nature methods

Article

https://doi.org/10.1038/s41592-024-02340-4

CryoDRGN-ET: deep reconstructing generative networks for visualizing dynamic biomolecules inside cells

Received: 8 October 2023

Accepted: 6 June 2024

Published online: 18 July 2024

Check for updates

Ramya Rangan[1,7], Ryan Feathers[1,7], Sagar Khavnekar[2], Adam Lerer[3], Jake D. Johnston[4,5], Ron Kelley[6], Martin Obr[6], Abhay Kotecha[6] ✉ & Ellen D. Zhong[1] ✉

Advances in cryo-electron tomography (cryo-ET) have produced new opportunities to visualize the structures of dynamic macromolecules in native cellular environments. While cryo-ET can reveal structures at molecular resolution, image processing algorithms remain a bottleneck in resolving the heterogeneity of biomolecular structures in situ. Here, we introduce cryoDRGN-ET for heterogeneous reconstruction of cryo-ET subtomograms. CryoDRGN-ET learns a deep generative model of three-dimensional density maps directly from subtomogram tilt-series images and can capture states diverse in both composition and conformation. We validate this approach by recovering the known translational states in *Mycoplasma pneumoniae* ribosomes in situ. We then perform cryo-ET on cryogenic focused ion beam–milled *Saccharomyces cerevisiae* cells. CryoDRGN-ET reveals the structural landscape of *S. cerevisiae* ribosomes during translation and captures continuous motions of fatty acid synthase complexes inside cells. This method is openly available in the cryoDRGN software.

Cryo-ET is an imaging technique that provides structural insights spanning cellular to molecular length scales[1–3]. By computationally combining a series of tilt images of intact cells or thinly milled lamellae, cryo-ET can visualize the architecture of whole cells in three dimensions at nanometer resolution. Further computational processing of the resulting three-dimensional (3D) tomograms with algorithms for segmentation and subtomogram reconstruction can resolve structures at sub-nanometer resolution, providing detailed snapshots of macromolecular structures and their localization in native contexts[4–13].

A major challenge in image processing workflows for cryo-ET is the analysis of structural heterogeneity within subtomogram data. Subtomogram reconstruction algorithms must cope with imaging attributes specific to cryo-ET such as the extremely low signal-to-noise ratio (SNR) in exposure-limited individual tilt images, as well as the inherent complexity from variations in conformation and composition of biomolecular complexes within cellular samples taken without purification. While some advanced methods for heterogeneity analysis have been proposed[6,14–16], the majority of subtomogram processing workflows rely on 3D classification to cluster subtomograms into a few discrete conformational states. Although this approach has been successfully used to reveal distinct states of macromolecular machines in situ[17–19], current processing workflows remain unwieldy, with many manual steps and substantial computational requirements. Furthermore, these methods are not well suited for modeling continuous heterogeneity and require specifying the number of expected states a priori, often additionally requiring user-provided masks to focus classification on

[1]Department of Computer Science, Princeton University, Princeton, NJ, USA. [2]Max Planck Institute of Biochemistry, Martinsried, Germany. [3]OpenAI, San Francisco, CA, USA. [4]Physiology and Cellular Biophysics, Columbia University, New York, NY, USA. [5]Simons Electron Microscopy Center, New York Structural Biology Center, New York, NY, USA. [6]Materials and Structural Analysis Division, Thermo Fisher Scientific, Eindhoven, the Netherlands. [7]These authors contributed equally: Ramya Rangan, Ryan Feathers. ✉e-mail: abhay.kotecha@thermofisher.com; zhonge@princeton.edu

https://doi.org/10.1038/s41592-024-02340-4

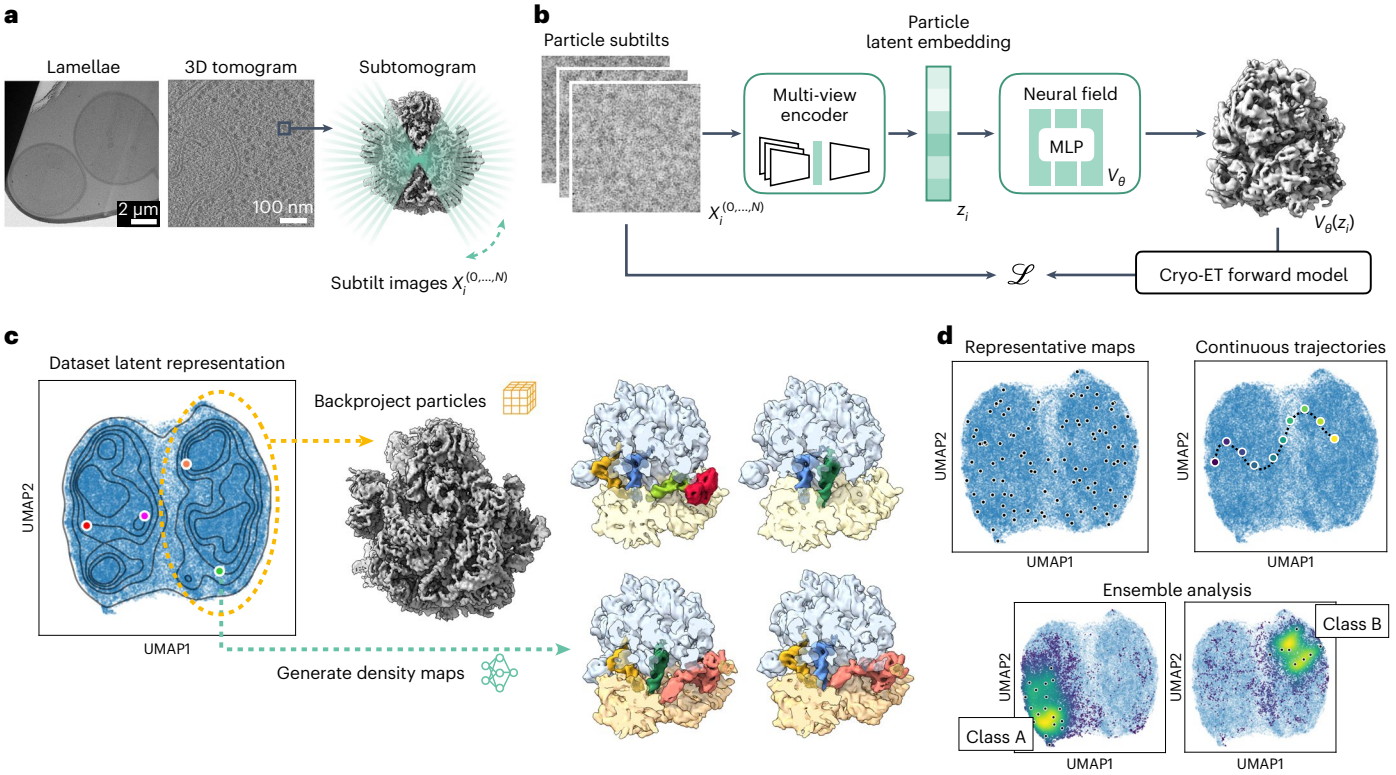


Fig. 1 | The cryoDRGN-ET method for heterogeneous reconstruction of cryo-ET subtomograms. a, Overview of tomography data acquisition and selection of subtomograms, including a schematic showing the series of tilt images that are obtained for each subtomogram. **b**, CryoDRGN-ET architecture. Particle subtilt images are transformed into a latent embedding through a multi-view encoder. The decoder includes a multilayer perceptron (MLP) that can reconstruct density maps given a particle's latent embedding. **c**, Density map generation. Once trained, particle latent embeddings can be visualized with UMAP and density maps can be generated via two approaches. First, density maps can be directly generated with the parameterized MLP given any latent embedding. For example, four density maps are generated from the colored points in the latent space, with density regions colored to show the ribosome's large subunit (blue), SSU (yellow), and different binding factors. Second, density maps can be generated by a voxel-based homogeneous reconstruction from particles selected based on the latent representation. **d**, Analysis of cryoDRGN-ET's generative model. Density maps can be systematically sampled from the latent representation (here we show *k*-means cluster centers, $k = 100$), and continuous trajectories between points in latent space can be explored. Density maps of representative states can be classified to further visualize the distributions of particle classes across the latent space.

regions with known variability. More fundamentally, 3D classification requires averaging subtomograms from thousands of particles to obtain well-resolved structures, leading to trade-offs between the number of states that can be reconstructed and the resolution of density maps of these states. While machine learning methods based on deep neural networks have shown recent successes in modeling structural variability in single-particle cryo-electron microscopy (cryo-EM)[20–22], their potential has yet to be realized in modeling heterogeneous structures from the cellular milieu.

Here, we introduce cryoDRGN-ET for heterogeneous reconstruction of cryo-ET data (Fig. 1). CryoDRGN-ET learns a deep generative model of 3D density maps directly from particle tilt images. Similar to the cryoDRGN method for single-particle analysis (SPA)[20], cryoDRGN-ET's generative model for structure is parameterized using a neural field[23] that is able to capture diverse sources of heterogeneity, including compositional changes, continuous conformational dynamics, and impurities and artifacts from imaging. Applied to a previously published in situ dataset of the bacterial ribosome, cryoDRGN-ET recapitulates the distribution of translational states in quantitative agreement with prior analyses[17], while visualizing continuous motions and membrane-associated conformations in a single model. We then performed cryo-ET on cryogenic focused ion beam–milled lamellae of *S. cerevisiae* cells and used cryoDRGN-ET to reveal the structural landscape of the *S. cerevisiae* eukaryotic ribosome. Finally, we showcased cryoDRGN-ET by analyzing the structure and conformational heterogeneity of the *S. cerevisiae* fatty acid synthase (FAS). CryoDRGN-ET is open-source software available in version 3.0 of the cryoDRGN software package (https://cryodrgn.cs.princeton.edu/).

Results

Heterogeneous subtomogram reconstruction with cryoDRGN-ET

CryoDRGN-ET is a generative neural network method for reconstructing a distribution of density maps from cryo-ET subtomogram data. As in the cryoDRGN method[20], cryoDRGN-ET uses a coordinate-based neural network to represent 3D density maps that is conditioned on a continuous latent variable $z \in \mathbb{R}^N$ for modeling heterogeneity. To train cryoDRGN-ET, we extend the standard cryo-EM image formation model to tomography (Methods). Unlike in SPA where a single projection image is captured for each particle, in tomography, multiple projections of the same particle are captured from different tilt angles (Fig. 1a). Instead of operating on 3D subtomograms extracted for each particle, which can suffer from challenges due to the missing wedge and higher computational requirements, cryoDRGN-ET processes the series of tilt images for each particle, referred to as 'subtilts', an approach followed by other recent methods[6,10,11,24]. CryoDRGN-ET aggregates the different tilt images for each particle with a multi-view encoder that outputs a vector z_i, also referred to as a latent embedding, representing the conformational state of particle i (Fig. 1b). Then, given this latent embedding z_i, cryoDRGN-ET's generative model outputs a 3D density map V_i. This map can then be rendered as two-dimensional (2D) projections

corresponding to each particle's tilt images given the image pose and estimated contrast transfer function (CTF) parameters. We model high-frequency signal attenuation as a function of electron exposure dose in the CTF for each tilt image[25]. A maximum likelihood objective is used to compare the rendered 2D projections against the observed tilt images. We additionally implement computational enhancements in the cryoDRGN software for handling large subtomogram datasets containing millions of particle images (Methods).

Once training is complete, cryoDRGN-ET provides a per-particle estimate of the dataset's heterogeneity that can be analyzed through multiple approaches (Fig. 1c,d). The distribution of latent embeddings of the particles in the dataset can be visualized in 2D, for example, with principal component analysis or uniform manifold approximation and projection (UMAP[26]; Fig. 1c). In cryoDRGN-ET's neural representation of 3D density, a density map can also be generated from any point in the latent space (Fig. 1c) and visualized to interpret the latent space and explore the conformational distribution. To more systematically analyze the observed heterogeneity, a large number of representative density maps can be used to sort particles into distinct 'classes', and trajectories can be generated by sampling maps along continuous paths in the latent space (Fig. 1d). Finally, observed states may be validated by selecting constituent particles in particle classes ('ensemble analysis', Fig. 1d) and performing a traditional homogeneous reconstruction using voxel-based backprojection, which we newly implement in the cryoDRGN software suite (Fig. 1c).

CryoDRGN-ET recovers bacterial ribosome translational states

To test cryoDRGN-ET's subtomogram analysis, we applied it to a previously published in situ dataset of the *Mycoplasma pneumoniae* bacterial ribosome after chloramphenicol (Cm) treatment[4], comparing against prior conventional 3D classification on this dataset[17]. Ribosome subtilts and their associated pose and CTF parameters were obtained from a 3.6 Å consensus refinement. We first assessed these particles' quality by training cryoDRGN-ET on all 18,466 particles, identifying outliers and particles that produce poor ribosome density maps to exclude in further training runs (Extended Data Fig. 1). We additionally assessed the resolution of the reconstruction when varying the number of tilt images used per particle, finding no further improvement when using more than eight images per particle (Extended Data Fig. 2). For subsequent cryoDRGN-ET analysis for this dataset, we trained on the first ten tilts in the dose-symmetric tilt acquisition scheme for each particle.

A cryoDRGN-ET network trained on the cleaned set of 16,655 particles reconstructed density maps displaying known compositional and conformational heterogeneity in the bacterial ribosome, including varying tRNA occupancy at the A-P-E sites, the appearance of elongation factors, subunit rotation and local motions (Fig. 2). Density maps from cryoDRGN-ET recapitulated the major translational states previously identified in this Cm-treated *M. pneumoniae* ribosome dataset[17]: the P state; EF-Tu-tRNA, P state; A, P state; and A*, P/E state (Fig. 2a). These states show density for tRNAs and elongation factor EF-Tu in the expected sites, with rotation of the small subunit (SSU) most visible in the A*, P/E state as expected[17]. We classified 100 representative density maps across the latent space to assign particles into these four states, dividing the latent space into four distinct regions (Fig. 2b and Supplementary Videos 1 and 2). This classification enabled quantifying the relative occupancy of ribosomes in these four states, with most particles representing the A, P state and other minor state populations similar to those found previously by conventional 3D classification[17] (Fig. 2c). These state distributions remained mostly consistent in earlier training epochs, with different latent variable dimensions, and when using input poses with small additional errors (Supplementary Figs. 1–3).

To validate cryoDRGN-ET density maps and our class assignments, we verified that we could reproduce the structures from homogeneous reconstruction of each state's particles (Supplementary Fig. 4). In addition, these reconstructions confirmed the rotation of the SSU, with the subunit rotated in the A*, P/E state relative to the A, P state (Extended Data Fig. 3). We compared the estimated resolution from homogeneous reconstruction for each particle class to those found previously by conventional 3D classification[17], finding similar relative resolution between states and higher resolution in three cases perhaps due to improved particle classification. Unlike the original analysis of this dataset, which relied on multiple rounds of 3D classification[17], cryoDRGN-ET can recover all states in a single round of training without the use of masks to focus on regions of expected variability in the tRNA channel. Since the A, P state included the most particles, a homogeneous reconstruction of this state produced the highest resolution map, with a global estimated resolution of 3.8 Å (Fig. 2d).

Beyond variation in the tRNA channel and factor-binding sites, cryoDRGN-ET density maps exhibited local dynamics, additional protein factor variability and larger-scale background variation (Supplementary Video 3). For instance, some cryoDRGN-ET maps showed the L1 stalk in the open state, while others included the closed state, with L1 stalk closed conformations overlapping with the A*, P/E state and partially with the A, P state (Fig. 2e). We validated the observed L1 stalk conformations with a conventional homogeneous reconstruction from each conformation's particles (Supplementary Fig. 5). Additionally, some cryoDRGN-ET maps showed density for the N-terminal domain (NTD) of the L7/L12 protein (Extended Data Fig. 4a,b), which is often challenging to resolve on ribosomes with SPA[27]. Finally, cryoDRGN-ET density maps exhibited larger-scale background variation when using all 41 tilts per particle, showing density for ribosomes bound to the cell membrane in the expected orientation (Extended Data Fig. 4a,c)[28], along with density for neighboring ribosomes in polysomes, also in a canonical orientation (Extended Data Fig. 4a,d)[29]. We expect that analysis of background variation will be further enhanced when training cryoDRGN-ET on particle sets with larger box sizes that include more surrounding context.

The native structural landscape of the *S. cerevisiae* ribosome

We further showcase cryoDRGN-ET by analyzing a large cryo-ET dataset of the *S. cerevisiae* eukaryotic ribosome using tomograms collected from lamellae that were milled with cryogenic plasma focused ion beam milling (cryo-PFIB). With cryoDRGN-ET, we analyzed heterogeneity in the *S. cerevisiae* ribosome from in situ cryo-ET, recapitulating known translational states and factor-binding events along with expected continuous conformational motions and spatial background variability. Since we again found that using a subset of tilt images enables high-resolution reconstructions (Extended Data Fig. 5a,b), we carried out all training runs using ten tilt images per particle for enhanced computational efficiency on this larger dataset of 119,031 particles. We began with a cryoDRGN-ET training run on the complete particle set, finding that the UMAP latent space representation separated particles into three classes, corresponding to rotated SSU, non-rotated SSU and a group of outlier particles (Fig. 3a). A homogeneous reconstruction of this outlier group yielded a very noisy map resembling broken particles (Fig. 3a), and removing this group of 25,750 particles did not impact the resolution of the consensus reconstruction (Extended Data Fig. 5c). Homogeneous reconstruction of the two remaining particle classes verified that they corresponded to the SSU rotated and non-rotated states (overlaid in Fig. 3a), and a subsequent cryoDRGN-ET training run excluding outlier particles reproduced the separation of SSU rotated and SSU non-rotated particles (Fig. 3a). To focus model capacity on further delineating translational states, we trained separate cryoDRGN-ET models on the SSU non-rotated and SSU rotated particles (Fig. 3b).

