This peer-reviewed article has been accepted for publication but not yet copyedited or typeset, and so may be subject to change during the production process. The article is considered published and may be cited using its DOI.
10.1017/S2633903X24000060

TomoNet: A streamlined cryoET software pipeline with automatic particle picking on flexible lattices

Hui Wang[1,2,3,*], Shiqing Liao[2,3], Xinye Yu[3], Jiayan Zhang[2,3], and Z. Hong Zhou[1,2,3,*]

[1]Department of Bioengineering, University of California, Los Angeles (UCLA), Los Angeles, CA 90095, USA

[2]California NanoSystems Institute, UCLA, Los Angeles, CA 90095, USA

[3]Department of Microbiology, Immunology, and Molecular Genetics, UCLA, Los Angeles, CA 90095, USA

***Corresponding author:**

Z. Hong Zhou (Hong.Zhou@UCLA.edu, 1-310-694-7527)

ORCIDs:

Hui Wang: 0000-0002-9922-7170

Shiqing Liao: 0000-0002-9723-8968

Xinye Yu: 0000-0002-1764-1141

Jiayan Zhang: 0000-0003-3602-1199

Z. Hong Zhou: 0000-0002-8373-4717

Keywords: cryogeneic electron tomography, subtomogram averaging; automatic particle picking; lattice structure; deep learning; in situ structures

ABSTRACT

Cryogenic electron tomography (cryoET) is capable of determining *in situ* biological structures of molecular complexes at near atomic resolution by averaging half a million subtomograms. While abundant complexes/particles are often clustered in arrays, precisely locating and seamlessly averaging such particles across many tomograms present major challenges. Here, we developed TomoNet, a software package with a modern graphical user interface to carry out the entire pipeline of cryoET and subtomogram averaging to achieve high resolution. TomoNet features built-in automatic particle picking and 3D classification functions and integrates commonly used packages to streamline high-resolution subtomogram averaging for structures in one-, two- or three-dimensional arrays. Automatic particle picking is accomplished in two complementary ways: one based on template matching and the other employing deep learning. TomoNet's hierarchical file organization and visual display facilitate efficient data management as required for large cryoET datasets. Applications of TomoNet to three types of datasets demonstrate its capability of efficient and accurate particle picking on flexible and imperfect lattices to obtain high-resolution 3D biological structures: virus-like particles, bacterial surface layers within cellular lamellae, and membranes decorated with nuclear egress protein complexes. These results demonstrate TomoNet's potential for broad applications to various cryoET projects targeting high-resolution in situ structures.

IMPACT STATEMENT

Cryogenic electron tomography (cryoET) has become a powerful approach to visualize organization and high-resolution structures of biological complexes in their native environment. Subtomogram averaging (STA) of hundreds of thousands of particles (i.e., subtomograms) is necessary to obtain near atomic resolution structures for each such complex. While abundant biological complexes often cluster in arrays that manifest as one to three-dimensional lattices, flexibility and imperfection of such lattices pose challenges for efficient and accurate particle

https://doi.org/10.1017/S2633903X24000060 Published online by Cambridge University Press

picking. To overcome these challenges and to meet the growing demand for efficient data processing and management in the cryoET and STA workflow, we have developed TomoNet, a user-friendly software package with a modern graphical user interface that allows users to execute the entire data processing pipeline seamlessly with integration of commonly used software packages. TomoNet addresses the particle picking challenge with two solutions, one based on geometric template matching and the other employing artificial intelligence. Applications of TomoNet to three representative datasets demonstrate its capability for high-resolution structure determination of biological complexes on flexible and imperfect lattices.

https://doi.org/10.1017/S2633903X24000060 Published online by Cambridge University Press

INTRODUCTION

Single-particle cryogenic electron microscopy (cryoEM) is employed to elucidate atomic-level structures of purified biological complexes. This methodology adheres to a standardized and well-established workflow supported by advanced software packages such as Relion[1] and cryoSparc[2]. In parallel, cryogenic electron tomography (cryoET), coupled with subtomogram averaging (STA), expands the investigative scope to encompass heterogeneous macromolecules in their native context[3-10]. To enhance the resolution of subunits within *in situ* macromolecules, subtomograms (*i.e.,* particles) are extracted from each tomogram and then subjected to 3D alignment and averaging, thereby improving signal-to-noise ratio. Notably, STA has achieved resolutions up to sub-3 Å for *in situ* structures of large cellular complexes such as ribosomes, approaching the capabilities of single-particle cryoEM methodologies[11-14].

The workflow for cryoET and STA typically involves five key components across specific software packages. In cryoET preprocessing, dose fractionated frames are collected from an electron microscope, undergo motion correction, organized, and then assembled into individual tilt series. In tomogram reconstruction, three-dimensional reconstructions are generated from those tilt series. In particle picking, particles of interest are identified and extracted from tomograms. Complexity varies based on the diverse and intricate nature of *in situ* cellular samples and their unique configurations. Many packages include their own particle picking methods, such as oversampling using a supporting geometry in Dynamo[15], template matching in emClarity[16] and machine learning in crYOLO[17]. In 3D refinement and classification, particles are iteratively classified and refined to obtain a final structure at sub-nanometer or near atomic resolution, which has been demonstrated by software packages like Relion[13,18], emClarity[16], EMAN2[4] and Warp[19]. Finally, post-processing activities include map sharpening, Fourier shell correlation (FSC) calculation, visualization by placing averaged maps back into the original tomogram, etc. Users often need to navigate between several specialized software packages

https://doi.org/10.1017/S2633903X24000060 Published online by Cambridge University Press

for optimal results, which often demands a certain level of computational proficiency that poses a barrier for many.

The method for particle picking varies on a case-by-case basis, dictated by the characteristics of *in situ* cellular samples. In the early works of STA, manual particle picking was employed, particularly when aiming for resolutions between 20-50 Å with a maximum of several hundred particles[20-22]. However, for biological samples exhibiting periodic structures, oversampling on specified geometry was leveraged to significantly reduce the labor associated with acquiring enough particles for improved resolutions. For instance, HIV virus-like particles (VLPs) adopt a hexagonal Gag protein lattice in its sphere-like configuration[23]. Other examples include the Marburg Virus[24], Herpes simplex virus[25], and the Coat protein complex II[26], all of which contain lattice-like arrangements with repeating subunits that could benefit from particle picking automation when performing cryoET data processing. With an increasing demand for automation to enhance efficiency with minimal manual intervention, template matching has emerged as a popular method for automatic particle picking, relying on a user-provided reference map[16,27]. Simultaneously, convolutional neural networks have shown promising results for cryoET automatic particle picking given its capacity to analyze three-dimensional feature maps and autonomously identify prominent features within specific samples[28-31]. These machine learning approaches typically operate template-free and often obviate the need for human annotation[32].

The expanding array of specialized software tools designed for specific tasks posts a critical need for seamless software integration within the cryoET workflow. Transitioning between various software packages can be a cumbersome process. Remarkably, recent initiatives have made notable progress in tackling this integration challenge. For example, TomoBEAR[33] offers an integrated solution, while ScipionTomo[34] and nextPYP[35] provide a comprehensive web-based platform for managing various tasks in the cryoET pipeline. Notably,

https://doi.org/10.1017/S2633903X24000060 Published online by Cambridge University Press

none of these packages takes specific advantage of the fact that abundant complexes exist in arrays of some sort, albeit with imperfections, variability, or flexibility[36-41].

In this context, we have developed TomoNet, a software package designed for streamlining the cryoET and STA data processing workflow, with a modern GUI (Figure 1 and Figure 2). Our methodology employs a geometric template matching approach rooted in the concept of "Auto Expansion" which serves as a general particle picking solution for biological complexes organized in flexible, variable, or imperfect arrays. TomoNet is also powered by a deep learning-based solution to automate particle picking, which only needs 1-3 tomograms with known particle locations as ground truth for model training. Importantly, while TomoNet is particularly powerful for locating and averaging particles arranged on flexible or imperfect lattices, it can be applied to a broader range of particle types, offering a more generalizable trained model. These methods significantly diminish the need for manual inputs, and their outcomes can be seamlessly imported into Relion for subsequent high-resolution 3D classifications and refinements. We demonstrate the capabilities of TomoNet by applying it to three datasets with distinct protein lattice types, highlighting its accuracy and efficiency in identifying particles across diverse scenarios.

RESULTS

Overall design of TomoNet

TomoNet is a Python-based software package that integrates commonly used cryoET packages to streamline the cryoET and STA pipeline, with a particular emphasis on automating particle picking of lattice-configured structures and cryoET project management. As shown in the main menu and the entire TomoNet pipeline (Figure 1 and Figure 2), after data collection from electron microscopy, TomoNet can perform motion correction with integration of MotionCorr2[42]; tilt series assembly and tomogram reconstruction with integration of IMOD[43]; CTF estimation with integration of CTFFIND4[44]; manual particle picking with IMOD; particle picking using built-in

https://doi.org/10.1017/S2633903X24000060 Published online by Cambridge University Press

geometric template matching-based algorithms with integration of PEET[45]; automatic particle picking using built-in deep learning-based algorithms; 3D classification/particle cleaning and subtomograms placing back with built-in algorithms. This design also allows on-the-fly tomogram reconstruction processing during data collection, which facilitates a quick quality check. TomoNet generates particle picking results in STAR format[46], which can be incorporated into Relion for high-resolution 3D refinement. It can also read Relion results in STAR format for particle cleaning and subtomograms placing back (Figure 1).

Particle picking with "Auto Expansion"

The "Auto Expansion" module is based on template matching and uses cross-correlation coefficient as a selection criterion, with a design to pick particles on flexible lattices with minimal manual inputs; its basic concept is elucidated in Figure 3. These particles exist in array-like configurations and manifest as flexible, partial, and imperfect lattices in one, two and three dimensions (1-3D). Examples are abound: microtubule doublets, ubiquitous in most cells, consist of 96 nm axonemal 1D translational repeat units[21,47] (1D rotational lattice); HIV VLPs[40] and surface layer (S-layer) lattice of prokaryotic cells[48,49] are composed of hexametric subunits (2D lattice); paraflagellar rod of protozoan species is organized into para-crystalline arrays in its distal zone[20] (3D lattice). In TomoNet, each of these isolated lattice densities is called a patch, within which all subunits of the complex are connected. For instance, Figure 3 illustrates two patches with different sizes.

