

# UCLA

## UCLA Previously Published Works

### Title

TomoNet: A streamlined cryogenic electron tomography software pipeline with automatic particle picking on flexible lattices

### Permalink

<https://escholarship.org/uc/item/1qq0g6jn>

### Authors

Wang, Hui

Liao, Shiqing

Yu, Xinye

et al.

### Publication Date

2024



### DOI

10.1017/s2633903x24000060

Peer reviewed

RESEARCH ARTICLE

# TomoNet: A streamlined cryogenic electron tomography software pipeline with automatic particle picking on flexible lattices

Hui Wang<sup>1,2,3</sup> , Shiqing Liao<sup>2,3</sup>, Xinye Yu<sup>3</sup>, Jiayan Zhang<sup>2,3</sup> and Z. Hong Zhou<sup>1,2,3</sup> 

<sup>1</sup>Department of Bioengineering, University of California, Los Angeles (UCLA), Los Angeles, CA, USA

<sup>2</sup>California NanoSystems Institute, UCLA, Los Angeles, CA, USA

<sup>3</sup>Department of Microbiology, Immunology, and Molecular Genetics, UCLA, Los Angeles, CA, USA

**Corresponding author:** Z. Hong Zhou; Email: [Hong.Zhou@UCLA.edu](mailto:Hong.Zhou@UCLA.edu)

**Received:** 04 December 2023; **Revised:** 04 March 2024; **Accepted:** 25 March 2024

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

## 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 three-dimensional (3D) classification functions and integrates commonly used packages to streamline high-resolution subtomogram averaging for structures in 1D, 2D, or 3D arrays. Automatic particle picking is accomplished in two complementary ways: one based on template matching and the other using 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 the 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 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 the integration of commonly used software packages. TomoNet addresses the particle-picking challenge with two solutions, one based on geometric template matching and the other using 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.

## 1. Introduction

Single-particle cryogenic electron microscopy (cryoEM) is used 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<sup>1</sup> and cryoSparc.<sup>2</sup> In parallel, cryogenic electron tomography (cryoET), coupled with subtomogram averaging (STA), expands the investigative scope to encompass heterogeneous macromolecules in their native context.<sup>3–10</sup> To enhance the resolution of subunits within *in situ* macromolecules, subtomograms (i.e., particles) are extracted from each tomogram and then subjected to three-dimensional (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.<sup>11–14</sup>

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, 3D 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,<sup>15</sup> template matching in emClarity<sup>16</sup> and machine learning in crYOLO.<sup>17</sup> In 3D refinement and classification, particles are iteratively classified and refined to obtain a final structure at subnanometer or near-atomic resolution, which has been demonstrated by software packages such as Relion<sup>13,18</sup>, emClarity,<sup>16</sup> EMAN2,<sup>4</sup> and Warp.<sup>19</sup> Finally, post-processing activities include map sharpening, Fourier shell correlation (FSC) calculation, visualization by placing averaged maps back into the original tomogram, and so forth users often need to navigate between several specialized software packages 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 used, particularly when aiming for resolutions between 20 and 50 Å with a maximum of several hundred particles.<sup>20–22</sup> 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.<sup>23</sup> Other examples include the Marburg Virus,<sup>24</sup> Herpes simplex virus,<sup>25</sup> and the Coat protein complex II,<sup>26</sup> 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.<sup>16,27</sup> Simultaneously, convolutional neural networks have shown promising results for cryoET automatic particle picking given its capacity to analyze 3D feature maps and autonomously identify prominent features within specific samples.<sup>28–31</sup> These machine-learning approaches typically operate template-free and often obviate the need for human annotation.<sup>32</sup>

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<sup>33</sup> offers an integrated solution, while ScipionTomo<sup>34</sup> and nextPYP<sup>35</sup> provide a comprehensive web-based platform for managing various tasks in the cryoET pipeline. Notably, 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.<sup>36–41</sup>

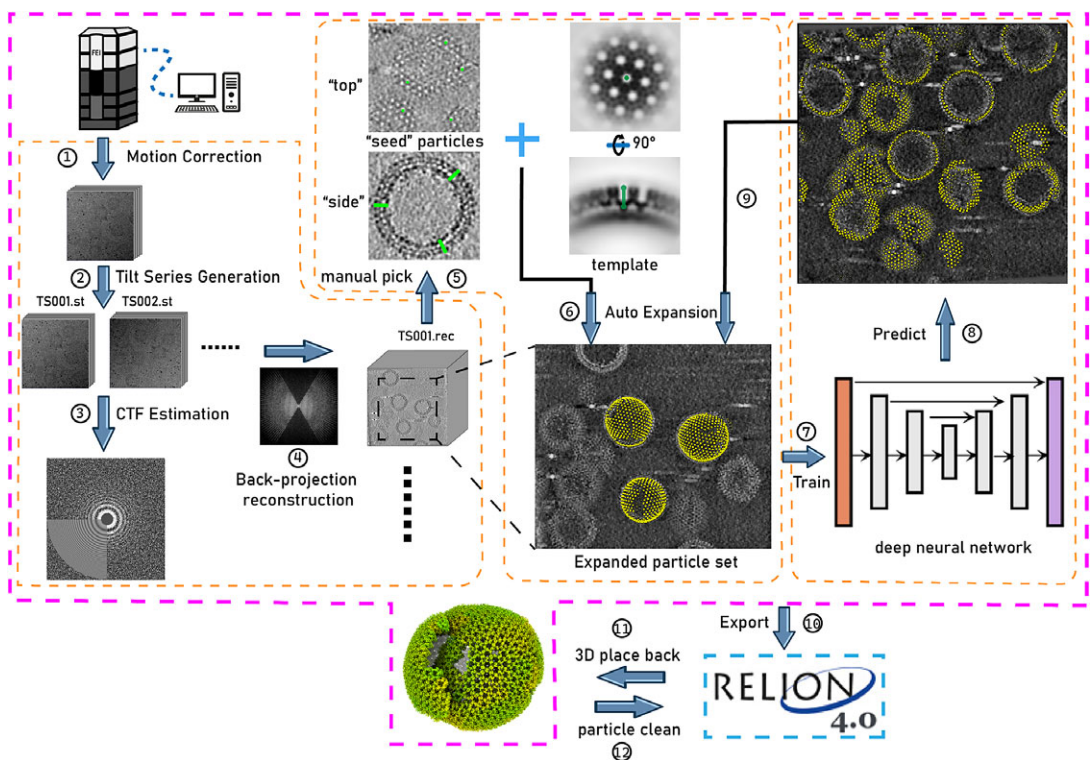
In this context, we have developed TomoNet, a software package designed for streamlining the cryoET and STA data processing workflow, with a modern graphical user interface (GUI) (Figures 1 and 2). Our

methodology uses 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.