Density maps sampled from cryoDRGN-ET training on SSU non-rotated and rotated particles could be classified primarily into four

https://doi.org/10.1038/s41592-024-02340-4

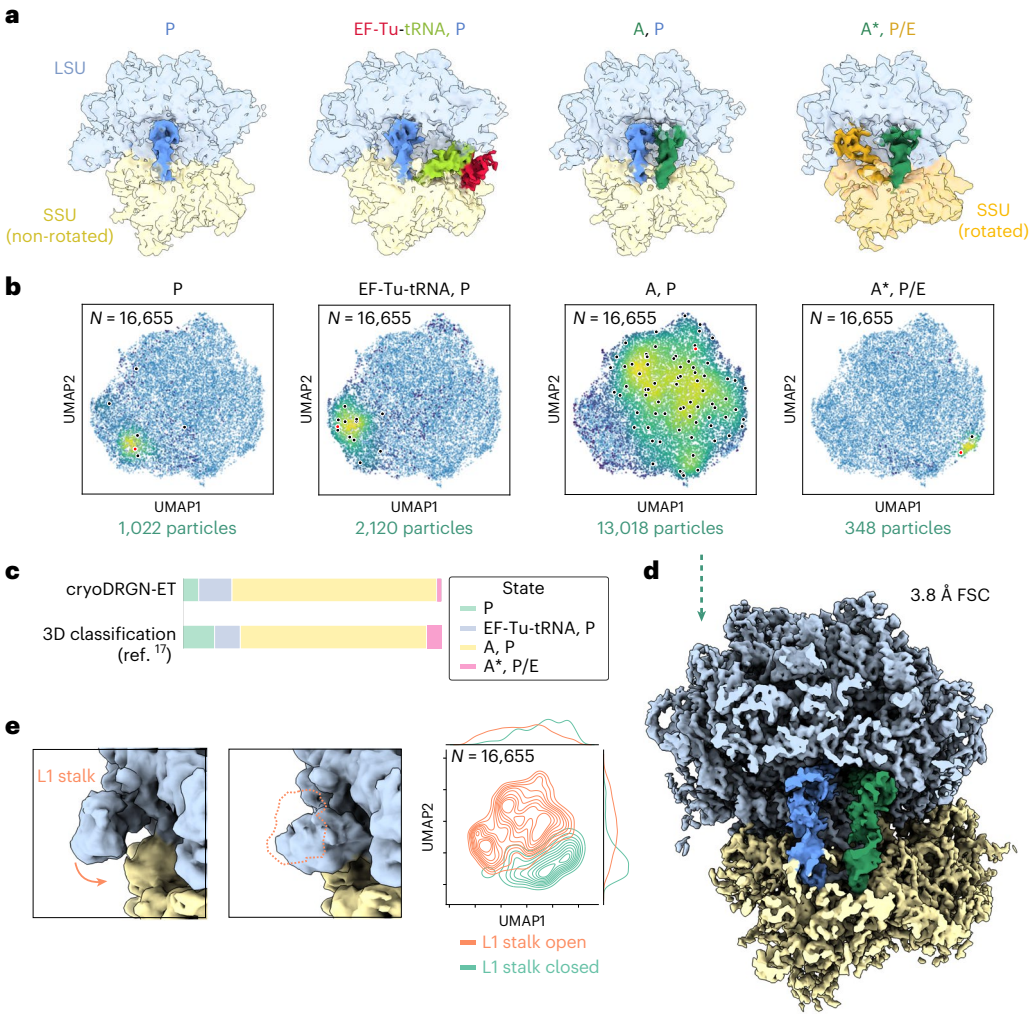


Fig. 2 | CryoDRGN-ET models translational states from an in situ subtomogram dataset of the chloramphenicol-treated *M. pneumoniae* 70S ribosome (EMPIAR-10499)[4]. **a**, Representative density maps from cryoDRGN-ET depicting four translational states. **b**, UMAP visualization of latent embeddings for all particles included in cryoDRGN-ET training, with overlaid heat maps highlighting the particles belonging to each translational state. Latent embeddings producing representative density maps in **a** are indicated in red. **c**, Quantification of particle populations in each observed translational state, compared between cryoDRGN-ET and prior conventional 3D classification[17]. **d**, High-resolution reconstruction from voxel-based backprojection in cryoDRGN-ET for particles assigned to the A, P state. **e**, Representative maps from cryoDRGN-ET displaying the L1 stalk open (left) and closed (middle) conformations. A kernel density estimate plot displaying the distribution of the two L1 stalk conformations in the latent space (right).

translational states (Fig. 3b), providing in situ evidence for these functional states and their relative populations in *S. cerevisiae* ribosomes (Supplementary Video 4). Most representative maps from the SSU non-rotated particles corresponded to the eEF1A, A/T, P state, a stage before peptidyl transfer. Indeed, the eEF1A, A/T, P state was the most populated across this *S. cerevisiae* in situ ribosome dataset (Extended Data Fig. 6a), agreeing with recent characterization of eukaryotic ribosomes in situ from *Dictyostelium discoideum* and human cells[18,19]. The conformation of eEF1A and the A/T tRNA in this state is aligned with codon sampling rather than codon recognition (Extended Data Fig. 6b)[19]. Next, we noted a class of representative maps from the SSU rotated particle set corresponding to the A, P state, accounting for the second largest particle population (Fig. 3b and Extended Data Fig. 6a). Finally, from the SSU rotated particle set, we noted representative maps corresponding to two post-translocation states: the eEF2, P, E state and the eEF2, P state. We validated particle classification into these four states through homogeneous reconstructions, finding that the resulting reconstruction for each class reproduced expected tRNA and factor density (Fig. 3c, Extended Data Fig. 6 and Supplementary Fig. 6). When fitting atomic models into these reconstructions, we observed expected SSU motion across these states, with SSU rolling and rotation visible in the A, P state and SSU rotation visible in the post-translocation states (Extended Data Fig. 7). Because the eEF1A, A/T, P state and the A, P state had the highest particle populations, reconstructions from these states produced the highest resolution maps at a global estimated resolution of 4.4 Å and 4.7 Å, respectively (Fig. 3d, Extended Data Fig. 6 and Supplementary Fig. 6).

As in the case of the bacterial ribosome, cryoDRGN-ET was able to further uncover larger-scale background variation and compositional heterogeneity for additional protein factors beyond these canonical translational states (Supplementary Video 5). For instance, representative density maps from cryoDRGN-ET included membrane-bound ribosomes (Fig. 3e) and polysomes (Extended Data Fig. 8), with polysome density visible in maps sampled from both SSU rotated and non-rotated states. While in the case of the *M. pneumoniae* ribosome using additional tilt images enabled better visualization of background heterogeneity (Extended Data Fig. 4), in this case additional tilts did not improve resolved polysomes (Supplementary Fig. 7). In addition to membrane-bound ribosomes and polysomes, we found that some representative maps included density for the initiation factor eIF5A

https://doi.org/10.1038/s41592-024-02340-4

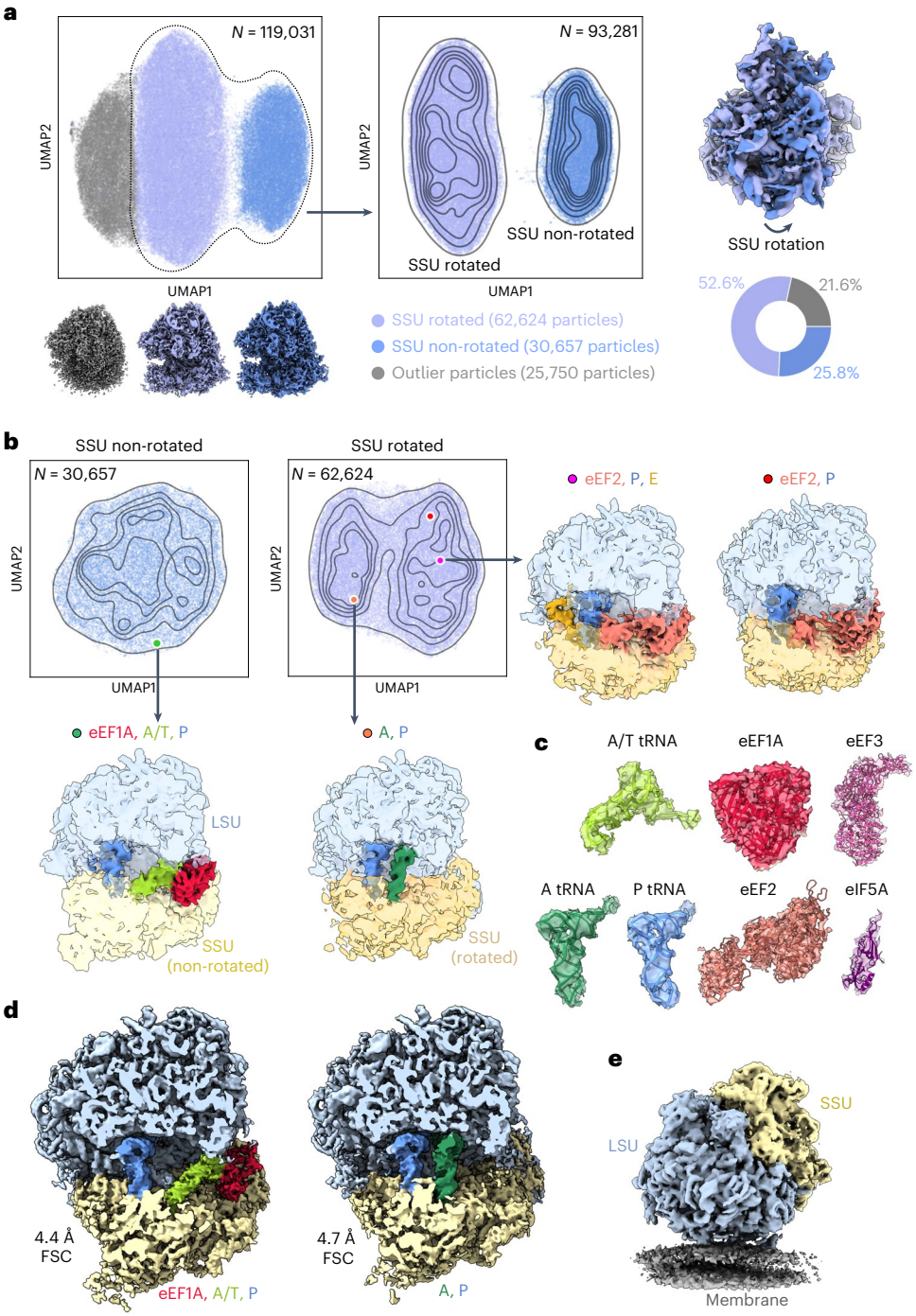


Fig. 3 | CryoDRGN-ET models in situ translational states of the *S. cerevisiae* 80S ribosome. a, UMAP visualization of cryoDRGN-ET's latent space representation of all particles (left) and after excluding a cluster of bad particles (center). Visualizations are shown as scatterplots of particle latent embeddings with the kernel density estimate overlaid. Density maps were obtained from a homogeneous reconstruction of particles from the three main clusters (bottom) and overlaid (right) to show the SSU rotation. **b**, UMAP visualization after cryoDRGN-ET training on SSU non-rotated particles (left) and SSU rotated particles (center) with representative cryoDRGN-ET density maps depicting four translational states. Latent embeddings of the representative maps are highlighted in the UMAP visualization. **c**, Atomic models were rigid-body fitted into backprojected densities for the A/T tRNA and eEF1A (in eEF1A, A/T, P state), A and P tRNA (in A, P state), eEF2 (in the eEF2, P, E state), eEF3 (in the eEF1A, A/T, P, eEF3 state) and eIF5A (in the A, P, eIF5A state). **d**, High-resolution backprojection of particles mapped to the eEF1A, A/T, P state (left) and the A, P state (right). **e**, Representative cryoDRGN-ET density map for a membrane-bound ribosome.

(Extended Data Fig. 8a)[30], and other maps included density for the uL10(P1-P2)$_2$ stalk (Extended Data Fig. 8b)[31]. Finally, we noted that some representative maps showed the presence of fungal-specific elongation factor eEF3 (Extended Data Fig. 8b). eEF3 was present only in representative density maps sampled from SSU non-rotated particles in the eEF1A, A/T, P state, aligning with prior suggestions that eEF3 binding is stabilized in non-rotated states[32]. Density for these additional factors was validated with homogeneous reconstructions of identified particles (Supplementary Figs. 8–10), for instance, with density for eIF5A and eEF3 agreeing with prior atomic models for these factors

https://doi.org/10.1038/s41592-024-02340-4

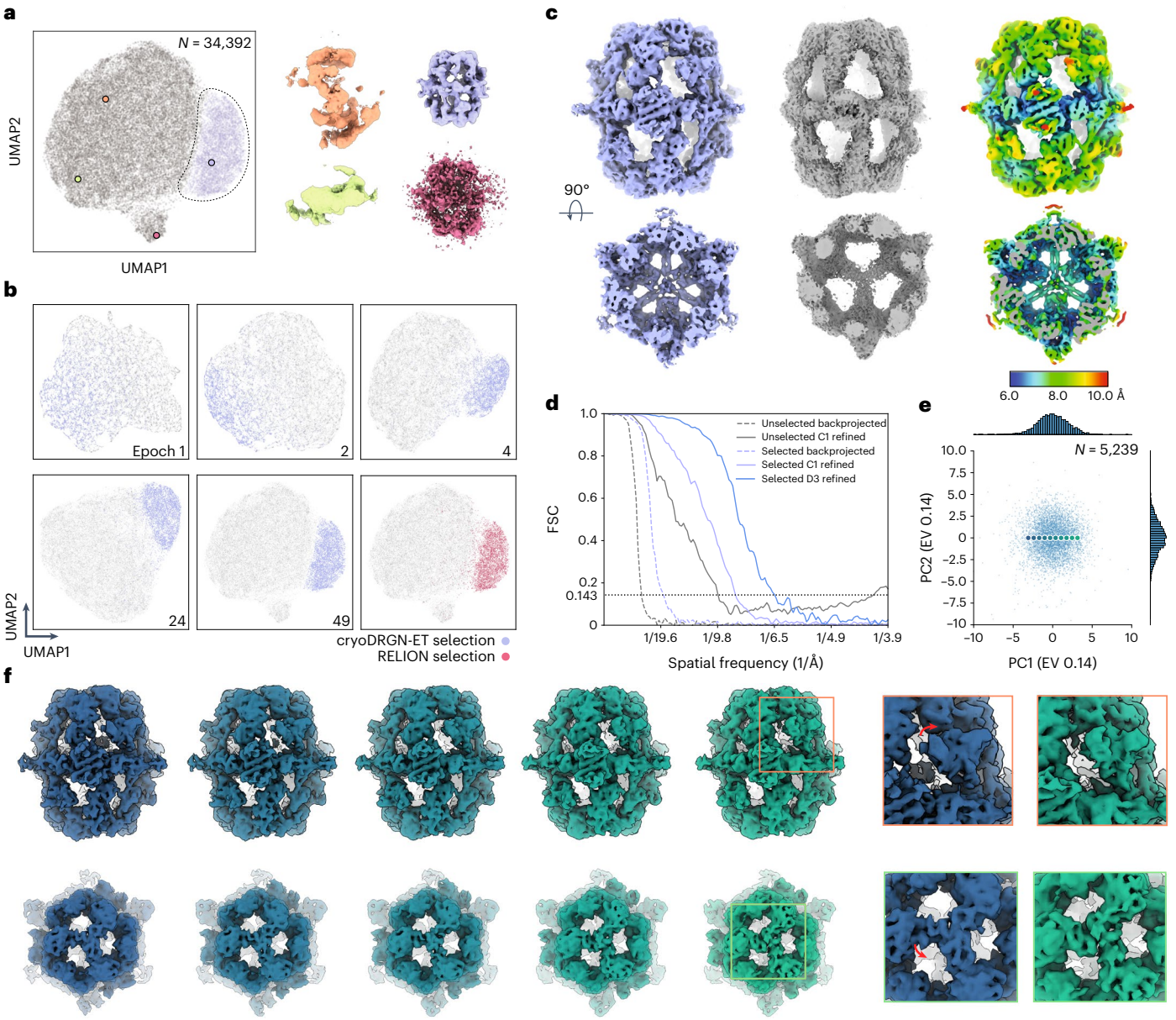


Fig. 4 | CryoDRGN-ET models in situ heterogeneity in the *S. cerevisiae* FAS. **a**, UMAP visualization of cryoDRGN-ET's latent space for all FAS particles, with four representative density maps shown sampled from points with corresponding colors in the latent space. A cluster of high-quality selected particles is highlighted in purple, and unselected particles are shown in gray. **b**, UMAP visualization of cryoDRGN-ET training run on all FAS particles over epochs of training, with particles selected from **a** highlighted in purple. In the bottom right, particles selected from RELION 3D classification are overlaid on the UMAP visualization from epoch 49. **c**, Side view (top row) and central slice (bottom row) of refined structures from RELION. We show the C1 symmetry refined map from selected particles (left), the C1 symmetry-refined map from unselected particles (middle) and the D3 symmetry-refined map from selected particles, colored by local resolution (right). **d**, GSFSC curves for refinements from RELION for selected and unselected particles (solid lines) and FSC curves between half-maps for backprojected maps from cryoDRGN-ET (dashed lines). **e**, Latent space for cryoDRGN-ET trained on selected particles from **a**, with colored points showing the trajectory across PC1. **f**, Sampled maps from cryoDRGN-ET across the PC1 trajectory from **d**, with insets showing motion in the β subunits between the first and last maps.

that were determined from purified samples (Fig. 3c). Through these analyses, we show that cryoDRGN-ET can model numerous sources of heterogeneity in *S. cerevisiae* ribosomes, providing a new analytical approach for interrogating structural distributions in situ.