"Auto Expansion" is an iterative process; each iteration expands the particle set by adding more unpicked ones. To initiate "Auto Expansion", users need to prepare a few "seed" particles that sparsely distribute across all observed patches. Typically, the numbers of such "seed" particles per tomogram range from 20 to 200, which depends on the number and size of patches in the input tomogram. Then, "Auto Expansion" iteratively expands the "seed" particle set to a final particle set that contains all particles on given flexible lattices, following three steps

https://doi.org/10.1017/S2633903X24000060 Published online by Cambridge University Press

for each iteration (Figure 3). Firstly, potential particles adjacent to each “seed” particle are calculated and selected as “candidate” particles. Secondly, these “candidate” particles undergo alignments to a user-provided reference and are evaluated based on cross-correlation coefficient, such that “wrong” particles with low cross-correlations are excluded. Thirdly, qualified “candidate” particles are added to the particle set and become “seed” particles for the next iteration. During this process, only unpicked ones can be considered as “candidate” particles, and “Auto Expansion” stops either when no “candidate” particles are detected or when the user-defined maximum iteration number is reached. Doing this allows for an exhaustive exploration of particles on given lattices following their assembly topology with no restriction on geometry and outputs a final particle picking result (Figure 2).

Compared with conventional template matching methods, “Auto Expansion” incorporates prior knowledge of lattice configuration to iteratively guide the search for “candidate” particles, *i.e.*, unpicked particles following user-defined paths, as detailed in the Method section and TomoNet’s user manual. Thus, “Auto Expansion” significantly reduces computational complexity by searching in the regions of interest only, with restricted angular and translational search ranges defined by users. As a result, it reduces the number of incorrectly picked particles. Notably, “Auto Expansion” potentially works for any flexible, imperfect, or variable lattices in 1D, 2D and 3D and has no intrinsic size limit of subunits.

Automatic particle picking by deep learning

The “AI AutoPicking” module is designed for automatic particle picking using supervised machine learning, which employs a U-net convolutional neural network for model training. There are three main steps in “AI AutoPicking”: training data preparation, neural network training, and particle coordinate prediction, as detailed in the Method section (Figure 4). It only requires an input training dataset consisting of 1-3 tomograms paired with their corresponding particles coordinate files. The trained model can then be applied on the entire tomography dataset and output predicted particles for each tomogram.

https://doi.org/10.1017/S2633903X24000060 Published online by Cambridge University Press

Essentially, the neural network in “AI AutoPicking” is trained as a voxel-wise binary classifier, which determines whether a voxel in density maps is part of a particle (Figure 4b). To prepare for training, data pairs (ground truth) consist of extracted subtomograms coupled with their associated segmentation maps, within where each particle is labeled by a cube near its center (Figure 4a). The trained neural network model can be applied on other tomograms to perform particle segmentation. Finally, the particles coordinate information can be retrieved from the predicted segmentation maps (Figure 4c).

3D classification using TomoNet

In addition to the above two commentary modules for particle picking, TomoNet allows users to eliminate “bad” particles based on user-defined geometric constraints, which could serve as 3D classification during high-resolution particle refinements. Lattice variation in cryoET data has multiple plausible causes. Biologically, particles may be incomplete near the lattice edge due to paused biology assembly process[50]. Experimentally, lattices tend to become flattened near the air-water interface of the sample during imaging. These variabilities pose challenges for 3D classification in the process of high-resolution STA, making it difficult to exclude “bad” particles that exhibit unexpected coordinates and orientations assignment as subunits of lattices (Supplementary Movie 1).

Removing these “bad” particles is necessary for achieving better resolutions[51]. To accomplish this, TomoNet assesses each particle by counting its neighboring particles and calculating the averaged tilt angle to these neighbors to represent local surface curvature of a lattice. TomoNet identifies particles with too few neighbors or large tilt angles to their neighbors as “bad” particles since they potentially deviate from the lattice configuration. This step can be integrated into high-resolution refinement in Relion, providing an alternative 3D classification method based on analyzing spatial relationships between particles.

Application to *in situ* viral protein arrays: the matrix protein lattice in HIV VLPs

https://doi.org/10.1017/S2633903X24000060 Published online by Cambridge University Press

To validate TomoNet as an integrated high-resolution cryoET and STA pipeline and an efficient particle picking tool, four tomograms were processed from the HIV-1 Gag dataset which resolved the Gag hexamer structure at 3.2 Å resolution. Motion corrected images underwent tilt series assembly, CTF estimation, and tomographic reconstruction using TomoNet. Within these tomograms, the VLP hexagonal lattice and its building blocks were observed, and some of these observed VLPs exhibited sphere-like geometry (Figure 5a).

As detailed in the Method section, a combination of "Auto Expansion" and "AI AutoPicking" was applied to the above four tomograms. The result shows that particles were readily picked on all the observed lattice patches (Figure 5b, c). Then, these picked particles were imported to Relion to perform high-resolution particle refinements, the resulting reconstruction of the Gag hexamer structure (Figure 6) looks identical to the published high-resolution structure[11,13], demonstrating particle picking accuracy and efficiency of TomoNet – capable of obtaining more particles from fewer tomograms.

Using the "3D subtomogram place back" function in TomoNet, 3D visualizations were generated to illustrate the *in situ* assembly of the VLP lattices (Figure 5d and Figure 7). All VLP lattices with various sizes and shapes were captured even with irregular shapes (Figure 7e and Supplementary Movie 2), demonstrated TomoNet's particle picking ability on flexible lattices. Lattice defects on each VLP were also identified consistent with previous studies[52], enhancing the understanding of lattice assembly mechanisms[53].

Application to cellular organelle sample: eukaryotic axoneme

We validated TomoNet's particle picking capability for one-dimensional lattices by processing one tomogram of extracted flagellum of *T. brucei*. The axoneme consists of 9 outer doublet microtubules (DMTs) and a pair of central singlet microtubules, where each DMT is a one-dimensional polymer of 96nm axonemal building blocks (Figure 8a). This typical 1-D lattice often exhibits imperfections like bends and twists, posing challenges for precise particle picking

https://doi.org/10.1017/S2633903X24000060 Published online by Cambridge University Press

(Figure 8a). Using “Auto Expansion”, TomoNet accurately picked the 96nm-spaced axonemal subunits from all DMTs, effectively adapting to lattice imperfections (Figure 8b).

Application to focused ion beam (FIB)-milled cellular sample: the S-layer lattice of prokaryotic cell

We validated TomoNet’s particle picking capability by processing one tomogram of FIB-milled *Caulobacter crescentus* cells from EMD-23622[54]. The S-layer functions as a component of the cell wall covering the cell body. Thus, its lattice geometry is typically defined by the shape of cells (Figure 9a). The pleomorphic shape of *C. crescentus* cell in variable sizes, with the low contrast shown in this tomogram, hindered locating subunits on the S-layer lattice and raised difficulty for efficient particle picking on its S-layer lattice (Figure 9a).

TomoNet overcame the above challenges by utilizing the hexagonal configuration of S-layer lattices. With minimal manual input, “Auto Expansion” picked over a thousand hexamer S-layer subunits. The binned STA result clearly reveals the S-layer inner domain, and docking previously resolved high-resolution structure[55] (EMD-10388) into it confirms the correct hexagonal distribution with well fitted major domains (Figure 9b, c). Visualization of S-layer lattices also shows that the picked particles were arranged in the expected hexagonal pattern, confirming the reliability and applicability of TomoNet as a particle picking tool (Figure 9d) and its broad application to structure determination of prokaryotic and archaeal cell walls[49,56].

Application to *in vitro* assembled arrays: nuclear egress complex (NEC) lattice

We further validated TomoNet as an integrated high-resolution STA pipeline and an efficient particle picking tool by processing samples containing NEC lattices within budded vehicles. Nuclear egress is a pivotal step in herpes virus replication, driven by NEC and responsible for translocating nascent viral particles from nucleus to cytoplasm. In our reported dataset[57], NEC heterodimers budded into large vesicles with diameters ranging from 100 nm to 500 nm, forming beehive-like lattices on the inner surface of these vesicles (Figure 10a, b). Because of their large sizes, noticeable compressions were observed during the sample freezing, reshaping the

https://doi.org/10.1017/S2633903X24000060 Published online by Cambridge University Press

vesicles and NEC lattices from spherical to flattened disk shapes (Figure 10a, b). This conformational change was a consequence of the limitation in ice thickness imposed by cryoET, which restricts the sample thickness to approximately 250 nm, consequently posing challenges for particle picking.

TomoNet successfully picked NEC hexamer subunits following the topology of lattices. The intermediate STA result generated in TomoNet already showed the six heterodimers within one hexamer subunit (Figure 10c). With these picked particles, high-resolution 3D classifications and refinements were carried out to obtain a final reconstruction of NEC hexamer subunit at 5.4 Å resolution, without preferred orientation bias (Figure 10c, d), and all the helices were well resolved (Figure 10e). Visualization of subtomograms placing back shows that the large vesicle was compressed during sample freezing which stretched the NEC lattice, making it appears flat and split at the air-water interface, while the middle part of the lattice appears to be more curved.

Application to other types of arrays and free-floating particles

The above examples show how TomoNet's ability to locate particles arrays arranged on flexible spheres (HIV), cell surfaces (S-layer) and nuclear membranes (NEC), which can be considered as topologically 2D lattices. In our published work of various cryoET structures, TomoNet has also been used to locate subtomograms arranged on flexible filaments (*i.e.*, 1D arrays) such as the flagella of *Trypanosoma brucei*[21,47] and the amyloid-like sheath protein on β-hoops of the prototypical archaeon, *Methanospirillum hungatei*[58]. In the case of 3D lattices, TomoNet has been also used to obtain the paraflagellar rod structure of *T. brucei*[20]. Since TomoNet has integrated packages and is designed for the entire cryoET and STA data processing pipeline, it can also be used as a general-purpose package for subtomogram averaging towards high resolution when particles are free floating and without local order. In the latter case, TomoNet would have the same limitation recognized for all other cryoET software packages, that is, high resolution is currently only achieved for large complexes, such as ribosomes.

https://doi.org/10.1017/S2633903X24000060 Published online by Cambridge University Press

DISCUSSION

In this paper, we report the implementation and application of TomoNet and demonstrate its efficacy in particle picking across three distinct datasets featuring particles with varying lattice configurations. TomoNet stands out as the first software to exhaustively trace lattices following its inherent topology. This unique approach ensures that the particle picking results faithfully reflect *in situ* or *in vitro* lattice shape, providing valuable insights into how these lattices are formed by their constituent subunits. For HIV VLPs, TomoNet application enabled us to directly visualize the VLPs lattices and their defects potentially caused by the absence of pentamer subunits. Similarly, for the NEC dataset, TomoNet facilitated a more direct observation of lattice conformation changes resulting from the sample freezing process. Since vesicles in this dataset were too large to be compressed from a sphere into a disk-like shape, the lattice regions near the air-water interface became stretched and subsequently divided into smaller fragments. Moreover, TomoNet demonstrated its exceptional performance, even when dealing with datasets characterized by extremely low contrast. For instance, in the cellular S-layer tomogram of a lamella, S-layer subunits were nearly imperceptible to human observations. Therefore, "Auto Expansion" excelled in particle picking without requiring denoising or contrast-enhancement algorithms.

