

# CryoDRGN: reconstruction of heterogeneous cryo-EM structures using neural networks

Ellen D. Zhong<sup>1,2</sup>, Tristan Bepler<sup>1,2</sup>, Bonnie Berger<sup>1,2,3</sup>✉ and Joseph H. Davis<sup>1,4</sup>✉

**Cryo-electron microscopy (cryo-EM) single-particle analysis has proven powerful in determining the structures of rigid macromolecules. However, many imaged protein complexes exhibit conformational and compositional heterogeneity that poses a major challenge to existing three-dimensional reconstruction methods. Here, we present cryoDRGN, an algorithm that leverages the representation power of deep neural networks to directly reconstruct continuous distributions of 3D density maps and map per-particle heterogeneity of single-particle cryo-EM datasets. Using cryoDRGN, we uncovered residual heterogeneity in high-resolution datasets of the 80S ribosome and the RAG complex, revealed a new structural state of the assembling 50S ribosome, and visualized large-scale continuous motions of a spliceosome complex. CryoDRGN contains interactive tools to visualize a dataset's distribution of per-particle variability, generate density maps for exploratory analysis, extract particle subsets for use with other tools and generate trajectories to visualize molecular motions. CryoDRGN is open-source software freely available at <http://cryodrgn.csail.mit.edu>.**

Proteins and their complexes are dynamic macromolecular machines that carry out the essential biological processes responsible for life. Although the mechanism of these macromolecular machines is often deduced from a static 3D structure, a more complete understanding could be achieved if one could analyze the full distribution of conformations relevant to function.

Single-particle cryo-EM is a rapidly maturing method for determining high-resolution structure of large macromolecular complexes<sup>1,2</sup>. Major advances in hardware<sup>3–5</sup> and software<sup>4–9</sup> have streamlined the collection and analysis of cryo-EM datasets, such that structures of rigid macromolecules can routinely be solved at near-atomic resolution<sup>10,11</sup>. Increasingly, cryo-EM has been applied to study heterogeneous complexes, as the experimental procedure is less sensitive to sample heterogeneity than other methods for structure determination<sup>12,13</sup>. Additionally, because single-particle cryo-EM can capture millions of snapshots of the molecule of interest, each carrying a unique molecule in its own conformational state<sup>14</sup>, cryo-EM holds promise in revealing the conformational landscape of dynamic macromolecular complexes. However, reconstructing ensembles of 3D volumes from such snapshots remains a major computational challenge.

Existing tools for heterogeneous reconstruction often make limiting assumptions on the observed structural heterogeneity. Most commonly, heterogeneity is modeled as though it originates from a small number of independent, discrete states, implemented as ‘3D classification’ or ‘heterogeneous refinement’ in many cryo-EM software packages<sup>15–18</sup>. However, these discrete classification approaches require specifying initial models for refinement, and because the number and nature of the underlying structural states is unknown *a priori*, this approach is error prone and often results in the omission of potentially relevant structures. More critically, such discrete approaches are ill-suited for reconstructing structures undergoing continuous conformational changes.

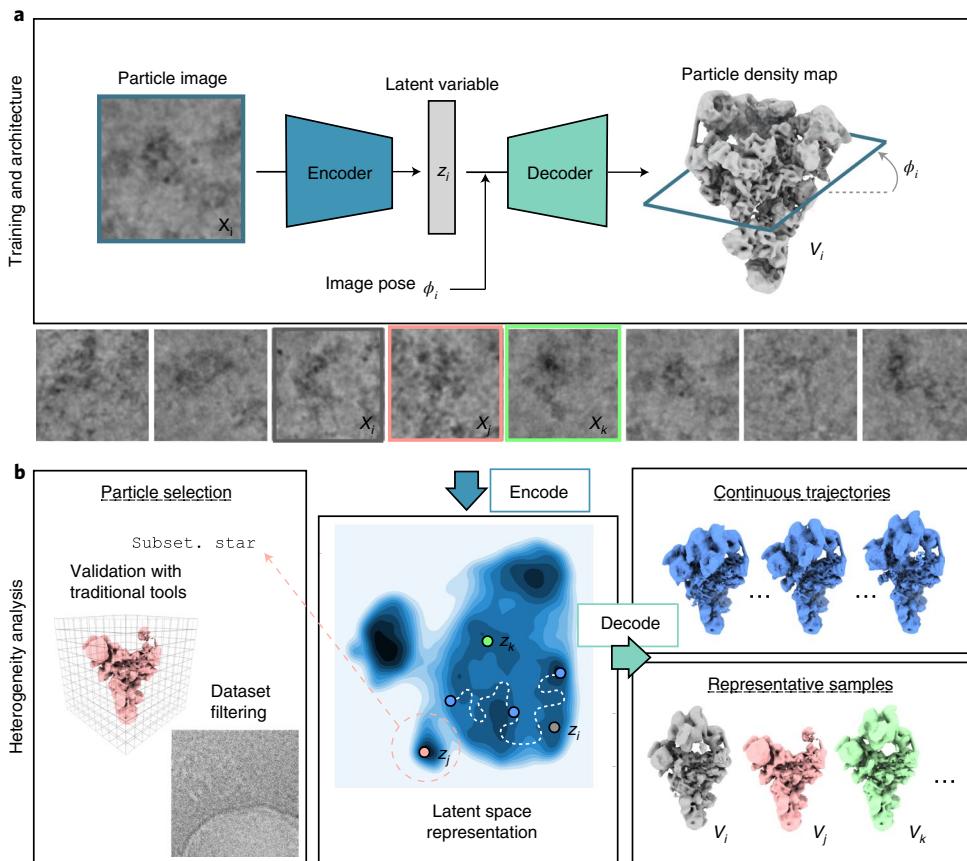
Advanced methods for heterogeneous reconstruction seek to more closely model the continuous nature of flexible molecules.

Multibody refinement, available in RELION, models the structure as the sum of user-defined rigid bodies that are allowed to rotate relative to one another, placing structural assumptions on the observed heterogeneity<sup>19</sup>. Continuous heterogeneity was also described using principal component analysis (PCA)-based approaches<sup>20–22</sup>, including the recent 3D Variability Analysis (3DVA) algorithm available in cryoSPARC<sup>23</sup>. Although the linear subspace model of these approaches can provide a summary of the overall variability within the dataset, the visualized heterogeneity contains artifacts when a molecule’s conformational deformations are poorly approximated by linear interpolations along basis volumes. In the manifold embedding approach proposed in refs. <sup>24,25</sup>, heterogeneous structures are recovered by binning particles along the data manifold, followed by traditional homogeneous reconstruction. Additional algorithms for continuous heterogeneous reconstruction were demonstrated on synthetic datasets<sup>26,27</sup>.

Here, we present cryoDRGN (Deep Reconstructing Generative Networks), a method for heterogeneous cryo-EM reconstruction based on deep neural networks. We hypothesized that neural networks, which are known for their ability to model complex, nonlinear functions<sup>28</sup>, could learn heterogeneous ensembles of cryo-EM density maps. We first show that our neural network representation of structure can model single-density maps at high resolution, before demonstrating the full cryoDRGN framework for unsupervised heterogeneous reconstruction.

We find that cryoDRGN is a powerful and general approach for analyzing structural heterogeneity in macromolecular complexes of varying size and expected sources of heterogeneity. We show that the cryoDRGN approach can uncover residual heterogeneity in ‘homogeneous’ datasets of the recombination-activating gene RAG1–RAG2 complex and the 80S ribosome and model large compositional changes of the assembling 50S ribosome and continuous conformational changes of the pre-catalytic spliceosome. Remarkably, cryoDRGN’s unsupervised approach for representation learning can readily identify and filter impurities in the dataset

<sup>1</sup>Computational and Systems Biology, Massachusetts Institute of Technology, Cambridge, MA, USA. <sup>2</sup>Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA, USA. <sup>3</sup>Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA, USA. <sup>4</sup>Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, USA. ✉e-mail: [bab@mit.edu](mailto:bab@mit.edu); [jhdavis@mit.edu](mailto:jhdavis@mit.edu)



**Fig. 1 | The cryoDRGN method for heterogeneous single-particle cryo-EM reconstruction.** **a**, The cryoDRGN model consists of two neural networks structured in an image-encoder-volume-decoder architecture with a continuous latent variable representation of heterogeneity. During training, each particle image is encoded into the low-dimensional latent space and then reconstructed as its corresponding model slice based on the Fourier slice theorem. Image and volume data are depicted in real space for visual clarity. **b**, Once a cryoDRGN model is trained, the full dataset of particle images is encoded into the latent space, which is visualized here as a contour map with darker regions corresponding to higher particle density (center). The decoder, which represents an ensemble of 3D density maps, can directly generate density maps from arbitrary values of the latent variable (right). The particle stack may also be filtered using the latent space representation for validation of specific structures with traditional tools or to remove impurities from the dataset (left). Example images are from EMPIAR-10180 (ref. 36).

and can identify rare structural states containing as few as ~1,000 particles. CryoDRGN is distributed as an open-source tool that can be easily integrated into existing pipelines and is freely available at [cryodrgn.csail.mit.edu](http://cryodrgn.csail.mit.edu).

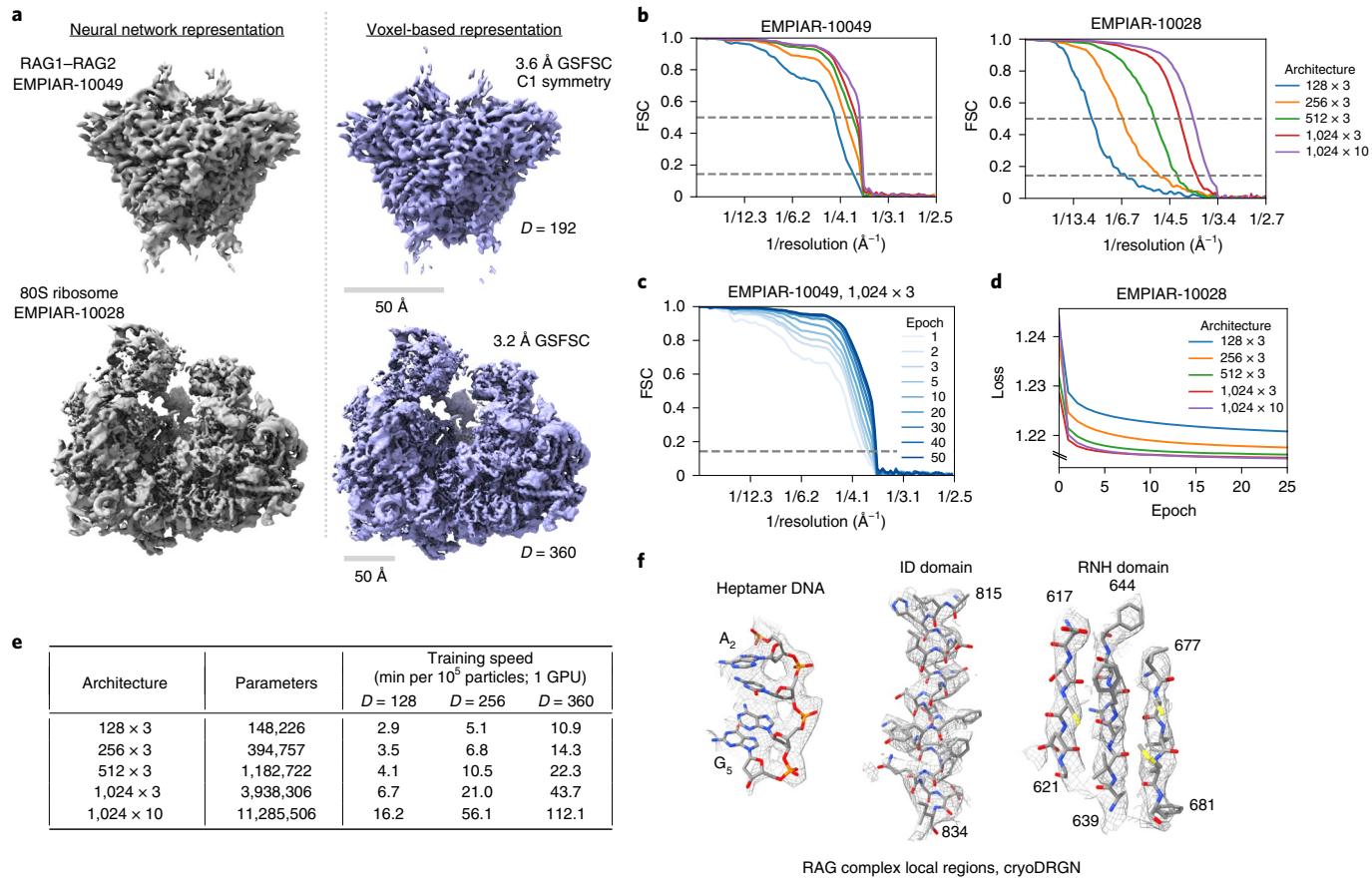
## Results

**The cryoDRGN method.** CryoDRGN performs heterogeneous reconstruction by learning a deep generative model of 3D structure from single-particle cryo-EM images. The method consists of a specialized image-encoder-volume-decoder architecture, which learns an encoding of two-dimensional (2D) particle images into a continuous vector space described by the latent variable  $z \in \mathbb{R}^n$  (that is, the latent space), and the concomitant reconstruction of 3D cryo-EM density maps from this latent space representation (Fig. 1a). This choice of model assumes that the heterogeneous structures can be embedded within a continuous, low-dimensional manifold in the latent space, where the dimensionality of the latent space is defined by the user. The model is specified in the Fourier domain to relate 2D images as planar slices of the 3D volume<sup>29</sup>, the orientation of which is previously determined from a consensus reconstruction. The neural networks are jointly trained from random initialization using stochastic gradient descent on an objective function that seeks to maximize (a variational lower bound on) the data likelihood as in standard variational autoencoders (VAEs)<sup>30</sup>. Additional

architectural and training details of cryoDRGN are provided in the Methods.

After training, the output of cryoDRGN analysis includes (1) per-particle latent encodings,  $z_i$ , describing the dataset's heterogeneity and (2) a neural network model of 3D density maps that can directly reconstruct a density map given  $z_i$ . Specifically, the encoder network encodes particle images into the continuous latent space, which allows for visualization and inspection of particle distribution (Fig. 1b, center). The trained decoder network can then generate 3D density maps given arbitrary values of the latent variable. For example, representative structures can be generated from regions of latent space with high particle density, and continuous conformational trajectories can be reconstructed by sampling points along a trajectory through latent space (Fig. 1b, right). Notably, any cryoDRGN-generated volume can be orthogonally validated by traditional reconstruction approaches<sup>15</sup> using nearby particles in the latent space (Fig. 1b, left). Lastly, any regions of the latent space that are enriched in impurities or imaging artifacts may be selected, and the encompassed particles may be filtered from subsequent analysis (Fig. 1b, left).

**Neural networks can represent cryo-EM density maps.** We first evaluated the ability of the cryoDRGN volume decoder to represent high-resolution cryo-EM density maps. To learn the homogeneous



**Fig. 2 | Neural network representation of cryo-EM density maps.** **a**, Density maps of the RAG1–RAG2 complex (EMPIAR-10049)<sup>31</sup> and of the eukaryotic Pf80S ribosome (EMPIAR-10028)<sup>32</sup> reconstructed by cryoDRGN’s decoder neural network (left) and a traditional, voxel-based reconstruction in cryoSPARC (right). The cryoDRGN volumes were generated from decoder networks with three hidden layers and 1,024 nodes per hidden layer (denoted as 1,024 × 3) trained for 25 epochs. **b**, FSC curves between density maps produced by the cryoDRGN decoder with varying architectures and the traditional reconstruction in **a**. **c,d**, Evolution of the FSC curve in **b** and the training curve over multiple epochs of cryoDRGN model training. **e**, Training speed in min per 10<sup>5</sup> images for cryoDRGN decoder networks of different architectures on different image sizes (**d**, in pixels) on a single Nvidia V100 graphics processing unit (GPU). The number of trainable parameters is specified for decoder networks trained on D = 256 images. **f**, Representative regions (insertion domain (ID), RNase H-like domain (RNH)) of the RAG1–RAG2 density map from cryoDRGN in **a** superimposed with the published atomic model (Protein Data Bank (PDB) 3JBX).