In situ structure of the *S. cerevisiae* FAS

As further demonstration of using cryoDRGN-ET for resolving structural heterogeneity in situ, we next applied cryoDRGN-ET to the *S. cerevisiae* FAS. The *S. cerevisiae* FAS is a 2.6-MDa complex consisting of six copies of α and β subunits that form a large barrel-shaped structure[33–35]. To determine the structure of *S. cerevisiae* FAS in situ, particles were selected from the same cryo-PFIB tomography dataset used to study *S. cerevisiae* ribosomes. After an initial round of template matching and subtomogram averaging, we obtained a 14.3 Å consensus refinement from which 1,269,832 subtilts for 33,492 particles were exported along with corresponding refined poses and CTF parameters for cryoDRGN-ET analysis.

We first trained a cryoDRGN-ET model on the full dataset of FAS particles. When inspecting density maps sampled from the resulting latent space, we found that cryoDRGN-ET identified a cluster of 5,239

https://doi.org/10.1038/s41592-024-02340-4

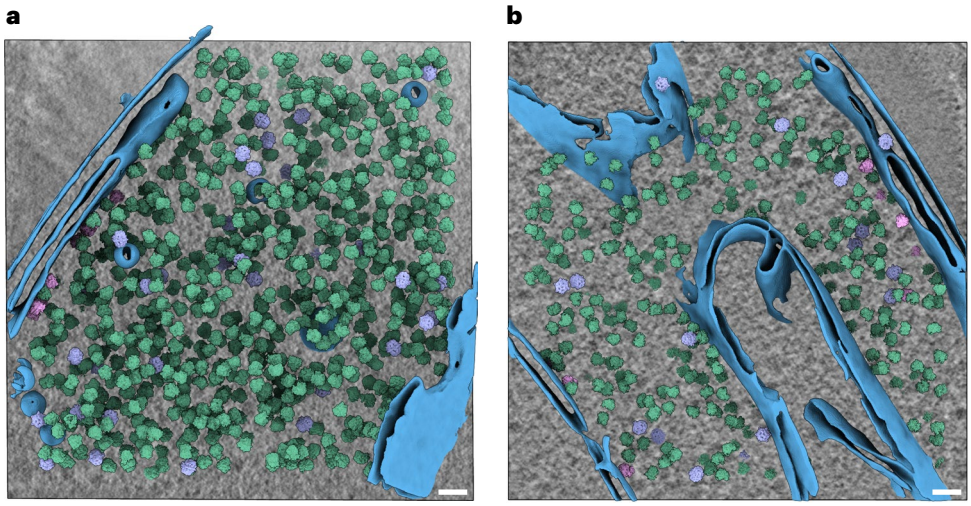


Fig. 5 | Visualization of two representative *S. cerevisiae* tomograms. a,**b**, Ribosome particles are shown in green, FAS particles in purple and membrane-bound ribosomes in pink. Membrane segmentation is shown in blue, highlighting plasma membranes and endoplasmic reticulum (**a** and **b**) and mitochondria (**b**). Scale bar, 50 nm.

particles resembling FAS (Fig. 4a). This separation of high-quality particles from noisy particles was evident even at early epochs of training (Fig. 4b). Homogeneous reconstructions with voxel-based backprojection produced maps of improved quality when using only the 5,239 filtered particles based on half-map Fourier shell correlation (FSC; Fig. 4d). To further validate this particle selection, we carried out 3D classification of the full particle set in RELION and isolated particles belonging to the well-resolved class (Extended Data Fig. 9a), finding that cryoDRGN-ET and RELION obtained a highly overlapping set of filtered particles (Fig. 4b). Next, to improve the reconstruction of the FAS map from selected particles, we refined poses in RELION for the filtered particle set from cryoDRGN-ET. Whereas an equivalent refinement of the unselected particles produced a noisy map (Fig. 4c,d), the refinement of selected particles showed improved global estimated resolution (gold-standard Fourier shell correlation (GSFSC) of 8.8 Å; Fig. 4c,d). We then carried out a D3 symmetry refinement of selected particles, further improving the estimated resolution (GSFSC of 6.8 Å; Fig. 4c,d).

The refined map for *S. cerevisiae* FAS in situ showed the density for key structural features of the complex, including the equatorial wheel created by the α subunits and the two domes made up of the β subunits[33–35]. Atomic models solved for FAS in a previously identified non-rotated state fit well with this map (Fig. 4c and Extended Data Fig. 9b)[36]. We observed low-resolution density in this map for the flexible acyl carrier protein, which carries substrates between active sites in FAS (Extended Data Fig. 9b). Here, acyl carrier protein is found by the FAS ketosynthase domain, taking the same conformation seen previously in yeast FAS structures in the non-rotated state[35–37].

We next applied cryoDRGN-ET on the selected subset of high-quality FAS particles to explore any further heterogeneity. When sampling density maps along the first principal component of the resulting latent embeddings, we observed conformational dynamics in the two domes of FAS, including rotation of the β subunits (Fig. 4e,f) and compaction of the domes (Supplementary Video 6). These motions align with previously observed rotated and non-rotated FAS structures[36] and provide a view of intermediate conformations between these states. Atomic models for the rotated and non-rotated states of FAS can be docked into maps sampled at extreme values along the first principal component axes of the latent embeddings (Extended Data Fig. 9c). Through analysis of this complex, we demonstrate that cryoDRGN-ET can isolate high-quality particles and resolve conformational heterogeneity for structures beyond ribosomes.

Visualizing molecular heterogeneity in cellular contexts

Cryo-ET provides an opportunity to visualize biomolecular complexes in their native environments. Since we processed datasets for the *S. cerevisiae* ribosome and FAS from the same tomograms, we were able to observe their relative positioning in situ by locating picked particles in tomograms (Fig. 5 and Supplementary Video 7). We focused on visualizing high-quality particles for both cases by using particle selections from cryoDRGN-ET filtering, selecting the 93,281 SSU rotated and non-rotated particles for the *S. cerevisiae* ribosome (Fig. 3a) and the 5,239 higher-quality particles from the FAS dataset (Fig. 4a). The resulting tomogram visualizations suggested that these filtered particle sets mostly avoided spurious picks, with particles remaining within the plasma membranes (Fig. 5a,b) and outside mitochondria (Fig. 5b). Ribosomes and FAS complexes are evenly distributed through the cytosol without apparent correlation between their spatial positions and orientations. When visualizing the membrane particle class corresponding to the representative volume sampled in Fig. 3e, we observed expected positioning adjacent to membranes, validating particle class selections from cryoDRGN-ET. By visualizing particle classes and continuous variation identified by cryoDRGN-ET on tomograms, it is thus possible to use cryoDRGN-ET to reveal spatial relationships between macromolecules' heterogeneity and their cellular environment.

Comparing methods for in situ heterogeneity analysis

We compared cryoDRGN-ET to more conventional approaches for 3D classification of subtomograms. As described above, 3D classification with RELION[38] reproduced selections of high-quality particles from cryoDRGN-ET (Fig. 4b and Extended Data Fig. 9a). While 3D classification was originally used to resolve the four major tRNA translational states along with other heterogeneity like opening of the L1 stalk and SSU rotation on the *M. pneumoniae* ribosome dataset[17], the analysis relied on multiple rounds of 3D classification and the use of masks around the tRNA channel and factor-binding sites. When performing 3D classification with RELION using a body mask around the full ribosome, all classes produced structures in the A, P state, the most abundant translational state (Supplementary Fig. 11a,b). In contrast, cryoDRGN-ET recovered multiple translational states in a single round of training without the use of masks to focus on specific regions with expected variability.

We additionally compared cryoDRGN-ET to a similar neural network architecture, tomoDRGN, that has been proposed to model heterogeneity in cryo-ET subtomograms[39]. This approach similarly

extends the cryoDRGN framework and thus leverages the same expressive neural field representation of density maps. However, tomoDRGN accounts for electron dose exposure using a weighting factor rather than a correction in its training objective and differs in its subtilt encoding scheme. When using tomoDRGN with default parameters, we did not observe tRNA or elongation factor heterogeneity in the *M. pneumoniae* ribosome, with all resulting density maps in the A, P state (Extended Data Fig. 10a). Additionally, tomoDRGN trained with default parameters on the *S. cerevisiae* FAS dataset shows artifactual isosurface variation rather than conformational heterogeneity between maps (Extended Data Fig. 10b). It is possible that alternate parameterizations of this network may exhibit more heterogeneity in these cases. Beyond performance on these datasets, we note that cryoDRGN-ET is integrated in the cryoDRGN software and thus inherits its features and improvements such as additional utilities for validation using voxel-based backprojection and efficient dataset loading to enable processing large datasets (Methods).

Discussion

In summary, with cryoDRGN-ET we provide new capabilities for analyzing heterogeneity within cryo-ET subtomograms. CryoDRGN-ET leverages the expressive representation power of deep neural networks to generate density maps with compositional and conformational variation from cryo-ET subtomograms. When applied to in situ datasets, we first characterized translational states of the bacterial ribosome in quantitative agreement with previous results[17]. We then newly visualized the native translational states of the *S. cerevisiae* ribosome, confirming the structural characterization from purified systems. Finally, we used cryoDRGN-ET to characterize the structure and heterogeneity of the *S. cerevisiae* FAS complex in situ. Notably, these *S. cerevisiae* structures were determined from the same cryo-ET dataset, showcasing a natural advantage of analyzing intact cells.

Instead of processing 3D subvolumes, cryoDRGN-ET operates on 2D tilt series of cropped particles (that is, subtilts), a choice also made by other recent approaches[6,10,11,24]. Although averaging subtilt-series data enables more efficient computation, these subtilts include background from the cellular environment. This background could hinder cryoDRGN-ET's performance by reducing the SNR. However, with a higher number of tilts and larger box sizes, we expect that cryoDRGN-ET is more likely to resolve some heterogeneity in neighboring particles in the cellular environment. Interestingly, in the examples here, we resolve structural heterogeneity using only a small subset of the collected tilt images (only the ten highest signal images). This aligns with prior observations that in some cases, it is possible to discard later tilt images without losing useful signal[40,41]. However, we do not assess the necessity of these additional tilts in the processing steps upstream of cryoDRGN-ET analysis.

We explored multiple hyperparameter choices for cryoDRGN-ET training, for instance, varying training lengths, latent space dimension and the number of tilt images used. We expect that the best choices for these parameters will vary between datasets depending on the number of particles, SNR in tilt-series images, the degree of heterogeneity and other factors. For instance, whereas using more tilt images per particle improved resolution of polysomes and membrane density surrounding the *M. pneumoniae* ribosome, we did not see this improvement when analyzing the *S. cerevisiae* ribosome, perhaps due to the larger number of *S. cerevisiae* ribosome particles.

Notably, cryoDRGN-ET yields a distinct estimate of structural heterogeneity for each particle (that is, a conformational state z_i and associated density map V_i) without focused masks on smaller domains to guide the search for heterogeneity during training. The relatively unbiased, per-particle heterogeneity estimate from cryoDRGN-ET can enable the joint analysis of inter-particle and intra-particle variation, potentially disentangling complex relationships between particles' conformational states, binding factor composition and spatial context. As an example, our analysis of compositional and conformational heterogeneity in the *M. pneumoniae* ribosome demonstrated that the opening of the L1 stalk correlated with factors present in the ribosome's tRNA channel.

We finally note that cryoDRGN-ET relies on image poses obtained from an upstream consensus refinement. While cryoDRGN-ET was robust to minor errors in pose estimation for individual particles, we found that larger pose deviations can limit resolved heterogeneity. Combining particle selections from cryoDRGN-ET with pose refinement in RELION could enable higher-resolution structure determination, as in the case of the FAS structure. In the future, the development of cryoDRGN-ET for subtomogram analysis can be coupled with recent developments in neural ab initio reconstruction[42,43] to avoid reliance on poses from consensus refinements. Such advances could enable studies of structural heterogeneity in more complex systems, expanding our understanding of dynamic macromolecular machinery within cells.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41592-024-02340-4.

References

1. Oikonomou, C. M. & Jensen, G. J. Cellular electron cryotomography: toward structural biology in situ. *Annu. Rev. Biochem.* **86**, 873–896 (2017).
2. Galaz-Montoya, J. G. & Ludtke, S. J. The advent of structural biology in situ by single particle cryo-electron tomography. *Biophys. Rep.* **3**, 17–35 (2017).
3. Nogales, E. & Mahamid, J. Bridging structural and cell biology with cryo-electron microscopy. *Nature* **628**, 47–56 (2024).
4. 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. *Nat. Methods* **18**, 186–193 (2021).
5. Zivanov, J. et al. A Bayesian approach to single-particle electron cryo-tomography in RELION-4.0. *Elife* **11**, e83724 (2022).
6. Himes, B. A. & Zhang, P. emClarity: software for high-resolution cryo-electron tomography and subtomogram averaging. *Nat. Methods* **15**, 955–961 (2018).
7. Chen, M. et al. A complete data processing workflow for cryo-ET and subtomogram averaging. *Nat. Methods* **16**, 1161–1168 (2019).
8. Wan, W., Khavnekar, S., Wagner, J., Erdmann, P. & Baumeister, W. STOPGAP: a software package for subtomogram averaging and refinement. *Microsc. Microanal.* **26**, 2516 (2020).
9. Khavnekar, S. et al. Optimizing Cryo-FIB lamellas for sub-5Å in situ structural biology Preprint at *bioRxiv* https://doi.org/10.1101/2022.06.16.496417 (2022).
10. Chen, M. et al. A complete data processing workflow for cryo-et and subtomogram averaging. *Nat. Methods* **16**, 1161–1168 (2019).
11. Jiménez de la Morena, J. et al. ScipionTomo: towards cryo-electron tomography software integration, reproducibility, and validation. *J. Struct. Biol.* **214**, 107872 (2022).
12. Balyschew, N. et al. Streamlined structure determination by cryo-electron tomography and subtomogram averaging using tomobear. *Nat. Commun.* https://doi.org/10.1038/s41467-023-42085-w (2023).
13. Liu, H.-F. et al. nextPYP: a comprehensive and scalable platform for characterizing protein variability in situ using single-particle cryo-electron tomography. *Nat. Methods* **20**, 1909–1919 (2023).

14. Harastani, M., Eltsov, M., Leforestier, A. & Jonic, S. HEMNMA-3D: cryo electron tomography method based on normal mode analysis to study continuous conformational variability of macromolecular complexes. *Front. Mol. Biosci.* **8**, 663121 (2021).
15. Harastani, M., Eltsov, M., Leforestier, A. & Jonic, S. TomoFlow: analysis of continuous conformational variability of macromolecules in cryogenic subtomograms based on 3D dense optical flow. *J. Mol. Biol.* **434**, 167381 (2022).
16. Erdmann, P. S. et al. In situ cryo-electron tomography reveals gradient organization of ribosome biogenesis in intact nucleoli. *Nat. Commun.* **12**, 5364 (2021).
17. Xue, L. et al. Visualizing translation dynamics at atomic detail inside a bacterial cell. *Nature* **610**, 205–211 (2022).
18. Hoffmann, P. C. et al. Structures of the eukaryotic ribosome and its translational states in situ. *Nat. Commun.* **13**, 7435 (2022).
19. Xing, H. et al. Translation dynamics in human cells visualized at high resolution reveal cancer drug action. *Science* **381**, 70–75 (2023).
20. Zhong, E. D., Bepler, T., Berger, B. & Davis, J. H. CryoDRGN: reconstruction of heterogeneous cryo-EM structures using neural networks. *Nat. Methods* **18**, 176–185 (2021).
21. Punjani, A. & Fleet, D. J. 3DFlex: determining structure and motion of flexible proteins from cryo-EM. *Nat. Methods* **20**, 860–870 (2023).
22. Chen, M. & Ludtke, S. J. Deep learning-based mixed-dimensional gaussian mixture model for characterizing variability in cryo-EM. *Nat. Methods* **18**, 930–936 (2021).
23. Xie, Y.et al. Neural fields in visual computing and beyond. Preprint at https://arxiv.org/abs/2111.11426 (2021).
24. Bartesaghi, A., Lecumberry, F., Sapiro, G. & Subramaniam, S. Protein secondary structure determination by constrained single-particle cryo-electron tomography. *Structure* **20**, 2003–2013 (2012).
25. Grant, T. & Grigorieff, N. Measuring the optimal exposure for single particle cryo-EM using a 2.6 Å reconstruction of rotavirus VP6. *Elife* **4**, e06980 (2015).
26. McInnes, L., Healy, J. & Melville, J. UMAP: uniform manifold approximation and projection for dimension reduction. Preprint at https://arxiv.org/abs/1802.03426 (2018).
27. Montesano-Roditis, L., Glitz, D. G., Traut, R. R. & Stewart, P. L. Cryo-electron microscopic localization of protein L7/L12 within the escherichia coli 70 S ribosome by difference mapping and nanogold labeling. *J. Biol. Chem.* **276**, 14117–14123 (2001).
28. Kater, L. et al. Partially inserted nascent chain unzips the lateral gate of the Sec translocon. *EMBO Rep.* **20**, e48191 (2019).
29. Brandt, F. et al. The native 3D organization of bacterial polysomes. *Cell* **136**, 261–271 (2009).
30. Shao, S. et al. Decoding mammalian ribosome-mRNA states by translational GTPase complexes. *Cell* **167**, 1229–1240 (2016).
31. Choi, A. K., Wong, E. C., Lee, K.-M. & Wong, K.-B. Structures of eukaryotic ribosomal stalk proteins and its complex with trichosanthin, and their implications in recruiting ribosome-inactivating proteins to the ribosomes. *Toxins* **7**, 638–647 (2015).
32. Ranjan, N. et al. Yeast translation elongation factor eEF3 promotes late stages of tRNA translocation. *EMBO J.* **40**, e106449 (2021).
33. Lomakin, I. B., Xiong, Y. & Steitz, T. A. The crystal structure of yeast fatty acid synthase, a cellular machine with eight active sites working together. *Cell* **129**, 319–332 (2007).
34. Leibundgut, M., Jenni, S., Frick, C. & Ban, N. Structural basis for substrate delivery by acyl carrier protein in the yeast fatty acid synthase. *Science* **316**, 288–290 (2007).
35. Jenni, S. et al. Structure of fungal fatty acid synthase and implications for iterative substrate shuttling. *Science* **316**, 254–261 (2007).
36. Singh, K. et al. Discovery of a regulatory subunit of the yeast fatty acid synthase. *Cell* **180**, 1130–1143 (2020).
37. Gipson, P. et al. Direct structural insight into the substrate-shuttling mechanism of yeast fatty acid synthase by electron cryomicroscopy. *Proc. Natl Acad. Sci. USA* **107**, 9164–9169 (2010).
38. Zivanov, J. et al. New tools for automated high-resolution cryo-em structure determination in RELION-3. *Elife* **7**, e42166 (2018).
39. Powell, B. M. & Davis, J. H. Learning structural heterogeneity from cryo-electron sub-tomograms with tomodrgn. *Nat. Methods* https://doi.org/10.1038/s41592-024-02210-z (2024).
40. Hagen, W. J., Wan, W. & Briggs, J. A. Implementation of a cryo-electron tomography tilt-scheme optimized for high resolution subtomogram averaging. *J. Struct. Biol.* **197**, 191–198 (2017).
41. Grant, T. & Grigorieff, N. Measuring the optimal exposure for single particle cryo-em using a 2.6 Å reconstruction of rotavirus VP6. *Elife* **4**, e06980 (2015).
42. Zhong, E. D., Lerer, A., Davis, J. H. & Berger, B. CryoDRGN2: ab initio neural reconstruction of 3D protein structures from real cryo-EM images. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 4066–4075 (2021).
43. Levy, A., Wetzstein, G., Martel, J. N., Poitevin, F. & Zhong, E. Amortized inference for heterogeneous reconstruction in cryo-EM. *Adv. Neural Inf. Process. Sys.* **35**, 13038–13049 (2022).