Additionally, "AI AutoPicking", the deep learning-based module, demonstrated excellent performance on automatic particle picking, showing potential in handling a wide range of particle types even beyond those with lattice-like arrangements. Compared to the template matching-based “Auto Expansion”, “AI AutoPicking” has several advantages in particle picking. Firstly, it applies to particles situated on flexible lattices and those arranged in scattered patterns, such as cellular ribosomes. The neural network learns to pick by discerning 3D features of individual particles, and it does not require prior knowledge about lattice configuration. Secondly, it utilizes GPUs for fast convolution operations, enabling particle prediction in just several minutes for

https://doi.org/10.1017/S2633903X24000060 Published online by Cambridge University Press

each tomogram. Thirdly, it does not require the “seed” particles used in “Auto Expansion”, which further reduces human efforts by approximately 5-15 minutes per tomogram. This is especially beneficial for processing extensive tomography datasets with hundreds of tomograms. However, comparing their final output particles, “AI AutoPicking” typically picks fewer particles than “Auto Expansion” because it misses certain particles on the flexible lattices. Thus, these two modules are complementary to each other and can be incorporated to further explore these missing particles.

Regarding the pipeline design, each module within TomoNet is designed to be highly independent, ensuring flexibility for integrating future methods and third-party packages. This adaptable framework positions TomoNet as a platform of choice for other developers to build their own innovations. At present, TomoNet is primarily tailored for integration with the Relion-related pipeline. However, it can accommodate specific demands and can be extended to integrate other pipelines, including emClarity[16], EMAN2[4], M[59], and others in the future. In summary, TomoNet significantly simplifies the overall process for users in managing and monitoring every step of the complete cryoET and STA pipeline. Its user-friendly GUI design notably reduces the entry barrier for newcomers to the fast-emerging cryoET field. The particle picking modules of TomoNet provide a general solution for particles organized in lattice-like arrangements, ensuring both accuracy and efficiency, thereby facilitating the high-resolution STA pipeline.

METHODS

TomoNet is an open-source software package developed using Python. It follows a highly modularized architecture with each module responsible for specific tasks in a typical cryoET and STA data processing pipeline. Modules in TomoNet mainly cover the upper stream of the cryoET and STA pipeline including procedures of motion correction, tilt series generation, tomogram reconstruction, CTF estimation and particle picking, while leave the high-resolution

https://doi.org/10.1017/S2633903X24000060 Published online by Cambridge University Press

3D refinement to established software package like Relion (Figure 1). The design of a modern GUI, established with PyQt5 platform, enhances user-friendliness, and helps with tracking the processing progress (Figure 2). With table views, users can obtain a comprehensive overview of the entire dataset, facilitating direct and intuitive management for each tomogram (Figure 2).

Implementation of modules for motion correction, tomogram reconstruction and CTF estimation

Motion correction, tomogram reconstruction, and CTF estimation related functions are organized into individual modules in TomoNet, with the integration of corresponding external software packages including MotionCorr2[42], IMOD[43] or AreTomo[60], and CTFFIND4[44], respectively. Since their codes are not rewritten in TomoNet, users have to install each of them before using the corresponding modules.

The “Motion Correction” module is used to correct bean-induced sample motion. It requires an input folder path that contains all the dose fractionated frames, then user can specify their MotionCorr2 parameters in the GUI. After clicking the “RUN” button, TomoNet will perform motion correction for all the input images and save the results in a separated directory. This module also allows on-the-fly motion correction during data collection.

The “3D Reconstruction” module comprises two sub-functions: “TS Generation” and “Reconstruction”. Within “TS Generation”, users can readily assemble tilt series for each tomogram from the previously generated motion corrected images. It provides advanced options for data cleaning, such as setting a minimum acceptable number of tilt images for a tomogram, removing duplicate images at the same tilt angle by excluding images with older time stamps. The “Reconstruction” tab automatically reads and lists all tomograms in a table view, with essential information, such as tilt image number and alignment errors, and action buttons for restarting, continuing, and deleting individual tomogram reconstruction processes. This simplifies the assessment of reconstruction results and facilitating tomogram reconstruction management.

https://doi.org/10.1017/S2633903X24000060 Published online by Cambridge University Press

The "CTF Estimation" module is used for the tilt series defocus estimation, with support of parallel processing using multiple CPUs. Its outcomes are also listed in a table view with visualization features, such as displaying defocus at 0 degree and plotting the defocus distribution across all tilt angles.

Implementation of the “Manual Picking” module

The “Manual Picking” module is designed for general management of manual particle picking, especially for the preparation of “seed” particles required in “Auto Expansion”. IMOD stalkInit picking criteria is implemented to define the Y-axis for each particle with 2 points, and the center in between them. In the example of HIV dataset, 5-10 particles were manually picked as the “seed” particles for each VLP lattice, which only takes several minutes per tomogram (Figure 5a).

Design and implementation of the “Auto Expansion” module

“Auto Expansion” consists of three steps as shown in Figure 2. “Generate tomograms.star” is used to generate a STAR format file that maintains information of tomograms and their associated “seed” particles to be applied in “Auto Expansion”. “Generate Picking Parameter” is used to set up parameters required for particle set expansion through the described iterative process. The parameters include angular search ranges and steps, translational search ranges and steps, a “transition list” (explained later), box size used in particle alignment, distance between neighboring repeating subunits, reference and mask map, cross-correlation threshold, etc. The “transition list” is customized by users to describe the targeting lattice configuration, with each transition denoted by *[sx, sy, sz]*, where *sx, sy and sz* are translational shifts from the center of “seed” particle to one of its neighbors along X, Y and Z-axis, respectively. Thus, “Auto Expansion” can use it to guide the search of “candidate” particles. These user defined parameters will then be saved into a JSON format file. “Run Particle Expansion” takes the above STAR and JSON format files as inputs to perform the iterative particle set expansion.

https://doi.org/10.1017/S2633903X24000060 Published online by Cambridge University Press

During the “Auto Expansion” processing, three directories will be generated for each tomogram. They are “*TomoName*” as the working directory for carrying out the current iteration, “*TomoName_cache*” that stores intermediate results from finished iterations, and “*TomoName_final*” that stores the final particle picking results. The iteration number of “Auto Expansion” is typically greater than one. However, “Auto Expansion” allows for some special usage cases. For example, in the scenario when users need to modify the particle picking setting such as a different cross-correlation threshold, user can generate the new picking parameter file, then execute “Run Particle Expansion” by setting the iteration number as 0. This prompts the program to skip the “candidate” searching steps, but just gather all intermediate results saved in “*TomoName_cache*” directories, then generate a new “*TomoName_final*” result.

Design and implementation of the “AI AutoPicking” module

The “AI AutoPicking” module comprise 3 main steps, “Prepare Training Dataset”, “Train Neural Network” and “Predict Particles coordinates”. It uses supervised machine learning that requires users to provide ground truth, *i.e.*, tomogram with the associated particle coordinates files, for the model training. In this study, the ground truth data were prepared by “Auto Expansion”.

In “Prepare Training Dataset”, extracted subtomograms are used as inputs to the network training model for two reasons. Firstly, the size of tomogram used for picking is typically around 1000x1000x1000 voxels which is not applicable to be loaded in the GPU memory, but the size of extracted subtomograms is under 100x100x100 voxels. Secondly, it helps with increasing the number of training data pairs to avoid over-fitting during the network training. For the model output, the particle coordinates information was embedded into 3D binary segmentation maps, where the voxels associated with particles were set to 1, otherwise set to 0 (Figure 4a).

In “Train Neural Network”, the above extracted subtomograms paired with their associated segmentation maps are used to train a neural network model to be a binary classifier that predict whether a voxel is near the center of a particle. The network architecture employed

https://doi.org/10.1017/S2633903X24000060 Published online by Cambridge University Press

is derived from the one used in IsoNet[50] as it is well-suited for capturing generalized features of 3D objects (Figure 4b). Since the learning task is voxel-wisely binary classification, cross entropy loss function is used instead of minimum squared error (MSE). Equipped with one RTX 3080Ti graphic card, the training process can be completed swiftly within 1-2 hours if using the default parameters.

In “Predict Particles coordinates”, users can apply the trained model on the entire tomography dataset for particle coordinate prediction (Figure 4c). For each tomogram, TomoNet generate a predicted segmentation map first, then its particle coordinates information can be retrieved from the segmentation map by utilizing the hierarchical clustering algorithm from *scipy* module in Python.

Implementation of tools within the “Other Utilities” module

The “Other Utilities” module consists of two sub-functions: "Recenter | Rotate | Assemble to .star file" and "3D Subtomogram Place Back" as useful tools for post particle picking processing. The first one allows users to assemble and convert the particle picking results into a STAR format file following the Relion4 convention, reset particles center to its symmetric center, and align the rotation axis to Relion Z-axis. The second one takes a user-provided STAR format file that contains particles information as input, then generates a ChimeraX[61] session file for 3D subtomograms placing back and a clean version of STAR format file with “bad” particles removed. This not only allows users to validate the accuracy of particle picking before importing into Relion, but also enables direct observation of the distribution and configuration of subunits after the high-resolution 3D refinements, providing overall *in situ* lattice observations (Figure 7).

Processing tomograms of HIV VLP dataset

The HIV VLP dataset was downloaded from the Electron Microscopy Public Image Archive (EMPIAR) with the accession code EMPIAR-10164[40]. Four tilt series, TS_01, TS_43, TS_45 and TS_54, were used in this study. Downloaded micrographs were loaded into the TomoNet

https://doi.org/10.1017/S2633903X24000060 Published online by Cambridge University Press

pipeline to perform tilt series assembly, CTF estimation, and tomogram reconstruction using the WBP algorithm.

Four-time binned tomograms with 5.4 Å pixel size were used for further particle picking. Firstly, tomograms TS_01 and TS_43 were used for “seed" particles preparation on 3 selected VLPs per tomogram, and an initial reference map was generated by averaging them in PEET. Secondly, one run of “Auto Expansion” was applied on the above two tomograms to get more particles, such as to refine the reference. Thirdly, with an improved reference, a new run of “Auto Expansion” was applied on the selected 3 VLPs in both tomogram (Figure 5b), then the particle picking result was used for neural network training in “AI AutoPicking”. Fourthly, after the particle prediction on all four tomograms with a trained model, “AI AutoPicking” produced 4,860, 3,704, 4,550 and 2,101 particles for tomograms TS_01, TS_43, TS_45 and TS_54, as shown in Figure 5c. Lastly, the predicted particles were input as “seed” particles for the final run of “Auto Expansion”, resulting in 5,765, 4,043, 5,006, and 2,838 particles for tomograms TS_01, TS_43, TS_45 and TS_54, which were imported into Relion to perform high-resolution refinements.

