial tomogram, with the corresponding 3D rendering of the predicted segmentation (right). Scale bar, 500 nm. **b**, Ground truth (left), Octopi predictions (middle), and 3D rendering (right) for a yeast tomogram annotated for four organellar classes: cytoplasm (red), mitochondrion (green), endomembrane (blue), and vesicle (orange). Scale bar, 250 nm. Although Octopi was originally developed for sparse macromolecular localization, the same self-configuring architecture search recovers competitive dense segmentation across both samples.

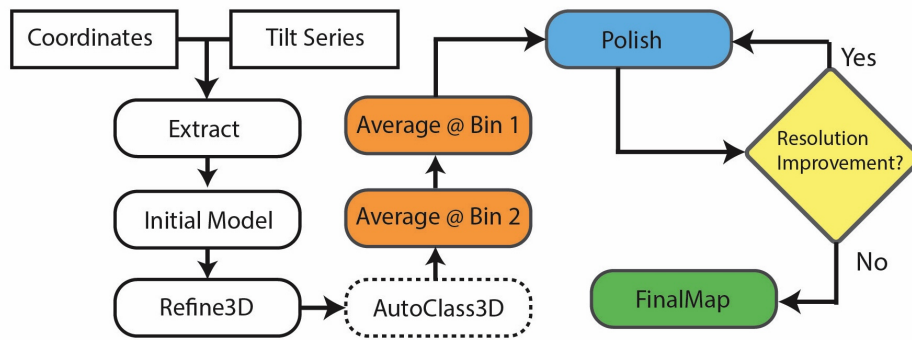


Extended Data Fig. 7 Comparison of Training Octopi with Manual and Template Matched Annotations for GroEL. **a**, Workflow comparison for Octopi trained on manually selected coordinates (left) versus template-matched coordinates (right). See Methods for details. **b**, GroEL reconstruction resolution as a function of seed annotation count from template matching, with representative maps shown below. **c**, 2D class averages from slab projections for both workflows, with retained (green) and discarded (red) classes. Scale bars, 10 nm.



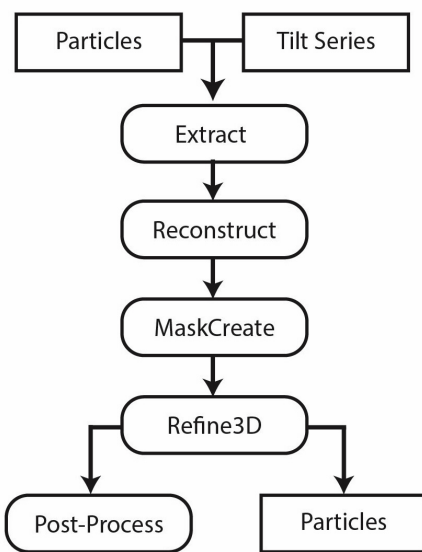
Extended Data Fig. 8 A single Octopi model generalizes across independent E. coli datasets. **a**, Training set used to fit the Octopi model: showing two out of three tomograms from a single plunge-frozen E. coli preparation. **b**, Bayesian architecture search over the Octopi hyperparameter space identified the model configuration used for all downstream predictions. **c**, Generalization of the trained Octopi model to three held-out E. coli datasets not seen during training: two independent plunge-frozen collections (25aug25a, 350 tomograms; 25oct25a, 402 tomograms) and a cryo-FIB lamellae dataset (26feb20d, 135 tomograms). For each dataset, a representative tomogram slice is shown with predicted VLP coordinates overlaid (tan), alongside the subtomogram average obtained by applying py2rely. Scale bars: 100 nm.