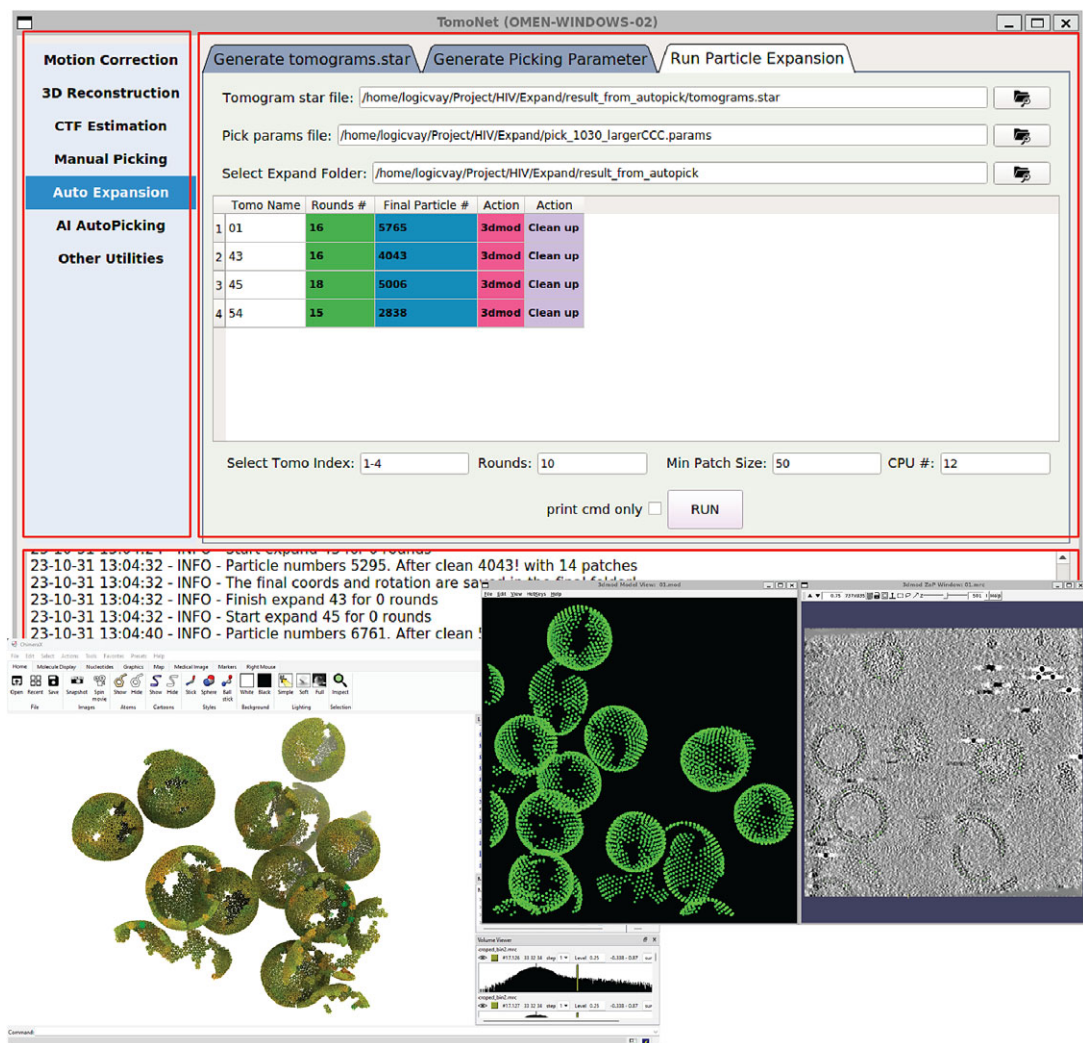
## 2. Results

### 2.1. 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 (Figures 1 and 2), after data collection from electron microscopy, TomoNet can perform motion correction with integration of MotionCorr2;<sup>42</sup> tilt series assembly and tomogram



**Figure 1. Illustration of TomoNet’s comprehensive pipeline for cryoET and STA.** The pink border encloses the sequential functions implemented in TomoNet, and they can be subdivided into three principal segments, delineated by the orange borders. These segments include tomogram preparation on the left, template matching-based particle picking “Auto Expansion” in the center, and deep learning-based automatic particle picking on the right.



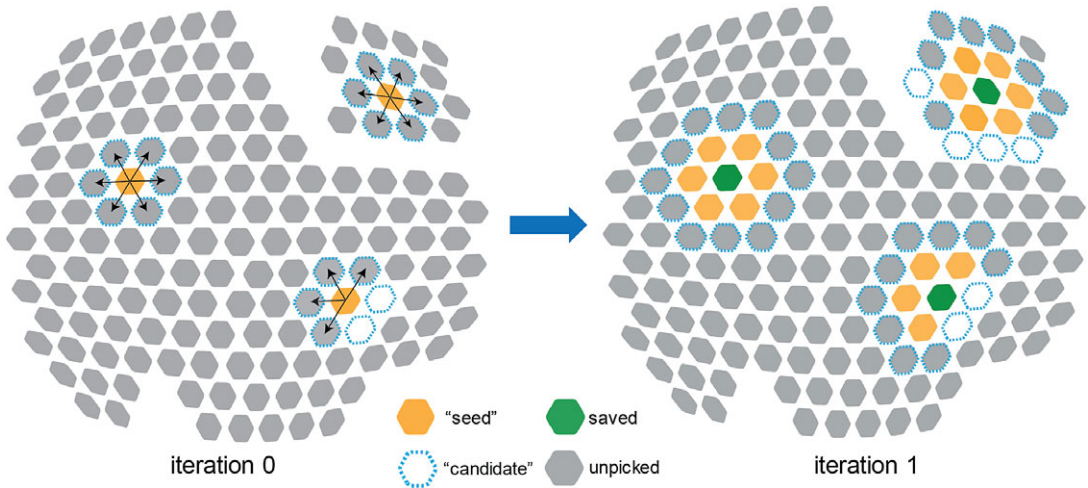
**Figure 2.** A screenshot of TomoNet GUI. The TomoNet GUI contains three main areas: the menu bar (top left), the input and operate area (top right), and the log window (bottom). Bottom left: results generated by the “3D Subtomogram Place Back” function can be visualized in ChimeraX. Bottom right: intermediate results of picked particles viewed with IMOD.

reconstruction with integration of IMOD,<sup>43</sup> CTF estimation with integration of CTFFIND4,<sup>44</sup> manual particle picking with IMOD; particle picking using built-in geometric template matching-based algorithms with integration of PEET,<sup>45</sup> 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,<sup>46</sup> 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).

## 2.2. 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





**Figure 3. Illustration of the first two iterations of “Auto Expansion” particle picking.** There are two patches of a hexagonal lattice with individual particles represented by solid hexagons. At iteration 0, 18 “candidate” particles (dashed blue) were selected from the neighbors of 3 “seed” particles (orange). 14 good particles remained and will serve as “seed” particles in iteration 1, and 3 “seed” particles in iteration 0 were saved in the final particle set (green). At iteration 1, 35 “candidate” particles were selected from the neighbors of 14 “seed” particles. 29 good particles remained and will serve as “seed” particles in iteration 2, and 14 “seed” particles were saved in the final particle set. “Auto Expansion” is an iterative process and will stop when no “candidate” can be detected.

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 (1D–3D). Examples are abound: microtubule doublets, ubiquitous in most cells, consist of 96 nm axonemal 1D translational repeat units<sup>21,47</sup> (1D rotational lattice); HIV VLPs<sup>40</sup> and surface layer (S-layer) lattice of prokaryotic cells<sup>48,49</sup> are composed of hexameric subunits (2D lattice); paraflagellar rod of protozoan species is organized into para-crystalline arrays in its distal zone<sup>20</sup> (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 shows two patches of 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 for each iteration (Figure 3). First, potential particles adjacent to each “seed” particle are calculated and selected as “candidate” particles. Second, 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. Third, 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.

### 2.3. Automatic particle picking by deep learning

The “AI AutoPicking” module is designed for automatic particle picking using supervised machine learning, which uses 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 particle coordinate files. The trained model can then be applied to the entire tomography dataset and output predicted particles for each tomogram.

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 which each particle is labeled by a cube near its center (Figure 4a). The trained neural network model can be applied to other tomograms to perform particle segmentation. Finally, the particle coordinate information can be retrieved from the predicted segmentation maps (Figure 4c).

### 2.4. 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.<sup>50</sup> 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 Material S1).