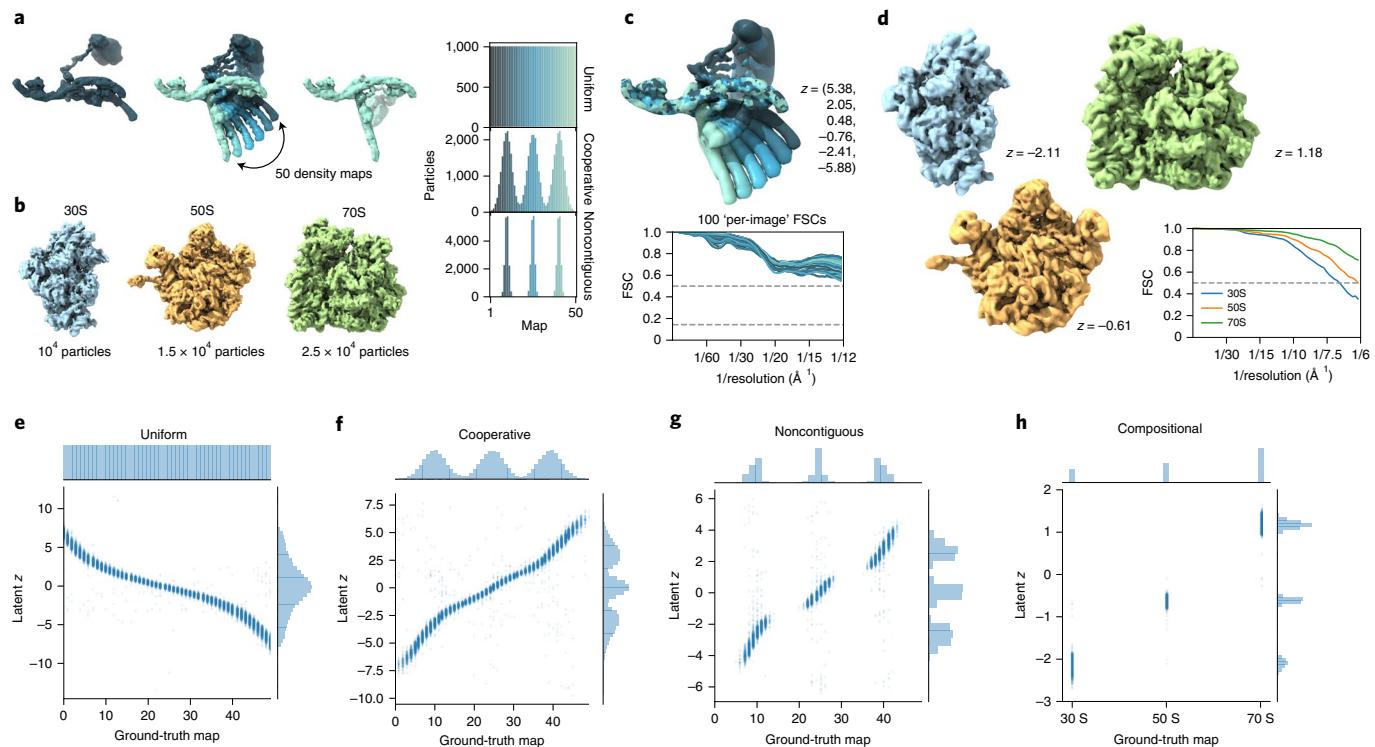
structure of the RAG1–RAG2 signal-end complex (RAG, 369 kDa)<sup>31</sup> and the *Plasmodium falciparum* 80S ribosome (Pf80S, 4.2 MDa)<sup>32</sup>, we trained the volume-decoder network with no latent variable input, using image poses obtained from C1 homogeneous refinements in cryoSPARC<sup>16</sup> (Methods). Trained on full-resolution images, the cryoDRGN decoder produced structures that correlated with the traditional, voxel-based reconstruction (Fig. 2a) at resolutions up to 3.6 Å for RAG and 3.9 Å for Pf80S at a threshold of Fourier shell correlation (FSC) = 0.5 (Fig. 2b), demonstrating the efficacy of this neural network-based representation of 3D structure.

As neural networks have a fixed capacity for representation that is constrained by their architecture, we next compared decoder architectures of different sizes to evaluate the tradeoff between representation power and training speed. We found that larger architectures, which have more trainable parameters, result in density maps that correlate with the traditionally reconstructed map at higher resolutions (Fig. 2b). The networks were trained by multiple passes through the dataset (that is, epochs) (Fig. 2c), with lower values of the objective function (Fig. 2d) as training progressed. Notably, while the resolution of the learned structure increased with neural network size, we found that larger models were slower to train (Fig. 2e). These tradeoffs suggest that the architecture and image size should be tuned to suit the desired balance of speed and achievable

resolution. Lastly, we found that the cryoDRGN architecture was capable of learning density maps at sufficiently high resolution to visualize structural features, such as bulky side chains, that were consistent with our FSC-based resolution estimates (Fig. 2f).

**CryoDRGN models both discrete and continuous structural heterogeneity.** We next sought to evaluate the complete cryoDRGN framework for heterogeneous reconstruction using simulated datasets (Fig. 3a,b). Datasets modeling continuous heterogeneity were produced by rotating a single dihedral angle of a hypothetical protein complex to simulate a conformational transition along a one-dimensional (1D) reaction coordinate. Single-particle cryo-EM images were then simulated uniformly along this reaction coordinate ('uniform'), with bias toward particular conformations exemplary of cooperative transitions ('cooperative') or with strong bias leading to unobserved transition states ('noncontiguous'). A dataset simulating discrete compositional heterogeneity was produced by mixing particles of the bacterial 30S, 50S and 70S ribosome ('compositional'). We then provided each of these four simulated datasets and their corresponding poses to cryoDRGN and trained a 1D latent variable model (Methods).

We found that cryoDRGN was capable of reconstructing both continuous and discrete heterogeneous ensembles (Fig. 3c–h). On



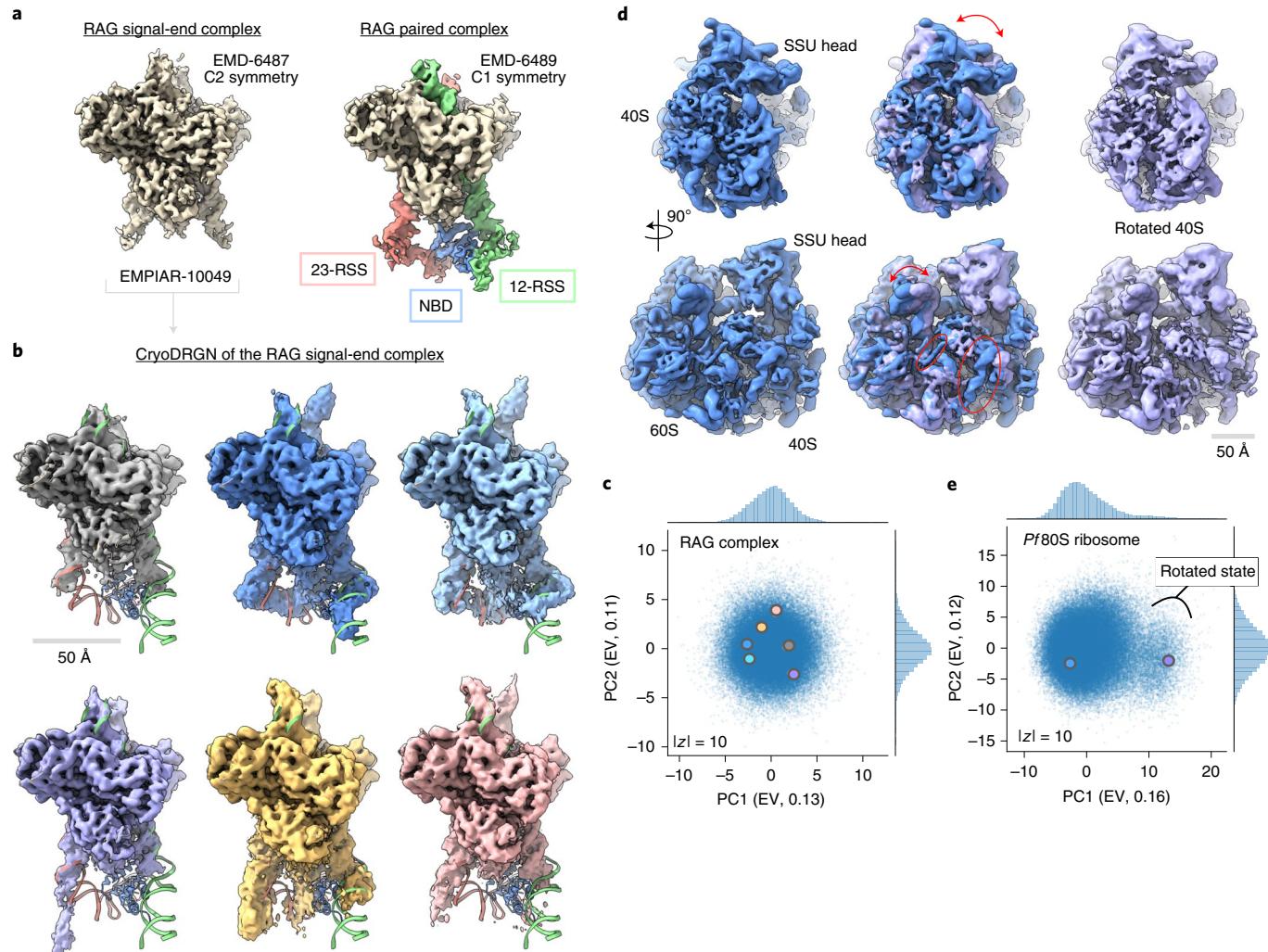
**Fig. 3 | CryoDRGN heterogeneous reconstruction of simulated datasets.** **a**, Ground-truth density maps simulating continuous heterogeneity generated by sampling conformations along a 1D conformational transition from the leftmost to the rightmost structure (left). Particles along this conformation transition were sampled uniformly (top) or from a mixture of Gaussian distributions of varying widths (middle, bottom) to simulate various degrees of cooperative transitions between three states. **b**, Compositional heterogeneity simulated by mixing particles of the 30S, 50S and 70S bacterial ribosomal complexes. **c**, Density maps reconstructed by cryoDRGN trained on the uniformly sampled dataset in **a**. Six structures were sampled from the specified values of the latent variable (top). ‘Per-image’ FSC curves are shown, where for 100 images equally spaced along the reaction coordinate, we computed the FSC between a map generated by cryoDRGN at the predicted latent encoding for each image and the ground-truth density map for that image (bottom). See Methods for description of the ‘per-image’ FSC approach. **d**, Density maps reconstructed by cryoDRGN from the compositional dataset in **b** and their FSCs to the corresponding ground-truth density map. **e-h**, Predicted latent space encoding for each particle image of different simulated datasets versus the ground-truth reaction coordinate describing the motion (**e-g**) or the ground-truth class assignment (**h**). All cryoDRGN reconstructions use a 1D latent variable model.

the uniform conformational heterogeneity dataset, cryoDRGN reconstructed density maps that reproduced the ground-truth continuous motion of the complex (Fig. 3c). When trained on the compositional dataset, cryoDRGN reconstructed density maps of the 30S, 50S and 70S ribosomes at distinct values of the latent variable (Fig. 3d).

In addition to reconstructing heterogeneous density maps, cryoDRGN produces a latent encoding for each particle that can be compared to the ground-truth reaction coordinate (Fig. 3e–h). For the datasets with continuous conformational changes, the latent encoding of each image correlated with the position along the reaction coordinate given by the dihedral angle of the underlying model (Spearman  $r = -0.996$ ,  $0.992$  and  $0.988$  for uniform, cooperative and noncontiguous, respectively) (Fig. 3e). We observed that the qualitative features of the distribution of latent encodings matched the ground truth, with three modes in the latent encoding distribution for the cooperative dataset (Fig. 3f) and distinct clusters for the noncontiguous and compositional datasets (Fig. 3g,h). We note that, in general, the parameterization of a reaction coordinate is non-unique (for example, when described by the learned latent variable or by the dihedral angle, leading to different marginal distributions in Fig. 3e). To quantitatively assess whether cryoDRGN learned the correct distribution of structures, we computed a ‘per-image’ FSC, which compares reconstructed density maps with the ground truth on images across the reaction coordinate (Methods), and found that the reconstructed structures of all four

datasets correlated well with the ground-truth distribution (Fig. 3c and Extended Data Fig. 1).

**CryoDRGN uncovers residual heterogeneity from ‘homogeneous’ cryo-EM datasets.** We next evaluated cryoDRGN’s ability to perform heterogeneous reconstruction on real cryo-EM datasets of the RAG complex<sup>31</sup> and the P<sub>f</sub>80S<sup>32</sup> ribosome from above. Ru et al. reported RAG complex structures from two distinct datasets: the ‘signal-end complex’, which failed to resolve the distal ends of the 12-recombination signal sequence (RSS) and 23-RSS DNA elements or the nonamer binding domain (NBD) of RAG1, and the ‘paired complex’, which resolved these elements at sufficient resolution for atomic model building (Fig. 4a). To test whether cryoDRGN could newly uncover heterogeneity of these distal elements in the ‘signal-end complex’, we trained a cryoDRGN ten-dimensional (10D) latent variable model on the deposited particle images (EMPIAR-10049)<sup>31</sup>. We found that cryoDRGN revealed substantial heterogeneity of the 12-RSS, 23-RSS and NBD (Fig. 4b). In addition to maps that only resolve the symmetric core (light gray in Fig. 4b), cryoDRGN revealed structures with RSS positioning that aligned with the canonical conformation found in the ‘paired complex’ atomic model<sup>31</sup> (dark blue), tilting of the RSS strands (light blue), linear 23-RSS DNA (purple), as well as the presence (yellow) and absence (coral) of the NBD. These representative maps were selected out of a large ensemble of generated structures (Methods) from different regions of the latent space (Fig. 4c). A trajectory