a Main STA Pipeline



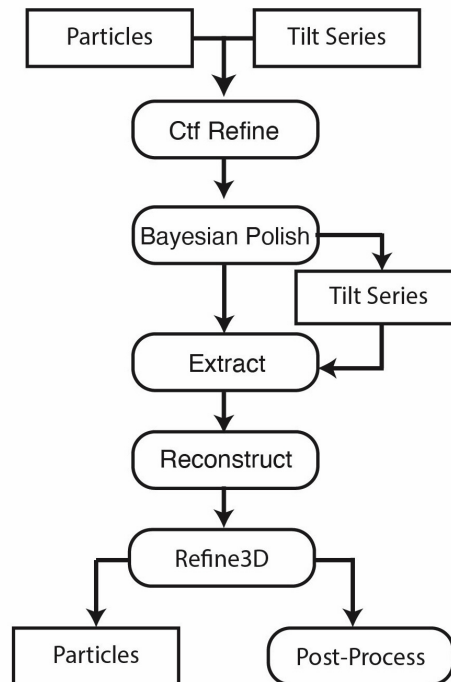
b

Averaging Block

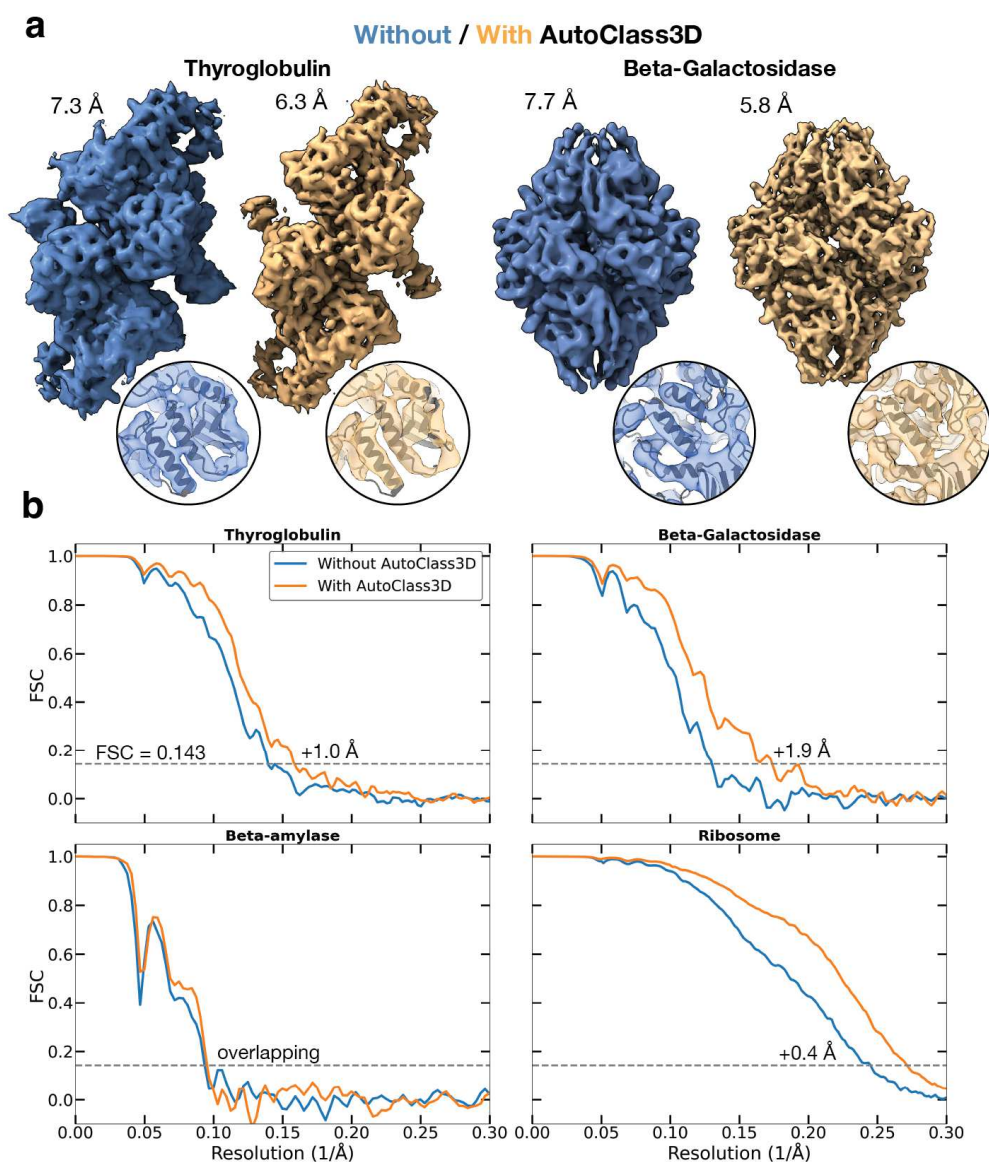


c

Polishing Block



Extended Data Fig. 9 The py2rely STA Workflow. **a**, Main STA pipeline. Particle coordinates and tilt series enter an iterative loop in which subtomograms are extracted, an initial model is generated, and 3D refinement is performed at progressively finer binning. Each averaging block (orange) invokes the procedure shown in **b**. An optional Class3D step (dashed outline) selects a homogeneous particle subset before the second averaging round. After the bin-1 average, particles enter a polishing block (blue) implementing the procedure in **c**. **b**, Averaging block. Particles and tilt series are used to extract subtomograms and reconstruct an average at the current binning level. **c**, Polishing block.



Extended Data Fig. 10 AutoClass3D Demonstrated on the top-ranking (Daddies) Phantom Challenge Coordinates. **a**, Representative reconstructions of thyroglobulin (left) and β -galactosidase (right) obtained without AutoClass3D (blue) and with AutoClass3D (orange). Insets show zoomed views of a central region to highlight differences in local map quality. Gold-standard FSC curves between independently refined half-maps for four particle species: thyroglobulin, β -galactosidase, β -amylase, and ribosome. Resolution improvements achieved by AutoClass3D are annotated for each species; FSC curves for β -amylase are overlapping, indicating no resolution gain for this particle.