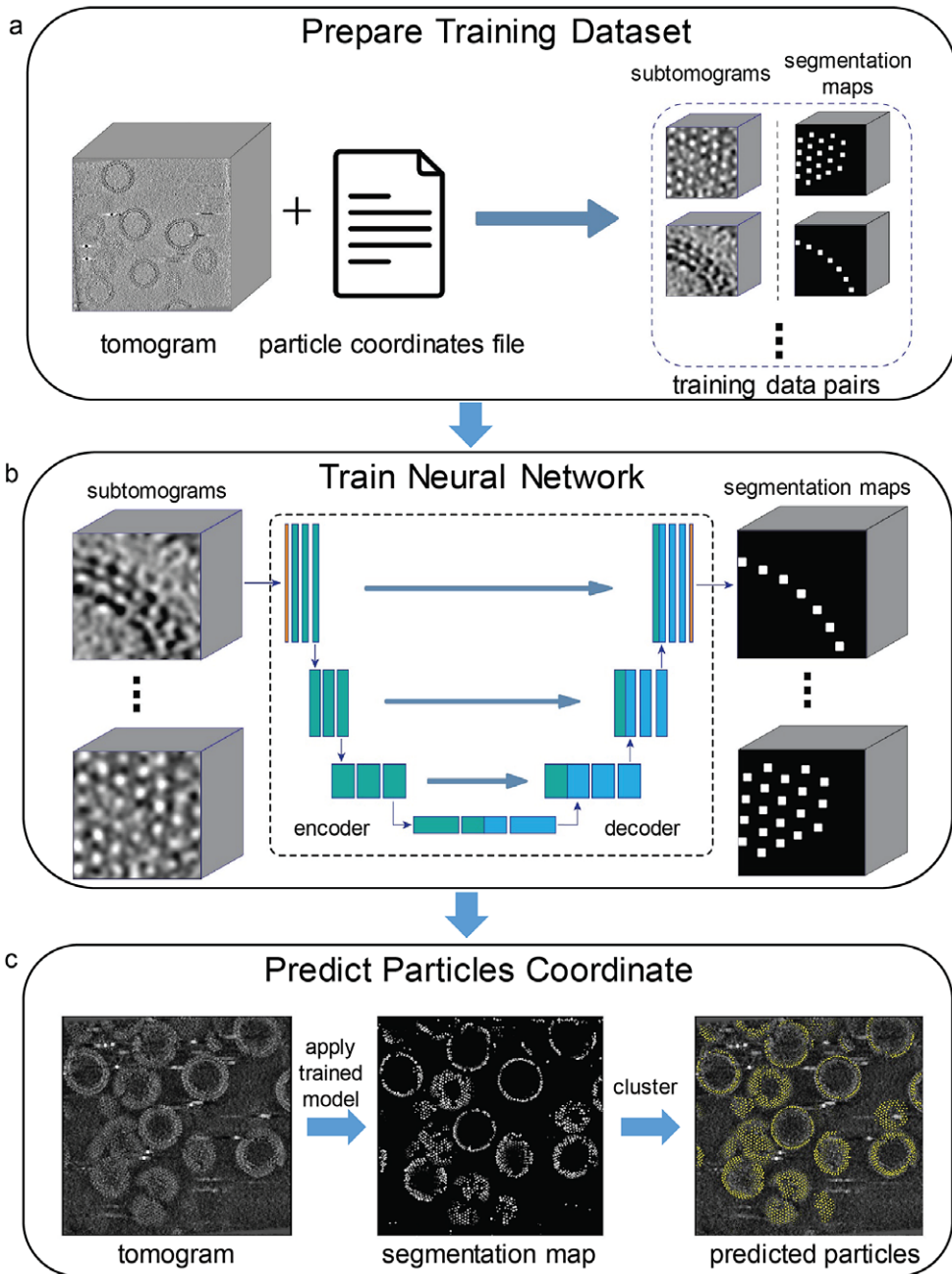
Removing these “bad” particles is necessary for achieving better resolutions.<sup>51</sup> To accomplish this, TomoNet assesses each particle by counting its neighboring particles and calculating the averaged tilt angle to these neighbors to represent the local surface curvature of a lattice. TomoNet 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.

### 2.5. Application to *in situ* viral protein arrays: The matrix protein lattice in HIV VLPs

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 showed sphere-like geometry (Figure 5a).

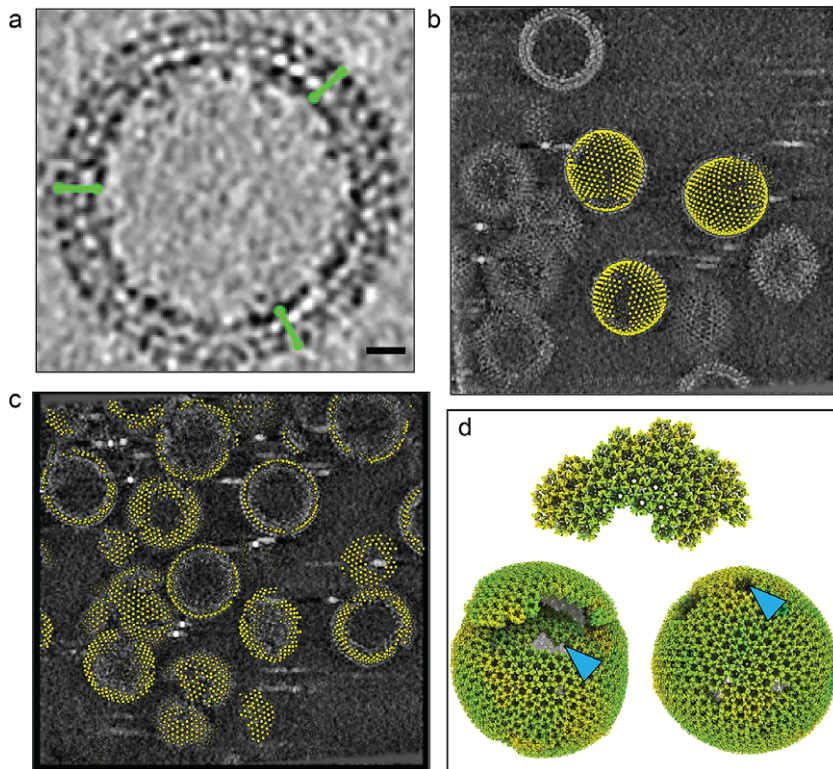
As detailed in the Method section, a combination of “Auto Expansion” and “AI AutoPicking” was applied to the above four tomograms. The result shows that particles were readily picked on all the observed lattice patches (Figure 5b,c). Then, these picked particles were imported to Relion to perform high-resolution particle refinements, the resulting reconstruction of the Gag hexamer structure (Figure 6) looks identical to the published high-resolution structure,<sup>11,13</sup> 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 (Figures 5d and 7). All VLP lattices with various sizes and



**Figure 4. Illustration of “AI AutoPicking” process consisting of three steps.** The HIV dataset was used for this illustration, and the particles refer to Gag hexamers. (a) Training dataset preparation. Using the user-provided tomograms with associated particle coordinate files, subtomograms containing particle densities were extracted. For each subtomogram, TomoNet generated a segmentation map based on the coordinates of particles, where the voxels near a particle’s center are shown as white and the others as black. (b) Neural network training. The generated subtomograms and segmentation maps were used as the input and output to train the convolutional neural network in learning how to segment out particle densities. (c) Particle coordinate prediction. Firstly, TomoNet applied the trained neural network model to unseen tomograms and generated associated predicted segmentation maps. Then, the particle coordinate information was obtained from the segmentation maps using clustering algorithms.





**Figure 5. TomoNet application to arrays of matrix protein in HIV VLPs.** (a) Illustration of picked “seed” particles on a spherical VLP. Green segments represent the particles’ Y-axis. Scale bar is 20 nm. (b) “Auto Expansion” result on three VLPs within tomogram TS\_01, with yellow dots representing the center of the hexamer subunits. (c) “AI AutoPicking” particle prediction result of tomogram TS\_45 shows its ability to pick particles on all lattices of different sizes and shapes. (d) Visualization of three different variations of the HIV Gag lattices generated by placing back averaged structures, two exhibiting a spherical shape, and one presented as a fragment. Blue arrows indicate defects in the lattice.