Following the same procedure carried out in the Relion4 tutorial together with TomoNet 3D classification, the Gag hexamer structure was resolved at 3.2 Å resolution with 13,558 particles from four tomograms. Resolution was calculated in Relion and on 3DFSC Processing Server[62]. The global resolution reported is based on the “gold standard” refinement procedures and the 0.143 Fourier shell correlation (FSC) criterion (Figure 6c).

Processing one tomogram of *T. brucei* Axoneme

The tomogram of *T. brucei* axoneme is from our previous work[21]. Initially, one “seed” particle was manually picked for each DMT, following by 4 iterations of “Auto Expansion” applied to 9 “seed” particles, resulting in a total of 75 particles. EMD-20012 was used for subtomogram placing back to validate our picking results and visualize the entire axoneme architecture.

Processing one tomogram of *C. Crescentus* S-layer

https://doi.org/10.1017/S2633903X24000060 Published online by Cambridge University Press

458 The FIB-milled *C. crescentus* data of one reconstructed tomogram was downloaded from
459 Electron Microscopy Data Bank (EMDB) with the accession code EMD-23622[54]. This tomogram
460 was directly used for “seed” particles preparation on two of the cells. Around 30 “seed” particles
461 were manually picked and averaged using PEET to generate an initial reference map. “Auto
462 Expansion” was applied on the “seed” particles for 5 iterations to get more particles such as to
463 refine the reference map. With the improved reference map, another run of “Auto Expansion”
464 was applied to the same “seed” particles for 15 iterations to search all particles on the outer
465 surface of the cells, and finally yielded ~1,500 S-layer particles of hexamer subunits (Figure 9d).

466 **Processing tomograms of NEC budding *in vitro***

467 The cryoET grid preparation and data collection were previously described[57]. Motion correction,
468 tomogram reconstruction and CTF estimation were performed using TomoNet. Around 50-150
469 “seed” particles were manually picked for each tomogram. “Auto Expansion” were applied on a
470 total of 35 tomograms and yield the ~48,000 particles before Relion refinements. Following one
471 round of 3D auto-refine job under four-binned pixel size and several rounds of 3D auto-refine
472 jobs under two-binned pixel size and one round of 3D auto-refine under unbinned pixel size,
473 together with TomoNet 3D classifications, the NEC hexamer structure was resolved at 5.4 Å
474 resolution with totally 35,039 particles.

475 **3D visualization**

476 IMOD[43] was used to visualize the 2D tomographic and segmentation map slices. UCSF
477 ChimeraX[61] was used to visualize the STA results and the lattices generated by 3D
478 subtomogram place back. The atomic models were fitted into the density map using the “fit in
479 map” tool in ChimeraX.

480 # AVAILABILITY

481 TomoNet code is available on Github website at https://github.com/logicvay2010/TomoNet, with
482 a user manual. For the HIV VLPs dataset, the raw data was downloaded from the Electron

https://doi.org/10.1017/S2633903X24000060 Published online by Cambridge University Press

Microscopy Public Image Archive (EMPIAR) with accession code EMPIAR-10164[40], the Gag atomic model was downloaded from the Protein Data Bank (PDB) with accession code 5L93[40]. For the *C. Crescentus* S-layer dataset, the reconstructed tomogram was downloaded from the Electron Microscopy Data Bank (EMDB) with accession code EMD-23622[54], and the subunit model was generated using atomic model with PDB accession code 6P5T[63]. The STA results of NEC hexamer[57] and HIV can be obtained from EMDB with accession code EMD-40224 and EMD-43869, respectively.

ACKNOWLEDGEMENTS

We thank Elizabeth Draganova and Ekaterina Heldwein for the NEC dataset.

FUNDING STATEMENT

We acknowledge funding from the US National Institutes of Health (GM071940 to Z.H.Z.) and the National Science Foundation (DMR-1548924 to Z.H.Z.).

AUTHORSHIP CONTRIBUTIONS

HW and ZHZ initialized and ZHZ supervised research; HW wrote the code and developed the software GUI with help from SL; HW, SL and XY tested the software on different datasets; HW, and ZHZ wrote the manuscript; JZ and XY assisted the manuscript writing; all authors reviewed and approved the paper.

COMPETING INTERESTS STATEMENT

The authors declare that there is no conflict of interest.

REFERENCES

1 Kimanius, D., Dong, L., Sharov, G., Nakane, T. & Scheres, S. H. W. New tools for automated cryo-EM single-particle analysis in RELION-4.0. *Biochem J* **478**, 4169-4185, doi:10.1042/BCJ20210708 (2021).

2 Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods* **14**, 290-296, doi:10.1038/nmeth.4169 (2017).

3 Wan, W. & Briggs, J. A. Cryo-Electron Tomography and Subtomogram Averaging. *Methods Enzymol* **579**, 329-367, doi:10.1016/bs.mie.2016.04.014 (2016).

4 Chen, M. *et al.* A complete data processing workflow for cryo-ET and subtomogram averaging. *Nat Methods* **16**, 1161-1168, doi:10.1038/s41592-019-0591-8 (2019).

5 Zhang, P. Advances in cryo-electron tomography and subtomogram averaging and classification. *Curr Opin Struct Biol* **58**, 249-258, doi:10.1016/j.sbi.2019.05.021 (2019).

6 Castano-Diez, D. & Zanetti, G. In situ structure determination by subtomogram averaging. *Curr Opin Struct Biol* **58**, 68-75, doi:10.1016/j.sbi.2019.05.011 (2019).

7 Hong, Y., Song, Y., Zhang, Z. & Li, S. Cryo-Electron Tomography: The Resolution Revolution and a Surge of In Situ Virological Discoveries. *Annu Rev Biophys* **52**, 339-360, doi:10.1146/annurev-biophys-092022-100958 (2023).

8 Huang, Y., Zhang, Y. & Ni, T. Towards in situ high-resolution imaging of viruses and macromolecular complexes using cryo-electron tomography. *J Struct Biol* **215**, 108000, doi:10.1016/j.jsb.2023.108000 (2023).

9 Sibert, B. S. *et al.* Workflow for High-resolution Sub-volume Averaging from Heterogenous Viral and Virus-like Assemblies. *Microsc Microanal* **29**, 943-944, doi:10.1093/micmic/ozad067.470 (2023).

10 Kopylov, M., Bobe, D., Johnston, J. D. & Paraan, R. M. Modern Tools for In-situ Tomography. *Microsc Microanal* **29**, 954-955, doi:10.1093/micmic/ozad067.476 (2023).

11 Ni, T. *et al.* High-resolution in situ structure determination by cryo-electron tomography and subtomogram averaging using emClarity. *Nat Protoc* **17**, 421-444, doi:10.1038/s41596-021-00648-5 (2022).

12 Xue, L. *et al.* Visualizing translation dynamics at atomic detail inside a bacterial cell. *Nature* **610**, 205-211, doi:10.1038/s41586-022-05255-2 (2022).

13 Zivanov, J. *et al.* A Bayesian approach to single-particle electron cryo-tomography in RELION-4.0. *Elife* **11**, doi:10.7554/eLife.83724 (2022).

14 Obr, M. & Schur, F. K. M. in *Advances in Virus Research* Vol. 105 (ed Félix A. Rey) 117-159 (Academic Press, 2019).

15 Castano-Diez, D., Kudryashev, M., Arheit, M. & Stahlberg, H. Dynamo: a flexible, user-friendly development tool for subtomogram averaging of cryo-EM data in high-performance computing environments. *J Struct Biol* **178**, 139-151, doi:10.1016/j.jsb.2011.12.017 (2012).

16 Himes, B. A. & Zhang, P. emClarity: software for high-resolution cryo-electron tomography and subtomogram averaging. *Nature Methods* **15**, 955-961, doi:10.1038/s41592-018-0167-z (2018).

17 Wagner, T. *et al.* SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. *Commun Biol* **2**, 218, doi:10.1038/s42003-019-0437-z (2019).

18 Bharat, T. A. M. & Scheres, S. H. W. Resolving macromolecular structures from electron cryo-tomography data using subtomogram averaging in RELION. *Nature Protocols* **11**, 2054-2065, doi:10.1038/nprot.2016.124 (2016).

19 Tegunov, D. & Cramer, P. Real-time cryo-electron microscopy data preprocessing with Warp. *Nature Methods* **16**, 1146-1152, doi:10.1038/s41592-019-0580-y (2019).

20 Zhang, J. *et al.* Structure of the trypanosome paraflagellar rod and insights into non-planar motility of eukaryotic cells. *Cell Discov* **7**, 51, doi:10.1038/s41421-021-00281-2 (2021).

21 Imhof, S. *et al.* Cryo electron tomography with volta phase plate reveals novel structural foundations of the 96-nm axonemal repeat in the pathogen Trypanosoma brucei. *eLife* **8**, e52058, doi:10.7554/eLife.52058 (2019).

22 Si, Z. *et al.* Different functional states of fusion protein gB revealed on human cytomegalovirus by cryo electron tomography with Volta phase plate. *PLOS Pathogens* **14**, e1007452, doi:10.1371/journal.ppat.1007452 (2018).

23 Scaramuzza, S. & Castaño-Díez, D. Step-by-step guide to efficient subtomogram averaging of virus-like particles with Dynamo. *PLOS Biology* **19**, e3001318, doi:10.1371/journal.pbio.3001318 (2021).

24 Bharat, T. A. *et al.* Cryo-electron tomography of Marburg virus particles and their morphogenesis within infected cells. *PLoS Biol* **9**, e1001196, doi:10.1371/journal.pbio.1001196 (2011).

25 Grünewald, K. *et al.* Three-dimensional structure of herpes simplex virus from cryo-electron tomography. *Science* **302**, 1396-1398, doi:10.1126/science.1090284 (2003).

26 Zanetti, G. *et al.* The structure of the COPII transport-vesicle coat assembled on membranes. *Elife* **2**, e00951, doi:10.7554/eLife.00951 (2013).

27 Böhm, J. *et al.* Toward detecting and identifying macromolecules in a cellular context: Template matching applied to electron tomograms. *Proceedings of the National Academy of Sciences* **97**, 14245-14250, doi:doi:10.1073/pnas.230282097 (2000).

28 de Teresa-Trueba, I. *et al.* Convolutional networks for supervised mining of molecular patterns within cellular context. *Nature Methods* **20**, 284-294, doi:10.1038/s41592-022-01746-2 (2023).

29 Moebel, E. *et al.* Deep learning improves macromolecule identification in 3D cellular cryo-electron tomograms. *Nature Methods* **18**, 1386-1394, doi:10.1038/s41592-021-01275-4 (2021).

30 Wu, S., Liu, G. & Yang, G. in *2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI).* 1-5.

31 Hao, Y. *et al.* VP-Detector: A 3D multi-scale dense convolutional neural network for macromolecule localization and classification in cryo-electron tomograms. *Computer Methods and Programs in Biomedicine* **221**, 106871, doi:https://doi.org/10.1016/j.cmpb.2022.106871 (2022).