Methods

CryoDRGN-ET generative model

CryoDRGN-ET performs heterogeneous reconstruction using a neural network representation for cryo-EM structures. In particular, the central task in cryoDRGN-ET is to learn a function $\mathcal{V} : \mathbb{R}^{3+N} \rightarrow \mathbb{R}$ describing an N dimensional continuous distribution over 3D cryo-EM density maps. We use a generic latent variable $z \in \mathbb{R}^N$ to model the conformational distribution and parameterize the generative model with a coordinate-based neural network, $V_\theta(\gamma(\mathbf{k}), z)$, where θ are parameters of a multi-layer perceptron (MLP). In cryoDRGN-ET, the density map is specified in the Fourier (or the closely related Hartley) domain; thus, $\mathbf{k} \in \mathbb{R}^3$ are Cartesian coordinates representing Fourier space wavevectors. Similar to recent development in neural fields for modeling 3D signals[23], input coordinates $\mathbf{k}$ are expanded in a sinusoidal basis; instead of geometrically-spaced axis-aligned frequencies in earlier versions of cryoDRGN[20], we use frequencies sampled from a Gaussian distribution:

$$\gamma(k) = [\cos(2\pi \mathbf{B}\mathbf{k}), \sin(2\pi \mathbf{B}\mathbf{k})]$$

where entries of $\mathbf{B} \in \mathbb{R}^{M\times 3}$ are sampled from $\mathcal{N}(0, \sigma^2)$ and M is a hyperparameter. Without loss of generality, we model density maps from V_θ on the domain $[-0.5, 0.5]^3$ in our coordinate-based neural network. By default, we set $\sigma = 0.5$ for our Fourier featurization and set M to be the box size of the training images, that is, the number of pixels along each dimension of the image.

The generation of cryo-ET subtomogram tilt images follows the standard cryo-EM image formation model with modifications for tomography. We note that many existing methods for subtomogram averaging align and average many subtomogram volumes of the same particle. Alternatively, some newer approaches perform subtomogram averaging by directly aligning and averaging the 2D tilt images rather than the subvolumes[4,5], which avoids artifacts due to the missing Fourier space wedge in individual subtomograms and can be more memory efficient. In cryoDRGN-ET, we treat the subtomograms as cropped 2D tilt series. Thus, the image formation of N tilted images $X_i^{(0,\ldots,N)}$ for particle i and tilt j closely follows that from single-particle cryo-EM:

$$X_i^{(j)} = \mathrm{CTF}_{i,j} T_{t_{i,j}} S_{i,j}[V_\theta(\cdot, z_i)] + \epsilon$$

where $\mathrm{CTF}_{i,j}$ applies the CTF, ϵ is additive Gaussian white noise, $T_{t_{i,j}}$ applies a phase shift corresponding to translation by $t \in \mathbb{R}^2$ in real space, and S applies a 2D slicing operator at orientation $R \in SO(3)$ on a volume $V : \mathbb{R}^3 \rightarrow \mathbb{R}$:

$$S_i[V] : (k_x, k_y) \mapsto V\left(R_i \cdot (k_x, k_y, 0)^T\right)$$

Thus, to generate an image with our coordinate-based neural network, we first obtain oriented 3D coordinates of the 2D central slice corresponding to each pixel from the image, taking a grid of 3D pixel coordinates originally spanning $[-0.5, 0.5]^2$ on the x–y plane and rotating by the pose of each tilt image. Then given these coordinates and the latent embedding predicted for the particle, the volume decoder V_θ can render a 2D slice in the Fourier (or Hartley) domain. The phase shift corresponding to the 2D real-space translation is applied before multiplying by the CTF.

To account for accumulated radiation damage in tomography, we additionally extend the CTF to account for lower SNR in tilts collected at later time points and higher angles. First, we include an electron dose exposure correction to account for frequency-dependent signal attenuation in later tilt images, as the sample has been exposed to higher electron doses when these tilts are collected. As described previously[41], for each tilt image we compute this dose exposure correction as $e^{-\frac{N}{N_e}}$, where N is the cumulative dose accrued in the sample when this tilt image was collected, and N_e is the dose at which the SNR is $1/e$ of its starting value. Based on previous calibration[41], N_e is computed as $2.81 + 0.245e^{-1.665s}$, dependent on spatial frequency s. These dose exposure corrections are then masked to 0 for frequencies where the cumulative dose exceeds the frequency-dependent optimal exposure value $(2.51284N_e)$[41]. We multiply the CTF by these dose exposure corrections during training. Additionally, since sample thickness effectively increases at higher tilt angles leading to decreasing SNR for these tilts, we further multiply the CTF by the cosine of the tilt angle[44]. Our current implementation assumes that data are collected with a dose-symmetric tilt scheme[40].

CryoDRGN-ET training system

The overall cryoDRGN-ET architecture consists of an image encoder–volume decoder based on the variational autoencoder[45]. The above coordinate-based neural network V_θ serves as the probabilistic decoder. The image encoder, q_ϕ, embeds cryo-EM image(s) associated with each particle into a lower-dimensional latent representation. In cryoDRGN for SPA, an MLP embeds a single image into an N-dimensional latent vector. In cryoDRGN-ET for tilt-series data, the encoder aggregates multiple images of each particle from the tilt series into a single latent vector. The encoder parameterizes a diagonal Gaussian approximate posterior over the latent variable z, which we sample from during training, but take the mean value during inference. To embed a series of tilt images, the encoder is split into two MLPs, where the first learns an intermediate embedding of each image, and the second maps the concatenation of the embeddings to the latent space. When experimenting with the number of tilt images that are needed for representation learning and reconstruction, tilt images are ordered by exposure so that the highest signal images are always included.

The training objective is based on the standard variational autoencoder objective consisting of a reconstruction error as the squared error between the observed image and a rendered slice from the model and a weighted regularization term on the predicted latent representation as the Kullback–Leibler divergence between the variational posterior and a standard normal prior on z. Models are optimized with stochastic gradient descent in minibatches of tilt images from eight particles using the Adam optimizer[46] with a learning rate of 0.0001. By default, the encoder and decoder MLPs have three hidden layers of width 1,024 and ReLU activations. For the multi-view image encoder, the intermediate embedding dimension for tilt images is 64 by default. We used an eight-dimensional latent variable in all experiments. We used a constant weighting factor β of 0.025 on the Kullback–Leibler divergence term. For a summary of training and architecture hyperparameters and runtimes in all computational experiments, see Supplementary Table 2.

Computational enhancements in cryoDRGN-ET

Starting in cryoDRGN software version 3.0, we implemented new dataset loading for memory-efficient training that can scale up to large multi-million-image datasets. Cryo-ET datasets often consist of millions of tilt-series images, which are too large to fit in system memory. As minibatch stochastic gradient descent requires random subsets of particle images, cryoDRGN operates most efficiently when the entire dataset can be loaded into memory. Loading batches of images from disk in a random order induces IO bottlenecks since non-sequential reads from (non-SSD) disk typically have overheads on the order of 10 ms per read. This would impose a minimum epoch time of $4 \times 10^6 \times 10 \times 10^{-3} = 11$ h for a 4 million-image dataset.

In version 3.0, we introduce an in-memory 'shuffle buffer'. The data loader fills this buffer with buffer_size elements, and then at each batch samples batch_size elements from the buffer, replacing them with a random contiguous chunk from disk. We typically use a buffer size of 32,000 images. This approach (a) avoids loading the full dataset to memory, (b) loads data in contiguous chunks, reducing filesystem overhead, and (c) achieves good shuffling of the data into batches.

The shuffling approach is similar to the one implemented in TensorFlow[47].

For a dataset of 4.4 million 128 × 128 images, the epoch time on four A100 GPUs was 45 min using either an in-memory dataset or shuffle-buffer loading from disk, compared to an epoch time of over 4 h with naively loading images from disk in a random order.

Voxel-based homogeneous reconstruction

To enable validation of particle selections, we implemented conventional voxel-based backprojection to reconstruct density maps given particles' tilt images and poses. We populate volume slices in Fourier space with the Fourier transform of each image based on its pose, applying the CTF using the Wiener filter. The 3D reconstruction is computed from the backprojected slices as previously described in[41,48] as

$$F^W(k) = \frac{\sum_{i=1}^{N} \mathrm{CTF}_i(k) F_i(k)}{\sum_{i=1}^{N} \mathrm{CTF}_i(k)^2 + \lambda_k}$$

The optimal value of λ_k is 1/SNR(k), which can be estimated from the data[48]. However, we found that this led to over-regularization in the absence of solvent masking, and we achieved acceptable results with a constant regularization across frequencies equal to the average of the unregularized denominator across voxels

$$\lambda_k = \mathbb{E}_{k'}\left[\sum_{i=1}^{N} \mathrm{CTF}_i(k')^2\right].$$

As with cryoDRGN-ET training, we apply dose exposure and tilt angle corrections to the CTF when carrying out voxel-based backprojection.

FSC calculation

To calculate FSC curves[49] between half-maps, we use a custom script implemented in the cryoDRGN-ET software, with backprojected maps from two random halves of the dataset. Before calculating FSC curves, we apply soft real-space masks that were obtained as previously described[20]. In particular, masks are defined by first thresholding the full dataset's consensus density map at half of the 99.99th percentile density value. The mask is then dilated by 25 Å from the original boundary, and a soft cosine edge is used to taper the mask to 0 at 15 Å from the dilated boundary. To ensure that these soft masks did not lead to artefacts, we computed corrected FSC curves using maps with high-resolution noise substitution[50] for the curves shown in Extended Data Fig. 5, following the approach used in RELION[38] and cryoSPARC[51]. Phase-randomization is carried out at frequencies above 80% of the 0.143 threshold value[38]. We found that FSC curves corrected with high-resolution noise substitution were nearly identical to uncorrected curves when using these soft masks; therefore, we show the original FSC curves throughout the paper.

Bacterial ribosome dataset preprocessing (EMPIAR-10499)

Raw tilt movies were processed in Warp[4], where motion correction and patch CTF estimation were performed. The tilt-series stack was generated from Warp and the tilt series were aligned using AreTomo[52]. The tilt-series CTFs were estimated in Warp and tomograms were reconstructed in Warp at a pixel size of 10 Å, where the tomograms were denoised to enhance the contrast for particle picking. Nine denoised tomograms were manually picked in crYOLO and used to train a crYOLO model[53]. In total, an initial 32,253 particle locations were found, and the subtomograms were extracted at a pixel size 10 Å with a box size of 64 pixels. Approximately 500 subtomograms were extracted at 10 Å, and an initial model was generated using the VDAM algorithm in RELION (v.4.0)[5]. Multiple rounds of 3D classification were performed using the generated initial model to remove obvious bad particles, filtering the dataset to 25,102 particles. These subtomograms were then extracted in Warp at a pixel size of 5 Å with a box size of 128 pixels. One more round of 3D classification was performed, where 18,326 subtomograms were selected and subjected to an initial alignment in RELION 4.0 3D-autorefine with a mask surrounding the large and small subunits. These subtomograms were then extracted in Warp at 1.705 Å, with a box size of 294, where multi-particle refinement was performed in M[4] with a binary mask encompassing the large and small subunits of the ribosome. Global movement and rotation with a 5 × 5 × 41 image-space warping grid, a 8 × 8 × 2 × 10 volume-space warping grid and particle pose trajectories with three temporal sampling points were refined with five iterations. Starting at the third iteration, CTF parameters were also refined, and at iteration 4, reference-based tilt-movie alignment was performed in M. This resulted in a 3.6 Å reconstruction of the *M. pneumoniae* 70S ribosome.

Bacterial ribosome cryoDRGN-ET analysis

Particle filtering. In the initial analysis of this dataset, a standard single-particle cryoDRGN model (software version 2.3.0) was trained on the 18,655 0-degree tilt images (D = 128, 3.9 Å/pix)[20]. The encoder and decoder architectures had three hidden layers of width 256 (denoted 256 × 3), and the latent variable dimension was 8. The model was trained for 50 epochs across four A100 GPUs, taking 13 min in total. Once trained, cryoDRGN's analysis pipeline ('cryodrgn analyze') was used to visualize the latent space and produce representative density maps. Outliers were removed using cryoDRGN's interactive lasso tool on the UMAP visualization of the latent embeddings, leading to a filtered dataset of 16,655 particles. A consensus refinement of the filtered dataset yielded the same global resolution map.

Reconstruction with varying number of tilts. We carried out separate voxel-based backprojections for the filtered dataset of 16,655 particles when using 1, 2, 5, 8, 10, 16, 32 and 41 tilts per particle. When using a subset of tilts, tilts were chosen to be those with the lowest dose exposure (collected earliest in the tilt series). Local resolution estimates were performed in RELION (v4.0)[54].

CryoDRGN-ET training. A cryoDRGN-ET model was trained on the filtered dataset of 16,655 particles for 50 epochs taking 3 h and 35 min on one GPU, with the top ten tilts used during training (D = 128, 3.9 Å/pix). The encoder and decoder architectures were 1,024 × 3, and the latent variable dimension was 8. We additionally trained a cryoDRGN-ET model with all 41 tilts per particle used during training (D = 128, 3.9 Å/pix) taking 12 h and 55 min on one GPU, using the same filtered particle set and architecture settings. To explore the effects of various hyperparameters on training, we additionally trained cryoDRGN-ET models with latent variable dimensions 2, 4 and 16, and we further trained cryoDRGN-ET models with encoder and decoder architectures 256 × 3 and 512 × 3.

CryoDRGN-ET ensemble analysis. After cryoDRGN-ET training, the distribution of structures from each training run was systematically sampled by using the 'cryodrgn analyze' pipeline with k = 100, where 100 representative density maps are generated at k-means cluster centers of the latent embeddings as described previously[20]. These 100 density maps were then manually classified into four states based on tRNA and elongation factor occupancy. As each of these 100 maps is representative of a cluster of particles, we assign the particles in each k-means cluster to the same class as its k-means cluster center. Additionally, all 100 density maps were classified into either L1 open or L1 closed conformations. A representative structure of each state was manually selected for visualization in Fig. 2. Additional representative density maps with membrane-bound ribosomes, polysomes and the NTD of L7/L12 visible were selected from the 100 representative density maps for the 41-tilt training run (Extended Data Fig. 4).

High-resolution reconstruction and validation. To validate each state, the particles corresponding to each selected cluster from *k*-means clustering were combined. We then backprojected the tilt images from the high-resolution dataset (D = 294, 1.7 Å/pix) using 'cryodrgn backproject_voxel'. FSC curves between half-maps were obtained as described above to assess resolution. High-resolution backprojections were low-pass filtered to the resolution from FSC curves for visualization.

Convergence analysis. We evaluated the convergence of cryoDRGN-ET training across 50 epochs by monitoring the distribution of states and inspecting the quality of sampled volumes from different training epochs. For epochs 20, 30, 40 and 50, we independently classified 100 representative density maps into each of the four translational states. We observed a consistent distribution of states after 30 epochs of training (Supplementary Fig. 1). Sampled volumes at 20, 30, 40 and 50 epochs showed visibly similar resolution and included density for key factors (for example, A, P tRNA). To assess map quality, we computed map-to-model FSC curves between sampled volumes from each epoch of training and a previously determined in situ model of the A, P state (PDB 7PHB)[17]. Compared to epochs 10–20, sampled maps in the A, P state from epochs 30–50 were more distant from this previously solved structure, perhaps due to overfitting to noise in the dataset or additional learned heterogeneity.