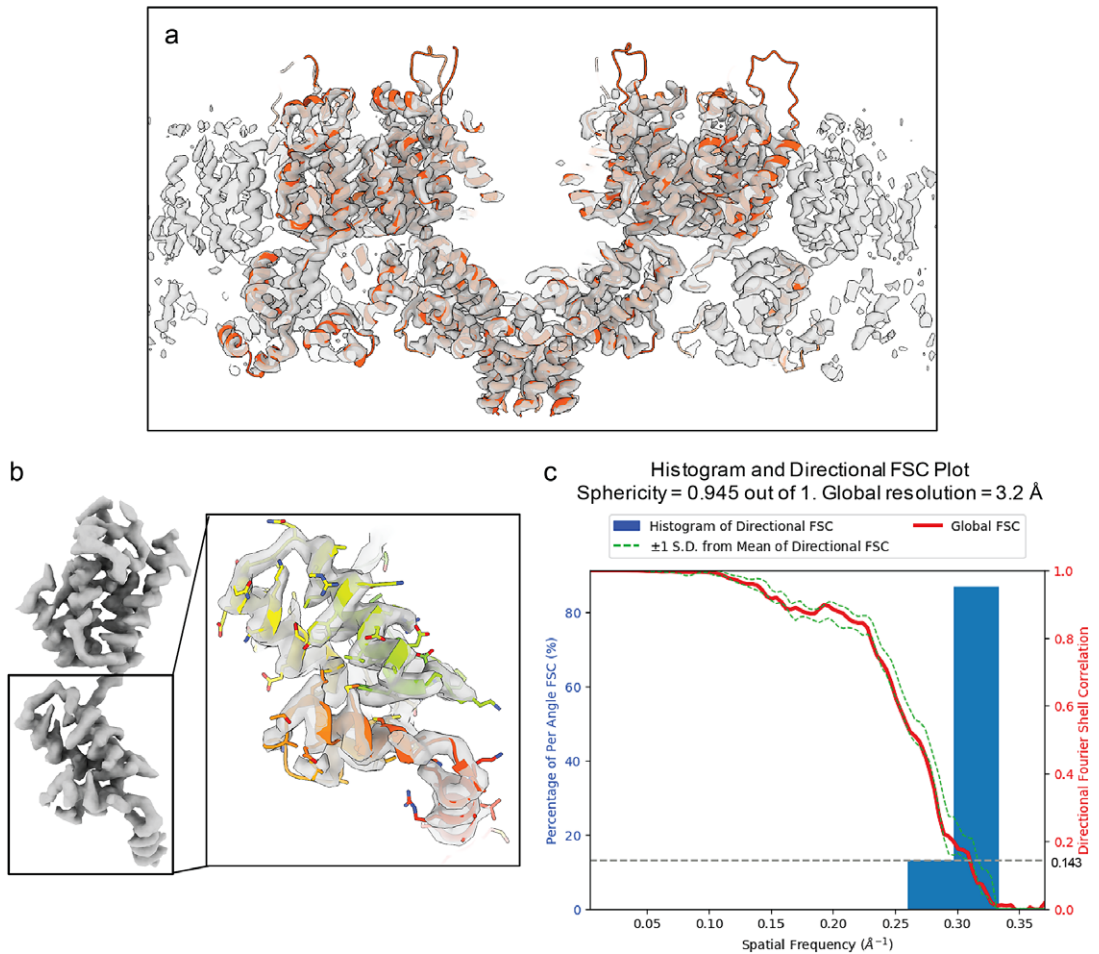
shapes were captured even with irregular shapes (Figure 7e and Supplementary Material S2), demonstrating TomoNet’s particle-picking ability on flexible lattices. Lattice defects on each VLP were also identified consistent with previous studies,<sup>52</sup> enhancing the understanding of lattice assembly mechanisms.<sup>53</sup>

## 2.6. Application to cellular organelle sample: Eukaryotic axoneme

We validated TomoNet’s particle-picking capability for 1D lattices by processing one tomogram of extracted flagellum of *Trypanosoma brucei*. The axoneme consists of 9 outer doublet microtubules (DMTs) and a pair of central singlet microtubules, where each DMT is a 1D polymer of 96 nm axonemal building blocks (Figure 8a). This typical 1-D lattice often exhibits imperfections like bends and twists, posing challenges for precise particle picking (Figure 8a). Using “Auto Expansion”, TomoNet accurately picked the 96 nm-spaced axonemal subunits from all DMTs, effectively adapting to lattice imperfections (Figure 8b).

## 2.7. Application to focused ion beam (FIB)-milled cellular sample: The S-layer lattice of prokaryotic cell

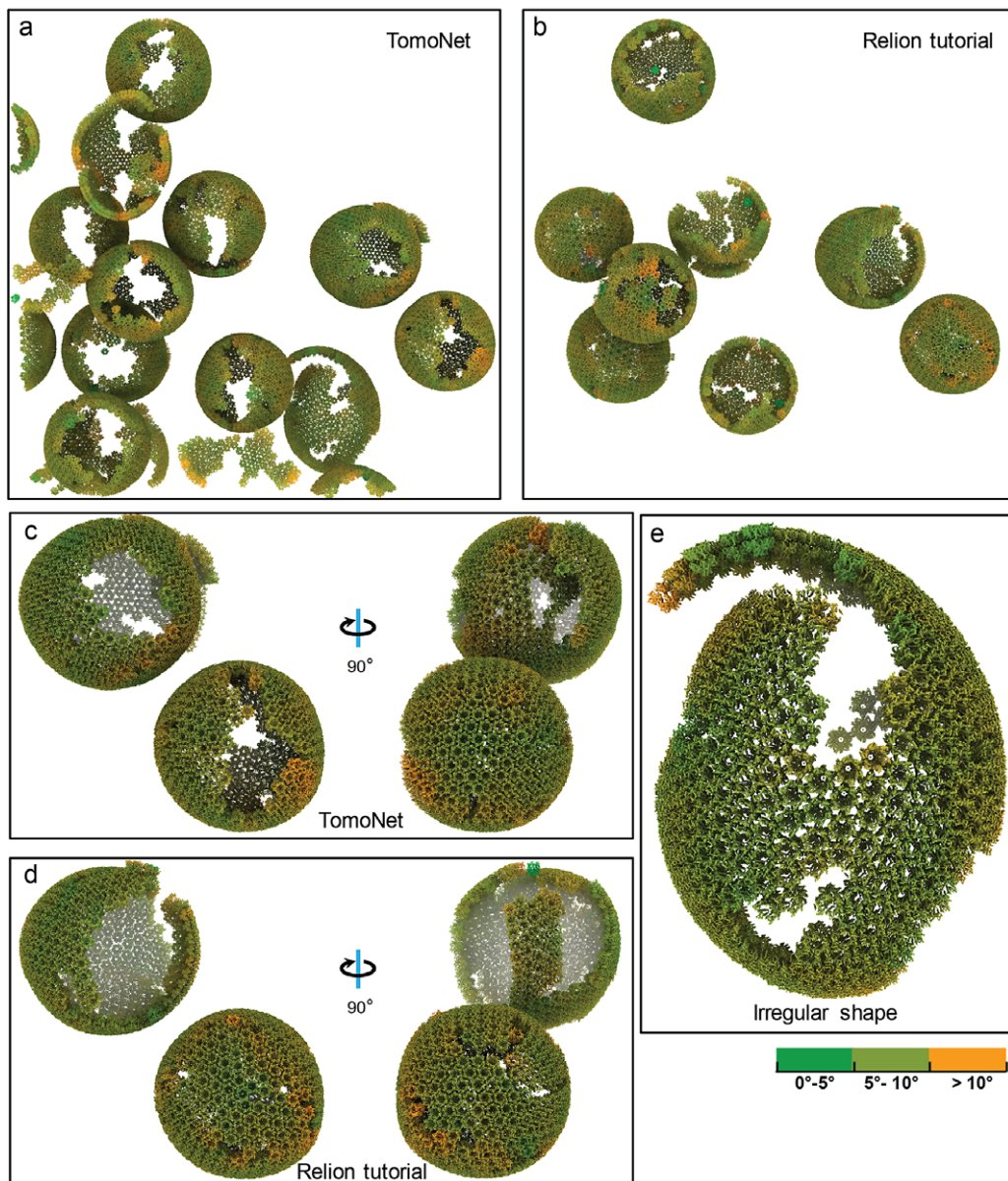
We validated TomoNet’s particle-picking capability by processing one tomogram of FIB-milled *Caulobacter crescentus* cells from EMD-23622.<sup>54</sup> The S-layer functions as a component of the cell wall



**Figure 6. Final map resolution of HIV Gag hexamer.** (a) Final reconstruction of Gag hexamer (grey) fitted with the atomic model (PDB: 5I93). (b) One segmented Gag monomer structure, inset shows a closer view of carboxy-terminal domain overlay with the atomic model. (c) Directional Fourier shell correlation (FSC) curves for the STA of Gag hexamer structure, with a global resolution at 3.2 Å.