32 Rice, G. *et al.* TomoTwin: generalized 3D localization of macromolecules in cryo-electron tomograms with structural data mining. *Nature Methods* **20**, 871-880, doi:10.1038/s41592-023-01878-z (2023).

33 Balyschew, N. *et al.* Streamlined structure determination by cryo-electron tomography and subtomogram averaging using TomoBEAR. *Nature Communications* **14**, 6543, doi:10.1038/s41467-023-42085-w (2023).

34 Jimenez de la Morena, J. *et al.* ScipionTomo: Towards cryo-electron tomography software integration, reproducibility, and validation. *J Struct Biol* **214**, 107872, doi:10.1016/j.jsb.2022.107872 (2022).

35 Liu, H.-F. *et al.* nextPYP: a comprehensive and scalable platform for characterizing protein variability in situ using single-particle cryo-electron tomography. *Nature Methods*, doi:10.1038/s41592-023-02045-0 (2023).

36 Schur, F. K. M. *et al.* Structure of the immature HIV-1 capsid in intact virus particles at 8.8 Å resolution. *Nature* **517**, 505-508, doi:10.1038/nature13838 (2015).

37 Mendonça, L. *et al.* CryoET structures of immature HIV Gag reveal six-helix bundle. *Communications Biology* **4**, 481, doi:10.1038/s42003-021-01999-1 (2021).

38 Zhao, G. *et al.* Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. *Nature* **497**, 643-646, doi:10.1038/nature12162 (2013).

39 Krebs, A.-S. *et al.* Molecular architecture and conservation of an immature human endogenous retrovirus. *Nature Communications* **14**, 5149, doi:10.1038/s41467-023-40786-w (2023).

40 Schur, F. K. M. *et al.* An atomic model of HIV-1 capsid-SP1 reveals structures regulating assembly and maturation. *Science* **353**, 506-508, doi:doi:10.1126/science.aaf9620 (2016).

41 Ni, T. *et al.* Structure of native HIV-1 cores and their interactions with IP6 and CypA. *Science Advances* **7**, eabj5715, doi:doi:10.1126/sciadv.abj5715 (2021).

42 Zheng, S. Q. *et al.* MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat Methods* **14**, 331-332, doi:10.1038/nmeth.4193 (2017).

43 Kremer, J. R., Mastronarde, D. N. & McIntosh, J. R. Computer visualization of three-dimensional image data using IMOD. *J Struct Biol* **116**, 71-76, doi:10.1006/jsbi.1996.0013 (1996).

https://doi.org/10.1017/S2633903X24000060 Published online by Cambridge University Press

44 Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J Struct Biol* **192**, 216-221, doi:10.1016/j.jsb.2015.08.008 (2015).

45 Heumann, J. M., Hoenger, A. & Mastronarde, D. N. Clustering and variance maps for cryo-electron tomography using wedge-masked differences. *Journal of Structural Biology* **175**, 288-299, doi:https://doi.org/10.1016/j.jsb.2011.05.011 (2011).

46 Hall, S. R. The STAR file: a new format for electronic data transfer and archiving. *Journal of Chemical Information and Computer Sciences* **31**, 326-333, doi:10.1021/ci00002a020 (1991).

47 Shimogawa, M. M. *et al.* FAP106 is an interaction hub for assembling microtubule inner proteins at the cilium inner junction. *Nature Communications* **14**, 5225, doi:10.1038/s41467-023-40230-z (2023).

48 von Kügelgen, A., Alva, V. & Bharat, T. A. M. Complete atomic structure of a native archaeal cell surface. *Cell Reports* **37**, 110052, doi:https://doi.org/10.1016/j.celrep.2021.110052 (2021).

49 Pum, D., Breitwieser, A. & Sleytr, U. B. Patterns in Nature—S-Layer Lattices of Bacterial and Archaeal Cells. *Crystals* **11**, 869 (2021).

50 Liu, Y.-T. *et al.* Isotropic reconstruction for electron tomography with deep learning. *Nature Communications* **13**, 6482, doi:10.1038/s41467-022-33957-8 (2022).

51 Tan, A., Pak, A. J., Morado, D. R., Voth, G. A. & Briggs, J. A. G. Immature HIV-1 assembles from Gag dimers leaving partial hexamers at lattice edges as potential substrates for proteolytic maturation. *Proceedings of the National Academy of Sciences* **118**, e2020054118, doi:doi:10.1073/pnas.2020054118 (2021).

52 Guo, S., Saha, I., Saffarian, S. & Johnson, M. E. Structure of the HIV immature lattice allows for essential lattice remodeling within budded virions. *eLife* **12**, e84881, doi:10.7554/eLife.84881 (2023).

53 Talledge, N. *et al.* HIV-2 Immature Particle Morphology Provides Insights into Gag Lattice Stability and Virus Maturation. *Journal of Molecular Biology* **435**, 168143, doi:https://doi.org/10.1016/j.jmb.2023.168143 (2023).

54 Lasker, K. *et al.* The material properties of a bacterial-derived biomolecular condensate tune biological function in natural and synthetic systems. *Nature Communications* **13**, 5643, doi:10.1038/s41467-022-33221-z (2022).

55 von Kügelgen, A. *et al.* In Situ Structure of an Intact Lipopolysaccharide-Bound Bacterial Surface Layer. *Cell* **180**, 348-358.e315, doi:https://doi.org/10.1016/j.cell.2019.12.006 (2020).

56 Sleytr, U. B., Schuster, B., Egelseer, E. M. & Pum, D. S-layers: principles and applications. *FEMS Microbiol Rev* **38**, 823-864, doi:10.1111/1574-6976.12063 (2014).

57 Draganova, E. B. *et al.* The universal suppressor mutation in the HSV-1 nuclear egress complex restores membrane budding defects by stabilizing the oligomeric lattice. *bioRxiv*, 2023.2006.2022.546118, doi:10.1101/2023.06.22.546118 (2023).

58 Wang, H. *et al.* Hierarchical organization and assembly of the archaeal cell sheath from an amyloid-like protein. *Nature Communications* **14**, 6720, doi:10.1038/s41467-023-42368-2 (2023).

59 Tegunov, D., Xue, L., Dienemann, C., Cramer, P. & Mahamid, J. Multi-particle cryo-EM refinement with M visualizes ribosome-antibiotic complex at 3.5 Å in cells. *Nature Methods* **18**, 186-193, doi:10.1038/s41592-020-01054-7 (2021).

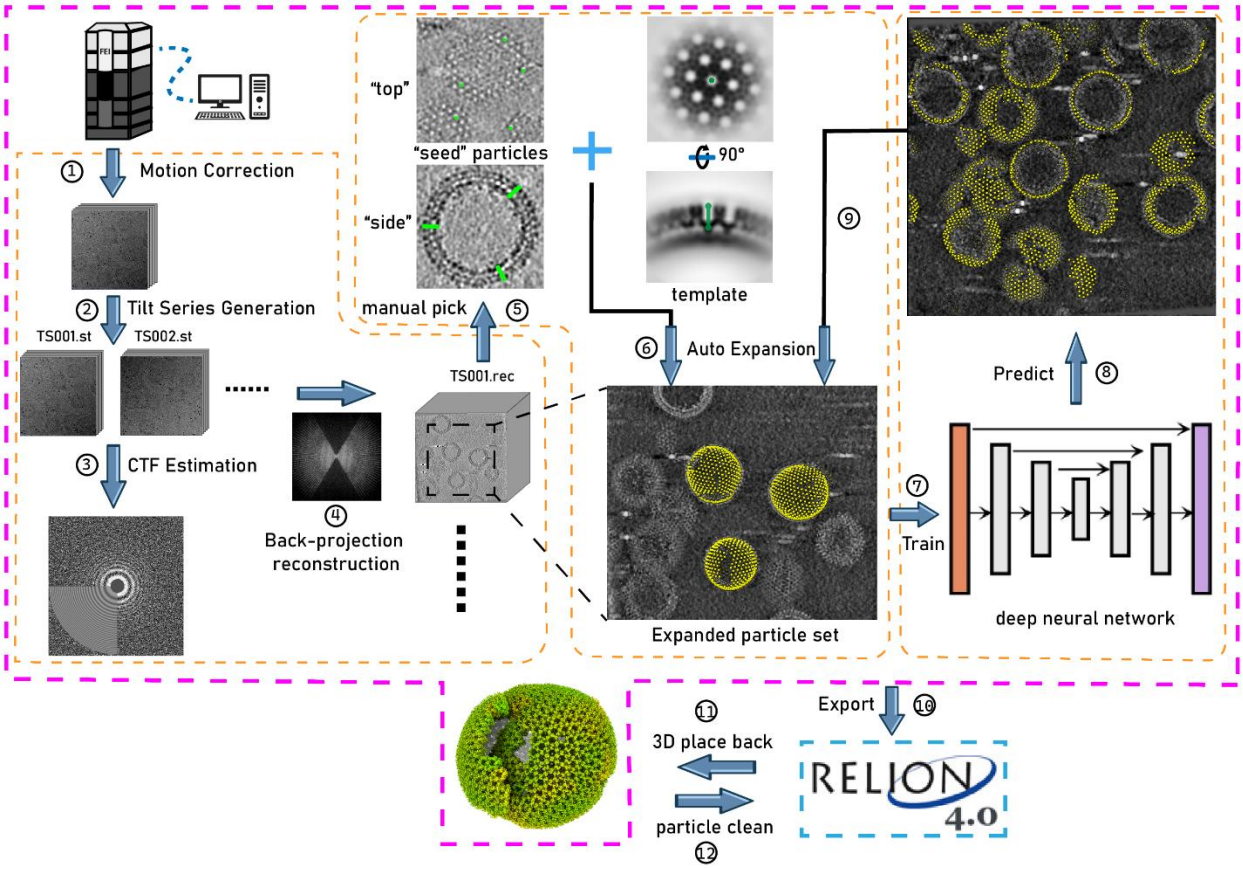
60 Zheng, S. *et al.* AreTomo: An integrated software package for automated marker-free, motion-corrected cryo-electron tomographic alignment and reconstruction. *Journal of Structural Biology: X* **6**, 100068, doi:https://doi.org/10.1016/j.yjsbx.2022.100068 (2022).

61 Meng, E. C. *et al.* UCSF ChimeraX: Tools for Structure Building and Analysis. *Protein Sci*, e4792, doi:10.1002/pro.4792 (2023).

62 Tan, Y. Z. *et al.* Addressing preferred specimen orientation in single-particle cryo-EM through tilting. *Nat Methods* **14**, 793-796, doi:10.1038/nmeth.4347 (2017).