Effects of cryoDRGN-ET training hyperparameters. We evaluated the distribution of translational states obtained when training with different latent embedding dimension sizes (z = 8 in Fig. 2). While cryoDRGN-ET identified less heterogeneity and missed low-occupancy translational states in runs with smaller latent embedding dimensions (z = 2 and z = 4), similar heterogeneity was obtained with a higher-dimension latent space (z = 8 and z = 16; Supplementary Fig. 2). It is possible that smaller latent space dimensions have insufficient representational capacity to model low-population translational states. We additionally evaluated the maps obtained from cryoDRGN-ET training runs with different architecture sizes (256 × 3 for the encoder and decoder architecture, and 512 × 3 for the encoder and decoder architecture). Volumes sampled from these cryoDRGN-ET training runs were qualitatively similar, but these smaller networks did not identify the minor A*, P/E ribosome population.

Analyzing effects of pose errors. To evaluate the effects of inaccuracies in particle poses required as inputs by cryoDRGN-ET, we trained cryoDRGN-ET models using poses from the consensus refinement with random perturbations of either 1°, 2°, 3°, 5° or 10°. For each experiment, we perturbed particles' poses by a fixed amount at a random orientation, and the same perturbation was applied to all tilt images for each particle. We analyzed homogeneous reconstructions and representative volumes from training runs obtained when using each set of perturbed poses. As expected, increasing rotational errors in poses led to homogeneous reconstructions with worsening resolution (Supplementary Fig. 3a,b). At larger deviations, low-occupancy translational states were not visible and more volumes were too noisy to be classified into any of the translational states (Supplementary Fig. 3c–e). However, cryoDRGN-ET was robust to smaller pose errors of 1°, producing similar heterogeneity and sampled densities (Supplementary Fig. 3c,e,f).

3D classification benchmark

Benchmarking for 3D classification was performed with RELION (v3.1)[38], analyzing subtomograms exported from the M software[4]. These subtomograms, derived from a 3.6 Å map obtained after refinement, had a box size of 294 pixels and a pixel size of 1.7005 Å. To compare against cryoDRGN-ET, classification was conducted without alignment, using the 3.6 Å map and corresponding poses from M. The map underwent a low-pass filter of 20 Å and classification was performed with a soft body mask, generating four classes per classification round. The regularization parameter varied from 1 to 10, and the resolution of E-steps was constrained between 5 and 10. All classification rounds produced maps in the A, P state, and volumes from a representative classification round are shown in Supplementary Fig. 11.

We compared the computational cost of classification with RELION with the cost of training a cryoDRGN-ET model. One round of classification at high resolution (D = 294, 1.7 Å/pix) in RELION required 43.20 h (5 MPIs, 8 threads). To compare runtime against cryoDRGN-ET at 2× binning, we also carried out classifications in RELION without alignment on subtomograms with a box size of 128 pixels at 3.9 Å/pixel. For these downsampled images (D = 128, 3.9 Å/pix), both cryoDRGN-ET and RELION are substantially faster requiring 3.58 h for cryoDRGN-ET (1 GPU) and 7.58 h for RELION (4 MPIs, 8 threads; Supplementary Table 2).

TomoDRGN analysis

TomoDRGN models were trained using software version 0.2.2 on the 70S ribosome and FAS complex subtomogram datasets. We used architecture and training settings matching the experiments reported in work by Powell et al.[39], including additional flags for dose exposure weighting, lattice masking, random tilt sampling and a latent variable dimension of 128. Resulting latent embeddings and volumes were analyzed using the same cryodrgn analyze pipeline. Networks with default architecture settings of 256 × 3 were trained for 50 epochs.

S. cerevisiae sample preparation

S. cerevisiae cells were grown in log phase conditions to an optical density at 600 nm of 0.8. In total, 4 µl of the cells was applied to a glow-discharged 200-mesh holey carbon grid copper grid (Quantifoil R1.2/3) and vitrified in a liquid ethane using Vitrobot Mark IV (Thermo Scientific) set at 4 °C and 100% humidity. The settings used were: blot force, 10; blot time, 10 s; wait time, 1 s. Samples were stored under liquid nitrogen until use. Grids were clipped in slotted Autogrids (Thermo Fisher Scientific) and subjected to automated lamella preparation using an Arctis cryo-plasma FIB (Thermo Fisher Scientific) with AutoTEM Cryo software (Thermo Fisher Scientific) as described elsewhere. Before milling, grids were coated with a layer of ion-sputtered, metallic platinum (Pt) for 30 s (Xe+, 12 kV, 70 nA). This was followed by 400-nm cryo-deposition of organometallic Pt using the gas injection system, then an additional ion-sputtered platinum layer (Xe+, 12 kV, 70 nA, 120 s). Next, grids were surveyed using Maps software (Thermo Fisher Scientific) for lamella site identification followed by automated lamella preparation using AutoTEM Cryo with a final thickness range set of 100–250 nm. All FIB milling was performed using xenon. After the final milling step, the lamellae were again sputter coated with a thin layer of ion-sputtered metallic Pt (Xe+, 12 kV, 30 nA, 8 s).

S. cerevisiae dataset acquisition

Datasets were collected using a Krios G4 equipped with XFEG, Selectris X energy filter and Falcon 4 direct electron detector (Thermo Fisher Scientific). Tilt series were collected with a dose-symmetric tilt scheme[40] using TEM Tomography 5 software (Thermo Fisher Scientific). The tilt span of ± 60° was used with 3° steps, starting at ± 10° to compensate for the lamella pre-tilt. The target focus was changed for each tilt-series in steps of 0.25 µm over a range of −1.5 µm to −3.5 µm. Data were acquired in EER mode of Falcon 4 with a calibrated physical pixel size of 1.96 Å and a total dose of 3.5 e-/Å^2 per tilt over ten frames. A 10-eV slit was used for the entire data collection. Eucentric height estimation was performed once for each lamella using the stage tilt method in TEM Tomography 5 software. Regions of interest were added manually, and positions saved. Tracking and focusing were applied before and after acquisition of each tilt step. The energy filter zero-loss peak was tuned only once before starting the data acquisition.

S. cerevisiae dataset preprocessing

The data were preprocessed using TOMOgram MANager (TOMOMAN)[55], which calls the following external packages. EER images were motion corrected using a modified implementation of RELION's motioncor[56]. The defocus was estimated using tiltCTF as implemented within TOMOMAN (tiltCTF uses CTFFIND4 for some steps[57]). Tilt series were aligned using fiducial-less alignment in AreTomo[52]. Initial tomograms without CTF correction were reconstructed using IMOD's tilt package[58] with IMOD reconstruction module in TOMOMAN. The 3D CTF corrected tomograms at 8× binning were reconstructed using novaCTF[59] module in TOMOMAN and used for template matching.

S. cerevisiae ribosome subvolume averaging

Initial particle positions for 80S ribosomes were determined using the noise correlation template matching approach implemented in STOPGAP[60]. PDB 6GQV (ref. 61 for 80S ribosomes was used to generate a template using the simulate command in cisTEM[62]. Approximately 1,000 particles per tomogram were picked from 260 tilt series. Subsequent subtomogram averaging and classification were performed using STOPGAP[60]. Then, 3D classification was performed using simulated annealing stochastic hill climbing multi-reference alignment[16,60]. The resulting 130,000 particles were then exported to Warp[63] using TOMOMAN[55]. Subtomograms were reconstructed in RELION (v3.0)[38] convention using Warp at 2× binning (3.92 Å/pix). An iterative approach with subtomogram alignment and additional 3D classification in RELION and tilt-series refinement in M was performed. First, subvolumes were aligned in RELION 3.1 using a ribosome-shaped mask. Aligned subvolumes were then classified while only performing local search for alignment with a ribosome-shaped mask. This resulted in 119,000 particles that were imported to M for multi-particle tilt-series refinement while solving for sample deformations. For final averages, 119,000 particles were reconstructed at an unbinned pixel size of 1.96 Å, and another round of subtomogram alignment in RELION was performed with a mask focused on the LSU. These particle positions were then imported to M for multi-particle tilt-series refinement[4] to solve for sample deformations and CTF. For the LSU focused reconstruction and the subsequent analysis of the structural heterogeneity of the 80S ribosome, an additional round of subtomogram alignment in RELION and subsequent tilt-series refinement in M were performed using a focused mask around the LSU. This resulted in a 4.3 Å reconstruction of the 80S ribosome. The final set of 119,000 particles was then extracted as 2D subtilts at a binning of 1× and 2× using Warp, and used for analyzing conformational heterogeneity with cryoDRGN-ET.

S. cerevisiae ribosome cryoDRGN-ET analysis

CryoDRGN-ET training: full dataset. A cryoDRGN-ET model was trained on the full dataset of 119,031 particles for 50 epochs, with the top ten tilts used during training ($D = 128$, 3.92 Å/pix). The architectures of the two encoder MLPs and decoder MLP were 1,024 × 3, and the latent variable dimension was 8. The model was trained for 26 epochs across four A100 GPUs, taking 10 h and 40 min total. Once trained, cryoDRGN-ET's analysis pipeline ('cryodrgn analyze') was used to visualize the latent space and produce representative density maps. We sampled both 20 structures for initial visualization and 100 density maps for a more comprehensive assessment. The UMAP visualization of the latent space revealed three clusters of particles, which were assigned as (1) outliers, (2) the SSU rotated particles and (3) the SSU non-rotated particles by visual inspection of representative density maps from each cluster. Particles corresponding to each cluster were selected using cryoDRGN-ET's interactive lasso tool on the UMAP visualization of the latent embeddings. A homogeneous reconstruction of each set of particles was then performed with 'cryodrgn backproject_voxel' (Fig. 3a).

Voxel-based reconstruction. We carried out voxel-based backprojections for the dataset of 93,281 SSU rotated and non-rotated particles when using 1, 2, 5, 8, 10, 16 and 32 tilts per particle. We did not explore using all 41 tilts for these comparisons and further experiments on this dataset, as many particles did not have all 41 tilt images available. We additionally carried out voxel-based backprojections with all available tilts per particle for both the full dataset of 119,031 particles and the filtered set with 93,281 particles to assess the effects of particle filtering. As before, when using a subset of tilts, tilts were chosen to be those with lowest dose exposure (collected earliest in the tilt series). Local resolution estimates were made in RELION (v4.0)[54].

CryoDRGN-ET training: hierarchical analysis. Three additional cryoDRGN-ET models were trained on the remaining good particles (93,281 particles; Fig. 3a), the SSU rotated state (62,624 particles) and SSU non-rotated state (30,657 particles; Fig. 3b). All training runs were carried out for 50 epochs, with latent variable dimension 8 and encoder and decoder MLP dimensions of 1,024 × 3. The training run on all SSU rotated and non-rotated particles took 18 h and 36 min on one A100 GPU, the training run on the SSU rotated particles alone took 12 h and 12 min on one A100 GPU, and the training run on the SSU non-rotated particles alone took 6 h and 8 min on one A100 GPU.

CryoDRGN-ET ensemble analysis. After cryoDRGN-ET training, the distribution of structures from each training run was systematically sampled by using the 'cryodrgn analyze' pipeline with $k = 100$, where 100 representative density maps were generated at k-means cluster centers of the latent embeddings. For the two training runs that separately processed SSU rotated particles and SSU non-rotated particles, we classified all 100 representative density maps into corresponding translational states. We searched density maps for the presence of factors seen previously in in situ eukaryotic ribosome datasets including the A tRNA, P tRNA, E tRNA, A/P tRNA, P/E tRNA, A/T tRNA, eEF2 and eEF1A[18,19]. More specifically, to classify density maps, we docked in ribosome structures (PDB IDs: 3J7R ref. 64, 5LZS ref. 30, 6GQV ref. 61 and 6TNU ref. 65) that included these tRNA and elongation factors, and we then inspected maps to identify the presence of factors. A representative structure for each state was manually selected for visualization in Fig. 3. We additionally identified all representative density maps that included density for eIF5A, as previous density was visible in the eIF5A factor-binding site in a prior in situ eukaryotic ribosome dataset[18]. We further pinpointed maps that included eEF3 (an essential fungal elongation factor[32]), along with maps that included uL10 and the NTD of P1 and P2. From both of these training runs, we additionally identified representative density maps that included partial density for polysomes. Finally, from the training run that included both SSU rotated and non-rotated particles together, we identified a membrane-bound representative ribosome map.

High-resolution reconstruction and validation. To validate each state, the particles corresponding to each selected cluster center from k-means clustering were combined. We then backprojected the tilt images from the high-resolution dataset ($D = 256$, 1.96 Å/pix) using 'cryodrgn backproject_voxel'. We compute FSC curves between half-maps to assess resolution. We apply a real-space mask to each volume before computing the FSC via 'cryodrgn gen_mask', generated by thresholding a density map at half of its max density, dilating the mask by 25 Å and tapering the mask to 0 over 15 Å from the dilated boundary via a soft cosine edge. High-resolution backprojections were low-pass filtered to the $FSC_{0.143}$ resolution for visualization.

To further validate translational states and the placement of factors in these states, we fit individual factors into backprojected volumes as rigid bodies and evaluated the match with surrounding density. To color factors in representative density maps and reconstructions (Fig. 3b), and to visualize the fit of individual factors in density

(Fig. 3c), atomic models for each state were assembled by docking into high-resolution reconstructions. For each translational state, we obtained atomic models for the elongation factors and tRNAs separately, along with separate atomic models for the LSU and SSU, and we docked each of these models as rigid bodies into the high-resolution reconstruction density maps with the 'fitinmap` tool in ChimeraX[66]. For the eEF1A, A/T, P state, we obtained atomic models for eEF1A, the A/T tRNA and the P tRNA from PDB 5LZS ref. 30 and the large subunit (LSU) and SSU from PDB 3J78 ref. 67. For the A, P state, the A tRNA, P tRNA, LSU and SSU were obtained from PDB 6TNU ref. 65. For the post-translocation states, the eEF2, E tRNA, LSU and SSU were obtained from PDB 6GQV ref. 61, and the P tRNA was obtained from PDB 6TNU ref. 65. For fitting in an atomic model for eIF5A, we used eIF5A model from PDB 5LZS ref. 30. For fitting in an atomic model for eEF3, we used the eEF3 model from PDB 7B7D ref. 32.

To observe SSU rotation in Extended Data Fig. 7, we fit atomic models for the SSU head, SSU body and LSU separately into high-resolution reconstructions for each visualized translation state. These three rigid bodies were obtained from an atomic model of the *S. cerevisiae* ribosome with the SSU in a non-rotated state (PDB 3J78)[67]. After these three rigid bodies were sequentially fit into each density map in ChimeraX[66], the resulting atomic models were aligned with PDB 3J78 on the large subunit to visualize SSU rotation and rolling. To visualize the SSU head swivel, models containing the SSU head and body alone were aligned on the SSU body.

S. cerevisiae FAS subvolume averaging

Initial particle positions for FAS were determined using the noise correlation template matching approach implemented in STOPGAP[60]. PDB 6TA1 (ref. 68) for FAS ribosomes was used to generate a template using the simulate command in cisTEM[62]. Approximately 200 particles per tomogram were picked from 260 tilt series. Subsequent subtomogram averaging and classification were performed using STOPGAP[60]. 3D classification was performed using simulated annealing stochastic hill climbing multi-reference alignment[60]. The resulting 33,492 particles were then exported to the Warp[63], RELION, M pipeline using TOMOMAN[55].

Subvolumes were reconstructed and aligned using Warp and RELION 3D-autorefine, first with a 64-pixel box size and 7.84 Å per pixel followed by further refinement with a 128-pixel box and 3.92 Å pixel size. The particles were then subjected to multi-particle refinement in M for two iterations using a 3 × 3 image warp grid and 3 × 3 × 2 × 10 volume warp grid. Defocus and movie frame alignment were refined only in the second iteration. The refined particles were exported as both reconstructed subvolumes and aligned image series with a box size of 300 pixels at 1.96 Å per pixel.

S. cerevisiae FAS cryoDRGN-ET analysis

For cryoDRGN-ET heterogeneity analysis, subtilts were Fourier cropped to a box size of 96 pixels resulting in a pixel size of 6.13 Å. A cryoDRGN-ET model was trained using the first ten tilts for the full dataset of 33,492 particles. Training for 50 epochs using an eight-dimensional latent variable model was completed in 3 h and 50 min on a single A100 GPU.

After training, the 'cryoDRGN analyze' pipeline was used to perform *k*-means clustering of the latent embeddings with $k = 20$ and visualize representative density maps at the cluster centers. Particles were assessed based on the manual inspection of *k*-means centers maps and selected using cryoDRGN's interactive lasso tool on the UMAP visualization of the latent embeddings. Particle indices were then used to filter the 3D subvolume star file yielding a subset of 5,239 selected 'good' particles and 28,253 unselected 'bad' particles.

Subvolumes with a box size of 300 pixels and 1.96 Å pixel size corresponding to the cryoDRGN-ET selected and unselected subsets were subjected to RELION 3D autorefine, producing a reconstruction with a global resolution of 8.8 Å estimated by RELION after processing without imposing symmetry. Particles from the refinement were then imported into M and exported as subtilts with a box size of 300 at 1.96 Å/pix. The subtilts were downsampled to a box size of 128 and a pixel size of 4.6 Å for a final round of cryoDRGN heterogeneity analysis using the same training parameters as previously described (Supplementary Table 2).