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<sup>55</sup> (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.<sup>49,56</sup>

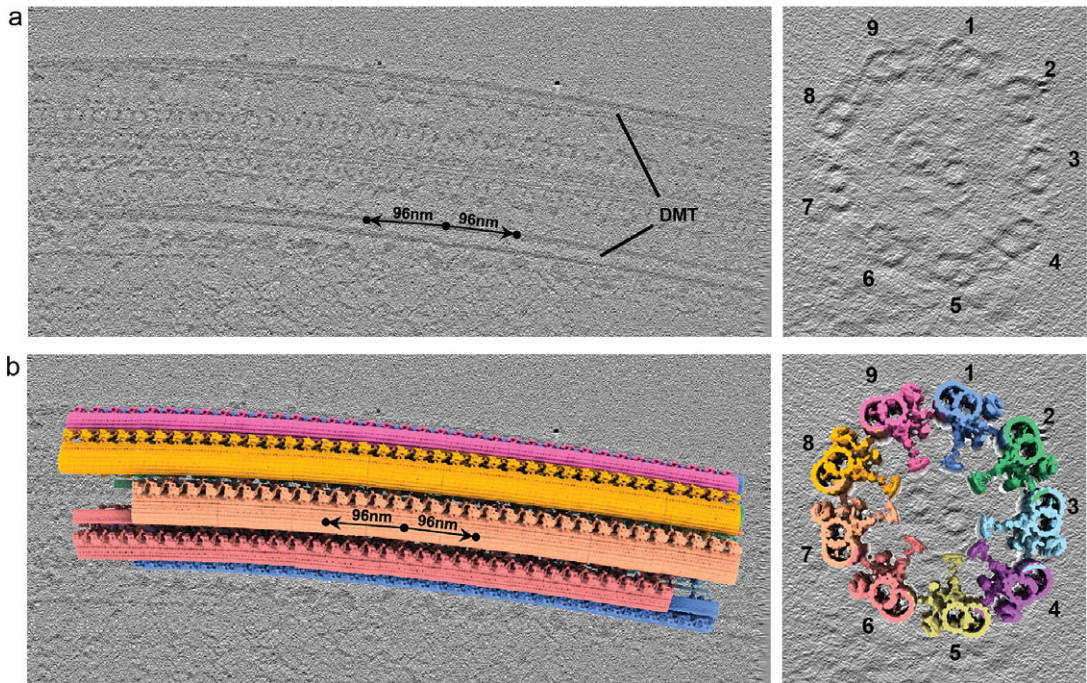


**Figure 7. Comparative visualization of lattices obtained from TomoNet and Relion tutorial.** (a, b) Visualized comparison of particles used in TomoNet and Relion tutorial within tomogram TS\_01. TomoNet can pick particles not only on a sphere-like lattice but also on others with random shapes. (c, d) A comparison of particle picking results on two sphere-like shape VLPs from TomoNet and Relion tutorial. (e) A zoom-in view of an irregularly shaped lattice. Coloring is based on surface curvatures at the point of each subunit.

### 2.8. 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,<sup>57</sup> NEC heterodimers budded into large





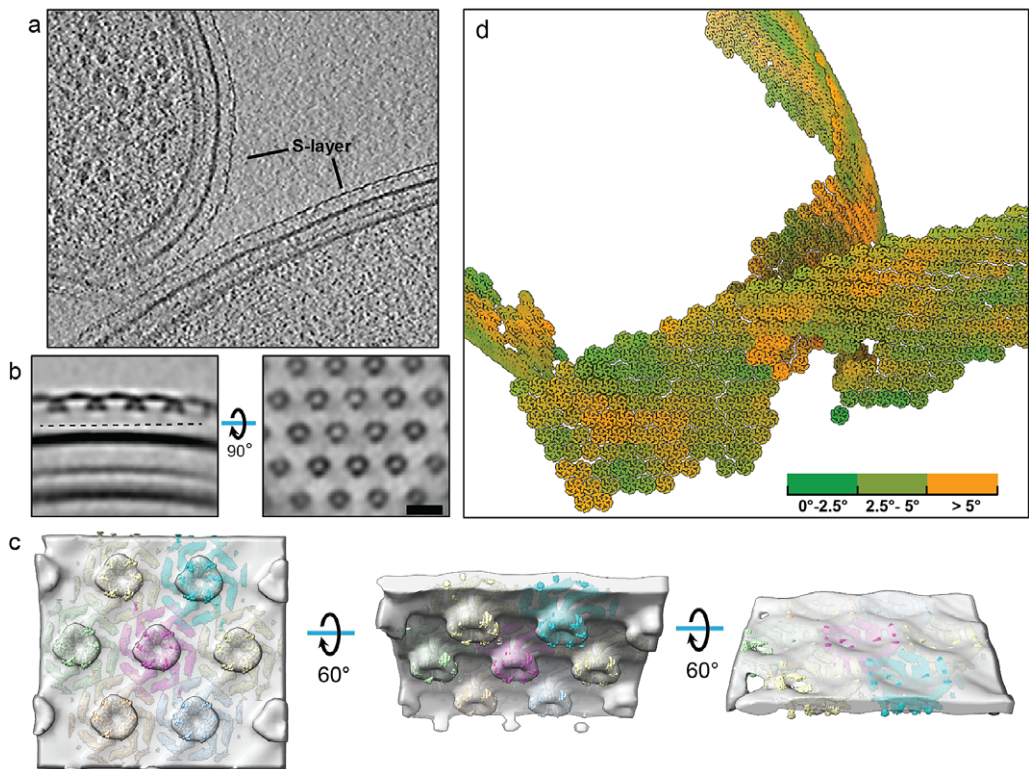
**Figure 8. TomoNet application to eukaryotic axoneme.** (a) Orthogonal slice views of axoneme structure of *T. brucei*, showing each DMT is a one-dimensional lattice. (b) Subtomogram placing back according to TomoNet “Auto Expansion” picking result with each DMT colored differently.

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 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 appear flat and split at the air-water interface, while the middle part of the lattice appears to be more curved.

### 2.9. 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 *T. brucei*<sup>21,47</sup> and the amyloid-like sheath protein on  $\beta$ -hoops of the prototypical archaeon, *Methanospirillum hungatei*.<sup>58</sup> In the case of 3D lattices, TomoNet has been also used to obtain the paraflagellar rod structure of *T. brucei*.<sup>20</sup>



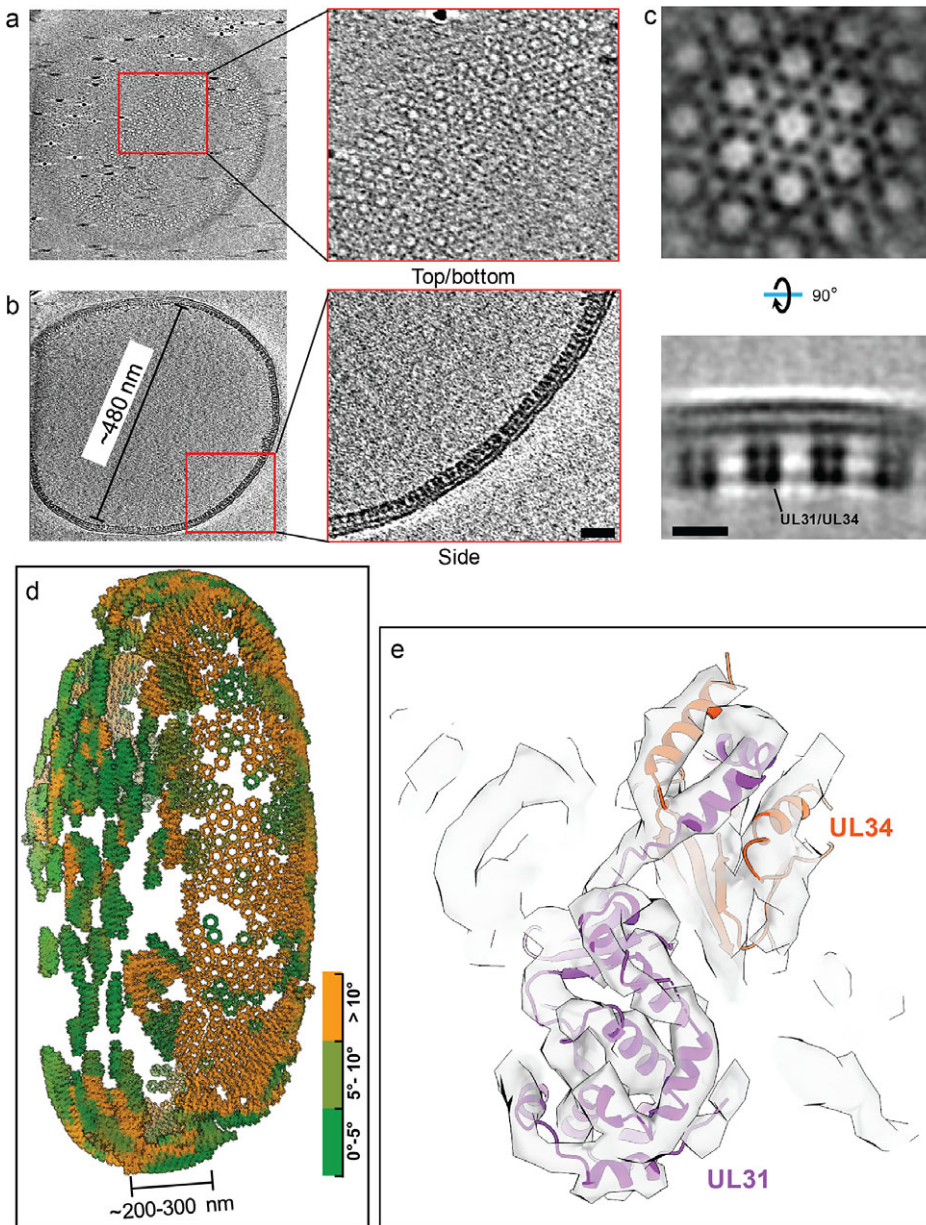
**Figure 9. TomoNet application to S-layer structure in FIB-milled cellular sample.** (a) A tomographic slice view shows two *C. crescentus* cells in a FIB-milled lamella. (b) Orthogonal slice views of the averaged density map generated in TomoNet, showing the hexagonal distribution of S-layer inner domains. Scale bar is 20 nm. (c) Our binned averaged map (transparent) docked with 7 copies of EMD-10388 (colored). (d) Visualization of S-layer lattices generated by placing back hexamer subunit maps simulated from PDB: 6P5T. Coloring is based on surface curvatures at the center of each subunit.