63 Herrmann, J. *et al.* A bacterial surface layer protein exploits multistep crystallization for rapid self-assembly. *Proceedings of the National Academy of Sciences* **117**, 388-394, doi:doi:10.1073/pnas.1909798116 (2020).

660 Figure 1

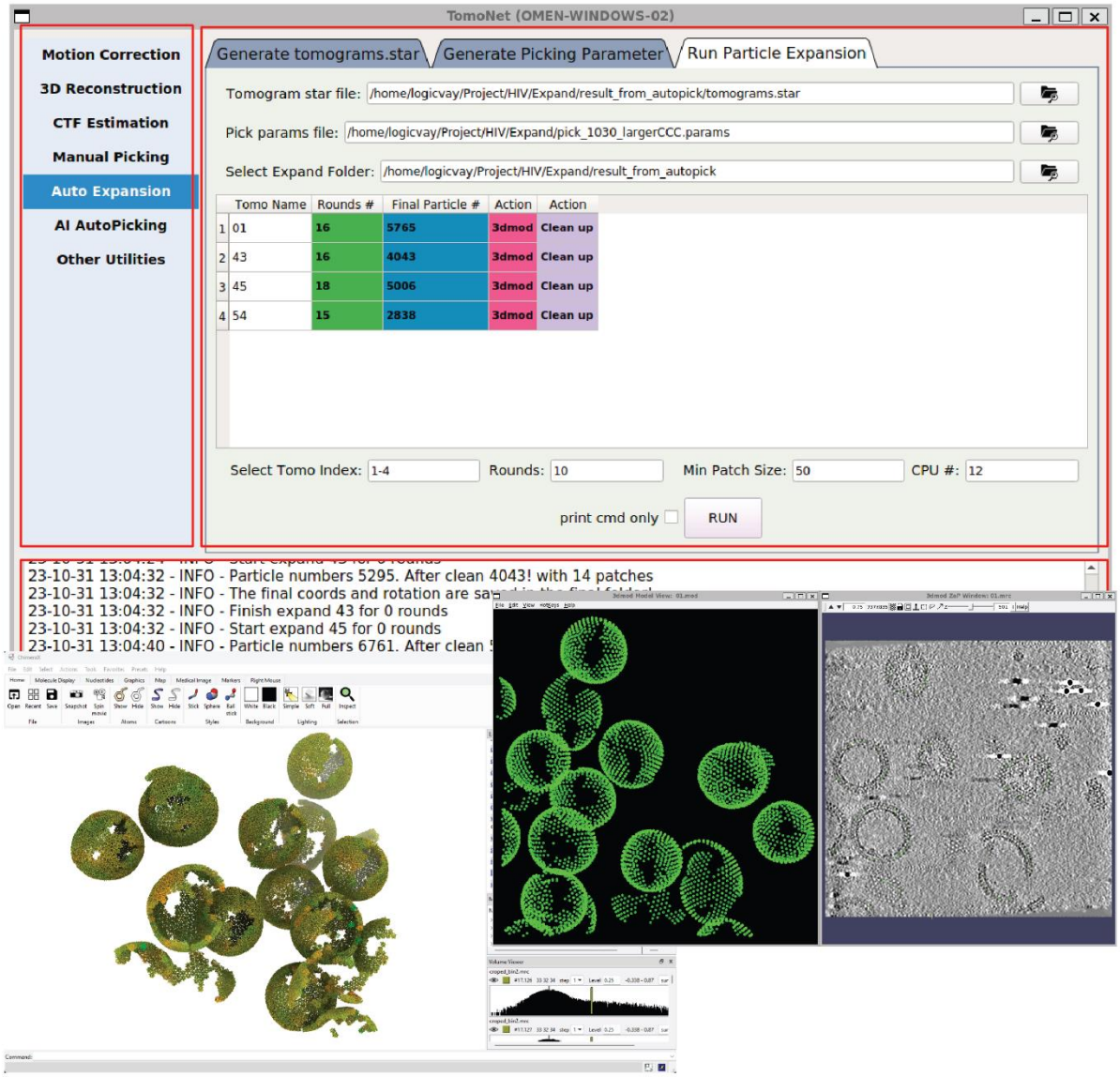


661

662

https://doi.org/10.1017/S2633903X24000060 Published online by Cambridge University Press

663 Figure 2

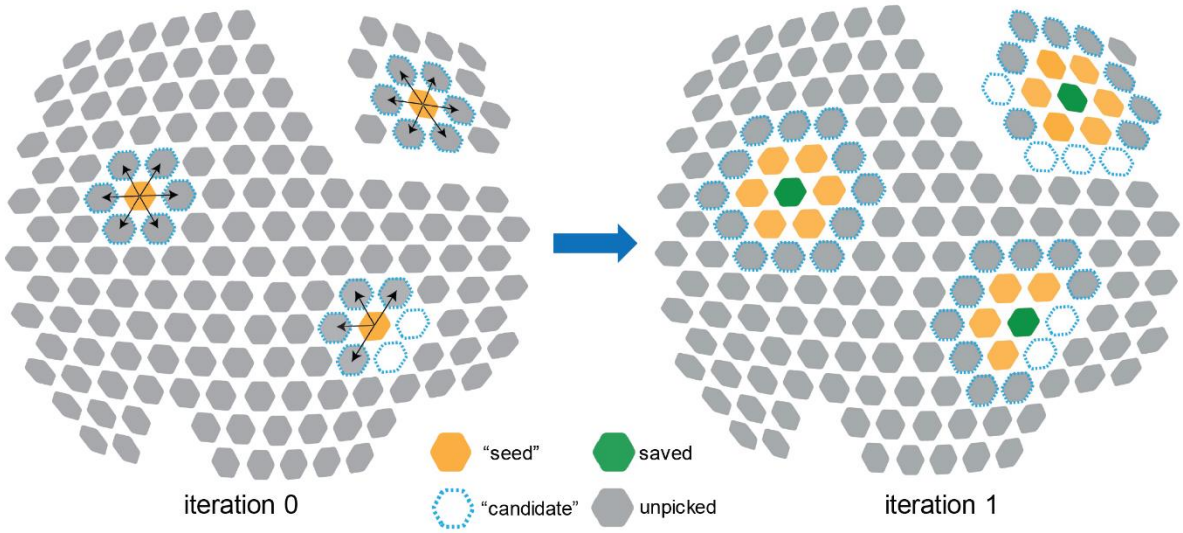


664

665

https://doi.org/10.1017/S2633903X24000060 Published online by Cambridge University Press

666 Figure 3

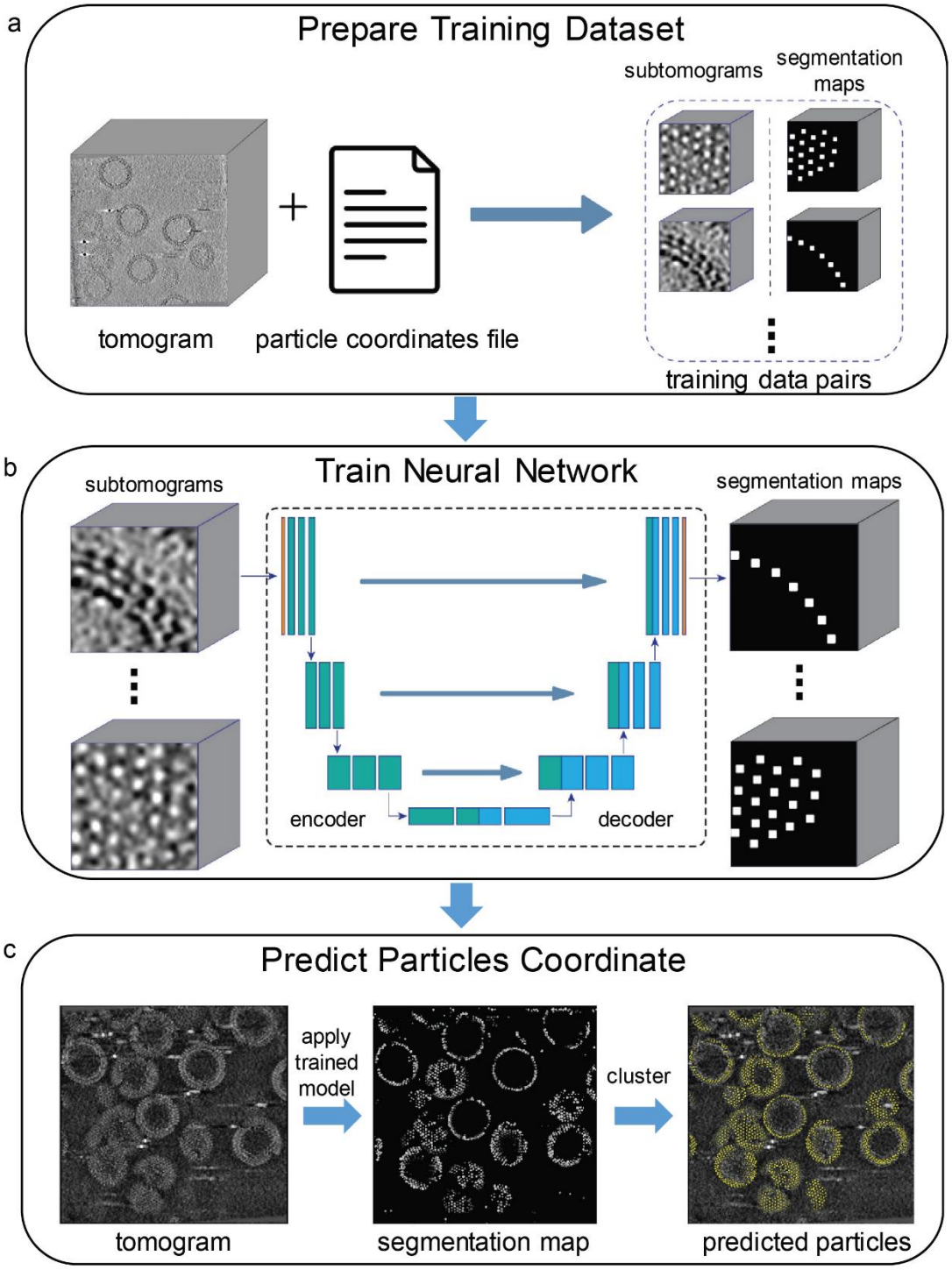


667

668

https://doi.org/10.1017/S2633903X24000060 Published online by Cambridge University Press