S. cerevisiae FAS subvolume classification

Subvolume 3D classification on the full FAS dataset was performed in RELION after exporting subvolumes from M using the same pose information, box size and pixel size as the subtilts used for initial cryoDRGN-ET training ($D = 96$, 6.15 Å/pix). A volume reconstructed from the downsampled particles from M was low-pass filtered to 30 Å for the initial reference and mask generation. RELION 3D classification was performed with $K = 3$, $T = 2$ to identify a subset of 4,633 particles.

Tomogram visualization

Deconvolved tomograms were constructed in Warp with a pixel size of 15.68 Å. Particle indices identified by cryoDRGN-ET analysis were used to filter the associated 3D subvolume star for membrane-associated ribosomes, selected ribosome particles (rotated and non-rotated SSU) and filtered FAS particles. Particle mapping and tomogram visualization was carried out using the plugin ArtiaX in ChimeraX[66]. Membrane surface density was generated using Membrain-Seg[69] and Segger[70].

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

S. cerevisiae raw data have been deposited to EMPIAR-11658. Maps have been deposited to the Electron Microscopy Data Bank (EMD-18197, EMD-18231, EMD-18232, EMD-45235 and EMD-45236) and cryoDRGN model weights, density maps and any associated files needed to reproduce this analysis are available on Zenodo via https://doi.org/10.5281/zenodo.11399378 (ref. 71). Atomic models used from previous studies were obtained from the PDB (7PHA, 7PHB, 7PHC, 7PH9, 3J7R, 5LZS, 6GQV, 6TNU, 3J78, 7B7D, 6TA1, 6QL5 and 6QL6). The dataset from EMPIAR-10499 was analyzed in this study.

Code availability

Software is available at https://github.com/ml-struct-bio/cryodrgn/ in version 3.0.0-beta.

References

44. Bharat, T. A., Russo, C. J., Löwe, J., Passmore, L. A. & Scheres, S. H. Advances in single-particle electron cryomicroscopy structure determination applied to sub-tomogram averaging. *Structure* **23**, 1743–1753 (2015).
45. Kingma, D. P. & Welling, M. Auto-encoding variational Bayes. Preprint at https://arxiv.org/abs/1312.6114 (2022).
46. Kingma, D. P. & Ba, J. Adam: a method for stochastic optimization. Preprint at https://arxiv.org/abs/1412.6980 (2014).
47. Abadi, M. et al. TensorFlow: large-scale machine learning on heterogeneous distributed systems. Preprint at https://arxiv.org/abs/1603.04467 (2015).
48. Sindelar, C. V. & Grigorieff, N. Optimal noise reduction in 3d reconstructions of single particles using a volume-normalized filter. *J. Struct. Biol.* **180**, 26–38 (2012).
49. Rosenthal, P. B. & Henderson, R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* **333**, 721–745 (2003).
50. Chen, S. et al. High-resolution noise substitution to measure overfitting and validate resolution in 3d structure determination by single particle electron cryomicroscopy. *Ultramicroscopy* **135**, 24–35 (2013).

51. 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 (2017).
52. Zheng, S. et al. AreTomo: An integrated software package for automated marker-free, motion-corrected cryo-electron tomographic alignment and reconstruction. *J. Struct. Biol.* **6**, 100068 (2022).
53. Wagner, T. et al. SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. *Commun. Biol.* **2**, 218 (2019).
54. 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 (2021).
55. Khavnekar, S., Erdmann, P. & Wan, W. TOMOMAN: streamlining cryo-electron tomography and subtomogram averaging workflows using TOMOgram MANager. *Microsc. Microanal.* **29**, 1020 (2023).
56. Bai, X.-c, Fernandez, I. S., McMullan, G. & Scheres, S. H. Ribosome structures to near-atomic resolution from thirty thousand cryo-EM particles. *Elife* **2**, e00461 (2013).
57. Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).
58. Mastronarde, D. N. & Held, S. R. Automated tilt series alignment and tomographic reconstruction in IMOD. *J. Struct. Biol.* **197**, 102–113 (2017).
59. Turoňová, B., Schur, F. K., Wan, W. & Briggs, J. A. Efficient 3D-CTF correction for cryo-electron tomography using NovaCTF improves subtomogram averaging resolution to 3.4 Å. *J. Struct. Biol.* **199**, 187–195 (2017).
60. Wan, W., Khavnekar, S. & Wagner, J. STOPGAP, an open-source package for template matching, subtomogram alignment, and classification. *Acta Crystallogr. D Struct. Biol.* **80**, 336–349–12 (2024).
61. Pellegrino, S. et al. Structural insights into the role of diphthamide on elongation factor 2 in mRNA reading-frame maintenance. *J. Mol. Biol.* **430**, 2677–2687 (2018).
62. Himes, B. & Grigorieff, N. Cryo-TEM simulations of amorphous radiation-sensitive samples using multislice wave propagation. *IUCrJ* **8**, 943–953 (2021).
63. Tegunov, D. & Cramer, P. Real-time cryo-electron microscopy data preprocessing with Warp. *Nat. Methods* **16**, 1146–1152 (2019).
64. Voorhees, R. M., Fernández, I. S., Scheres, S. H. & Hegde, R. S. Structure of the mammalian ribosome-Sec61 complex to 3.4 Å resolution. *Cell* **157**, 1632–1643 (2014).
65. Buschauer, R. et al. The Ccr4-Not complex monitors the translating ribosome for codon optimality. *Science* **368**, eaay6912 (2020).
66. Pettersen, E. F. et al. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci.* **30**, 70–82 (2021).
67. Svidritskiy, E., Brilot, A. F., San Koh, C., Grigorieff, N. & Korostelev, A. A. Structures of yeast 80S ribosome-tRNA complexes in the rotated and nonrotated conformations. *Structure* **22**, 1210–1218 (2014).
68. Joppe, M. et al. The resolution revolution in cryoem requires high-quality sample preparation: a rapid pipeline to a high-resolution map of yeast fatty acid synthase. *IUCrJ* **7**, 220–227 (2020).
69. Lamm, L. et al. MemBrain v2: an end-to-end tool for the analysis of membranes in cryo-electron tomography. Preprint at *bioRxiv* https://doi.org/10.1101/2024.01.05.574336 (2024).
70. Pintilie, G. D., Zhang, J., Goddard, T. D., Chiu, W. & Gossard, D. C. Quantitative analysis of cryo-EM density map segmentation by watershed and scale-space filtering, and fitting of structures by alignment to regions. *J. Struct. Biol.* **170**, 427–438 (2010).
71. Rangan, R. et al. Data for 'CryoDRGN-ET: Deep reconstructing generative networks for visualizing dynamic biomolecules inside cells'. *Zenodo* https://doi.org/10.5281/zenodo.11399378 (2024).

Acknowledgements

We thank V. Bansal and M. Grzadkowski for software engineering support, N. Fischer for feedback and insight on ribosome structures, F. Hughson for feedback on the manuscript, and F. Beck for discussions and assistance with computing resources. Some of this work was performed at the Max Planck Institute of Biochemistry with support from J. Plitzko and J. Briggs for funding and infrastructure and at the Simons Electron Microscopy Center at the New York Structural Biology Center, with support from the Simons Foundation (SF349247), National Institutes of Health (U24 GM129539) and NY State Assembly. We acknowledge that the work reported herein was substantially performed using the Princeton Research Computing resources at Princeton University, which is a consortium of groups led by the Princeton Institute for Computational Science and Engineering (PICSciE) and Office of Information Technology's Research Computing. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

A.K. and E.Z. conceived of the work and supervised; R.R., A.L. and E.Z. implemented the methods; R.R., R.F. and E.Z. performed the computational experiments; R.K. and A.K. prepared yeast samples and collected cryo-ET data; R.F., S.K., J.J. and M.O. processed raw data; R.R. and E.Z. wrote the manuscript with feedback from all authors.

Competing interests

S.K., R.K., M.O. and A.K. are employees of Thermo Fisher Scientific, a commercial entity that sells instrumentation used in this study. The other authors declare no competing interests.

Additional information

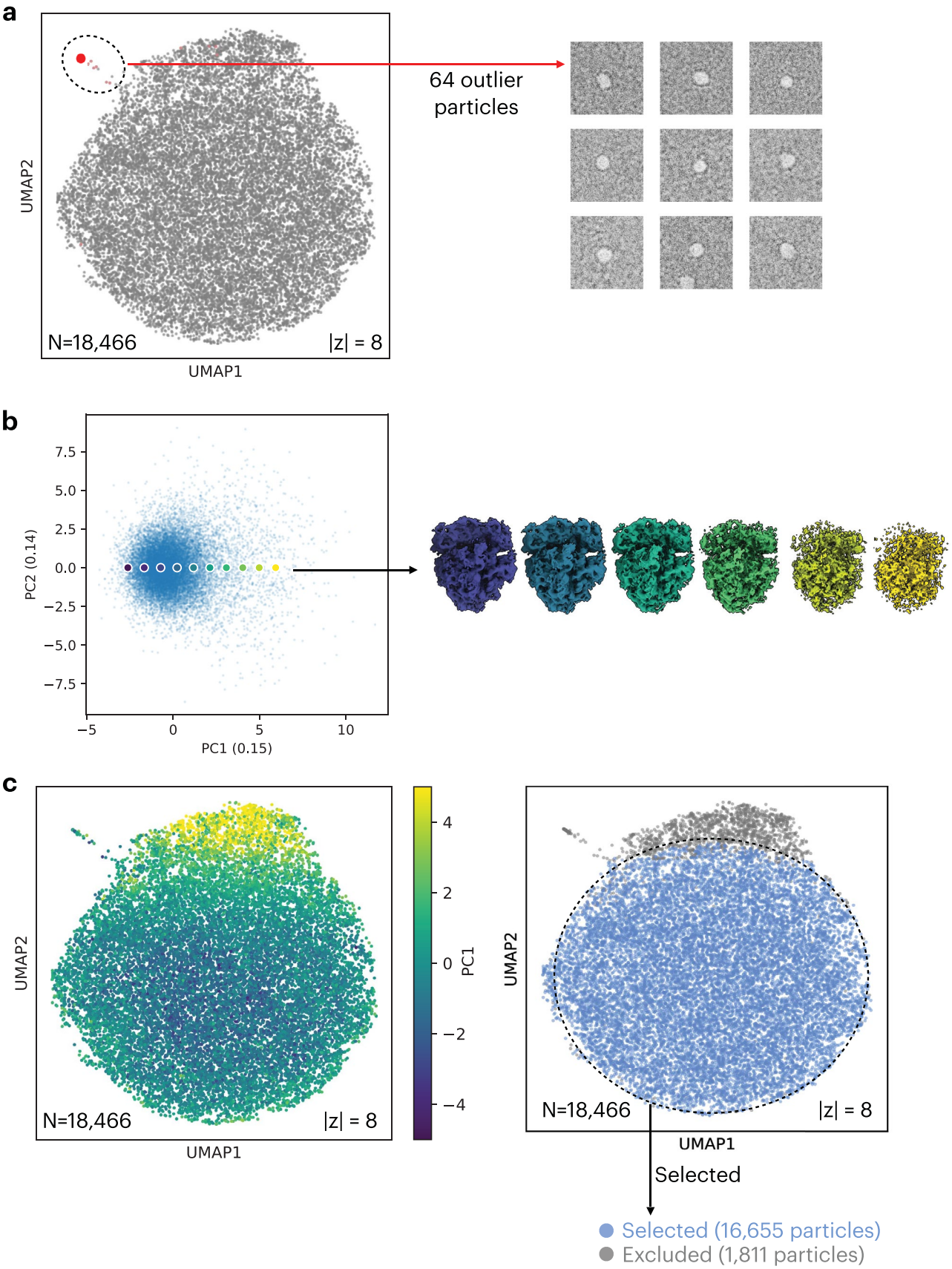
Extended data is available for this paper at https://doi.org/10.1038/s41592-024-02340-4.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41592-024-02340-4.

Correspondence and requests for materials should be addressed to Abhay Kotecha or Ellen D. Zhong.

Peer review information *Nature Methods* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editor: Arunima Singh, in collaboration with the *Nature Methods* team. Peer reviewer reports are available.

https://doi.org/10.1038/s41592-024-02340-4

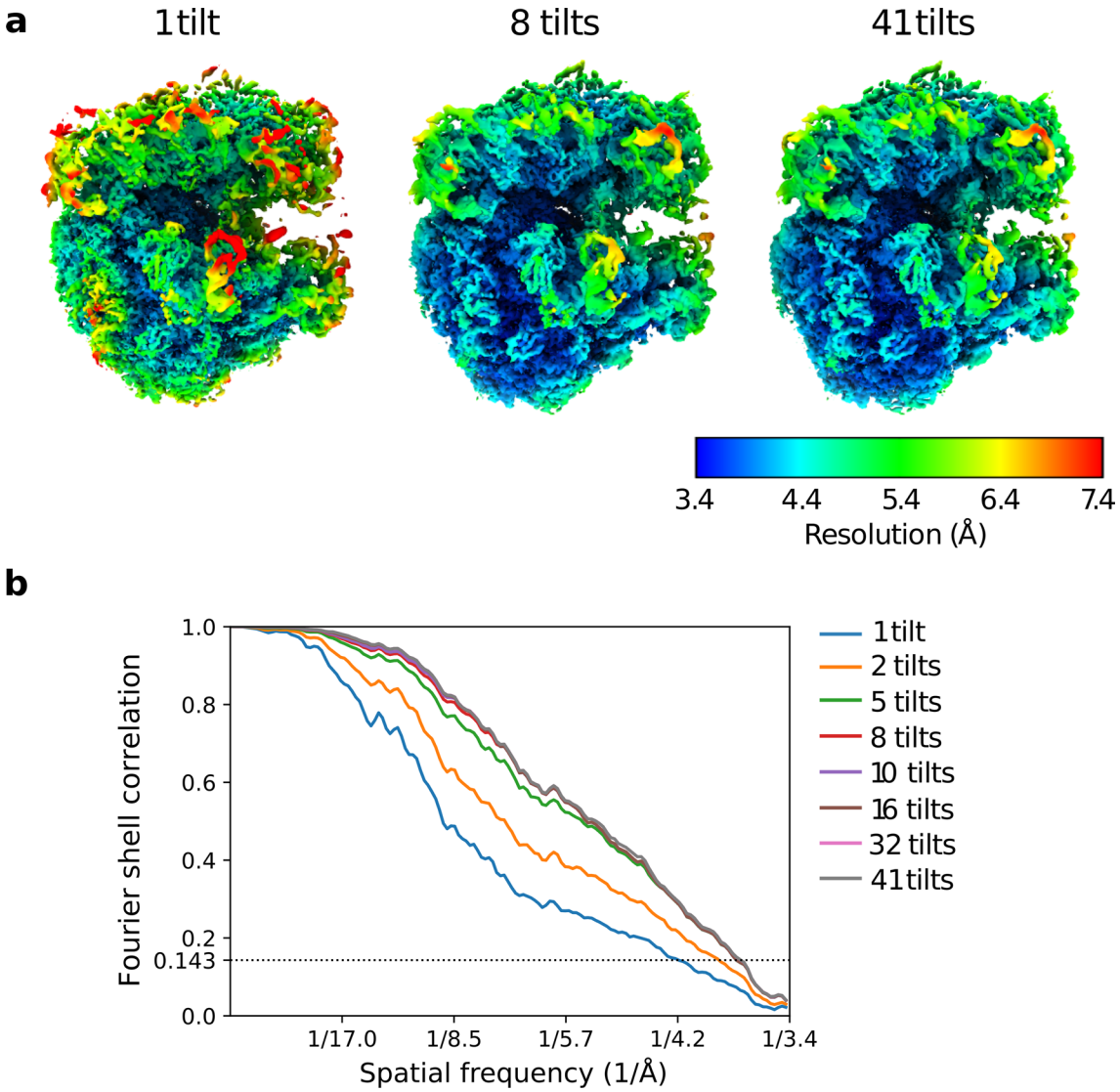


Extended Data Fig. 1 | See next page for caption.