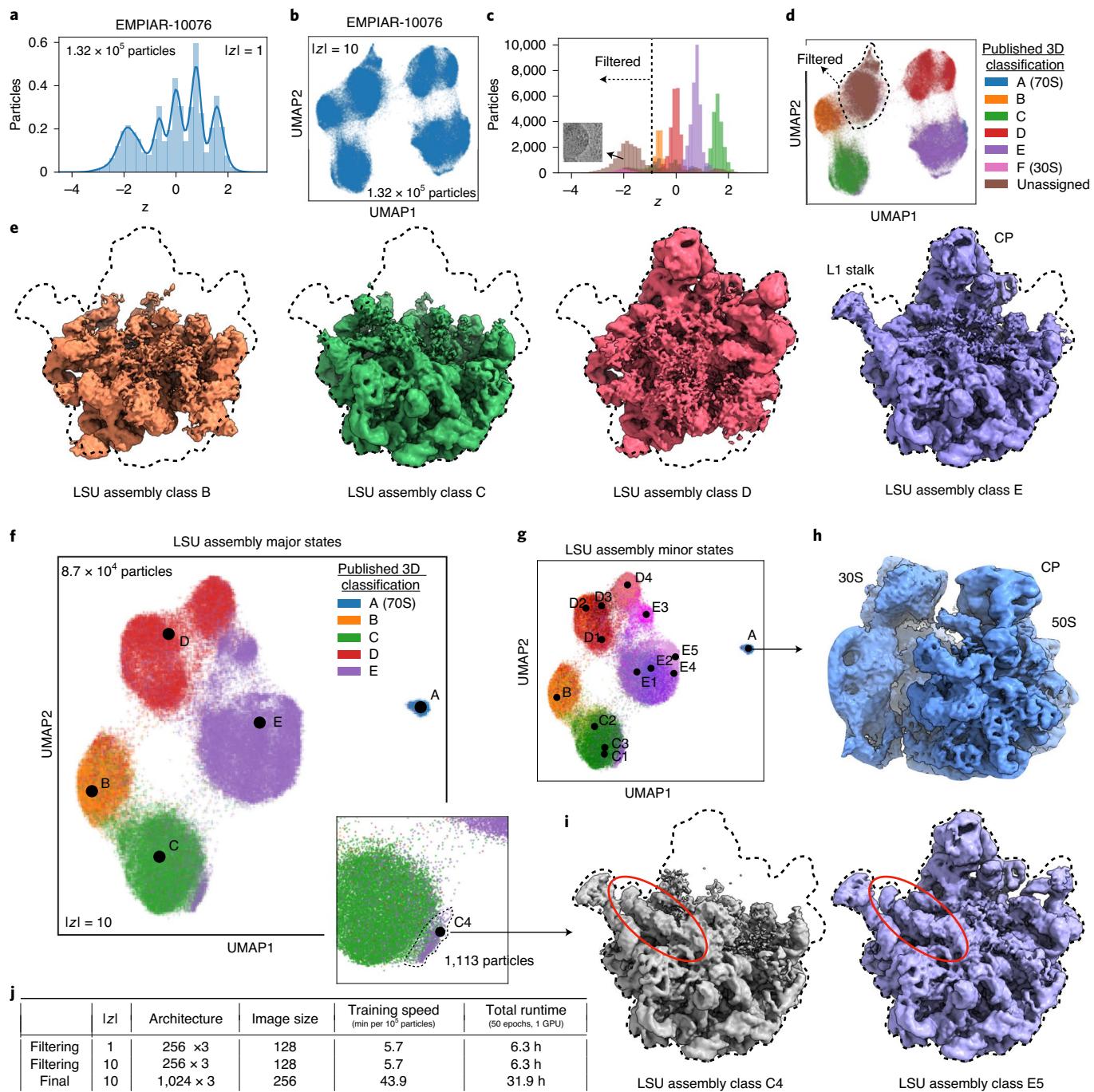


**Fig. 4 | Discovery of residual heterogeneity in ‘homogeneous’ datasets.** **a**, Published density maps of the 369-kDa RAG1-RAG2 complex. The signal-end complex (left) shows the C2 symmetric core, and the paired complex (right) resolves additional asymmetric 12- and 23-RSS DNA elements and the RAG1 NBD that extend below the core. **b**, Representative density maps of the RAG signal-end complex (EMPIAR-10049)<sup>31</sup> reconstructed by cryoDRGN. Density maps resolve variable conformations of the 12- and 23-RSS DNA elements and the NBD, which are missing from the homogeneous refinement. The docked atomic model (PDB 3JBW) of the RAG paired complex includes an asymmetric conformation of the RSS and NBD elements that extend from the core RAG complex. **c**, Latent space representation of particle images from the EMPIAR-10049 dataset (ref. <sup>31</sup>), visualized using PCA with explained variance (EV) noted. Structures from **b** are marked with the corresponding color. **d**, Density map of the 4.2-MDa Pf80S ribosome (EMPIAR-10028)<sup>32</sup> in an unrotated (blue) and rotated (purple) state reconstructed by cryoDRGN. Arrows indicate rotation of the 40S subunit relative to the 60S subunit (top) and motion of the L1 stalk (bottom). Circles indicate differential occupancy of the C-terminal helix of eL8 and an rRNA helix between the two states. **e**, Latent space representation of particle images from the EMPIAR-10028 dataset (ref. <sup>32</sup>), visualized using PCA with explained variance noted. Structures from **d** are marked with the corresponding color. A cluster of particles separated along PC1 of **d** that corresponds to the rotated state of the Pf80S ribosome is noted. Additional density maps from these datasets are shown in Extended Data Fig. 3 and Supplementary Videos 1 and 2.

sampling from the continuous distribution modeled by cryoDRGN is shown in Supplementary Video 1. To validate the presence of these heterogeneous states in the dataset, we performed heterogeneous 3D refinement in cryoSPARC<sup>16</sup> using the cryoDRGN density maps as initial models, which reproduced the heterogeneity of the RSS elements (Extended Data Fig. 2). Subsequent work by Ru et al. suggested that the conformational dynamics and asymmetric positioning of the 12- and 23-RSS by the NBD in the pre-cleavage form are fundamental to the structural mechanism underlying the 12-23 rule of variable, diversity, and joining gene segment (V(D)J) recombination<sup>33</sup>. Our results newly demonstrate that such heterogeneity was also present in the post-cleavage ‘signal-end’ RAG complex.

While analyzing a homogeneous reconstruction of the *Pf*80S ribosome, Wong et al. observed flexibility in the small subunit (SSU)

head region and missing density for peripheral rRNA expansion segment elements<sup>32</sup>. To explore if this unresolved density resulted from residual heterogeneity, we trained a cryoDRGN 10D latent variable model on their deposited dataset (EMPIAR-10028)<sup>32</sup> and reconstructed an ensemble of density maps that not only contained structures consistent with the homogeneous reconstruction but also revealed rotation of the 40S SSU (Fig. 4d), heterogeneity within the SSU head (Extended Data Fig. 3) and motion of many peripheral rRNA expansion segments (Supplementary Video 2). By visualizing representative 40S-rotated and unrotated density maps, we found that cryoDRGN was able to simultaneously capture the large-scale intersubunit rotation and coordinated smaller-scale structural rearrangements, including motion of the L1 stalk, disappearance of an rRNA helix and the disappearance of the intersubunit bridge



**Fig. 5 | CryoDRGN heterogeneous reconstruction of the assembly landscape of the bacterial LSU.** **a,b**, Latent space representation of particle images of the assembling LSU (EMPIAR-10076)<sup>12</sup> as a normalized histogram or UMAP embeddings after training a cryoDRGN 1D or 10D latent variable model, respectively. **c,d**, Latent space representation of particles colored by major LSU assembly state assigned from the 3D classification in Davis et al.<sup>12</sup>. Impurities in the dataset were assigned and subsequently filtered based on a cutoff of  $z = -1$  in the 1D case (dotted line) and cluster assignment from a five-component Gaussian mixture model (GMM) in the 10D case. The dashed line in **d** indicates a rough outline of cluster assignment, shown in Extended Data Fig. 5. **e**, Density maps of the four major assembly states of the LSU reconstructed by cryoDRGN after training on the filtered dataset, colored by major and minor assembly states assigned from the 3D classification in Davis et al.<sup>12</sup>. Points denote cluster centers for the corresponding assembly state. Major assembly state labels correspond to the structures from **e**. Inset shows a magnified view of the state C cluster and a population of particles originally misclassified into state E. **h,i**, CryoDRGN reconstruction of additional density maps, showing the 70S ribosome, an impurity during purification and LSU minor states C4 and E5. The newly identified C4 state resembles major state C in maturation but contains rRNA helix 68, previously present only in mature assembly states E4 and E5. **j**, Hyperparameters and runtime of the initial pilot experiments for particle filtering (**a-d**) and the final cryoDRGN model (**e-i**) trained on the assembling LSU dataset. Additional density maps are shown in Extended Data Fig. 6 and Supplementary Video 3.

formed by the C-terminal helix of eL8, which is consistent with Sun et al.'s characterization of *P*/*f*80S dynamics<sup>34</sup> (Fig. 4d).

We then visualized the 10D latent space representation of the *P*/*f*80S particles with PCA (Fig. 4e) and with uniform manifold approximate and projection (UMAP)<sup>35</sup> (Extended Data Fig. 4). The 40S-rotated density map originated from a region of the particle distribution separated along the first PC of the latent space. To validate the presence of this state, we extracted 4,889 particles constituting the outlying cluster (Methods). Traditional homogeneous reconstruction of these particles in cryoSPARC produced a 6.4-Å reconstruction of the rotated 40S state that was consistent with the cryoDRGN structure (Extended Data Fig. 4). Additionally, by sampling many density maps from the latent space, we observed that structures with density missing from the SSU head group were located within a subregion of the main cluster of the UMAP visualization (Extended Data Fig. 3). We hypothesize that the 40S-rotated state appears as a visually distinct cluster because more mass changes to rotate the entire 40S subunit, as opposed to the missing SSU head group state, which involves changes in a smaller region of the 40S subunit.

**CryoDRGN automatically partitions assembly states of the bacterial ribosome.** Next, we sought to evaluate cryoDRGN on a highly heterogeneous cryo-EM dataset of the *Escherichia coli* large ribosomal subunit (LSU) undergoing assembly (EMPIAR-10076)<sup>12</sup>. This dataset is known to contain substantial compositional and conformational heterogeneity; in the original analysis, multiple expert-guided rounds of hierarchical 3D classification resulted in 13 discrete structures that were grouped into four major assembly states. Here, we aimed to assess if cryoDRGN could automatically reveal these heterogeneous states without user-guided 3D classification.

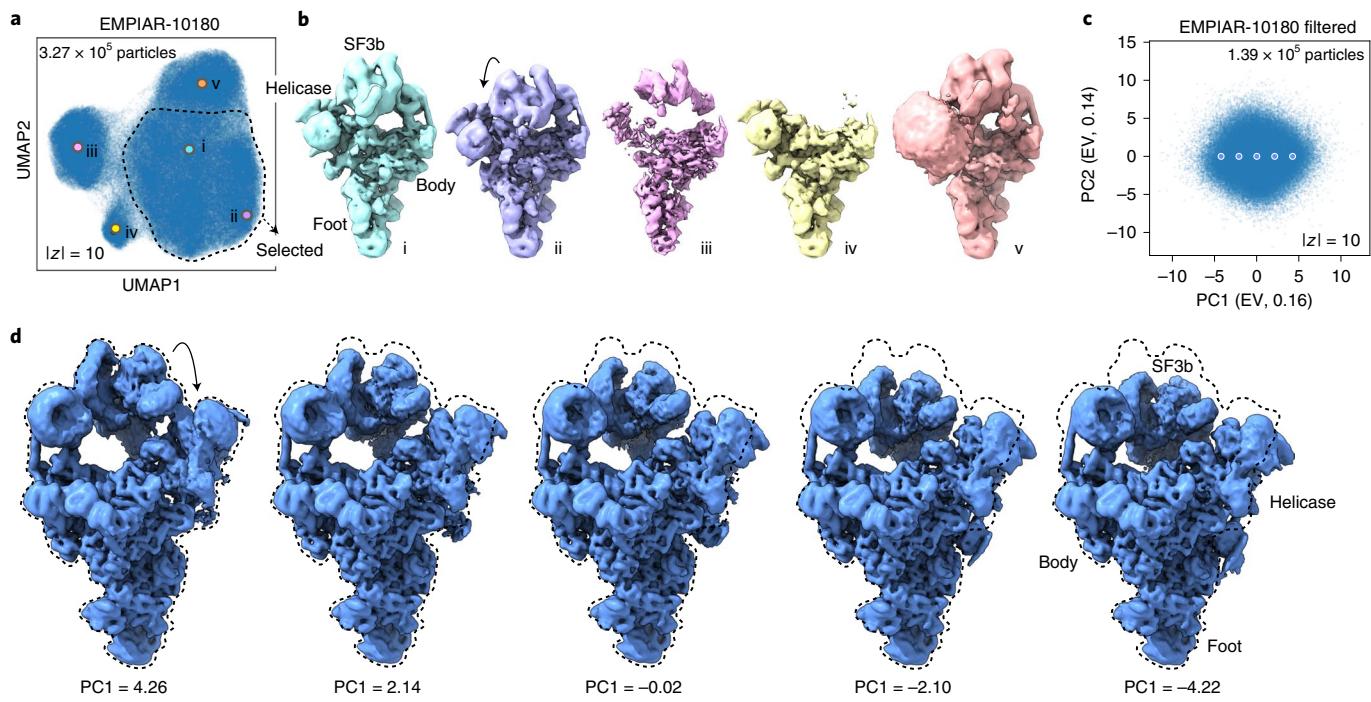
As initial pilot experiments, we first trained 1D and 10D latent variable models on downsampled images of the dataset (Methods). The dataset's latent space representation exhibited distinct peaks in the 1D case or clusters in the 10D case when visualized with UMAP<sup>35</sup> (Fig. 5a,b) that corresponded to the major assembly states when grouped by the published 3D classification labels (Fig. 5c,d). As the particles were obtained by crudely fractionating a lysate to capture the full ensemble of cellular assembly intermediates, a substantial fraction of the published particle stack corresponds to 30S or non-ribosomal impurities. These unassigned particles were outliers in the latent representation (Fig. 5c and Extended Data Fig. 5), and neither 2D class averages nor a traditional 3D reconstruction of these particles produced structures consistent with assembling LSU ribosomes (Extended Data Fig. 5). As we did not wish to devote representation capacity of the cryoDRGN neural networks to modeling these impurities, we used the latent representation to filter the dataset before further analysis, taking the intersection of the particle stack after filtering based on the 1D and 10D latent variable model (Methods).

To explore the heterogeneity within the LSU assembly states, we trained a cryoDRGN 10D latent variable model on the remaining images at higher resolution (Methods). The decoder network reconstructed density maps that matched the reported major (Fig. 5e) and minor (Extended Data Fig. 6) assembly states of the LSU. We visualized the encodings of particle images in the 10D latent space with UMAP and observed clusters corresponding to the major (Fig. 5f) and minor states (Fig. 5g and Supplementary Fig. 6) of LSU assembly after coloring by the published 3D classification. From the latent representation, we also noted a clearly separated cluster of particles assigned to class A, and structures sampled from this region of latent space reconstructed the 70S ribosome, an impurity in the dataset (Fig. 5h). Finally, we identified a small cluster of ~1,100 particles adjacent to the class C cluster, the particles of which were originally classified into class E (Fig. 5f, inset). The density map

reconstructed by the decoder from this region revealed a previously unreported assembly intermediate that we newly define as class C4 (Fig. 5i). Like the other class C structures, class C4 lacked the central protuberance but possessed clearly resolved density for rRNA helix 68, which was only present in the mature E4 and E5 classes from Davis et al.<sup>12</sup>. Traditional homogeneous reconstruction of the particle images constituting this cluster reproduced a similar, albeit lower-resolution structure, which confirmed the existence of this structural state in the original dataset (Extended Data Fig. 7). We found that the cryoDRGN latent representation was highly reproducible across replicates (Extended Data Fig. 8). CryoDRGN experiments and runtimes are summarized in Fig. 5j. In addition to illustrating cryoDRGN's ability to model extremely heterogeneous datasets without user-driven classification, this analysis further demonstrated that cryoDRGN can identify novel and rare (~1% of all particles) structural classes that would likely be overlooked by traditional hierarchical classification.

**CryoDRGN reveals dynamic continuous motions in the pre-catalytic spliceosome.** Finally, to assess cryoDRGN's ability to model large continuous conformational changes, we reanalyzed a dataset of the pre-catalytic spliceosome (EMPIAR-10180)<sup>36</sup>. Using expert-guided focused classifications, Plaschka et al. reconstructed a composite map for this complex and suggested that the complex sampled a continuum of conformations with large motions of the SF3b subcomplex<sup>36</sup>. In our analysis, we first trained a 10D latent variable model on the downsampled images using image poses derived from a consensus reconstruction (Methods). Multiple clusters were observed in the latent space encodings of the dataset's particle images (Fig. 6a). In sampling structures from the latent space, the generated density maps revealed expected spliceosome conformations from the largest cluster, poorly resolved structures (likely due to imaging artifacts from the leftmost cluster), structures lacking density for the SF3b subcomplex from a third cluster and extra density of the U2 core, which is thought to be highly dynamic<sup>13</sup>, from the uppermost cluster (Fig. 6b). To focus our analysis on bona fide pre-catalytic spliceosome particles, we leveraged the latent space representation to eliminate any particles that mapped to the undesired clusters from two replicate runs (Methods).

With the filtered particle stack, we trained a 10D model on higher resolution images and visualized the dataset's latent encodings in 2D using PCA (Fig. 6c). The visualized data manifold was unfeatured, consistent with a molecule undergoing non-cooperative conformational changes. By generating structures along the first principal component (PC) of the latent space encodings, we reconstructed a trajectory of the SF3b and helicase subcomplexes in motion, smoothly transitioning from an elongated state to one compressed against the body of the spliceosome (Fig. 6d). This large-scale motion was consistent with motions derived from the first PC of rigid-body orientations from multibody analysis (Extended Data Fig. 9) and in the first PC of the linear subspace model from 3DVA (Extended Data Fig 10). A similar traversal along the second PC produced a continuous trajectory of the SF3b and helicase subcomplexes moving in opposition (Supplementary Fig. 1 and Supplementary Video 4). The anticorrelated motion of the SF3b and helicase subcomplexes in PC2, together with their correlated motion in PC1, suggests that the two domains move independently in the imaged ensemble. Finally, although trajectories along latent space PCs provide a summary of the extent of variability in the structure, cryoDRGN can also generate structures at arbitrary points from the latent space. By traversing along the nearest neighbor graph of the latent encodings and generating structures at the visited nodes, cryoDRGN generated a plausible trajectory of the conformations adopted by the pre-catalytic spliceosome (Supplementary Video 4), highlighting the potential of single-particle cryo-EM to uncover the conformational dynamics of molecular machines.



**Fig. 6 | CryoDRGN heterogeneous reconstruction of the pre-catalytic spliceosome.** **a**, UMAP visualization of the latent space representation of particle images of the pre-catalytic spliceosome (EMPIAR-10180)<sup>36</sup> after training a 10D latent variable model with cryoDRGN. **b**, Representative structures generated at points shown in **a** that depict the expected structures of the pre-catalytic spliceosome (**i**,**ii**), structures likely corrupted by imaging artifacts (**iii**), the complex lacking the SF3b subcomplex (**iv**) and the complex with the U2 core (**v**). Density maps are shown at identical isosurface levels except for (**v**), which required a lower value to highlight the U2 core. **c**, PCA projection of latent space encodings after training a 10D latent variable model on the dataset filtered for the selected region in **a**. **d**, Structures generated by traversing along PC1 of the latent space representation at the points shown in **c**. Additional density maps are shown in Supplementary Fig. 1 and Supplementary Video 4.