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 STA toward 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.

### 3. 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





**Figure 10. TomoNet application to in vitro assembled NEC-bound membrane.** (a, b) Tomographic slice views show a large NEC lattice; the insets show different views of NEC hexamer subunits. Scale bar is 20 nm. (c) Orthogonal slice views of an averaged density map generated in TomoNet show that NEC hexamer subunits consist of UL31/UL34 heterodimers. Scale bar is 10 nm. (d) Visualization of an NEC lattice generated by placing back averaged maps shows that the large vesicle is compressed into a disk-like shape. The compression caused by sample freezing stretched the lattice, making it flat and split at the air-water surface. Coloring is based on surface curvatures at the center of each subunit. (e) Atomic model of the UL31/UL34 heterodimers fits into the final averaged map, with all helices well resolved.

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. First, 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. Second, it utilizes GPUs for fast convolution operations, enabling particle prediction in just several minutes for each tomogram. Third, 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,<sup>16</sup> EMAN2,<sup>4</sup> M,<sup>59</sup> 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.

## 4. 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 leaving the high-resolution 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).

### 4.1. 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,<sup>42</sup> IMOD<sup>43</sup> or AreTomo,<sup>60</sup> and CTFFIND4,<sup>44</sup> 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 beam-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.

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.

#### 4.2. 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).

#### 4.3. 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, and so forth. 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.

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.

#### 4.4. Design and implementation of the “AI AutoPicking” module

The “AI AutoPicking” module comprises three 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. First, the size of the 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. Second, 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 predicts whether a voxel is near the center of a particle. The network architecture used is derived from the one used in IsoNet<sup>50</sup> 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 generates 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.

#### 4.5. 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 particle 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 particle information as input, then generates a ChimeraX<sup>61</sup> 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).

#### 4.6. 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.<sup>40</sup> Four tilt series, TS\_01, TS\_43, TS\_45, and TS\_54, were used in this study. Downloaded micrographs were loaded into the TomoNet 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. First, 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. Second, one run of “Auto Expansion” was applied on the above two tomograms to get more particles, such as to refine the reference. Third, 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.” Fourth, 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.<sup>62</sup> The global resolution reported is based on the “gold standard” refinement procedures and the 0.143 FSC criterion (Figure 6c).

#### 4.7. Processing one tomogram of *T. brucei* Axoneme

The tomogram of *T. brucei* axoneme is from our previous work.<sup>21</sup> Initially, one “seed” particle was manually picked for each DMT, followed by four 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.

#### 4.8. Processing one tomogram of *C. crescentus* S-layer

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

#### 4.9. Processing tomograms of NEC budding in vitro

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

#### 4.10. 3D visualization

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

**Supplementary material.** The supplementary material for this article can be found at <http://doi.org/10.1017/S2633903X24000060>.

**Data availability statement.** TomoNet code is available on the Github website at <https://github.com/logicvay2010/TomoNet>, with a user manual. For the HIV VLPs dataset, the raw data was downloaded from the Electron Microscopy Public Image Archive (EMPIAR) with accession code EMPIAR-10164,<sup>40</sup> the Gag atomic model was downloaded from the Protein Data Bank (PDB) with accession code 5L93.<sup>40</sup> For the *C. crescentus* S-layer dataset, the reconstructed tomogram was downloaded from the Electron Microscopy Data Bank (EMDB) with accession code EMD-23622,<sup>54</sup> and the subunit model was generated using an atomic model with PDB accession code 6P5T.<sup>63</sup> The STA results of NEC hexamer<sup>57</sup> and HIV can be obtained from EMDB with accession codes EMD-40224 and EMD-43869, respectively.

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

**Author contribution.** H.W. and Z.H.Z. initialized and Z.H.Z. supervised research. H.W. wrote the code and developed the software GUI with help from S.L. H.W., S.L., and X.Y. tested the software on different datasets. H.W. and Z.H.Z. wrote the manuscript. X. Y. and J.Z. assisted in writing of article. All authors reviewed and approved the paper.