669 Figure 4

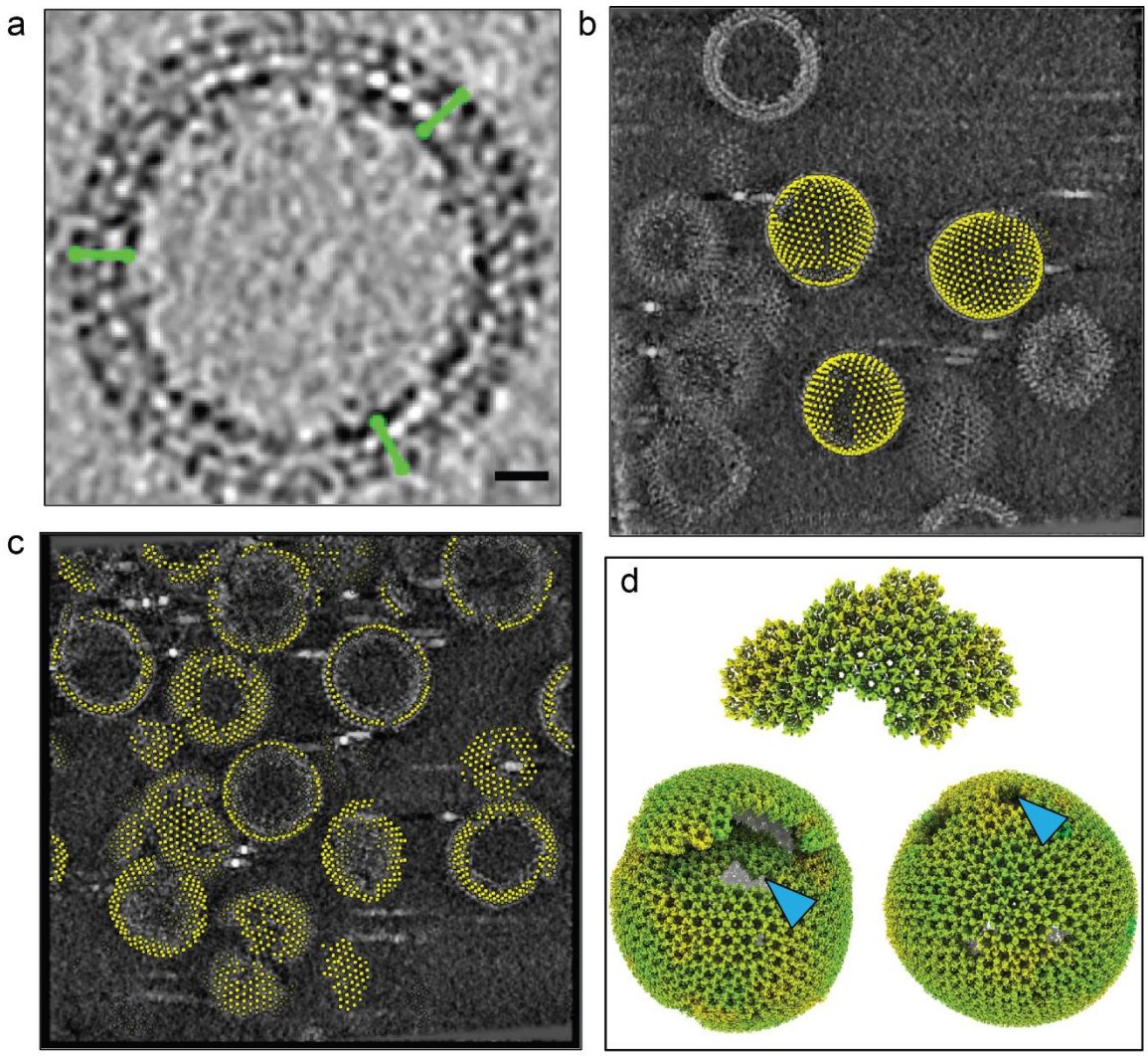


670

671

https://doi.org/10.1017/S2633903X24000060 Published online by Cambridge University Press

672 Figure 5

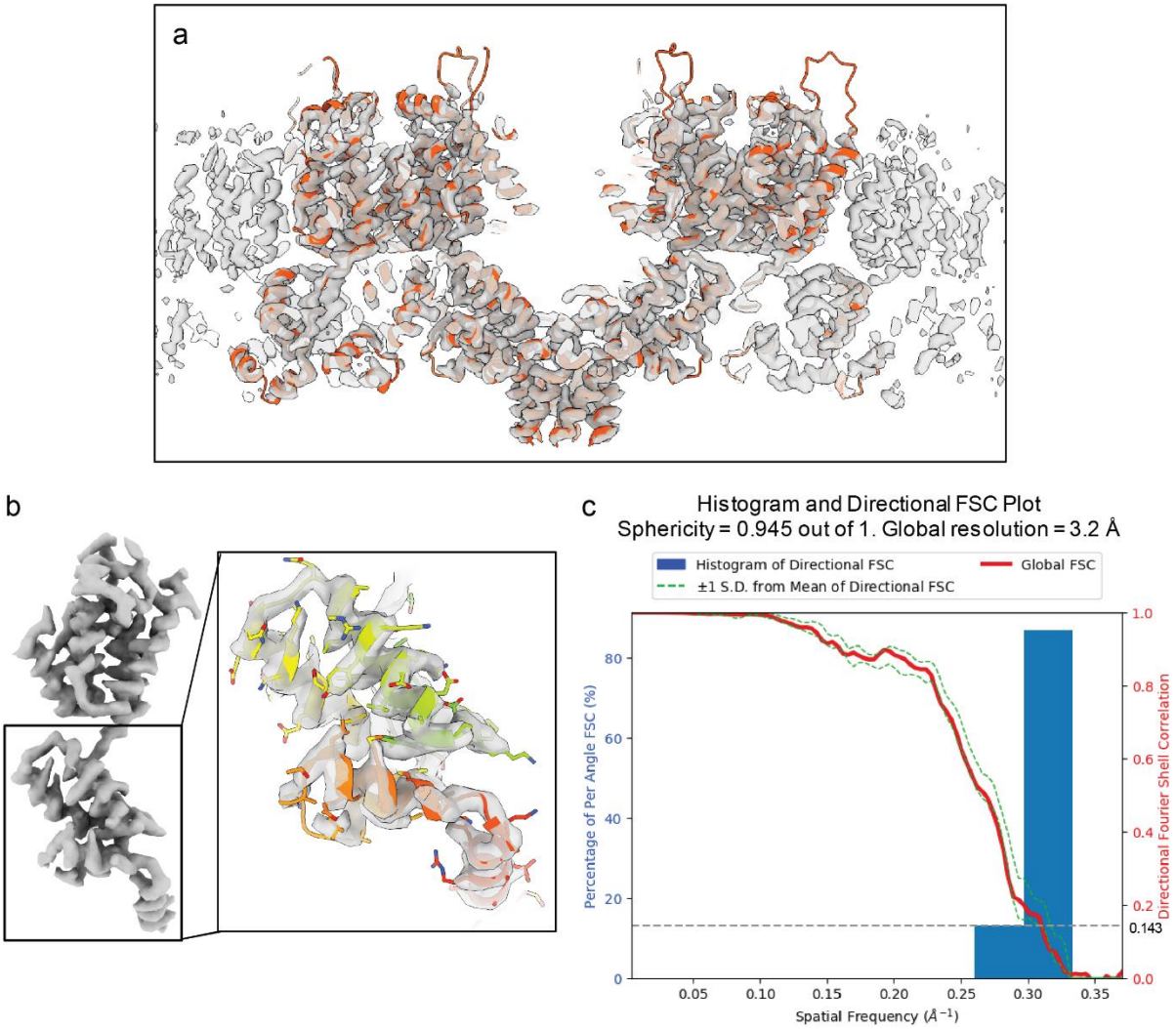


673

674

https://doi.org/10.1017/S2633903X24000060 Published online by Cambridge University Press

675 Figure 6

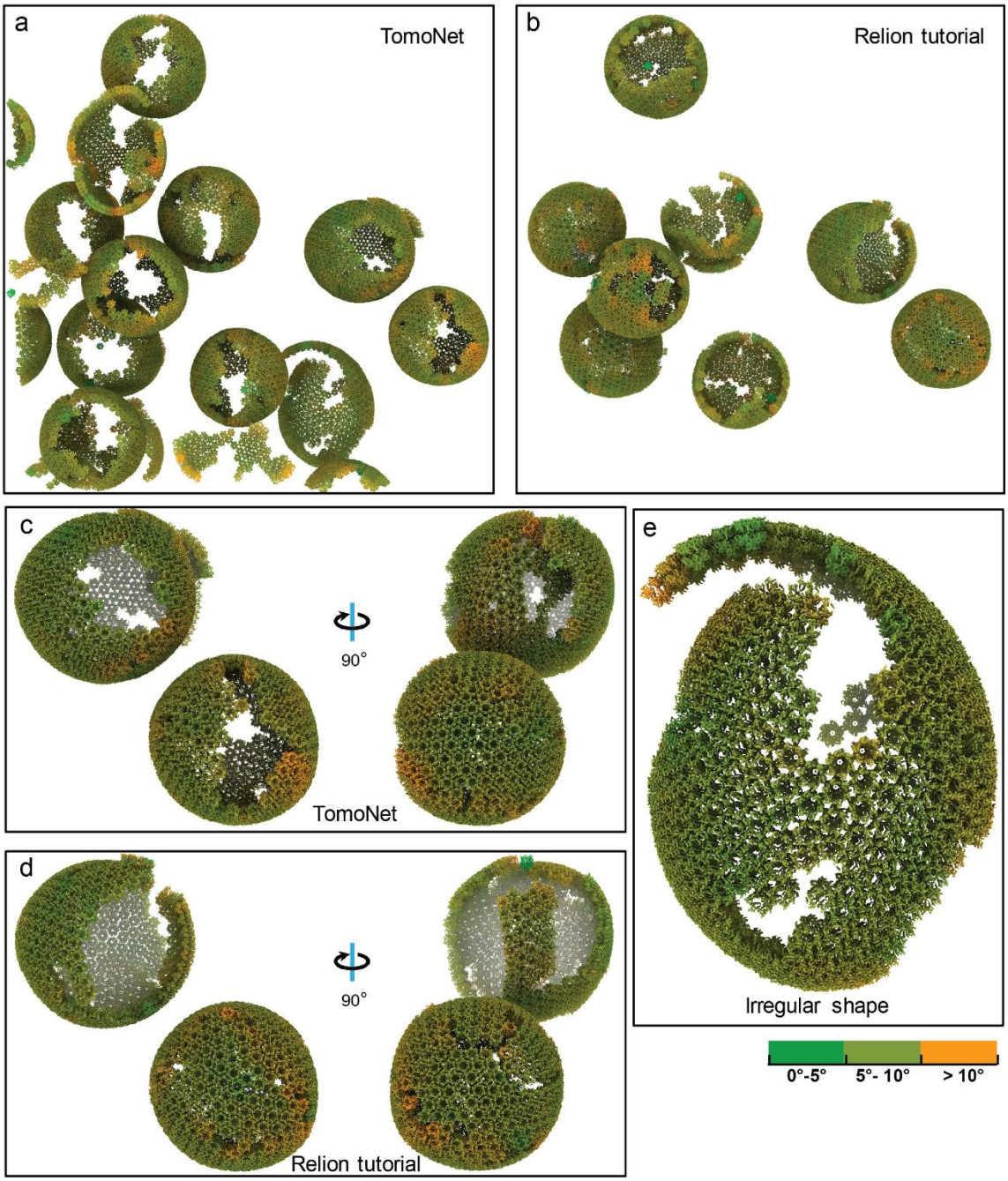


676

677

https://doi.org/10.1017/S2633903X24000060 Published online by Cambridge University Press