## Discussion

This work introduces cryoDRGN, a method using neural networks to reconstruct 3D density maps from heterogeneous single-particle cryo-EM datasets. The power of this approach lies in its ability to represent heterogeneous structures without simplifying assumptions on the type of heterogeneity. In principle, cryoDRGN is able to represent any distribution of structures that can be approximated by a deep neural network, a broad class of function approximators for continuous, nonlinear functions<sup>28</sup>. This flexibility contrasts with existing methods that impose limiting assumptions on the types of structural heterogeneity present in the sample. For example, 3D classification assumes a mixture of discrete structural classes, multibody refinement assumes conformational changes are composed of user-defined rigid-body motions, and 3DVA assumes that heterogeneity is generated from linear combinations of density maps. Although these approaches have proven useful, their model for heterogeneity is often mismatched with the true structural heterogeneity in many systems and thus can introduce bias into reconstructions. By contrast, we empirically show that the cryoDRGN architecture can model both discrete compositional heterogeneity and continuous conformational changes without the aforementioned structural assumptions. For example, we discovered heterogeneous states of the RAG complex and P<sub>f</sub>80S ribosome that were originally averaged out of the homogeneous reconstruction. When analyzing the dataset of the assembling *E. coli* LSU, cryoDRGN learned an ensemble of LSU assembly states without a priori specification of the number of states or initial models as is required for 3D classification. Finally, when analyzing the pre-catalytic spliceosome, we found that the continuous conformational changes reconstructed by cryoDRGN lacked the rigid-body boundary artifacts from the mask-based multibody

refinement approach<sup>19</sup> (Extended Data Fig. 9) or linear interpolation artifacts from the linear subspace 3DVA model<sup>23</sup> (Extended Data Fig. 10).

**Interpretation of the latent space.** A key feature of cryoDRGN is its ability to provide a low-dimensional representation of the dataset's heterogeneity given by each particle's latent encoding. Subject to optimization, cryoDRGN organizes the latent space such that structurally related particles are in close proximity. In simulated and real datasets, we find that continuous motions are embedded along a continuum in latent space (Figs. 3e–g and 6c) and that compositionally distinct states manifest as clusters (Figs. 3h and 5f). These empirical results demonstrate that visualization of the distribution of latent encodings can be informative in exploring the structural heterogeneity within the imaged ensemble and even suggest a possible interpretation of the latent space as a pseudo-conformational landscape. However, we note that cryoDRGN's objective function aims only to reproduce the distribution of structures and does not guarantee that the latent space layout (or its 2D visualization) will produce interpretable features of the underlying energy landscape. Furthermore, structures reconstructed from unoccupied regions of the latent space will not in general correspond to true physical structures, as cryoDRGN optimizes the likelihood of the observed data, and these structures are not observed.

Finally, in real datasets, there may exist images that do not originate from the standard single-particle image formation model, for example, false positives encountered during particle picking. We demonstrated the utility of the latent space representation in identifying such impurities, ice artifacts and other out-of-distribution particle images that may be filtered out in subsequent analyses (Figs. 5a–d and 6a).

We emphasize that different datasets have diverse sources of heterogeneity, and thus the interpretation of the cryoDRGN latent space is highly dataset dependent. We provide interactive analysis tools in the cryoDRGN software for exploring the learned latent space.

**Visualizing structural trajectories.** In addition to encoding particles in an unsupervised manner, cryoDRGN can reconstruct 3D density maps from user-defined positions in latent space. Because cryoDRGN learns a generative model for structure, an unlimited number of structures can be generated and analyzed, thus enabling visualization of structural trajectories. By leveraging the latent encodings of the particle images, users can directly traverse the data manifold and only sample structures from regions of latent space with substantial particle occupancy. Indeed, we applied a well-established graph-traversal algorithm<sup>37</sup> to visualize data-supported motions in the RAG complex, the P<sub>f</sub>80S ribosome, bL17-independent assembly of the bacterial ribosome and the pre-catalytic spliceosome (Supplementary Videos 1–4). We note that while this approach is useful in visualizing potential structural changes linking one state to another, they do not necessarily reveal the kinetically preferred path.

**Practical considerations in choosing training hyperparameters.** Although this method emphasizes an unsupervised approach for analyzing structural heterogeneity, cryoDRGN does require that the user define the dimensionality of the latent space and the architecture of both the encoder and decoder networks. We find that, in practice, training a smaller architecture on downsampled images is effective at distinguishing bona fide particles from contaminants and imaging artifacts (Figs. 5a–d and 6a), and we recommend users initially employ such pilot experiments to filter their dataset. Additionally, we find that, in our tested datasets, a 10D latent space provides sufficient representation capacity to effectively model structural heterogeneity and that this 10D space can be readily visualized with PCA or UMAP. Notably, we recommend the use of such a 10D latent space instead of lower-dimensional space, as we found that 10D spaces result in much more rapid overall training, which is consistent with similar observations of related overparameterized neural network architectures<sup>38</sup>. Finally, users must specify the number of nodes and layers in the neural networks, hyperparameters that limit the complexity of the learned function. Here, we find an inverse relationship between neural network size and the achievable resolution of a given structure (Fig. 2b). Training larger networks on larger images is substantially slower (Fig. 2e), and we recommend that users perform an initial assessment using downsampled images and relatively small networks before proceeding to high-resolution reconstructions. We note that use of excessively complex models (that is, large architectures or latent variable dimensions) can lead to overfitting, which may be alleviated by standard neural network regularization techniques, such as early stopping or using a simpler model<sup>38</sup>. We provide recommended training settings in the cryoDRGN software.

**Discovering new states using cryoDRGN.** CryoDRGN can be used to identify novel clusters of structurally related particles, which can then be visualized by generating density maps from that region of latent space. Indeed, in analyzing the LSU assembly dataset, we noted a new structural state, C4, that was missed in traditional hierarchical classification. C4 provides structural evidence that a functionally critical intersubunit helix (h68) can dock in a native conformation in the absence of the central protuberance (Fig. 5i). Notably, we could validate the existence of this state by performing traditional homogeneous refinement using ~1,000 particles from this cluster in the cryoDRGN latent space (Extended Data Fig. 7). Although we were able to identify this state from a distinct cluster

present in the UMAP visualization (Fig. 5g), in general, the definition of distinct ‘states’ may not be as readily apparent (for example, the ‘missing SSU head’ state in Extended Data Fig. 3), and we view the unsupervised identification of states from the cryoDRGN structural ensemble as an exciting area to pursue.

In future work, we envision using cryoDRGN to reveal the number of discrete classes and their constituent particles and to produce initial 3D models that could be used as inputs for a traditional 3D reconstruction. Given the mature state of such tools<sup>39,40</sup>, this data-driven classification approach, followed by traditional homogeneous reconstruction, particle polishing and higher-order image aberration correction, has the potential to produce high-resolution structures of the full spectrum of discrete structural states.

**De novo pose estimation.** As implemented, cryoDRGN uses pose estimates resulting from a traditional consensus 3D reconstruction. In analyzing four publicly available datasets, we found that such consensus pose estimates were sufficiently accurate to generate meaningful latent space encodings and to produce interpretable density maps of distinct structures. It is clear, however, that this approach will fail if the degree of structural heterogeneity in the dataset results in inaccurate pose estimates. For example, a mixture of structurally unrelated complexes will align poorly to a consensus structure, and thus produce poor pose estimates. Notably, our framework is differentiable with respect to pose variables, which, in principle, could allow on-the-fly pose refinement or de novo pose estimation. Future work will explore the efficacy of incorporating such features to enable fully unsupervised reconstruction of heterogeneous distributions of protein structure from cryo-EM images.

## Online content

Any methods, additional references, Nature Research 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/s41592-020-01049-4>.

Received: 25 March 2020; Accepted: 18 December 2020;

Published online: 4 February 2021

## References

1. Nogales, E. The development of cryo-EM into a mainstream structural biology technique. *Nat. Methods* **13**, 24–27 (2015).
2. Cheng, Y. Single-particle cryo-EM—how did it get here and where will it go. *Science* **361**, 876–880 (2018).
3. Bammes, B. E., Rochat, R. H., Jakana, J., Chen, D.-H. & Chiu, W. Direct electron detection yields cryo-EM reconstructions at resolutions beyond 3/4 Nyquist frequency. *J. Struct. Biol.* **177**, 589–601 (2012).
4. Suloway, C. et al. Automated molecular microscopy: the new Leginon system. *J. Struct. Biol.* **151**, 41–60 (2005).
5. Li, X. et al. Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat. Methods* **10**, 584–590 (2013).
6. Zhang, K. Gctf: real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).
7. Brubaker, M. A., Punjani, A. & Fleet, D. J. Building proteins in a day: efficient 3D molecular reconstruction. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 3099–3108 (CVPR, 2015).
8. Scheres, S. H. W. A Bayesian view on cryo-EM structure determination. *J. Mol. Biol.* **415**, 406–418 (2012).
9. Bepler, T. et al. Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nat. Methods* **16**, 1153–1160 (2019).
10. Ahmed, T., Yin, Z. & Bhushan, S. Cryo-EM structure of the large subunit of the spinach chloroplast ribosome. *Sci. Rep.* **6**, 35793 (2016).
11. Wrapp, D. et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* **367**, 1260–1263 (2020).
12. Davis, J. H. et al. Modular assembly of the bacterial large ribosomal subunit. *Cell* **167**, 1610–1622 (2016).

13. Haselbach, D. et al. Structure and conformational dynamics of the human spliceosomal B<sup>act</sup> complex. *Cell* **172**, 454–464 (2018).
14. Sigworth, F. J. Principles of cryo-EM single-particle image processing. *Microscopy* **65**, 57–67 (2016).
15. Scheres, S. H. W. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
16. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
17. Lyumkis, D., Brilot, A. F., Theobald, D. L. & Grigorieff, N. Likelihood-based classification of cryo-EM images using FREALIGN. *J. Struct. Biol.* **183**, 377–388 (2013).
18. Grant, T., Rohou, A. & Grigorieff, N. *cis*TEM, user-friendly software for single-particle image processing. *eLife* **7**, e14874 (2018).
19. Nakane, T., Kimanius, D., Lindahl, E. & Scheres, S. H. Characterisation of molecular motions in cryo-EM single-particle data by multi-body refinement in RELION. *eLife* **7**, e36861 (2018).
20. Liu, W. & Frank, J. Estimation of variance distribution in three-dimensional reconstruction. I. Theory. *J. Opt. Soc. Am. A* **12**, 2615–2627 (1995).
21. Penczek, P. A., Kimmel, M. & Spahn, C. M. T. Identifying conformational states of macromolecules by eigen-analysis of resampled cryo-EM images. *Structure* **19**, 1582–1590 (2011).
22. Tagare, H. D., Kucukelbir, A., Sigworth, F. J., Wang, H. & Rao, M. Directly reconstructing principal components of heterogeneous particles from cryo-EM images. *J. Struct. Biol.* **191**, 245–262 (2015).
23. Punjani, A. & Fleet, D. J. 3D Variability Analysis: directly resolving continuous flexibility and discrete heterogeneity from single particle cryo-EM images. Preprint at *bioRxiv* <https://doi.org/10.1101/2020.04.08.032466> (2020).
24. Dashti, A. et al. Trajectories of the ribosome as a Brownian nanomachine. *Proc. Natl Acad. Sci. USA* **111**, 17492–17497 (2014).
25. Frank, J. & Ourmazd, A. Continuous changes in structure mapped by manifold embedding of single-particle data in cryo-EM. *Methods* **100**, 61–67 (2016).
26. Moscovich, A., Halevi, A., Andén, J. & Singer, A. Cryo-EM reconstruction of continuous heterogeneity by Laplacian spectral volumes. *Inverse Probl.* **36**, 024003 (2020).
27. Lederman, R. R. & Singer, A. Continuously heterogeneous hyper-objects in cryo-EM and 3-D movies of many temporal dimensions. Preprint at <https://arxiv.org/abs/1704.02899> (2017).
28. Hornik, K., Stinchcombe, M. B. & White, H. Multilayer feedforward networks are universal approximators. *Neural Netw.* **2**, 359–366 (1989).
29. Bracewell, R. N. Strip integration in radio astronomy. *Aust. J. Phys.* **9**, 198–217 (1956).
30. Kingma, D. P. & Welling, M. Auto-encoding variational Bayes. In *International Conference on Learning Representations* (ICLR, 2014).
31. Ru, H. et al. Molecular mechanism of V(D)J recombination from synaptic RAG1-RAG2 complex structures. *Cell* **163**, 1138–1152 (2015).
32. Wong, W. et al. Cryo-EM structure of the *Plasmodium falciparum* 80S ribosome bound to the anti-protozoan drug emetine. *eLife* **3**, e01963 (2014).
33. Ru, H., Zhang, P. & Wu, H. Structural gymnastics of RAG-mediated DNA cleavage in V(D)J recombination. *Curr. Opin. Struct. Biol.* **53**, 178–186 (2018).
34. Sun, M. et al. Dynamical features of the *Plasmodium falciparum* ribosome during translation. *Nucleic Acids Res.* **43**, 10515–10524 (2015).
35. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for dimension reduction. Preprint at <https://arxiv.org/abs/1802.03426> (2018).
36. Plaschka, C., Lin, P.-C. & Nagai, K. Structure of a pre-catalytic spliceosome. *Nature* **546**, 617–621 (2017).
37. Cormen, T. H., Leiserson, C. E., Rivest, R. L. & Stein, C. *Introduction to Algorithms*. 595–601 (MIT Press and McGraw-Hill, 2009).
38. Zhang, C., Bengio, S., Hardt, M., Recht, B. & Vinyals, O. Understanding deep learning requires rethinking generalization. In *International Conference on Learning Representations* (ICLR, 2017).
39. Zivanov, J., Nakane, T. & Scheres, S. H. W. Estimation of high-order aberrations and anisotropic magnification from cryo-EM data sets in RELION-3.1. *IUCrJ* **7**, 253–267 (2020).
40. Punjani, A., Zhang, H. & Fleet, D. J. Non-uniform refinement: adaptive regularization improves single particle cryo-EM reconstruction. *Nat. Methods* **17**, 1214–1221 (2020).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2021

## Methods

**The cryoDRGN method.** *Coordinate-based networks to represent 3D structure.* The cryoDRGN method performs heterogeneous cryo-EM reconstruction by learning a neural network representation of 3D structure. In particular, we use a positionally encoded multilayer perceptron (MLP) to approximate the function  $V : \mathbb{R}^{3+n} \rightarrow \mathbb{R}$ , which models structures as generated from an  $n$ -dimensional continuous latent space. We refer to this architecture as a ‘coordinate-based neural network’ (refs. 41,42), as we explicitly model the volume  $V$  as a function of Cartesian coordinates.

Without loss of generality, we model volumes on the domain  $(-0.5, 0.5)^3$ . Instead of directly supplying the 3D Cartesian coordinates  $k$  to the deep coordinate network, each coordinate is featurized with a fixed positional encoding function<sup>43</sup> consisting of  $D$  sinusoids with wavelengths following a geometric progression from 1 to the Nyquist limit:

$$pe^{(2i)}(k_j) = \sin\left(k_j D \pi \left(\frac{2}{D}\right)^{\frac{j}{D-1}}\right), i = 0, \dots, \frac{D}{2} - 1; k_j \in k$$

$$pe^{(2i+1)}(k_j) = \cos\left(k_j D \pi \left(\frac{2}{D}\right)^{\frac{j}{D-1}}\right), i = 0, \dots, \frac{D}{2} - 1; k_j \in k$$

where  $D$  is set to the image size (number of pixels along one dimension of the image, that is, a  $D \times D$  image) used in training. Empirically, we found that excluding the highest frequencies of the positional encoding led to better performance when training on noisy data, and we provide an option to modify the positional encoding function by increasing all wavelengths by a factor of  $2\pi$ .

**Training system.** This neural representation of 3D structure is learned via an image-encoder-volume-decoder architecture based on the VAE<sup>30,44</sup>. We followed the standard image formation model in single-particle cryo-EM, in which observed images are generated from projections of a volume at a random unknown orientation,  $R \in SO(3)$ . We used an additive Gaussian white noise model. Volume heterogeneity is generated from a continuous latent space, modeled by the latent variable  $z$ , where the dimension of  $z$  is a hyperparameter of the model.