**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.).

**Competing interest.** The authors declare none.

## References

- Kimanius D, Dong L, Sharov G, Nakane T and Scheres SHW (2021) New tools for automated cryo-EM single-particle analysis in RELION-4.0. *Biochemical Journal* **478**, 4169–4185. <https://doi.org/10.1042/BCJ20210708>.
- Punjani A, Rubinstein JL, Fleet DJ and Brubaker MA (2017) cryoSPARC: Algorithms for rapid unsupervised cryo-EM structure determination. *Nature Methods* **14**, 290–296. <https://doi.org/10.1038/nmeth.4169>.
- Wan W and Briggs JA (2016) Cryo-electron tomography and subtomogram averaging. *Methods in Enzymology* **579**, 329–367. <https://doi.org/10.1016/bs.mie.2016.04.014>.
- Chen M, Bell JM, Shi X, Sun SY, Wang Z and Ludtke SJ (2019) A complete data processing workflow for cryo-ET and subtomogram averaging. *Nature Methods*, **16**(11), 1161–1168. <https://doi.org/10.1038/s41592-019-0591-8>.
- Zhang P (2019) Advances in cryo-electron tomography and subtomogram averaging and classification. *Current Opinion in Structural Biology* **58**, 249–258. <https://doi.org/10.1016/j.sbi.2019.05.021>.
- Castano-Diez D and Zanetti G (2019) In situ structure determination by subtomogram averaging. *Current Opinion in Structural Biology* **58**, 68–75. <https://doi.org/10.1016/j.sbi.2019.05.011>.
- Hong Y, Song Y, Zhang Z and Li S (2023) Cryo-electron tomography: the resolution revolution and a surge of in situ Virological discoveries. *Annual Review of Biophysics* **52**, 339–360. <https://doi.org/10.1146/annurev-biophys-092022-100958>.
- Huang Y, Zhang Y & Ni T (2023) Towards in situ high-resolution imaging of viruses and macromolecular complexes using cryo-electron tomography. *Journal of Structural Biology* **215**, 108000. <https://doi.org/10.1016/j.jsb.2023.108000>.
- Sibert BS, Kim JY, Yang JE, Hannon-Hatfield A, Ke Z, Garfinkel DJ and Wright ER (2023) Workflow for high-resolution sub-volume averaging from heterogenous viral and virus-like assemblies. *Microscopy and Microanalysis*, **29**(Suppl. 1), 943–944. <https://doi.org/10.1093/micmic/ozad067.470>.
- Kopylov M, Bobe D, Johnston JD and Paraan RM (2023) Modern tools for in-situ tomography. *Microscopy and Microanalysis* **29**, 954–955. <https://doi.org/10.1093/micmic/ozad067.476>.
- Ni T, Frosio T, Mendonça L, Sheng Y, Clare D, Himes BA and Zhang P (2022) High-resolution in situ structure determination by cryo-electron tomography and subtomogram averaging using emClarity. *Nature Protocols* **17**, 421–444. <https://doi.org/10.1038/s41596-021-00648-5>.
- Xue L, Lenz S, Zimmermann-Kogadeeva M, Tegunov D, Cramer P, Bork P, Rappsilber J and Mahamid J (2022) Visualizing translation dynamics at atomic detail inside a bacterial cell. *Nature* **610**, 205–211. <https://doi.org/10.1038/s41586-022-05255-2>.
- Zivanov J, Otón J, Ke Z, von Kügelgen A, Pyle E, Qu K, Morado D, Castaño-Diez D, Zanetti G, Bharat TA, Briggs JA and Scheres SH (2022) A Bayesian approach to single-particle electron cryo-tomography in RELION-4.0. *eLife*, **11**. CLOCKSS. <https://doi.org/10.7554/eLife.83724>.
- Obr M and Schur FKM (2019) Structural analysis of pleomorphic and asymmetric viruses using cryo-electron tomography and subtomogram averaging. In *Complementary Strategies to Understand Virus Structure and Function*, Elsevier, pp. 117–159. <https://doi.org/10.1016/bs.aivir.2019.07.008>.
- Castano-Diez D, Kudryashev M, Arbeit M and Stahlberg H (2012) Dynamo: A flexible, user-friendly development tool for subtomogram averaging of cryo-EM data in high-performance computing environments. *Journal of Structural Biology* **178**, 139–151. <https://doi.org/10.1016/j.jsb.2011.12.017>.
- Himes BA and Zhang P (2018) emClarity: Software for high-resolution cryo-electron tomography and subtomogram averaging. *Nature Methods* **15**, 955–961. <https://doi.org/10.1038/s41592-018-0167-z>.
- Wagner T, Merino F, Stabrin M, Moriya T, Antoni C, Apelbaum A, Hagel P, Sitsel O, Raisch T, Prumbaum D, Quentin D, Roderer D, Tacke S, Siebolds B, Schubert E, Shaikh TR, Lill P, Gatsogiannis C and Raunser S (2019) SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. *Communications Biology* **2**(1). <https://doi.org/10.1038/s42003-019-0437-z>.
- Bharat TAM and Scheres SHW (2016) Resolving macromolecular structures from electron cryo-tomography data using subtomogram averaging in RELION. *Nature Protocols* **11**, 2054–2065. <https://doi.org/10.1038/nprot.2016.124>.
- Tegunov D and Cramer P (2019) Real-time cryo-electron microscopy data preprocessing with warp. *Nature Methods* **16**, 1146–1152. <https://doi.org/10.1038/s41592-019-0580-y>.
- Zhang J, Wang H, Imhof S, Zhou X, Liao S, Atanasov I, Hui WH, Hill KL and Zhou ZH (2021) Structure of the trypanosome paraflagellar rod and insights into non-planar motility of eukaryotic cells. *Cell Discovery* **7**(1). <https://doi.org/10.1038/s41421-021-00281-2>.
- Imhof S, Zhang J, Wang H, Bui KH, Nguyen H, Atanasov I, Hui WH, Yang SK, Zhou ZH and Hill KL (2019) Cryo electron tomography with volta phase plate reveals novel structural foundations of the 96-nm axonemal repeat in the pathogen *Trypanosoma brucei*. *eLife* **8**. CLOCKSS. <https://doi.org/10.7554/eLife.52058>.

22. Si Z, Zhang J, Shivakoti S, Atanasov I, Tao C-L, Hui WH, Zhou K, Yu X, Li W, Luo M, Bi G-Q and Zhou ZH (2018) Different functional states of fusion protein gB revealed on human cytomegalovirus by cryo electron tomography with Volta phase plate. *PLoS Pathogens* **14**(12), e1007452. <https://doi.org/10.1371/journal.ppat.1007452>.
23. Scaramuzza S and Castaño-Díez D (2021) Step-by-step guide to efficient subtomogram averaging of virus-like particles with dynamo. *PLoS Biology* **19**, e3001318. <https://doi.org/10.1371/journal.pbio.3001318>.
24. Bharat TAM, Riches JD, Kolesnikova L, Welsch S, Krähling V, Davey N, Parsy M-L, Becker S and Briggs JAG (2011) Cryo-electron tomography of Marburg virus particles and their morphogenesis within infected cells. *PLoS Biology* **9**(11), e1001196. <https://doi.org/10.1371/journal.pbio.1001196>.
25. Grünewald K, Desai P, Winkler DC, Heymann JB, Belnap DM, Baumeister W and Steven AC (2003) Three-dimensional structure of herpes simplex virus from cryo-electron Tomography. *Science* **302**(5649), 1396–1398. <https://doi.org/10.1126/science.1090284>.
26. Zanetti G, Prinz S, Daum S, Meister A, Schekman R, Bacia K and Briggs JA (2013) The structure of the COPII transport-vesicle coat assembled on membranes. *eLife* **2**. CLOCKSS. <https://doi.org/10.7554/eLife.00951>.
27. Böhm J, Frangakis AS, Hegerl R, Nickell S, Typke D and Baumeister W (2000) Toward detecting and identifying macromolecules in a cellular context: template matching applied to electron tomograms. *Proceedings of the National Academy of Sciences* **97**(26), 14245–14250. <https://doi.org/10.1073/pnas.230282097>.
28. de Teresa-Trueba I, Goetz SK, Mattausch A, Stojanovska F, Zimmerli CE, Toro-Nahuelpan M, Cheng DWC, Tollervy F, Pape C, Beck M, Diz-Muñoz A, Kreshuk A, Mahamid J and Zaugg JB (2023) Convolutional networks for supervised mining of molecular patterns within cellular context. *Nature Methods* **20**(2), 284–294. <https://doi.org/10.1038/s41592-022-01746-2>.
29. Moebel E, Martínez-Sánchez A, Lamm L, Righetto RD, Wietrzynski W, Albert S, Larivière D, Fourmentin E, Pfeffer S, Ortiz J, Baumeister W, Peng T, Engel BD and Kervran C (2021) Deep learning improves macromolecule identification in 3D cellular cryo-electron tomograms. *Nature Methods*, **18**(11). <https://doi.org/10.1038/s41592-021-01275-4.30>.
30. Wu S, Liu G and Yang G (2022) Fast particle picking for cryo-electron tomography using one-stage detection. *2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI)*. <https://doi.org/10.1109/isbi52829.2022.9761580>.
31. Hao Y, Wan X, Yan R, Liu Z, Li J, Zhang S, Cui X and Zhang F (2022) 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. <https://doi.org/10.1016/j.cmpb.2022.106871>.
32. Rice G, Wagner T, Stabrin M, Sitsel O, Prumbaum D and Raunser S (2023) TomoTwin: generalized 3D localization of macromolecules in cryo-electron tomograms with structural data mining. *Nature Methods* **20**(6), 871–880. <https://doi.org/10.1038/s41592-023-01878-z>.
33. Balyschew N, Yushkevich A, Mikirtumov V, Sanchez RM, Sprink T and Kudryashev M (2023) Streamlined structure determination by cryo-electron tomography and subtomogram averaging using TomoBEAR. *Nature Communications*, **14**(1). <https://doi.org/10.1038/s41467-023-42085-w>.
34. Jiménez de la Morena J, Conesa P, Fonseca YC, de Isidro-Gómez FP, Herreros D, Fernández-Giménez E, Strelak D, Moebel E, Buchholz TO, Jug F, Martínez-Sánchez A, Harastani M, Jonic S, Conesa JJ, Cuervo A, Losana P, Sánchez I, Iceta M, del Cano L, Gragera M, Melero R, Sharov G, Castaño-Díez D, Koster A, Piccirillo JG, Vilas JL, Otón J, Marabini R, Sorzano COS and Carazo JM (2022) ScipionTomo: towards cryo-electron tomography software integration, reproducibility, and validation. *Journal of Structural Biology* **214**(3), 107872. <https://doi.org/10.1016/j.jsb.2022.107872>.
35. Liu H-F, Zhou Y, Huang Q, Piland J, Jin W, Mandel J, Du X, Martin J and Bartesaghi A (2023) nextPYP: a comprehensive and scalable platform for characterizing protein variability in situ using single-particle cryo-electron tomography. *Nature Methods* **20**(12), 1909–1919. <https://doi.org/10.1038/s41592-023-02045-0>
36. Schur FKM, Hagen WJH, Rumlová M, Ruml T, Müller B, Kräusslich H-G and Briggs JAG (2014) Structure of the immature HIV-1 capsid in intact virus particles at 8.8 Å resolution. *Nature* **517**(7535), 505–508. <https://doi.org/10.1038/nature13838>.
37. Mendonça L, Sun D, Ning J, Liu J, Kotecha A, Olek M, Frosio T, Fu X, Himes BA, Kleinpeter AB, Freed EO, Zhou J, Aiken C and Zhang P (2021) CryoET structures of immature HIV Gag reveal six-helix bundle. *Communications Biology*, **4**(1). <https://doi.org/10.1038/s42003-021-01999-1>.
38. Zhao G, Perilla JR, Yufenyuy EL, Meng X, Chen B, Ning J, Ahn J, Gronenborn AM, Schulten K, Aiken C and Zhang P (2013) Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. *Nature* **497**(7451), 643–646. <https://doi.org/10.1038/nature12162>.
39. Krebs A-S, Liu H-F, Zhou Y, Rey JS, Levintov L, Shen J, Howe A, Perilla JR, Bartesaghi A and Zhang P (2023) Molecular architecture and conservation of an immature human endogenous retrovirus. *Nature Communications*, **14**(1). <https://doi.org/10.1038/s41467-023-40786-w>.
40. Schur FKM, Obr M, Hagen WJH, Wan W, Jakobi AJ, Kirkpatrick JM, Sachse C, Kräusslich H-G and Briggs JAG (2016) An atomic model of HIV-1 capsid-SP1 reveals structures regulating assembly and maturation. *Science* **353**(6298), 506–508. <https://doi.org/10.1126/science.aaf9620>.
41. Ni T, Zhu Y, Yang Z, Xu C, Chaban Y, Nesterova T, Ning J, Böcking T, Parker MW, Monnie C, Ahn J, Perilla JR and Zhang P (2021) Structure of native HIV-1 cores and their interactions with IP6 and CypA. *Science Advances* **7**(47). <https://doi.org/10.1126/sciadv.abj5715>.
42. Zheng SQ, Palovcak E, Armache J-P, Verba KA, Cheng Y and Agard DA (2017) MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nature Methods* **14**(4), 331–332. <https://doi.org/10.1038/nmeth.4193>.