https://doi.org/10.1038/s41592-024-02340-4

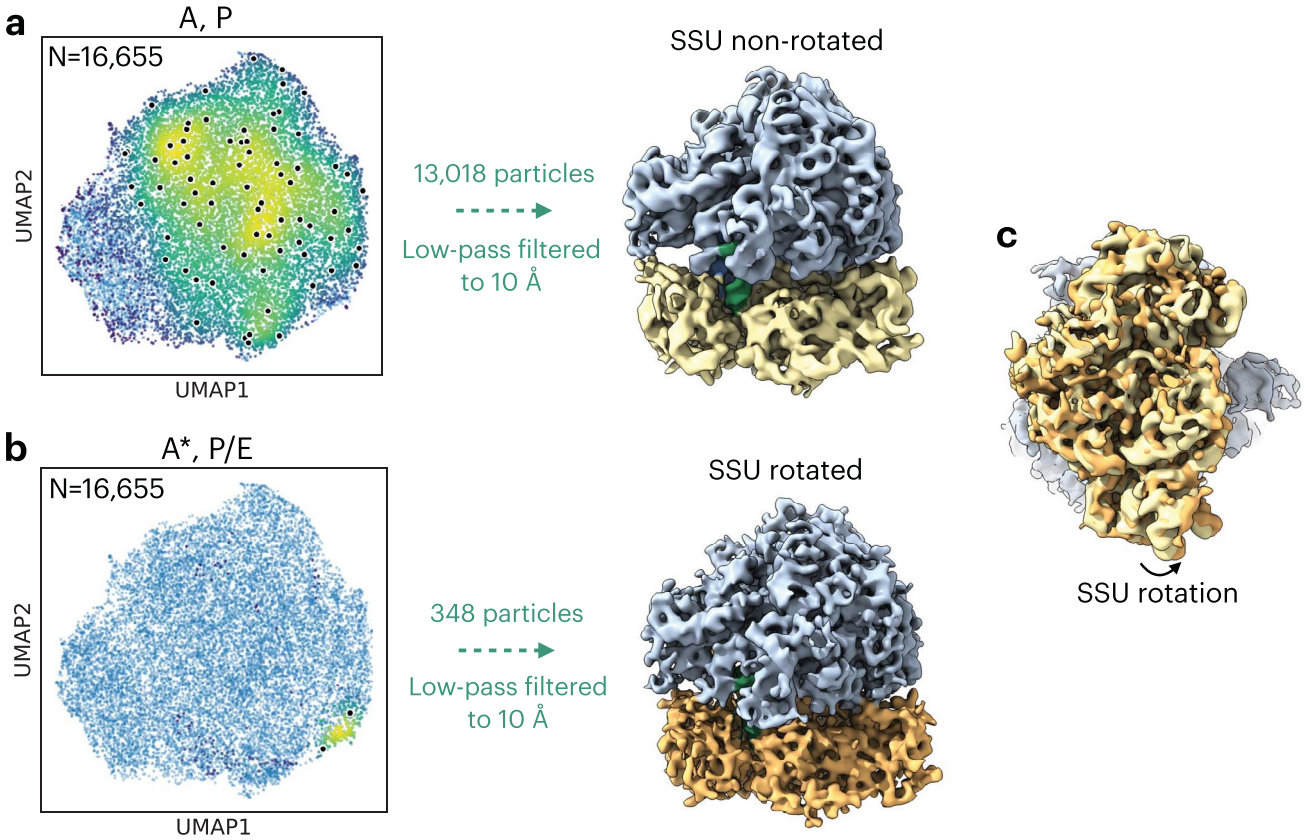
Extended Data Fig. 1 | Latent space-based filtering of the *M. pneumoniae* bacterial ribosome subtomograms. a. UMAP visualization of cryoDRGN's latent space representation from a training run using all particles (D=128, 3.9 Å/pixel), showing example particle images for the highlighted group of outlier particles. b. Visualization of the latent space along the first and second principal components (PCs), showing representative density maps for the highlighted traversal across PC1. c. UMAP visualization of the latent space for the same cryoDRGN training run colored by PC1 using the same coloring as in b) (left), and colored by a particle selection that excludes outlier particles (right).

https://doi.org/10.1038/s41592-024-02340-4



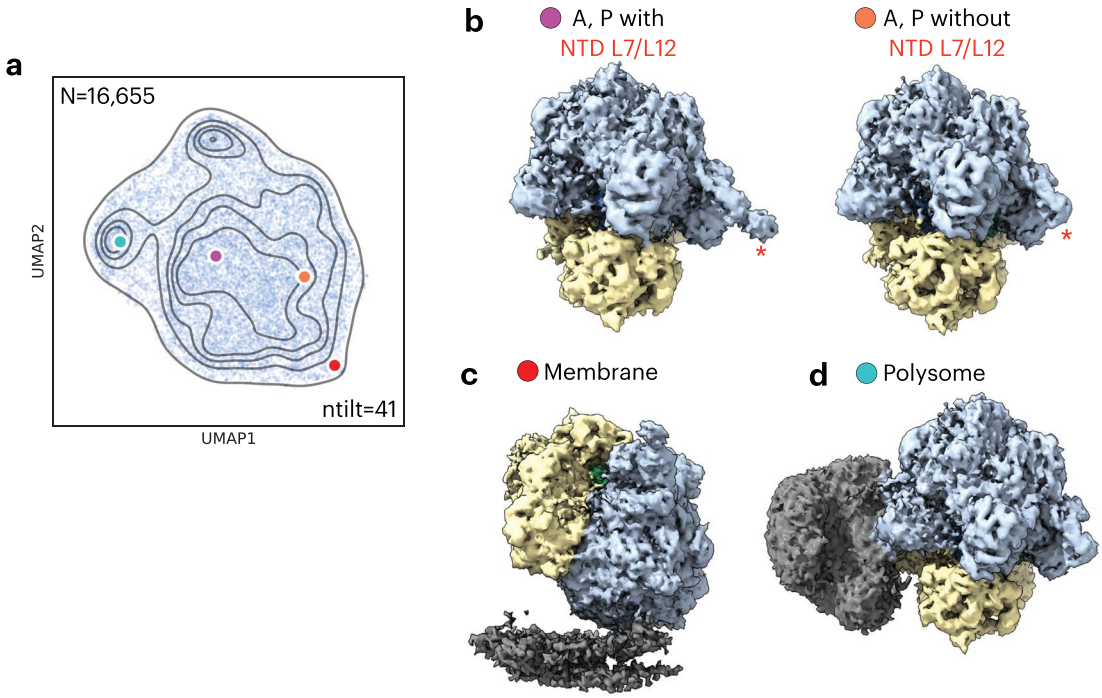
Extended Data Fig. 2 | Homogeneous reconstruction of the *M. pneumoniae* ribosome varying the number of tilt images (D=294, 1.7 Å/pixel). a. Local resolution estimated from RELION 4.0[55] for reconstructions using 1 tilt, 8 tilt, and 41 tilts per particle, with maps obtained through voxel-based backprojection in cryoDRGN-ET. b. FSC curves between half-maps for varying numbers of tilts per particle.

https://doi.org/10.1038/s41592-024-02340-4



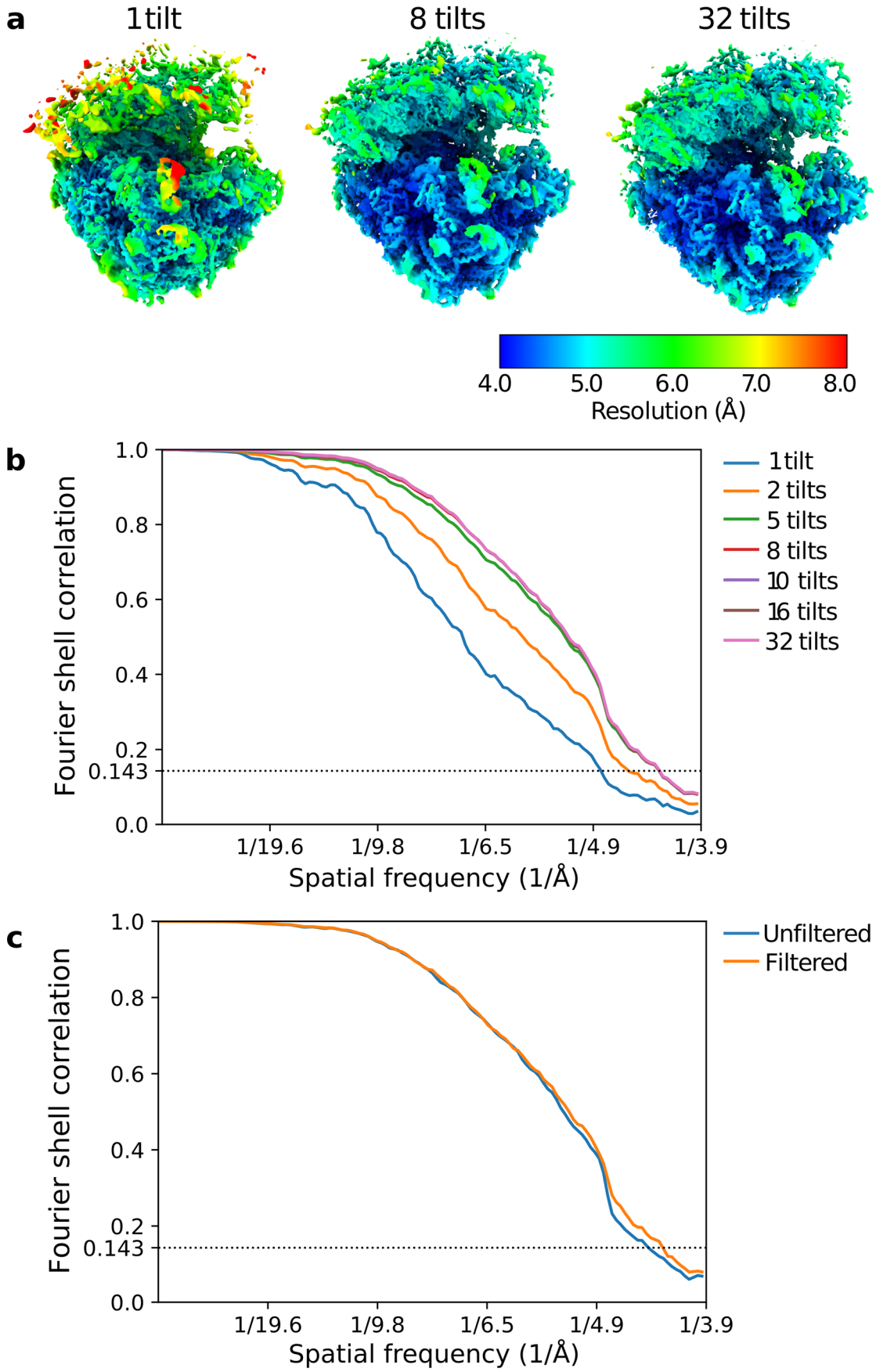
Extended Data Fig. 3 | SSU rotation of the *M. pneumoniae* ribosome identified in cryoDRGN-ET. a, b. In the left column is the UMAP visualization of the latent space from cryoDRGN-ET training on the filtered particle set, with a heatmap overlaid depicting the distribution of particles in a. the A, P state, and b. the A*, P/E state. In the middle column are density maps obtained by voxel-based backprojection of particles (D=294, 1.7 Å/pixel) from these two states low-pass filtered to 10 Å resolution. c. These two reconstructions are overlaid and viewed facing the SSU to depict SSU rotation.

https://doi.org/10.1038/s41592-024-02340-4



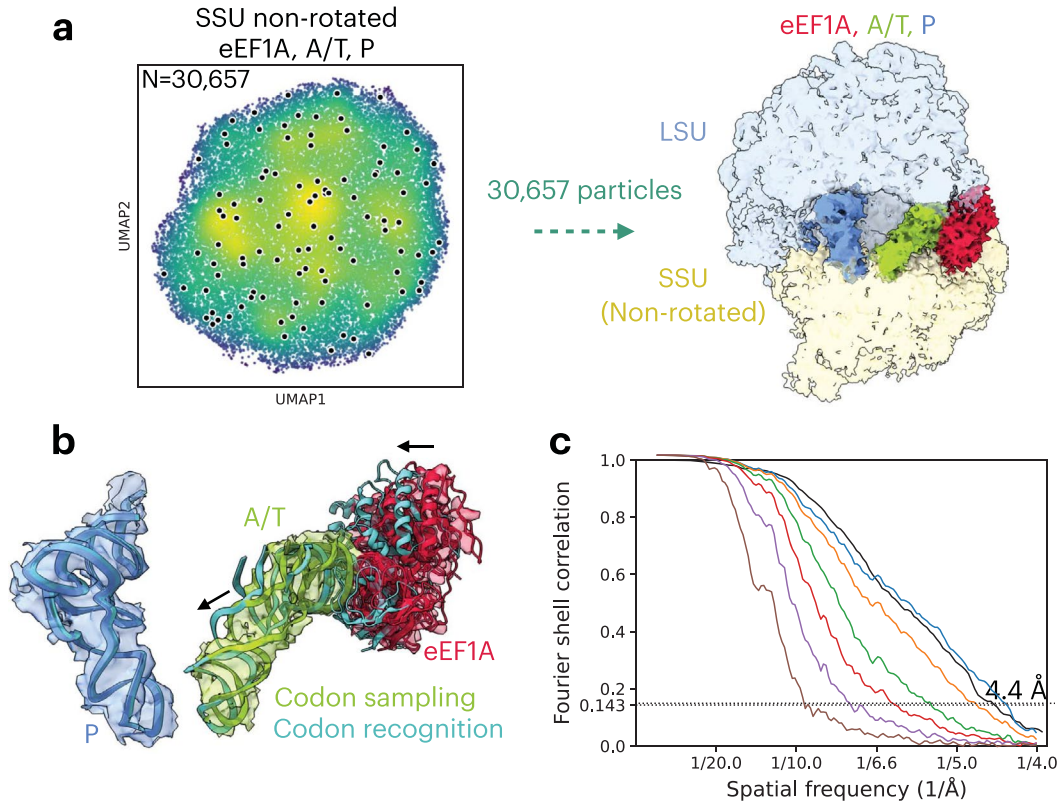
Extended Data Fig. 4 | Additional states of the *M. pneumoniae* ribosome identified in cryoDRGN-ET. a. UMAP visualization of the latent space from cryoDRGN-ET training (D=128, 3.9 Å/pixel) on the *M. pneumoniae* ribosome filtered particle set, using 41 tilts per particle during training. Latent embeddings for representative density maps are highlighted. b. Representative maps with (left) and without (right) density present for the NTD of L7/L12, as highlighted by the red asterisk. c. Representative map depicting a membrane-bound ribosome. d. Representative map depicting polysome density.

https://doi.org/10.1038/s41592-024-02340-4



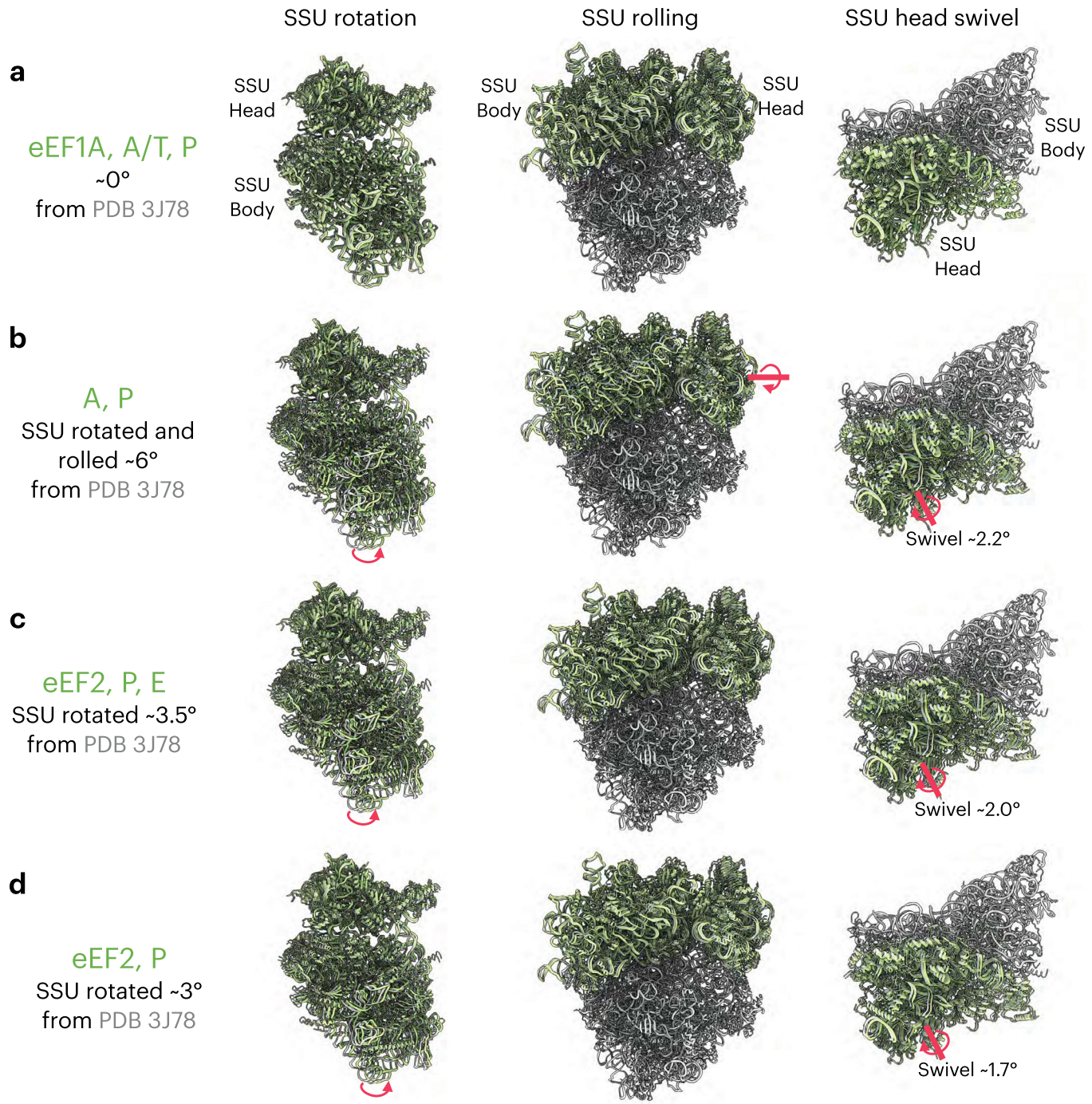
Extended Data Fig. 5 | Homogeneous reconstruction of the *S. cerevisiae* ribosome varying the number of tilt images (D=256, 1.96 Å/pixel). a. Local resolution estimated from RELION 4.0[55] for reconstructions using 1 tilt, 8 tilt, and 32 tilts per particle, with maps obtained through voxel-based backprojection in cryoDRGN. b. FSC curves between half-maps for varying numbers of tilts per particles. c. FSC curves for either the full particle set (119,031 particles) or the filtered particle set (93,281 particles) using all tilts per particle.