Given an image  $X$ , the variational encoder  $q_\xi(z|X)$  produces a mean and variance,  $\mu_{z|X}$  and  $\Sigma_{z|X}$ , statistics that parameterize a Gaussian distribution with diagonal covariance, as the variational approximation to the true posterior  $p(z|X)$ . The prior on the latent variable is a standard normal distribution  $\mathcal{N}(0, I)$ . The positionally encoded MLP is used as the probabilistic decoder,  $p_\theta(V|k, z)$  and models structures in frequency space. Given Cartesian coordinate  $k \in \mathbb{R}^3$  and latent variable  $z$ , the probabilistic decoder predicts a Gaussian distribution over  $V(k, z)$ . The encoder and decoder are parameterized with fully connected neural networks with parameters  $\xi$  and  $\theta$ , respectively.

Because 2D projection images can be related to volumes as 2D central slices in Fourier space<sup>29</sup>, oriented 3D coordinates for a given image can be obtained by rotating a  $D \times D$  lattice spanning  $(-0.5, 0.5)^2$  originally on the  $x$ - $y$  plane by  $R$ , the orientation of the volume during imaging. Then, given a sample out of  $q_\xi(z|X)$  and the oriented coordinates, an image can be reconstructed pixel by pixel through the decoder. The reconstructed image is then translated by the image’s in-plane shift and multiplied by the contrast transfer function (CTF) before it is compared to the input image. The negative log likelihood of a given image under our model is computed as the mean square error between the reconstructed image and the input image. Following the standard VAE framework, the optimization objective is a variational lower bound of the model evidence:

$$\mathcal{L}(X; \xi, \theta) = E_{q_\xi(z|X)}(\log p(X|z)) - \beta KL(q_\xi(z|X) || p(z))$$

where the first term is the reconstruction error estimated with one Monte Carlo sample, the second Kullback–Leibler (KL) divergence term is a regularization term on the latent representation, and  $\beta$  is an additional hyperparameter, which we set by default to  $1/|z|$ . By training on many 2D slices with sufficiently diverse orientations, the 3D volume can be learned through feedback from the 2D views. For further details, we refer the reader to a preliminary version of the method described in the proceedings of the International Conference for Learning Representations<sup>41</sup>. The results presented here employ the training regime described by Zhong et al. using previously determined poses from a consensus reconstruction<sup>41</sup>.

**Datasets.** *Simulated compositionally heterogeneous dataset generation.* To generate the compositionally heterogeneous dataset, the 30S, 50S and 70S subunits of the *E. coli* ribosome were extracted from PDB 4YBB in PyMOL<sup>45</sup>. A density map of each subunit was generated from the atomic model using the molmap<sup>46</sup> command in Chimera<sup>47</sup> at a grid spacing of  $1.5 \text{ \AA}$  per pixel and a resolution of  $4.5 \text{ \AA}$ . The resulting volume was padded to a box size of  $D = 256$ , where  $D$  is the width in pixels along one dimension. Simulated particle images were generated with a custom Python script available in the cryoDRGN software by rotating the density map with a random rotation sampled uniformly from  $SO(3)$ , projecting along the  $z$  axis and shifting the image with an in-plane translation sampled uniformly from  $(-20, 20)^2$  pixels. Images were then downsampled to  $D = 128$  by Fourier clipping using a custom Python script, corresponding to a Nyquist limit

of  $6 \text{ \AA}$ . Projection images were multiplied with the CTF in Fourier space, where the CTF was computed from defocus values randomly sampled from those given in the EMPIAR-10028 dataset (ref. 32), no astigmatism, an accelerating voltage of  $300 \text{ kV}$ , a spherical aberration of  $2 \text{ mm}$  and an amplitude contrast ratio of  $0.1$ . An envelope function with a  $B$  factor of  $100 \text{ \AA}^2$  was applied. Noise was added with a signal-to-noise ratio of  $0.1$ , where the noise-free signal images were defined as the entire  $D \times D$  image. After performing this procedure for each subunit, 10,000, 15,000 and 25,000 simulated particles of the 30S, 50S and 70S ribosome, respectively, were combined.

*Simulated conformationally heterogeneous dataset generation.* To simulate continuous conformational heterogeneity, 50 density maps were generated along a 1D reaction coordinate defined by rotating a dihedral angle in an atomic model of a hypothetical protein complex. Each model was generated at  $0.03\text{-radian}$  increments of the bond rotation, leading to a total range of  $1.5$  radians. Density maps were generated using the molmap<sup>46</sup> command in Chimera<sup>47</sup> at a grid spacing of  $6 \text{ \AA}$  per pixel and resolution of  $12 \text{ \AA}$  and padded to a box size of  $D = 128$ . For the uniform dataset, 1,000 projection images were generated for each density map at random orientations, and in-plane translations were sampled from  $(-10, 10)^2$  pixels. For the nonuniform datasets, particles were generated along the reaction according to a three-component GMM with means at the 10th, 25th and 40th density map and standard deviations of  $0.09$  and  $0.03$  radians for the cooperative and noncontiguous datasets, respectively. Sampled reaction coordinate values were binned to convert into a particle distribution among the 50 generated density maps and clipped at values of the reaction coordinate beyond the 50 maps. A total of 50,000 particles were generated for each dataset. CTF and noise at a signal-to-noise ratio of  $0.1$  were added to all datasets using the same procedure described above with CTF defocus values randomly sampled from the EMPIAR-10028 dataset (ref. 32).

*Real cryo-EM datasets.* Picked particles and the star file containing CTF parameters were downloaded from the Electron Microscopy Public Image Archive (EMPIAR)<sup>48</sup> for datasets EMPIAR-10049, EMPIAR-10028, EMPIAR-10076 and EMPIAR-10180. Particle images were downsampled to the image size used in training by clipping in Fourier space with a custom Python script available in the cryoDRGN software.

**Consensus reconstructions.** Homogeneous 3D reconstruction of the *Pj80S* ribosome (EMPIAR-10028) was performed in cryoSPARC version 2.4 (ref. 16) using the ab initio reconstruction job followed by the homogeneous refinement job with default parameters. The final reconstruction reported a gold-standard (GS) FSC<sub>0.143</sub> (ref. 49) resolution of  $3.1 \text{ \AA}$  with a tight mask and  $4.1 \text{ \AA}$  when unmasked.

Homogeneous 3D reconstruction of the bL17-depleted ribosome assembly intermediates (EMPIAR-10076) was performed as above, leading to a final structure with a GSFSC<sub>0.143</sub> resolution of  $3.2 \text{ \AA}$  with a tight mask and  $4.8 \text{ \AA}$  when unmasked.

Homogeneous 3D reconstruction of the RAG complex (EMPIAR-10049) was performed as a ‘Homogeneous Refinement (NEW!)’ job in cryoSPARC version 2.15 with all default settings, including C1 symmetry. The asymmetric PC map of the RAG complex was used as an initial model (EMDB-6489) low-pass filtered by  $30 \text{ \AA}$ . The final structure had a GSFSC<sub>0.143</sub> resolution of  $3.6 \text{ \AA}$  with a tight mask and  $4.6 \text{ \AA}$  when unmasked.

Poses from a consensus reconstruction of the pre-catalytic spliceosome were obtained from the star file deposited in the EMPIAR-10180 dataset.

**CryoDRGN homogeneous reconstruction.** CryoDRGN decoder networks with no input latent variable were trained for 50 epochs on full-resolution images of the RAG complex ( $D = 192, 1.23 \text{ \AA}$  per pixel) and the *Pj80S* ribosome ( $D = 360, 1.34 \text{ \AA}$  per pixel). The tested architectures were MLPs with ReLU activations, where the network size was three hidden layers with a width of 128 nodes (denoted  $128 \times 3$ ,  $256 \times 3$ ,  $512 \times 3$ ,  $1,024 \times 3$  or  $1,024 \times 10$ ). Image poses were set to poses obtained from a consensus reconstruction in cryoSPARC, as described above<sup>16</sup>. Networks were trained on minibatches of eight images using the Adam optimizer with a learning rate of 0.0001. Once training was complete, the decoder network was evaluated on the 3D coordinates of a  $D \times D \times D$  voxel array spanning  $(-0.5, 0.5)^3$ , where  $D$  is the image size in pixels along one dimension. For visualization in Fig. 2, the RAG complex density maps were sharpened by  $-54 \text{ \AA}^2$  and  $-127.4 \text{ \AA}^2$  for the cryoDRGN and cryoSPARC maps, respectively, based on a Guinier analysis<sup>40</sup> performed in a custom Python script; both the cryoSPARC and cryoDRGN density maps of the *Pj80S* ribosome were sharpened using the published  $B$  factor of  $-80.1 \text{ \AA}^2$ .

**Map-to-map FSC.** FSC curves were computed between the cryoSPARC density maps and the cryoDRGN density maps using a custom Python script available in the cryoDRGN software. Real-space masks were defined by first thresholding the cryoDRGN volume at half of the 99.99th percentile density value. The mask was then dilated by  $25 \text{ \AA}$  from the original boundary, and a soft cosine edge was used to taper the mask to 0 at  $15 \text{ \AA}$  from the dilated boundary.

**CryoDRGN heterogeneous reconstruction. Model training.** A summary of the datasets, hyperparameters and runtimes for all cryoDRGN heterogeneous

reconstruction experiments is given in Supplementary Table 1. CryoDRGN encoder-decoder networks were trained from their randomly initialized values for each single-particle cryo-EM dataset. Image poses used for training were either the ground-truth poses for simulated datasets or poses obtained from a consensus reconstruction as described above. All networks were trained on minibatches of eight images using the Adam optimizer with a learning rate of 0.0001. After training, the dataset images were evaluated through the encoder to obtain the latent encoding for each image. We define the latent encoding as the maximum a posteriori value of  $q_i(z|X)$  predicted by the encoder.

**Latent space visualization.** For latent spaces with dimension greater than 2, the distribution of latent encodings were visualized with standard dimensionality reduction techniques, such as PCA and UMAP<sup>35</sup>. PCA projections of latent space particle distributions were computed using the implementation provided by scikit-learn<sup>50</sup>. Two-dimensional UMAP<sup>35</sup> embeddings were computed using version 0.4.1 of the Python implementation (<https://github.com/lmcinnes/umap>) with the default settings of  $k=15$  for the  $k$ -nearest neighbors graph and a minimum distance parameter of 0.1. Automated tools to analyze and visualize the latent space given the outputs of model training are provided in the cryoDRGN software.

**Density map generation.** Density maps were generated for a given value of the latent variable  $z$  by evaluating the trained decoder on  $z$  and the 3D coordinates of a  $D \times D \times D$  voxel array spanning  $(-0.5, 0.5)^3$ . To generate representative samples from different regions of the latent space for higher dimensional latent spaces ( $|z| > 1$ ), we performed  $k$ -means clustering of the dataset's latent encodings to partition the latent space into  $k$  regions. A representative density map for each region was generated at the 'on-data' cluster center; the latent encoding closest in Euclidean distance to the  $k$ -means cluster center was defined as the 'on-data' cluster center. Automated tools to generate  $k$  representative density maps following this procedure are provided in the cryoDRGN software.

**Heterogeneous reconstruction of simulated datasets.** For each simulated heterogeneous dataset, a 1D latent variable model was trained for 100 epochs. The encoder architecture was  $256 \times 3$ , and the decoder architecture was  $512 \times 5$ . The image poses used for training were the ground-truth image poses. After training on the uniform simulated dataset, structures shown in Fig. 3c were generated at the 5th, 23rd, 41st, 59th, 77th and 95th percentile values of the latent encodings and sharpened by a  $B$  factor of  $-100 \text{ \AA}^2$ . After training on the compositional simulated dataset, the structures shown in Fig. 3d were generated at the  $k$ -means cluster centers after performing  $k$ -means clustering with  $k=3$  on the latent encodings and sharpened by a  $B$  factor of  $-100 \text{ \AA}^2$ . Spearman correlation was computed using the implementation provided in the scipy version 1.5.2 Python package (<https://www.scipy.org>).

**Per-image FSC.** For simulated datasets for which the ground-truth distribution of structures is known, 'per-image' FSC curves can be computed between cryoDRGN-reconstructed density maps and the ground-truth density maps to quantitatively evaluate the reconstructed ensemble. To compute a per-image FSC, an FSC curve was computed between the density map generated by the cryoDRGN decoder at the value of the latent variable predicted for a given particle image and the ground-truth density map used to generate the image. One hundred images, randomly sampled according to the ground-truth distribution of structures, were used in the assessment of each of the simulated datasets. No real-space mask was used to compute the FSC.

**Heterogeneous reconstruction of the RAG complex (EMPIAR-10049).** A 10D latent variable model was trained on full-resolution particle images from the EMPIAR-10049 dataset ( $D=192, 1.23 \text{ \AA}$  per pixel) and their consensus reconstruction poses for 25 epochs. The encoder and decoder architectures were  $1,024 \times 3$ .

**Density map generation.** After training,  $k$ -means clustering with  $k=100$  was performed on the predicted latent encodings for the dataset, and volumes were generated at the 'on-data' cluster centers using the decoder network. Six structurally diverse representative structures were manually selected for visualization in Fig. 4a.

**Traditional heterogeneous refinement.** To validate the heterogeneous RSS and NBD conformations observed in cryoDRGN, the six selected density maps, low-pass filtered by  $20 \text{ \AA}$ , were used as initial models for a heterogeneous refinement job in cryoSPARC version 2.15.

**Heterogeneous reconstruction of the 80S ribosome (EMPIAR-10028). Pilot experiments.** A 10D latent variable model was trained on downsampled images ( $D=128, 3.78 \text{ \AA}$  per pixel) from the EMPIAR-10028 dataset and their consensus reconstruction poses for 50 epochs. The encoder and decoder architectures were  $256 \times 3$ .

**Particle filtering.** After training,  $k$ -means clustering with  $k=20$  was performed on the predicted latent encodings for the dataset. One cluster contained 860 particles that

were outliers when viewing the projected encodings along the first and second PCs. This observation was reproducible, and the particles belonging to the outlier cluster from either of two replicates (960 particles in total) were removed from the dataset.

**High-resolution training.** After particle filtering, a 10D latent variable model was trained on a random 90% of the remaining 104,280 images ( $D=256, 1.88 \text{ \AA}$  per pixel) for 25 epochs. The encoder and decoder architectures were  $1,024 \times 3$ .

**Density map generation.** After training,  $k$ -means clustering with  $k=50$  was performed on the predicted latent encodings for the dataset, and volumes were generated at the 'on-data' cluster centers using the decoder network. Representative structures of the rotated state and the unrotated state were manually selected for visualization in Fig. 4b. A representative structure of the missing head group state was manually selected for visualization in Extended Data Fig. 3. The numbered  $k$ -means cluster centers shown in Extended Data Fig. 3a, originally arbitrarily ordered, were reordered based on hierarchical clustering of the latent encodings with Euclidean distance metric and average linkage.

**Validation with traditional reconstruction.** To validate the 40S-rotated state, we selected 4,889 particles as the cluster from  $k$ -means clustering with  $k=20$  that was separated along PC1 (Extended Data Fig. 4). These particles were then input to a homogeneous refinement job in cryoSPARC version 2.15. The cryoDRGN density map, low-pass filtered by  $30 \text{ \AA}$ , was used as the initial model.

**Heterogeneous reconstruction of the assembling 50S ribosome (EMPIAR-10076). Pilot experiments.** A 1D and a 10D latent variable model were trained on downsampled images ( $D=128, 3.3 \text{ \AA}$  per pixel) from the EMPIAR-10076 dataset with poses from a consensus reconstruction for 50 epochs. The encoder and decoder architectures were  $256 \times 3$ .

**Particle filtering.** For the 1D experiment, particles with  $z \leq -1$  were removed from subsequent analysis. For the 10D experiment, a five-component, full-covariance GMM was fit to the latent encodings using scikit-learn<sup>50</sup>, and particles from the outlier cluster were removed. The outlier cluster was identified by visualizing the magnitude of the latent encodings (Extended Data Fig. 4). The intersection of both filtered particles stacks was used for subsequent analysis. Two-dimensional classification of the kept and removed particles was performed in cryoSPARC version 2.4 (ref. <sup>16</sup>) using all default options except for the number of 2D classes, which was set to 20. Ab initio reconstruction of the kept and removed particles was performed in cryoSPARC version 2.4 (ref. <sup>16</sup>) using all default options.

**High-resolution training.** A 10D latent variable model was trained on a random 90% of the remaining 97,031 images ( $D=256, 1.7 \text{ \AA}$  per pixel) for 50 epochs. The encoder and decoder architectures were  $1,024 \times 3$ . Two additional replicates were run, one with the exact settings from a different random initialization and a second with latent variable dimension of  $|z|=8$ .