43. Kremer JR, Mastrorade DN AND McIntosh JR (1996) Computer visualization of three-dimensional image data using IMOD. *Journal of Structural Biology* **116**, 71–76. <https://doi.org/10.1006/jsbi.1996.0013>.
44. Rohou A and Grigorieff N (2015) CTFIND4: Fast and accurate defocus estimation from electron micrographs. *Journal of Structural Biology* **192**, 216–221. <https://doi.org/10.1016/j.jsb.2015.08.008>.
45. Heumann JM, Hoenger A and Mastrorade DN (2011) Clustering and variance maps for cryo-electron tomography using wedge-masked differences. *Journal of Structural Biology* **175**, 288–299. <https://doi.org/10.1016/j.jsb.2011.05.011>.
46. Hall SR (1991) The STAR file: A new format for electronic data transfer and archiving. *Journal of Chemical Information and Modeling* **31**, 326–333. <https://doi.org/10.1021/ci00002a020>.
47. Shimogawa MM, Wijono AS, Wang H, Zhang J, Sha J, Szombathy N, Vadakkan S, Pelayo P, Jonnalagadda K, Wohlschlegel J, Zhou ZH and Hill KL (2023) FAP106 is an interaction hub for assembling microtubule inner proteins at the cilium inner junction. *Nature Communications* **14**(1). <https://doi.org/10.1038/s41467-023-40230-z>.
48. von Kügelgen A, Alva V and Bharat TAM (2021) Complete atomic structure of a native archaeal cell surface. *Cell Reports* **37**, 110052. <https://doi.org/10.1016/j.celrep.2021.110052>.
49. Pum D, Breitwieser A and Sleytr UB (2021) Patterns in nature—S-layer lattices of bacterial and archaeal cells. *Crystals* **11**(8), 869. <https://doi.org/10.3390/cryst11080869>.
50. Liu Y-T, Zhang H, Wang H, Tao C-L, Bi G-Q and Zhou ZH (2022) Isotropic reconstruction for electron tomography with deep learning. *Nature Communications* **13**(1). <https://doi.org/10.1038/s41467-022-33957-8>.
51. Tan A, Pak AJ, Morado DR, Voth GA and Briggs JAG (2021) 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 of the United States of America* **118**, e2020054118. <https://doi.org/10.1073/pnas.2020054118>.
52. Guo S, Saha I, Saffarian S and Johnson ME (2023) Structure of the HIV immature lattice allows for essential lattice remodeling within budded virions. *elife* **12**, e84881. <https://doi.org/10.7554/eLife.84881>.
53. Talledge N, Yang H, Shi K, Coray R, Yu G, Arndt WG, Meng S, Baxter GC, Mendonça LM, Castaño-Díez D, Aihara H, Mansky LM and Zhang W (2023). HIV-2 immature particle morphology provides insights into Gag lattice stability and virus maturation. *Journal of Molecular Biology* **435**(15), 168143. <https://doi.org/10.1016/j.jmb.2023.168143>.
54. Lasker K, Boeynaems S, Lam V, Scholl D, Stainton E, Briner A, Jacquemyn M, Daelemans D, Deniz A, Villa E, Holehouse AS, Gitler AD and Shapiro L (2022) The material properties of a bacterial-derived biomolecular condensate tune biological function in natural and synthetic systems. *Nature Communications* **13**(1). <https://doi.org/10.1038/s41467-022-33221-z>.
55. von Kügelgen A, Tang H, Hardy GG, Kureisaite-Ciziene D, Brun YV, Stansfeld PJ, Robinson CV and Bharat TAM (2020) In situ structure of an intact lipopolysaccharide-bound bacterial surface layer. *Cell* **180**(2), 348–358.e15. <https://doi.org/10.1016/j.cell.2019.12.006>.
56. Sleytr UB, Schuster B, Egelseer EM and Pum D (2014) S-layers: Principles and applications. *FEMS Microbiology Reviews* **38**, 823–864. <https://doi.org/10.1111/1574-6976.12063>.
57. Draganova EB, Wang H, Wu M, Liao S, Vu A, Gonzalez-Del Pino GL, Zhou ZH, Roller RJ and Heldwein EE (2024) The universal suppressor mutation restores membrane budding defects in the HSV-1 nuclear egress complex by stabilizing the oligomeric lattice. *PLOS Pathogens* **20**, e1011936. <https://doi.org/10.1371/journal.ppat.1011936>.
58. Wang H, Zhang J, Toso D, Liao S, Sedighian F, Gunsalus R and Zhou ZH (2023) Hierarchical organization and assembly of the archaeal cell sheath from an amyloid-like protein. *Nature Communications* **14**(1). <https://doi.org/10.1038/s41467-023-42368-2>.
59. Tegunov D, Xue L, Dienemann C, Cramer P and Mahamid J (2021) Multi-particle cryo-EM refinement with M visualizes ribosome-antibiotic complex at 3.5 Å in cells. *Nature Methods* **18**, 186–193. <https://doi.org/10.1038/s41592-020-01054-7>.
60. Zheng S, Wolff G, Greenan G, Chen Z, Faas FGA, Bárcena M, Koster AJ, Cheng Y and Agard DA (2022) AreTomo: An integrated software package for automated marker-free, motion-corrected cryo-electron tomographic alignment and reconstruction. *Journal of Structural Biology: X* **6**, 100068. <https://doi.org/10.1016/j.yjsbx.2022.100068>.
61. Meng EC, Goddard TD, Pettersen EF, Couch GS, Pearson ZJ, Morris JH and Ferrin TE (2023) UCSF ChimeraX: tools for structure building and analysis. *Protein Science* **32**(11). Portico. <https://doi.org/10.1002/pro.4792>.
62. Tan YZ, Baldwin PR, Davis JH, Williamson JR, Potter CS, Carragher B and Lyumkis D (2017) Addressing preferred specimen orientation in single-particle cryo-EM through tilting. *Nature Methods* **14**(8), 793–796. <https://doi.org/10.1038/nmeth.4347>.
63. Herrmann J, Li P-N, Jabbarpour F, Chan ACK, Rajkovic I, Matsui T, Shapiro L, Smit J, Weiss TM, Murphy MEP and Wakatsuki S (2019) A bacterial surface layer protein exploits multistep crystallization for rapid self-assembly. *Proceedings of the National Academy of Sciences* **117**(1), 388–394. <https://doi.org/10.1073/pnas.1909798116>.