https://doi.org/10.1038/s41592-024-02340-4



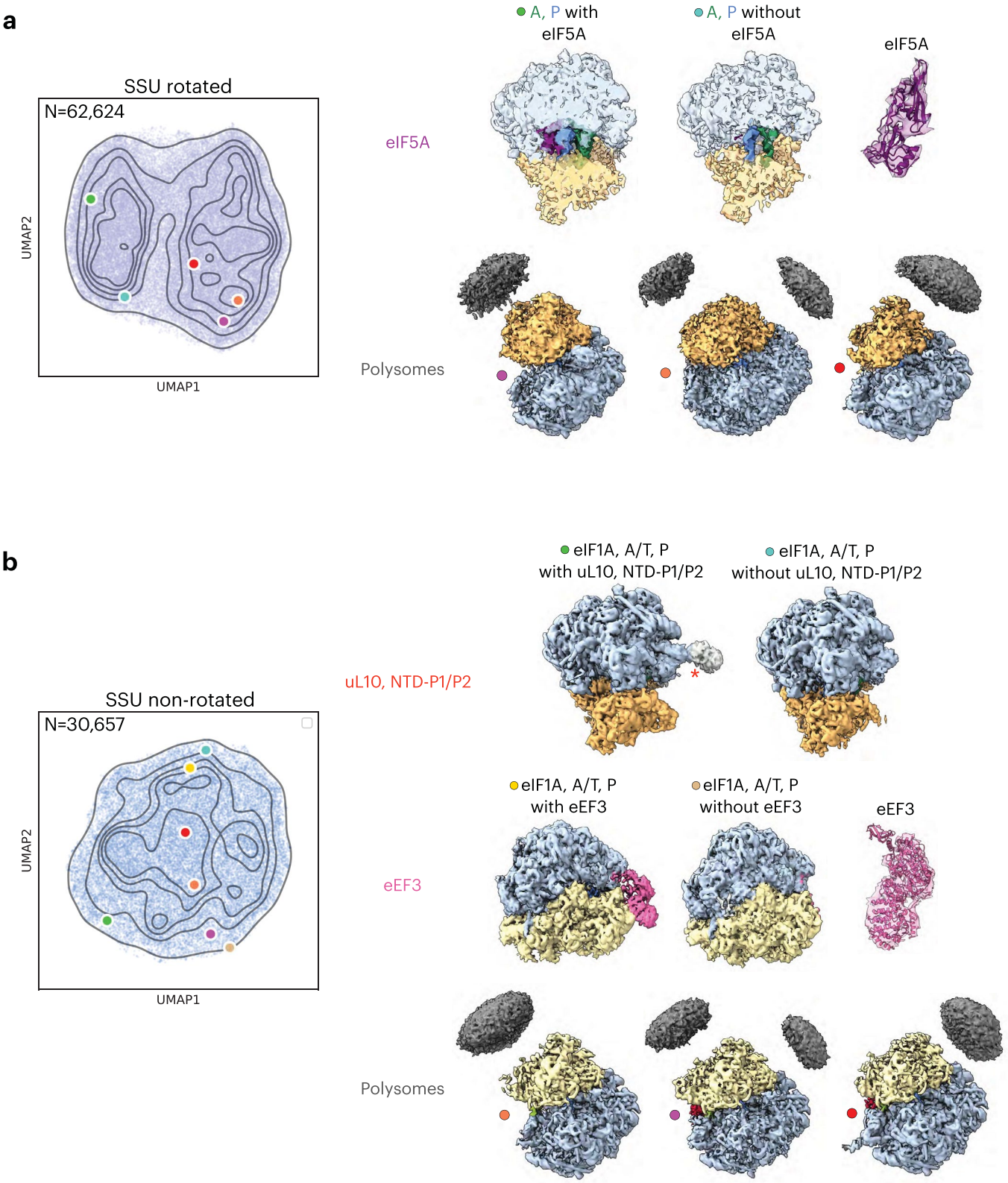
Extended Data Fig. 6 | Homogeneous reconstruction of translational states of the *S. cerevisiae* ribosome identified in cryoDRGN-ET in the eEF1A, A/T, P state. a. The left panel shows the UMAP visualization of the latent space from cryoDRGN-ET training on the indicated particle set, with overlaid heatmaps highlighting particles belonging to each state. The right column depicts the homogeneous reconstruction (D=256, 1.96 Å/pixel) from cryoDRGN-ET's voxel-based backprojection for particles selected in each state. Reconstructions are low-pass filtered to the FSC resolution and colored by corresponding factors. b. Superposition of the P tRNA, A/T tRNA, and eEF1A from this state vs the codon recognition state (cyan) from PDB ID: 5LZS ref. 30. The P tRNA, A/T tRNA, and eEF1A were separately docked into the reconstruction from a) for comparison to the codon recognition state. Density is shown from the reconstruction in a) around these factors. Arrows indicate the shift in position of the A/T tRNA and eEF1A between the codon sampling and codon recognition states. c. FSC curves bewteen half-maps for the eEF1A, A/T, P state.

https://doi.org/10.1038/s41592-024-02340-4



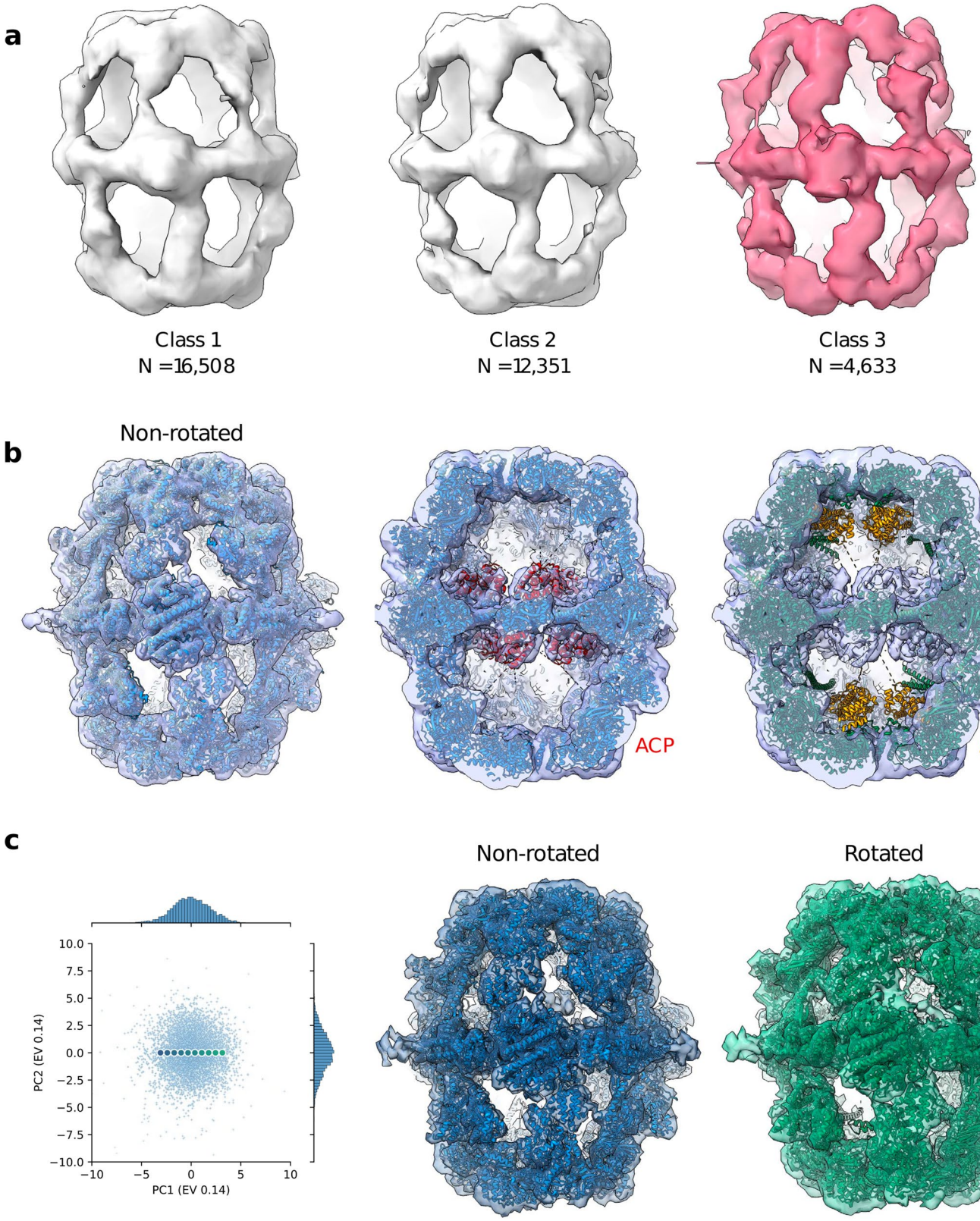
Extended Data Fig. 7 | Conformational motions of the *S. cerevisiae* ribosome SSU identified in cryoDRGN-ET in the following states: a. the eEF1A, A/T, P state; b. the A, P state; c. the eEF2, P, E state; and d. the eEF2, P state. In all panels, the non-rotated *S. cerevisiae* ribosome structure in PDB ID 3J78 ref. 68 is shown in grey, and docked models into high-resolution reconstructions from cryoDRGN-ET are shown in green. The left column shows the rotation of the SSU, with models aligned on the LSU (LSU not shown for clarity) and red arrows indicating rotations. The middle column shows rolling of the SSU, again with models aligned on the LSU and red arrows indicating rolling. Finally, the right column shows the SSU head swivel, with models aligned on the SSU body (LSU removed and not shown), and red arrows indicating cases where a minor head swivel is present. Rotations angles between docked models and coordinates from PDB ID 3J78 ref. 68 are measured in ChimeraX[67]. For each state we report two rotation angles. The angle representing either 'SSU rotation' or 'SSU rolling and rotation' is measured as the rotation angle required to superimpose the state's SSU onto the SSU of PDB 3J78, when structures aligned on the LSU. The angle representing the head swivel is measured as the rotation angle required to superimpose the SSU head between that state and PDB ID 3J78, when the SSU structures are aligned on the SSU body.

https://doi.org/10.1038/s41592-024-02340-4



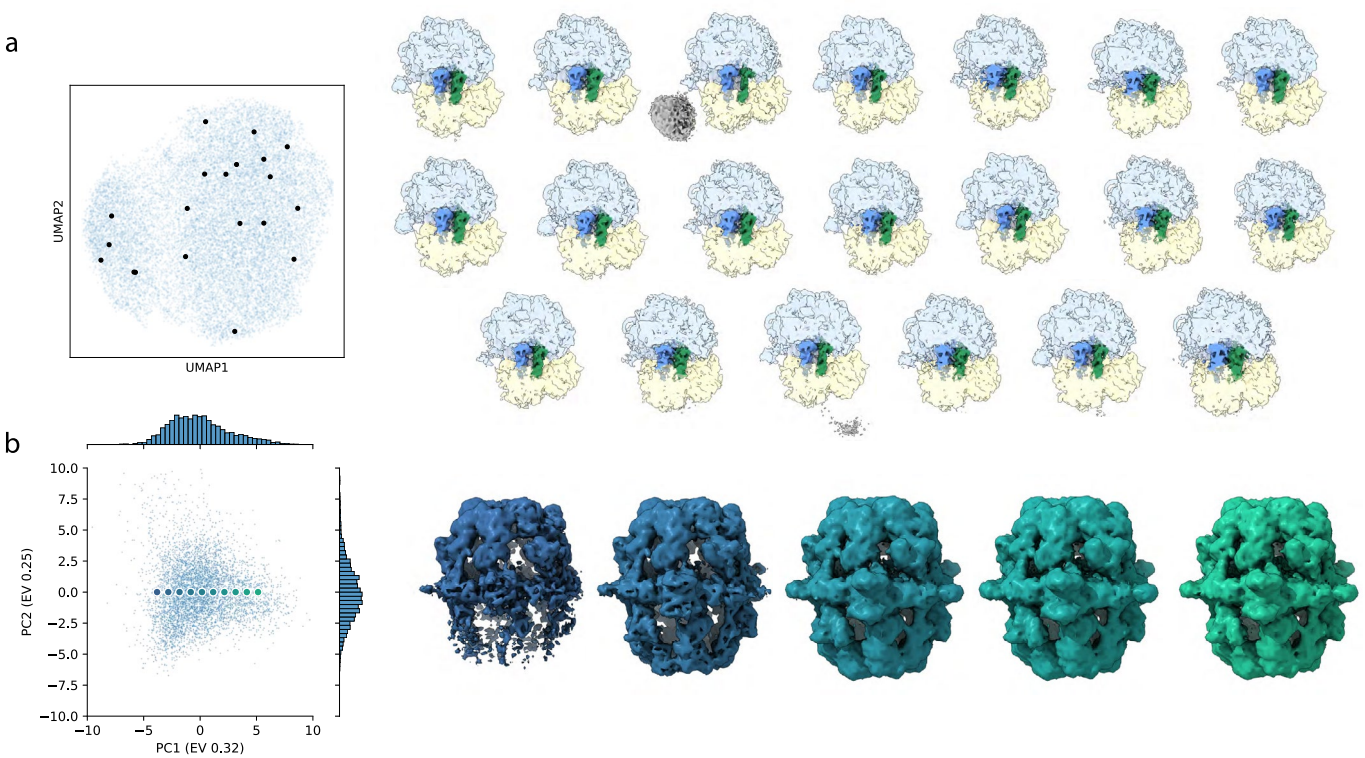
Extended Data Fig. 8 | Structures of polysomes, eIF5A, uL10 and NTD-P1/P2, and eEF3 identified in cryoDRGN-ET analysis of the *S. cerevisiae* ribosome from a. SSU rotated and b. SSU non-rotated particles. In the left column, UMAP visualization of the latent space from cryoDRGN-ET training (D=128, 3.92 Å/pixel) on the indicated particle set. Latent embeddings for representative density maps are highlighted. In the right column of a), representative density maps from cryoDRGN-ET are shown depicting ribosomes with and without eIF5A (top row) and with polysome density (bottom row). In the right column of b), representative maps from cryoDRGN-ET are shown depicting ribosomes with and without uL10(P1-P2)$_2$ density (top row), with and without eEF3 density (middle row), and with polysome density (bottom row). Atomic models for eIF5A and eEF3 are shown in density from the cryoDRGN-ET representative map.

https://doi.org/10.1038/s41592-024-02340-4



Extended Data Fig. 9 | *S. cerevisiae* fatty acid synthase classification and atomic models fit in maps. a. Maps and particle counts from RELION 3D classification (3 classes). The map from the best quality particles is shown in magenta. b. D3 symmetry RELION refinement of 5,239 high quality particles selected by cryoDRGN-ET (Fig. 4) with a fit atomic models for the non-rotated (blue, PDB ID: 6QL6) and rotated (green, PDB ID: 6QL5) states[36]. Left: FAS map with best fitting atomic model (non-rotated state). Middle: FAS map at low threshold showing density for ACP (red) in the non-rotated state. Right: FAS map at low threshold missing density for ACP (orange) in the rotated state. c. Sampled volumes from the extremes of a scan along PC1 of the latent space (left) with best fitting atomic models (non-rotated PDB ID: 6QL6 ref. 36 in blue; rotated PDB ID: 6QL5 ref. 36 in green).

https://doi.org/10.1038/s41592-024-02340-4



Extended Data Fig. 10 | TomoDRGN[39] performance on *M. pneumoniae* ribosome and *S. cerevisiae* fatty acid synthase particles. a. 20 representative sampled volumes across the latent space (black dots in UMAP representation, left) are all in the A, P state (right). b. Sampled volumes (right) across PC1 of the latent space (left) show isosurface changes rather than conformational heterogeneity. Maps are generated from points in the latent space depicted with corresponding colors.

nature portfolio

Corresponding author(s): Ellen Zhong, Abhay Kotecha

Last updated by author(s): May 30, 2025

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☒	☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☒	☐ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	The following tools were used for data collection: TEM Tomography 5 software (Thermo Fisher Scientific)
Data analysis	The following tools were used for data analysis: ChimeraX 1.6.1, Warp/M 1.0.9, AreTomo 1.3.4, RELION 4.0, RELION 3.0, RELION 3.1, TOMOMAN (https://github.com/wan-lab-vanderbilt/TOMOMAN), CTFFIND4 4.1.14, STOPGAP (https://github.com/wan-lab-vanderbilt/STOPGAP), cisTEM 1.0.0-beta, crYOLO 1.9.9, IMOD 4.11, Membrane-Seg (https://github.com/teamtomo/membrain-seg), Segger 1.9.2, cryoDRGN 3.0.0-beta (https://github.com/ml-struct-bio/cryodrgn)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

S. cerevisiae raw data has been deposited to EMPIAR-11658. Maps have been deposited to EMDB (EMD-18197, EMD-18231, EMD-18232) and cryoDRGN model

weights, density maps, and any associated files needed to reproduce this analysis are available on Zenodo (doi: 10.5281/zenodo.11399378). Atomic models used from previous studies were obtained from the PDB (7PHA, 7PHB, 7PHC, 7PH9, 3J7R, 5LZS, 6GQV, 6TNU, 3J78, 7B7D, 6TA1, 6QL5, and 6QL6). The dataset from EMPIAR-10499 was analyzed in this study.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	N/A
Reporting on race, ethnicity, or other socially relevant groupings	N/A
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	N/A

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample sizes for 3D reconstruction were determined by using all available data.
Data exclusions	No data were excluded
Replication	Replication is not currently reported; however replica training runs produced qualitatively similar results. Replicate runs were performed for at least one choice of hyperparameters for the three datasets. Further hyperparameter and convergence exploration was carried out with a single replicate.
Randomization	Image datasets were split into random halves to compute resolution metrics.
Blinding	Researchers were blinded to expectd results by construction as the study performs unsupervised representation learning. However, researchers were not blinded to previously published analyses of the dataset.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Plants

Seed stocks	*Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.*
Novel plant genotypes	*Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.*
Authentication	*Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.*