**Density map generation.** After training, the dataset's latent encodings were viewed in 2D with UMAP<sup>35</sup>. Density maps corresponding to the major and minor assembly states were generated at the 'on-data' mean latent encoding for each class, that is,  $\hat{z}_M = \frac{1}{|M|} \sum_{i \in M} z_i$ , where  $M$  is the set of particles assigned to a given class in the published 3D classification.

**Map-to-map FSC.** The map-to-map FSC was computed between the cryoDRGN and published density map for each minor class. Density maps were aligned in Chimera, and a loose real-space mask (obtained as described above) was applied before computing an FSC curve.

**Reproducibility analysis.** For each replicate, a five-component, full-covariance GMM was fit to the UMAP embeddings using scikit-learn<sup>50</sup>. UMAP axes were negated to facilitate visual comparison. Label assignments were permuted to ensure consistent assignments between replicates. Clustering consistency was computed as the percentage of particles with identical GMM labels.

**New assembly state C4.** Particles corresponding to the new assembly state were manually selected from the UMAP embeddings with an interactive lasso tool in a custom visualization script available in the cryoDRGN software, the outline of which is shown in the Fig. 5f inset. The mean latent encoding of the resulting 1,113 selected particles was used to generate the structure representative for this new assembly state.

**Validation of C4 with traditional refinement.** The particles associated with class C4 ( $D=128, 3.3 \text{ \AA}$  per pixel) were then input to a homogeneous refinement job in cryoSPARC version 2.15. The cryoDRGN density map, low-pass filtered to  $30 \text{ \AA}$ , was used as the initial model.

**Heterogeneous reconstruction of the pre-catalytic spliceosome (EMPIAR-10180). Pilot experiments.** A 10D latent variable model was trained on

downsampled images ( $D=128$ , 4.25 Å per pixel) from the [EMPIAR-10180](#) dataset for 50 epochs. The encoder and decoder architectures were 256 × 3. Poses were obtained from the consensus reconstruction values given in the consensus\_data.star deposited in [EMPIAR-10180](#).

**Particle filtering.** The UMAP embeddings showed multiple clusters where the largest cluster corresponded to fully formed pre-catalytic spliceosomes. Particles corresponding to other clusters were removed from subsequent analyses by first performing  $k$ -means clustering with  $k=20$  on the latent encodings, and by removing  $k$ -means clusters for which the structure did not resemble the fully formed pre-catalytic spliceosome (11 of 20  $k$ -means clusters in one replicate, and ten of 20 clusters in a second replicate).

**High-resolution training.** A 10D latent variable model was trained on a random 90% of the remaining 155,247 images ( $D=256$ , 2.1 Å per pixel) for 50 epochs. The encoder and decoder architectures were  $1,024 \times 3$ .

**Density map generation.** After training, the dataset's latent encoding was viewed in 2D with UMAP and PCA. Density maps in Fig. 6d were generated at the latent encoding values along the PC1 axis at five equally spaced points between the 5th and 95th percentile of PC1 values. Density maps in Supplementary Fig. 1 were generated at the latent encoding values along the PC2 axis at five equally spaced points between the 5th and 95th percentile of PC2 values. Density map generation along PC axes was implemented with a custom script in the cryoDRGN software.

**Latent space graph traversal for generating trajectories.** Trajectories were generated by first creating a nearest-neighbors graph from the latent encodings of the images, in which a neighbor was defined if the Euclidean distance was below a threshold computed from the statistics of all pairwise distances. We chose a value for each dataset such that the average number of neighbors across all nodes was five. Edges were then pruned such that a given node did not have more than ten neighbors. Then, Djikstra's algorithm<sup>37</sup> was used to find the shortest path along the graph connecting a series of anchor points, and density maps were generated at the z value of the visited nodes. Anchor points were either defined manually or set to be the 'on-data' cluster centers after performing  $k$ -means clustering of the latent encodings.

For the graph traversal of the RAG complex in Supplementary Video 1, we used the latent encodings of the six density maps shown in Fig. 4a as the anchor points. For the graph traversal of the P<sub>f</sub>80S ribosome in Supplementary Video 2, we used ten randomly chosen latent encodings as the anchor points out of the  $k$ -means cluster centers with  $k=20$  that are shown before the graph traversal. For the graph traversal of the assembling ribosome in Supplementary Video 3, we used the latent encodings of the minor assembly states following the three assembly pathways given in Fig. 7 of Davis et al.<sup>12</sup>. For the graph traversal of the pre-catalytic spliceosome in Supplementary Video 4, we used the latent encodings of the  $k$ -means cluster centers with  $k=20$  as the anchor points. All density map figures and trajectories were prepared with ChimeraX<sup>51</sup> and are viewed at identical isosurface levels for a given model unless otherwise specified. CryoDRGN's graph-traversal algorithm is provided in the cryoDRGN software.

**3D Variability Analysis.** 3DVA<sup>23</sup> was performed in cryoSPARC version 2.15 on the 139,722 particles and their consensus poses comprising the filtered [EMPIAR-10180](#) dataset used in cryoDRGN analysis. Three variability modes were solved with all default options, and the low-pass filter resolution was set to 7 Å. 3DVA's per-particle latent encodings were extracted from the cryoSPARC metadata file. Spearman correlation was computed using the implementation provided in the scipy version 1.5.2 Python package (<https://www.scipy.org>). To visualize the 3DVA component 1 trajectory in Extended Data Fig. 10, the consensus density map was combined with the component 1 eigen-volume at five equally spaced points between the first and 99th percentile value of the 3DVA component 1 latent encoding distribution.

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

## Data availability

Trained cryoDRGN models and generated volumes were deposited in Zenodo at <https://doi.org/10.5281/zenodo.4355284><sup>52</sup>. Input files for training (excluding particle stacks) were deposited in Zenodo at <https://doi.org/10.5281/zenodo.4412072> and are also available at <https://www.github.com/zhonge/cryodrgn.empiar><sup>53</sup>. We used the following publicly available datasets: [EMPIAR-10049](#) (cryo-EM structures of a synaptic RAG1–RAG2 complex), [EMPIAR-10028](#) (cryo-EM structure of a *P. falciparum* 80S ribosome bound to the anti-protozoan drug emetine), [EMPIAR-10076](#) (modular assembly of the large bacterial ribosome) and [EMPIAR-10180](#) (structure of a pre-catalytic spliceosome).

The simulated heterogeneous datasets were deposited in Zenodo at <https://doi.org/10.5281/zenodo.4355284><sup>52</sup>.

## Code availability

CryoDRGN software and analysis scripts are implemented in custom software deposited in Zenodo at <https://doi.org/10.5281/zenodo.4355743> and are also available at <https://www.github.com/zhonge/cryodrgn><sup>54</sup>.

## References

41. Zhong, E. D., Bepler, T., Davis, J. H. & Berger, B. Reconstructing continuous distributions of 3D protein structure from cryo-EM images. In *International Conference of Learning Representations* (ICLR, 2020).
42. Bepler, T., Zhong, E., Kelley, K., Brignole, E. & Berger, B. Explicitly disentangling image content from translation and rotation with spatial-VAE. In *Advances in Neural Information Processing Systems* (NeurIPS, 2019).
43. Vaswani, A. et al. Attention is all you need. In *Advances in Neural Information Processing Systems* (NIPS, 2017).
44. Rezende, D. J., Mohamed, S. & Wierstra, D. Stochastic backpropagation and approximate inference in deep generative models. In *International Conference on Machine Learning* (ICML, 2014).
45. The PyMOL Molecular Graphics System, version 2.3 (Schrodinger, 2019).
46. Tang, G. et al. EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157**, 38–46 (2007).
47. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
48. Iudin, A., Korir, P. K., Salavert-Torres, J., Kleywegt, G. J. & Patwardhan, A. EMPIAR: a public archive for raw electron microscopy image data. *Nat. Methods* **13**, 387–388 (2016).
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. Pedregosa, F. et al. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
51. Goddard, T. D. et al. UCSF ChimeraX: meeting modern challenges in visualization and analysis. *Protein Sci.* **27**, 14–25 (2018).
52. Zhong, E. D. Data for "CryoDRGN: Reconstruction of heterogeneous cryo-EM structures using neural networks". Zenodo <https://doi.org/10.5281/zenodo.4355284> (2021).
53. Zhong, E. D. zhonge/cryodrgn\_empiar: initial release. Zenodo <https://doi.org/10.5281/zenodo.4412072> (2021).
54. Zhong, E. D. zhonge/cryodrgn: version 0.3.0. Zenodo <https://doi.org/10.5281/zenodo.4355743> (2020).

## Acknowledgements

We thank A. Lerer, R. Lederman, B. Demeo, A. Narayan, K. Kelley, B. Sauer, P. Sharp, S. Rodrigues and D. Haselbach for helpful discussions and feedback. We are grateful to the MIT-IBM Satori team for GPU computing resources and support. This work was funded by the National Science Foundation Graduate Research Fellowship Program to E.D.Z., NIH grant R01-GM081871 to B.B., NIH grant R00-AG050749 to J.H.D., NVIDIA-GPU grant to J.H.D. and a grant from the MIT J-Clinic for Machine Learning and Health to J.H.D. and B.B.

## Author contributions

All authors conceived of the work and developed the method. E.D.Z. implemented the software and performed the experiments. E.D.Z., B.B. and J.H.D. wrote the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

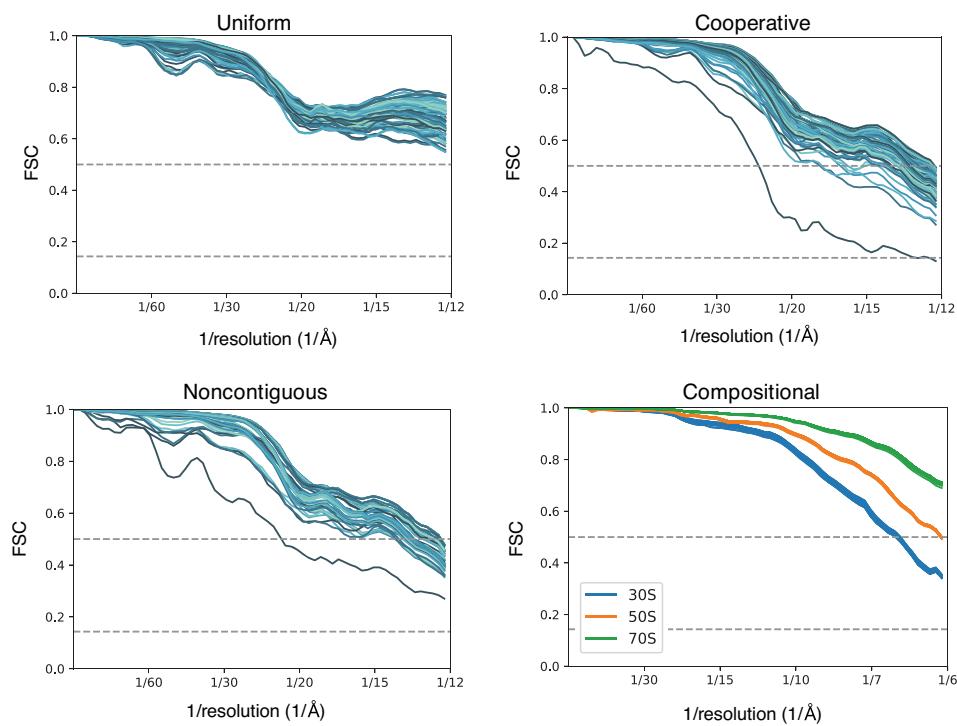
**Extended data** is available for this paper at <https://doi.org/10.1038/s41592-020-01049-4>.

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41592-020-01049-4>.

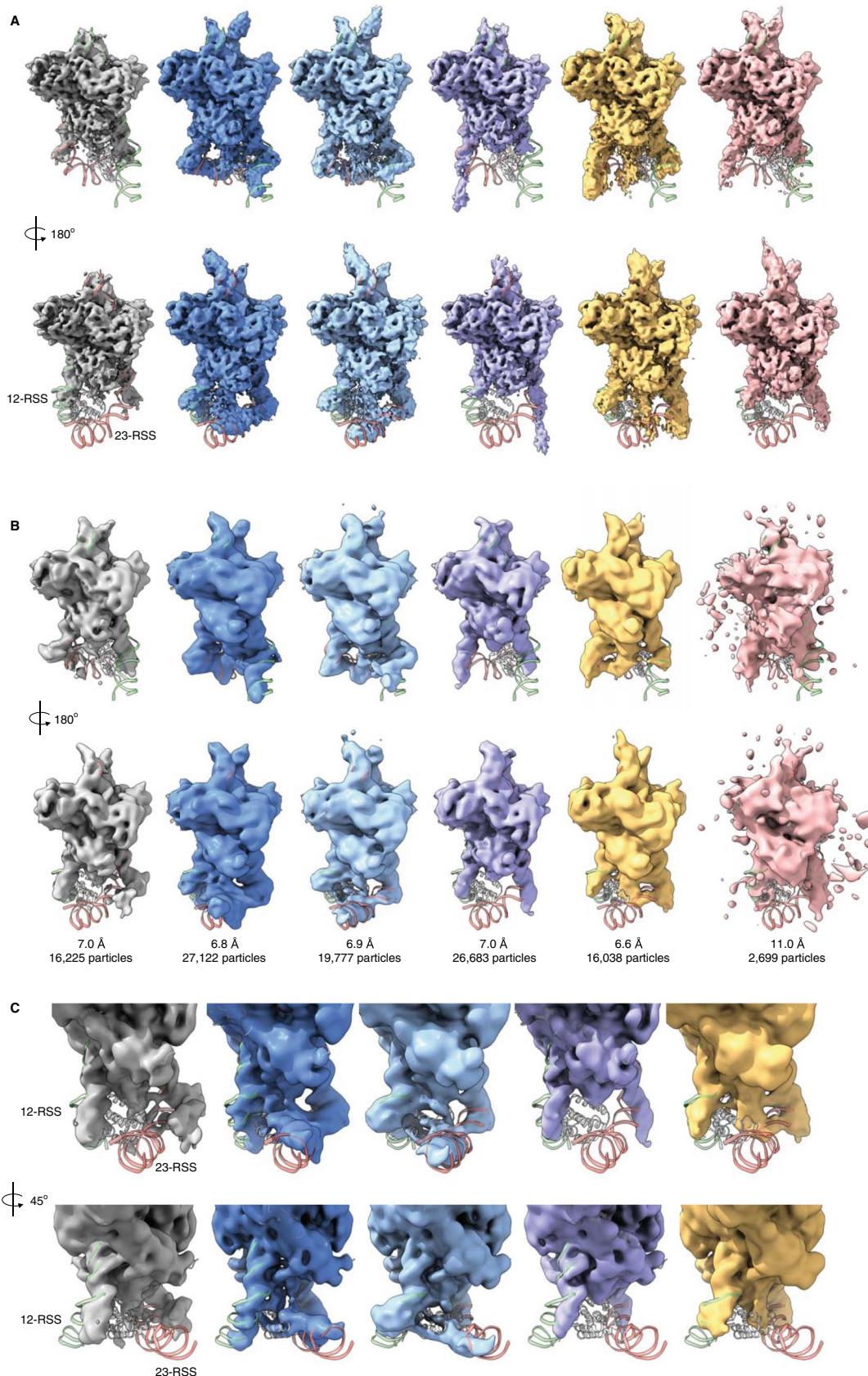
**Correspondence and requests for materials** should be addressed to B.B. or J.H.D.