678 Figure 7

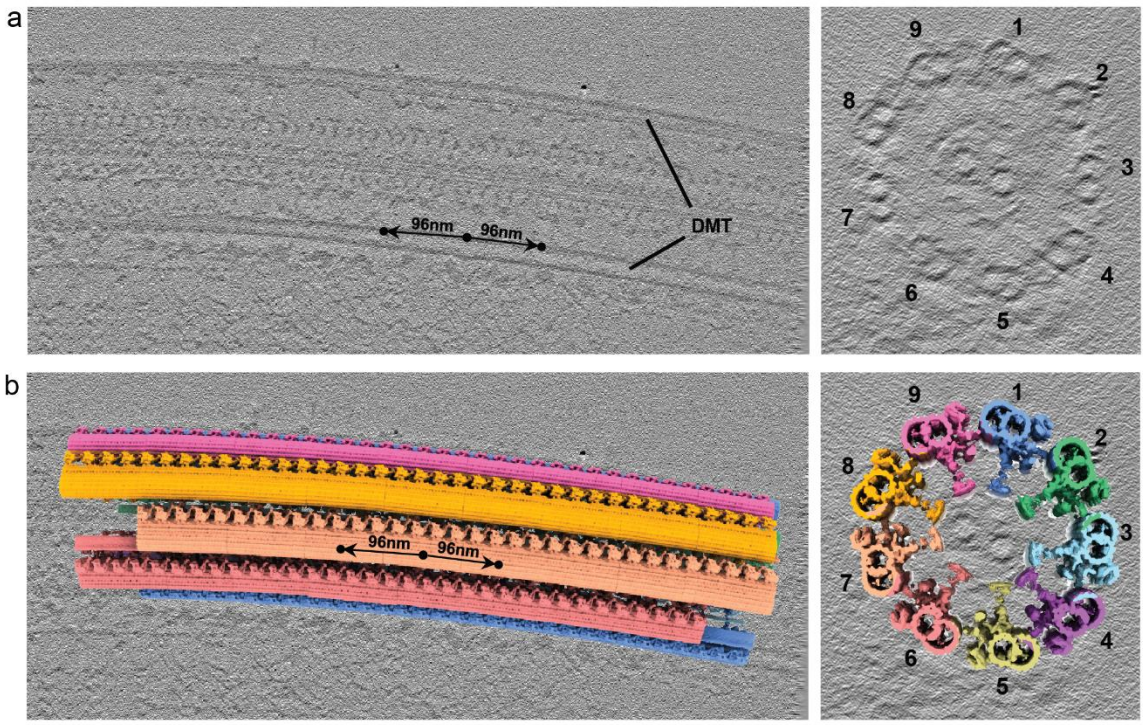


679

680

https://doi.org/10.1017/S2633903X24000060 Published online by Cambridge University Press

681 Figure 8

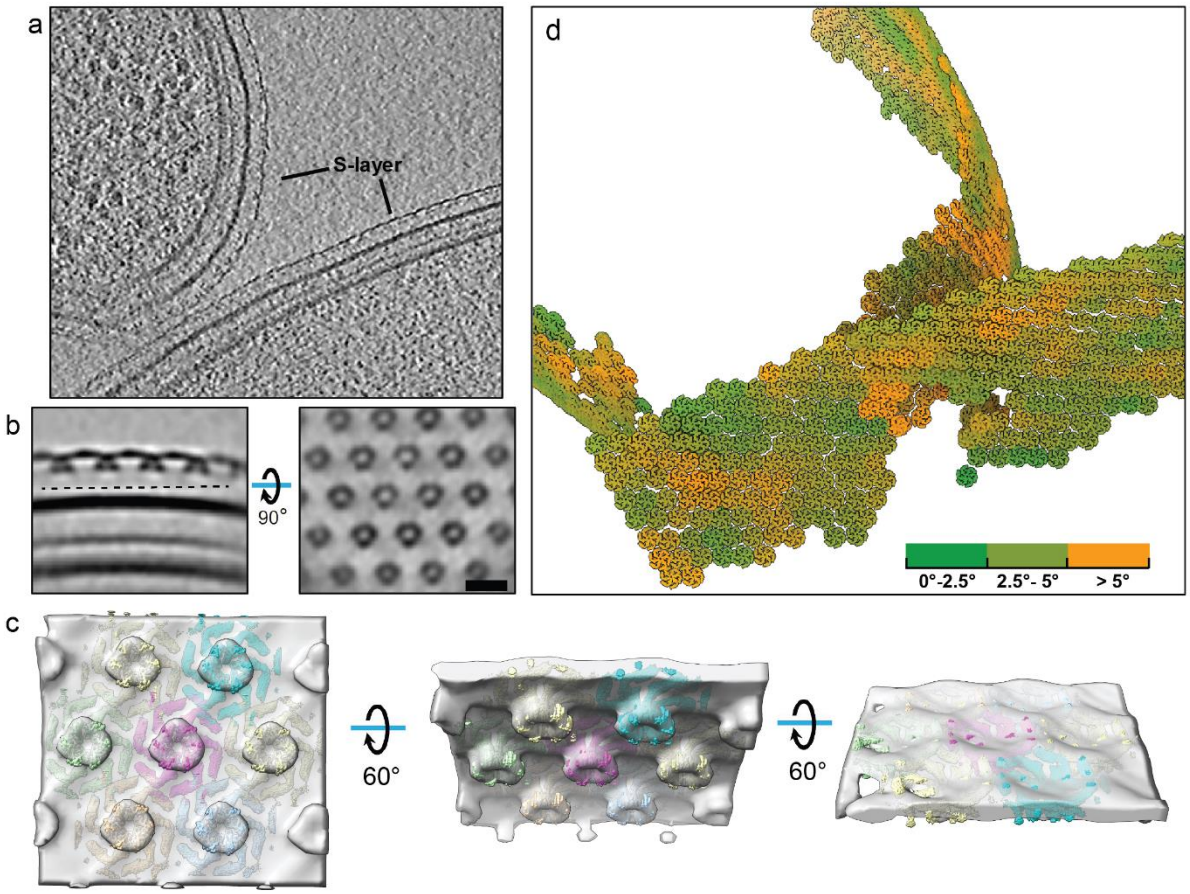


682

683

https://doi.org/10.1017/S2633903X24000060 Published online by Cambridge University Press

684 Figure 9

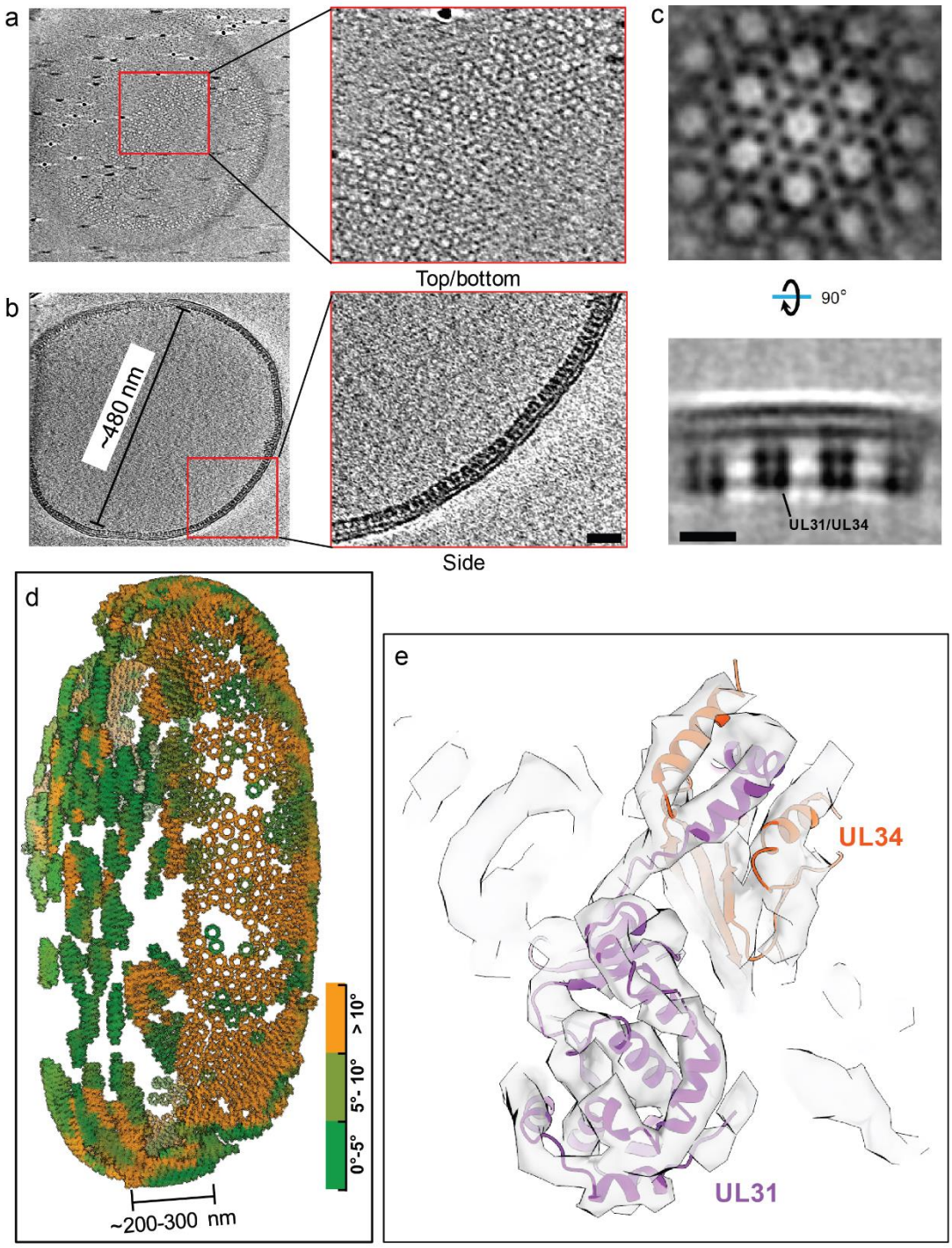


685

686

https://doi.org/10.1017/S2633903X24000060 Published online by Cambridge University Press

687 Figure 10



688

https://doi.org/10.1017/S2633903X24000060 Published online by Cambridge University Press