**Peer review information** Arunima Singh was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

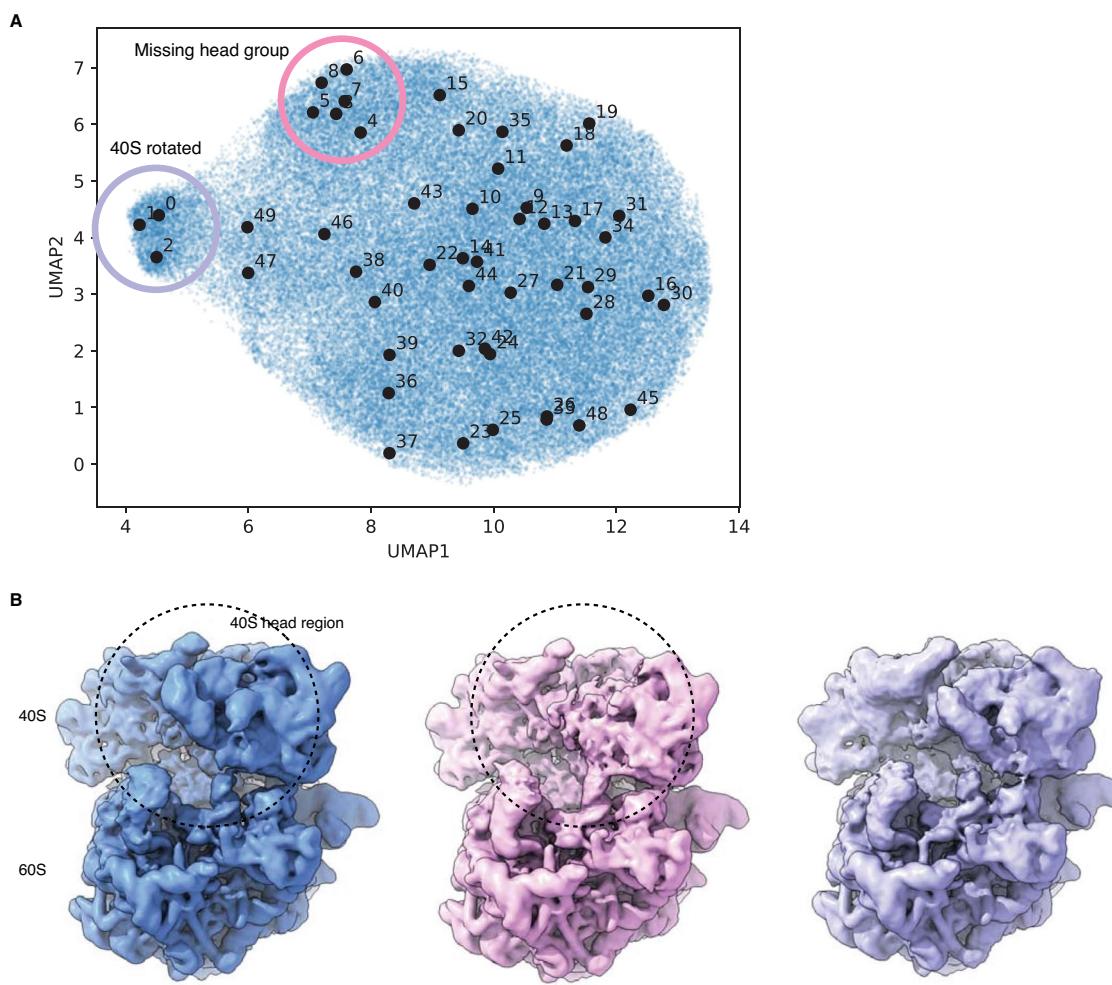
**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).



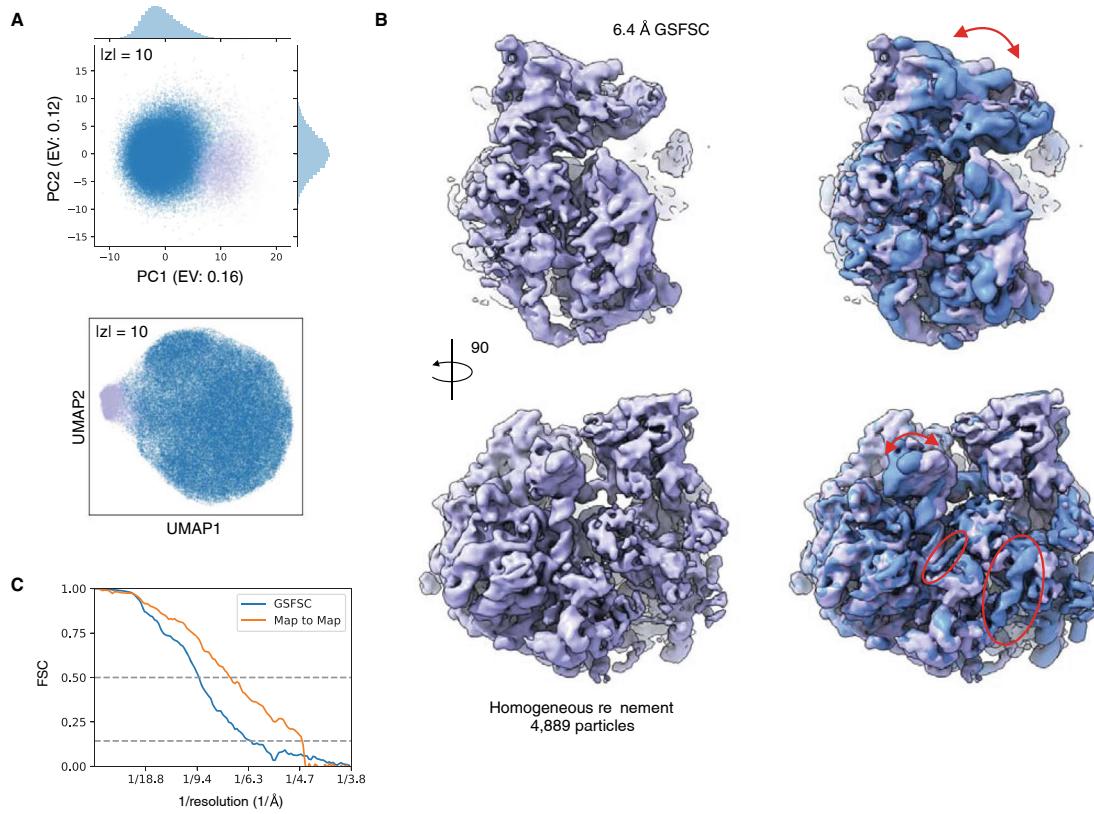
**Extended Data Fig. 1 | Per-image FSC curves between ground-truth maps and density maps from cryoDRGN trained on simulated heterogeneous datasets.** For each dataset, we compute 100 ‘per-image FSC curves’ between generated and ground-truth density maps (Methods). Images are sampled at equally spaced percentiles along the reaction coordinate for the *Uniform*, *Cooperative*, and *Noncontiguous* datasets. For the *Compositional* dataset, the per-image FSC for 20, 30, and 50 randomly sampled images of the 30S, 50S, and 70S ribosome, respectively, are shown. No mask is used in computing the FSC.



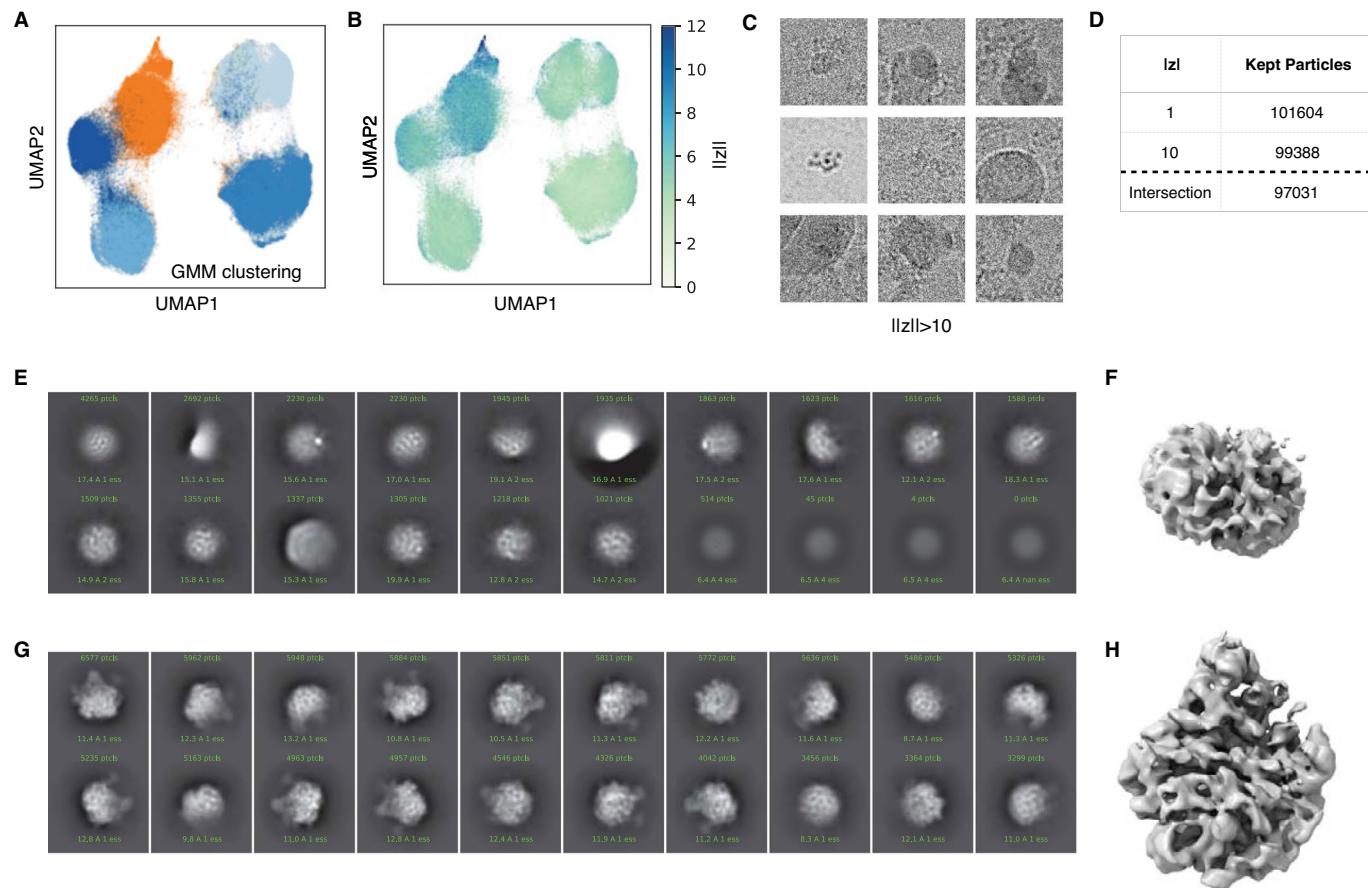
**Extended Data Fig. 2 | RAG complex density maps reconstructed by cryoDRGN and by heterogeneous refinement in cryoSPARC.** **a**, Front (top) and back (bottom) view of the six cryoDRGN density maps of the RAG complex from Fig. 4b. **b**, Density maps from 3D classification in cryoSPARC using the cryoDRGN density maps in (a) as initial models. Gold-standard FSC resolution and number of particles used in reconstruction are noted. **c**, Two side views of the density maps from 3D classification in (b), focusing on the RSS and NBD.



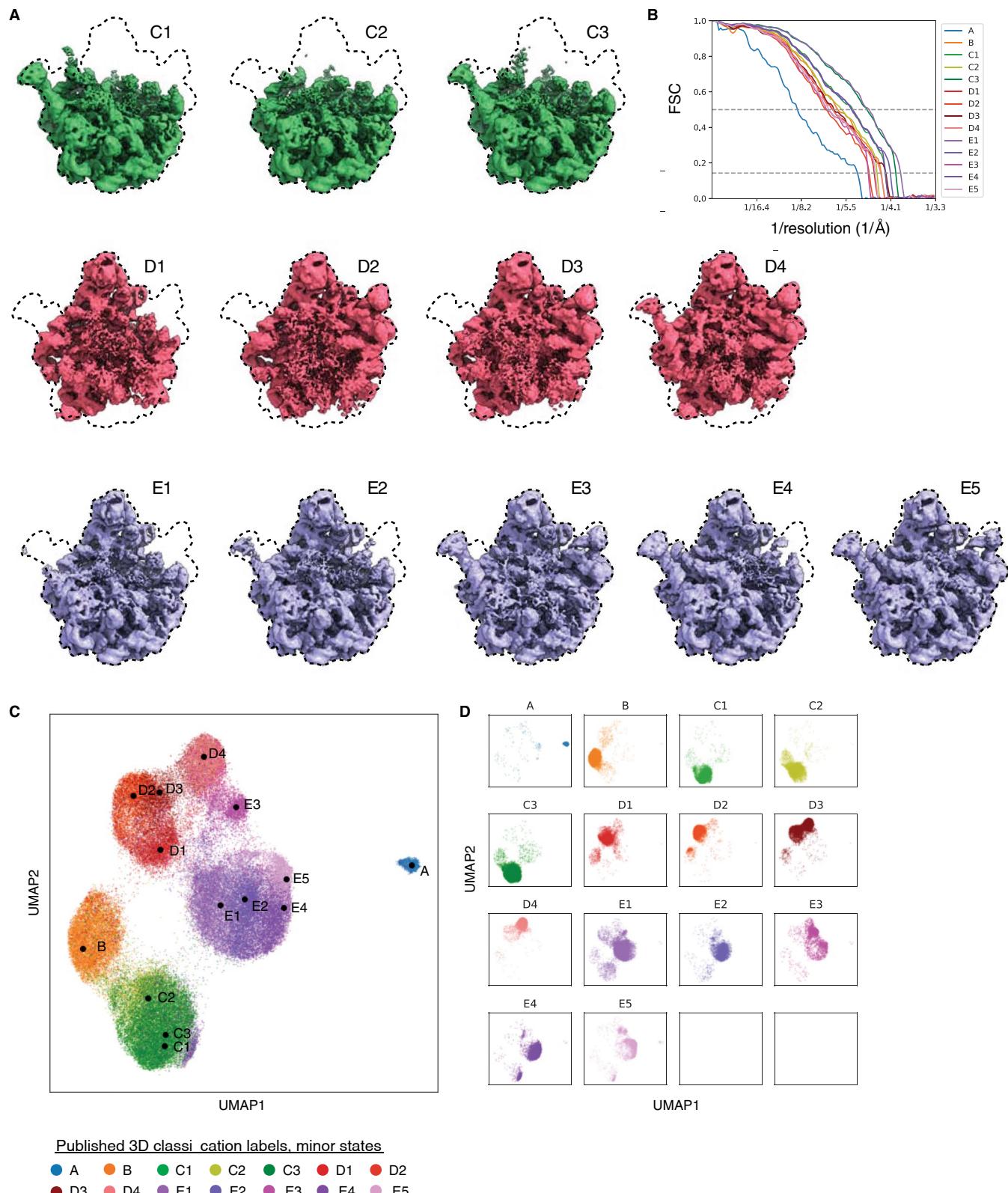
**Extended Data Fig. 3 | Missing head group of the Pf80S ribosome.** **a**, UMAP visualization of latent space encodings of EMPIAR-10028 particles with 50 sampled points shown in black. Sampled points are ordered according to distances in latent space (Methods). Visual inspection of the 50 volumes generated at the depicted points reveals 3 volumes with the 40S in a rotated state (purple) and 6 volumes with portions of the 40S head region missing (pink). **b**, Density map of the 80S ribosome with the missing head group reconstructed by cryoDRGN (pink) compared with the density maps from Fig. 4c showing the canonical (blue) and 40S-rotated (purple) forms of the 80S ribosome. The density maps are generated from points 32, 4, and 1 in panel A from left to right.



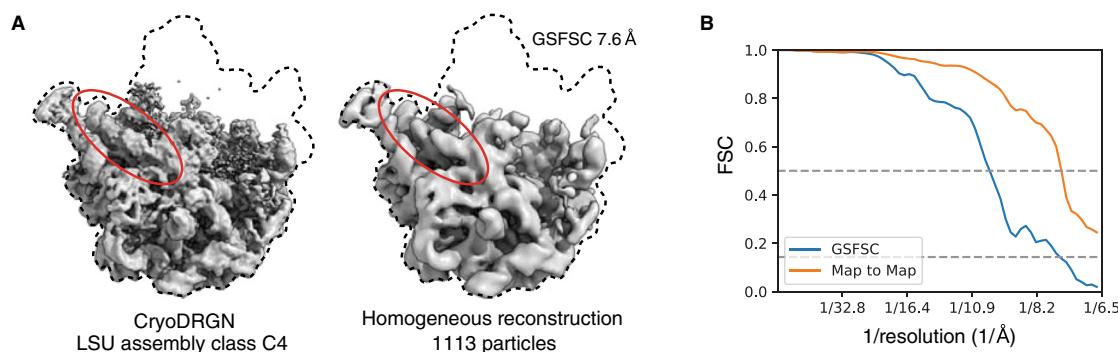
**Extended Data Fig. 4 | Validation of *Pf80S* rotated state with cryoSPARC.** **a**, PCA and UMAP visualization of the cryoDRGN latent space representation of *Pf80S* particle images with 4,889 particles separated along PC1, selected with k-means clustering, colored in purple (Methods). **b**, Density map from cryoSPARC homogeneous refinement (purple) using the 4,889 particles selected in **(a)**. The density map is also shown superimposed with the cryoDRGN unrotated state (blue) and annotated as in Fig. 4c. **c**, Gold standard FSC (GSFSC) curve between independent half-maps of the cryoSPARC refinement of the *Pf80S* rotated state and map-to-map FSC between the cryoDRGN and cryoSPARC density map of the *Pf80S* rotated state. Dotted lines indicate 0.5 and 0.143 cutoffs.



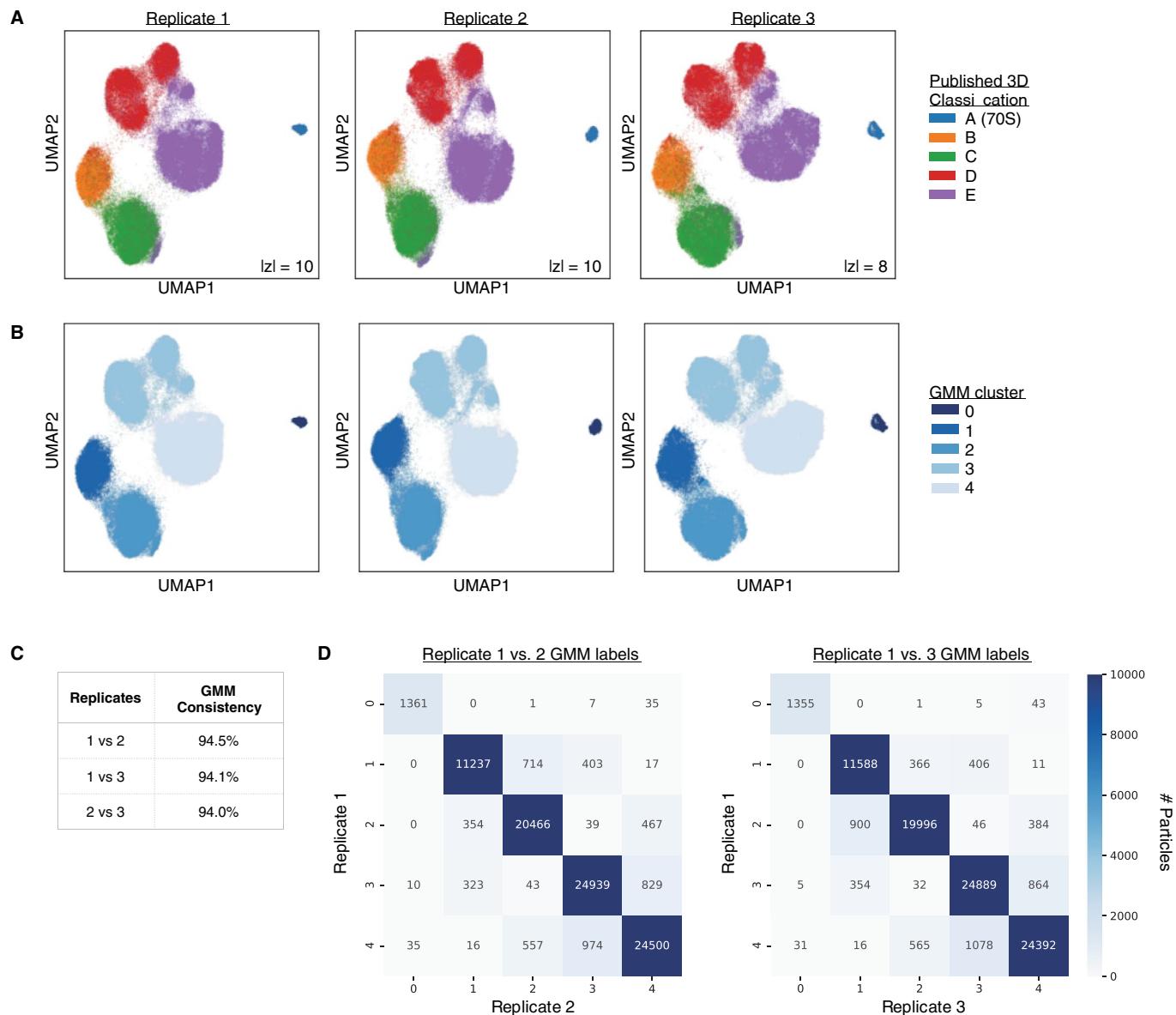
**Extended Data Fig. 5 | Filtering of particles from the assembling ribosome dataset.** **a**, UMAP visualization of the 10-D latent encodings from cryoDRGN as in Fig. 5b, colored by cluster after fitting a 5-component Gaussian mixture model. The cluster that was removed from subsequent analysis is colored orange. **b**, UMAP visualization of (a), colored by the magnitude of the latent encodings,  $\|z\|$ . **c**, Nine randomly sampled particle images from EMPIAR-10076 with latent encoding magnitude  $\|z\| > 10$  as predicted from cryoDRGN training in (a,b). Each image is 419.2 Å along each side. **d**, Table summarizing dataset filtering. **e,f**, 2D classification and *ab initio* reconstruction of the 34,868 removed particles. **g,h**, 2D classification and *ab initio* reconstruction of the 97,031 kept particles.



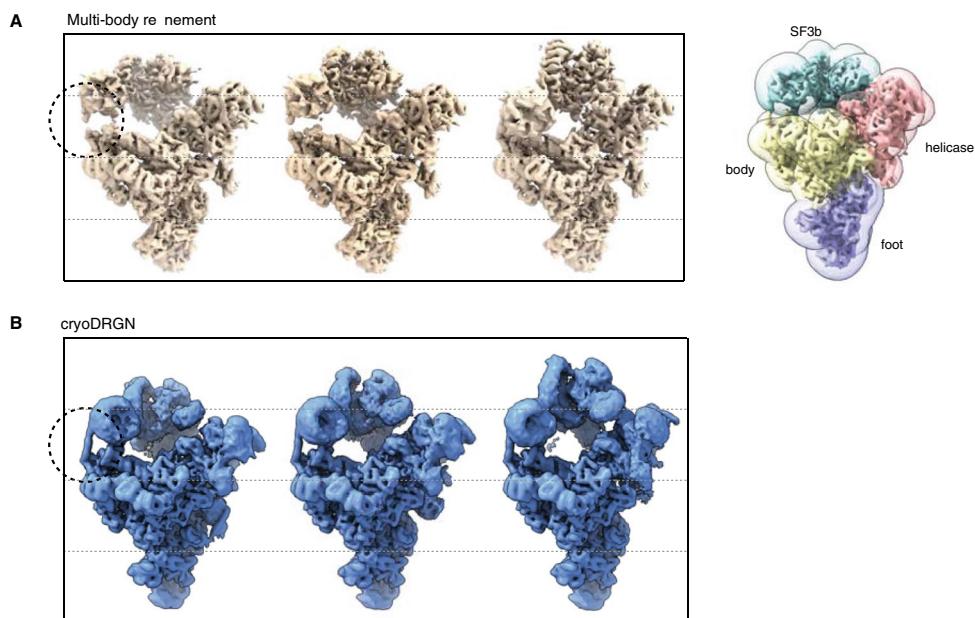
**Extended Data Fig. 6 | Minor LSU assembly states reconstructed by cryoDRGN.** **a**, Density maps of the LSU minor assembly states reconstructed by cryoDRGN. Each cryoDRGN structure is generated at mean of the latent encoding of particles with the corresponding class assignment from Davis *et al.*<sup>12</sup>. **b**, Map-to-map FSC curves between the generated cryoDRGN density maps and the published density map from Davis *et al.*<sup>12</sup>. Published resolutions for assembly states B-E ranged between ~4–5 Å. Dotted lines indicate 0.5 and 0.143 cutoffs. **c,d**, Reproduction of the cryoDRGN latent space shown in Fig. 5g, colored by minor assembly state (**c**), or viewed in separate panels (**d**).



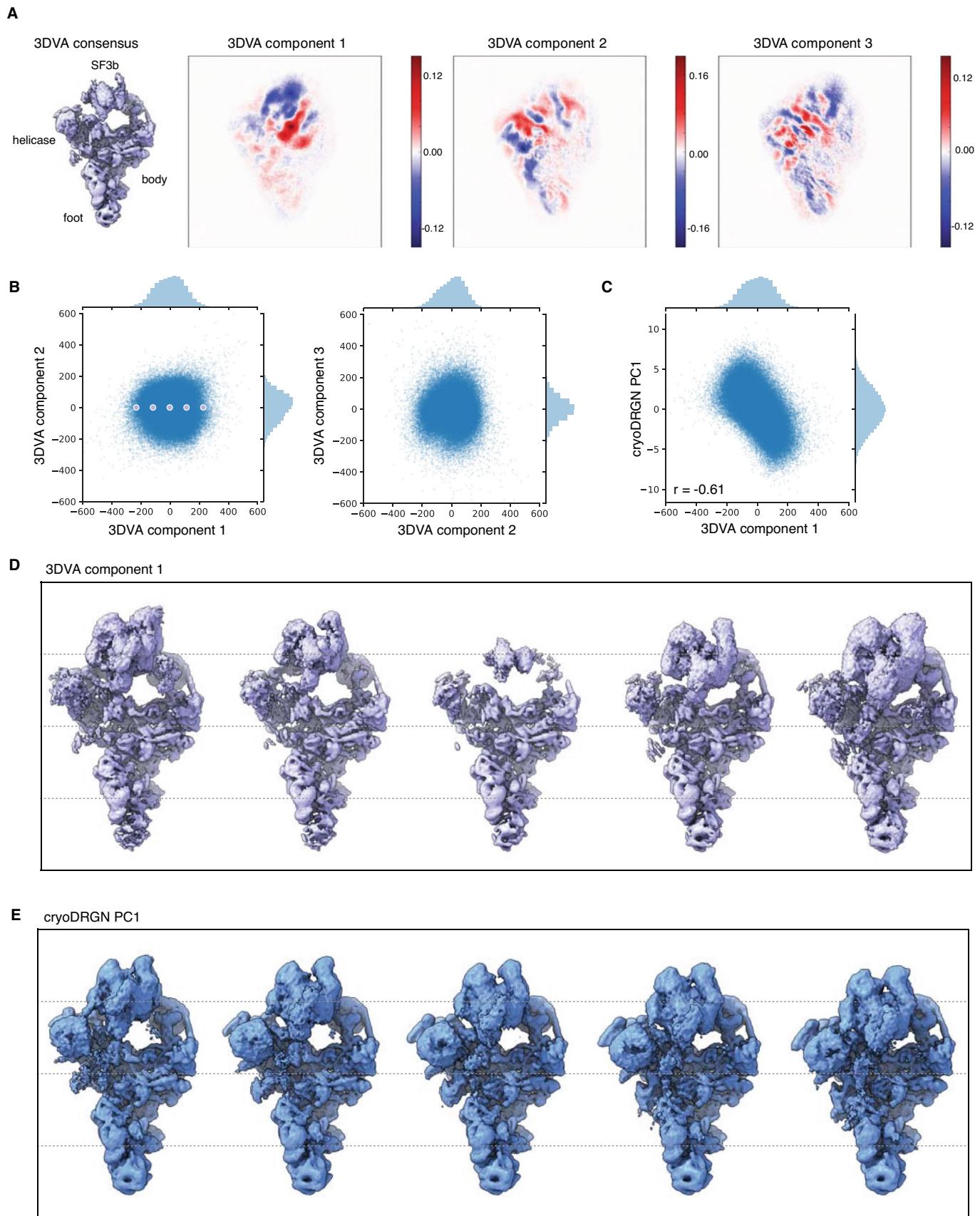
**Extended Data Fig. 7 | Validation of LSU class C4 with cryoSPARC.** **a**, Density map from cryoSPARC homogeneous refinement of the 1,113 particles selected from the cryoDRGN latent representation that constitute class C4 (right), compared with the density map generated by cryoDRGN (left) from Fig. 5i. rRNA helix 68 is circled in red. **b**, Gold standard FSC (GSFSC) curve between independent half-maps of the cryoSPARC reconstruction and map-to-map FSC between the cryoDRGN and cryoSPARC maps shown in (a). Dotted lines indicate 0.5 and 0.143 cutoffs.



**Extended Data Fig. 8 | Reproducibility of cryoDRGN's latent space representation of the assembling ribosome.** **a**, UMAP visualization of the latent encodings from replicate runs of cryoDRGN trained on the filtered particles of EMPIAR-10076. Particle embeddings are colored by major assembly state assigned from 3D classification in Davis *et al*<sup>2</sup>. **b**, UMAP visualization of (a), colored by cluster after fitting a 5-component Gaussian mixture model on the UMAP embeddings. **c,d**, Consistency of the GMM labeling between replicates reported as the percentage of particles with identical labels (c) and the confusion matrix of GMM cluster assignments (d).



**Extended Data Fig. 9 | Comparison of multi-body refinement and cryoDRGN of the pre-catalytic spliceosome.** **a**, Visualization of a rigid-body trajectory from multibody refinement of the pre-catalytic spliceosome. Snapshots are extracted from the trajectory along PC1 of rigid-body orientations, showing a large-scale motion of the SF3b subcomplex. The masks that define the rigid-body decomposition of the complex are shown on the right. The circle highlights a helix that breaks at the boundary between bodies where the rigid-body assumption no longer holds. Adapted from Video 3 of Nakane *et al.*<sup>19</sup> and density maps and masks deposited in EMPIAR-10180. **b**, Alternate view of cryoDRGN's PC1 traversal in Fig. 6. CryoDRGN learns the same overall motion of the SF3b subcomplex, however its neural network representation lacks the helix-breaking artifact.



Extended Data Fig. 10 | See next page for caption.

**Extended Data Fig. 10 | Comparison of cryoSPARC's 3D variability analysis and cryoDRGN.** **a**, Density map of the consensus reconstruction and 2D projections of the top three 3DVA variability components (that is, eigen-volumes) that form a linear basis describing structural heterogeneity of the pre-catalytic spliceosome. **b**, 3DVA latent encodings of particles from the filtered EMPIAR-10180 dataset. **c**, Comparison of 3DVA component 1 latent encodings and PC1 of the cryoDRGN 10D latent encodings from Fig. 6c. Correlation indicates Spearman correlation. **d**, 3DVA component 1 trajectory at the depicted points in **(b)**. **e**, Alternate view of the density maps from the cryoDRGN PC1 trajectory in Fig. 6d.

## Reporting Summary

Nature Research 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 Research 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 No software was used for data collection

Data analysis The following tools were used in data analysis: chimera v1.14.0, chimeraX v0.9, cryoSPARC v2.4, cryoSPARC v2.15; sklearn v0.19.1; umap v0.4.1; Python v3.6; cryoDRGN v0.0.17-0.3.0 [cryodrgn.csail.mit.edu]

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 Research [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 list of figures that have associated raw data
- A description of any restrictions on data availability

Trained cryoDRGN models for all experiments, simulated datasets, and indices of filtered particles of EMPIAR-10049, EMPIAR-10028, EMPIAR-10076, and EMPIAR-10180 will be deposited on Zenodo before publication. Simulated datasets will be deposited in EMPIAR before publication. Publicly available datasets used in this study include: EMPIAR-10049, EMPIAR-10028, EMPIAR-10076, and EMPIAR-10180; and PDB 3JBX, 3JBW, 4YBB.

# 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](https://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	For publicly available datasets, the sample sizes were determined by the published datasets. For the simulated datasets, we generated datasets of 50k images, which is of comparable size to published datasets.
Data exclusions	In some of the analyses, particles were excluded from further analysis based on visualization of outliers in the latent space or in volume space. Our approach is described in the methods section. We will deposit the filtered dataset to zenodo. Our approach was not pre-established. Data filtering was performed for EMPIAR-10028, EMPIAR-10076, and EMPIAR-10180.
Replication	All instances of replicate computational analysis are noted in the methods section. All attempts at replications were successful in producing qualitatively similar latent spaces/volume distributions. We quantify the reproducibility of replicate runs for EMPIAR-10076.
Randomization	Neural networks are initialized with random weights and the dataset is randomly ordered each iteration of training.
Blinding	Researchers were blinded to expected results by construction as the study performs unsupervised representation learning. However, researchers were not blinded to previous 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
<input checked="" type="checkbox"/>	Antibodies
<input checked="" type="checkbox"/>	Eukaryotic cell lines
<input checked="" type="checkbox"/>	Palaeontology and archaeology
<input checked="" type="checkbox"/>	Animals and other organisms
<input checked="" type="checkbox"/>	Human research participants
<input checked="" type="checkbox"/>	Clinical data
<input checked="" type="checkbox"/>	Dual use research of concern

### Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	ChIP-seq
<input checked="" type="checkbox"/>	Flow cytometry
<input checked="" type="checkbox"/>	MRI-based neuroimaging