End-to-end neural *ab initio* reconstruction of challenging cryo-EM and cryo-ET datasets

Axel Levy[1,2], Rishwanth Raghu[3], J. Ryan Feathers[3], Michal Grzadkowski[3], Frédéric Poitevin[2], Jake D. Johnston[4,5], Francesca Vallese[4,6], Oliver Biggs Clarke[4,6], Gordon Wetzstein[1], and Ellen D. Zhong[3]*

Proteins and other biomolecules form dynamic macromolecular machines that are tightly orchestrated to move, bind, and perform chemistry. Cryo-electron microscopy (cryo-EM) and cryo-electron tomography (cryo-ET) can access the intrinsic heterogeneity of these complexes and are therefore key tools for understanding mechanism and function. However, 3D reconstruction of the resulting imaging data presents a challenging computational problem, especially without any starting information, a setting termed *ab initio* reconstruction. Here, we introduce DRGN-AI, a method leveraging an expressive neural representation and combining an exhaustive search strategy with gradient-based optimization to process challenging heterogeneous datasets. Using DRGN-AI, we reveal new conformational states in large datasets, reconstruct previously unresolved motions from unfiltered datasets, and demonstrate *ab initio* reconstruction of biomolecular complexes from *in situ* data. With this expressive and scalable model for structure determination, we hope to unlock the full potential of cryo-EM and cryo-ET as a high-throughput tool for structural biology and discovery.

1 INTRODUCTION

Cryo-electron microscopy (cryo-EM) and cryo-electron tomography (cryo-ET) have transformed structural biology with their ability to visualize biomolecular structures at atomic resolution [47, 74] and uncover the conformations of dynamic molecular machines in solution [9, 10, 16, 46, 53, 75] or directly inside cells [17, 49, 56, 72]. However, the path between cryo-EM data and scientific discoveries is often hindered by complex processing pipelines involving many application-specific algorithms. Automating image processing and reconstruction methods and making them accessible to nonspecialists can therefore bridge the gap between raw data and scientific knowledge.

Understanding structural variability is essential for elucidating the mechanisms of biomolecular complexes. In cryo-EM and cryo-ET, these complexes, or "particles", are vitrified from ambient temperatures and thus adopt a range of conformations within the space of accessible low-energy states (i.e., the conformational space). Heterogeneous reconstruction methods aim to reveal this space from the observed set of noisy and randomly oriented projection images. Recent neural and non-neural algorithms have shown success in the unsupervised reconstruction of structural heterogeneity from cryo-EM images [9, 53, 54, 75, 76] (see **Section 3** for further details). Existing tools, however, require preliminary structural information of the imaged complex, either in the form of an initial density map or pre-assigned orientations (i.e., poses) of the particle images. These initial inputs are usually derived from a prior homogeneous reconstruction step, which inherently limits the extent of structural variability that can be effectively modeled. In the presence of heterogeneity, multiple rounds of 3D classification are commonly used to sample an initial structure that can serve as a reasonable starting point for further refinement. Although these approaches can theoretically operate *ab initio* (i.e., from a random or uninformative starting point), they assume heterogeneity arises from a small set of independent

[1]Department of Electrical Engineering, Stanford University, Stanford, CA, USA [2]SLAC National Laboratory, Menlo Park, CA, USA [3]Department of Computer Science, Princeton University, Princeton, NJ, USA [4]Department of Physiology and Cellular Biophysics, Columbia University, New York, NY, USA [5]Department of Anesthesiology, Columbia University Irving Medical Center, New York, NY, USA [6]Irving Institute for Clinical and Translational Research, Columbia University, New York, NY, USA *Correspondence to: zhonge@princeton.edu.

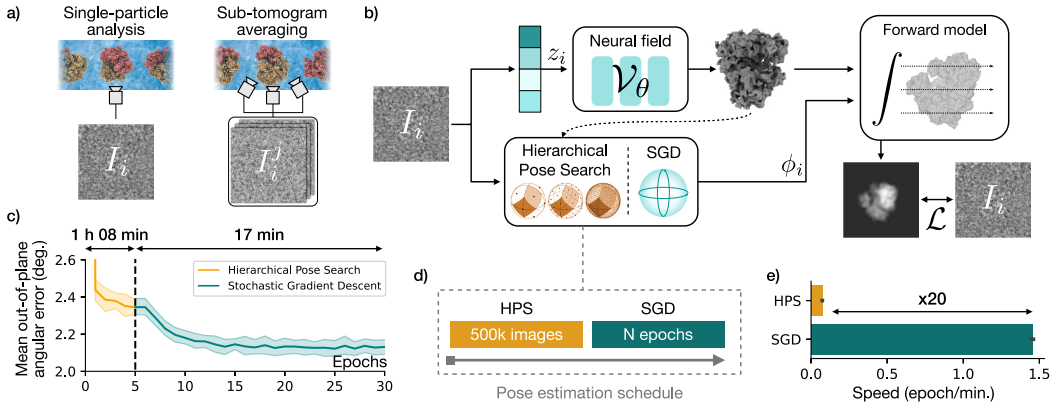


Fig. 1. **The DRGN-AI method for *ab initio* heterogeneous reconstruction. a)** DRGN-AI can process single particle images or subtilts for subtomogram reconstruction. **b)** Architecture overview. Objects in blue are optimized to minimize the loss $\mathcal{L}$. For each image, a projection is generated by a differentiable forward model. Poses ϕ_i are first estimated using hierarchical pose search (HPS) and then refined by stochastic gradient descent (SGD). Conformations z_i are optimized by SGD. **c)** Pose error during training on a synthetic 80S ribosome dataset (100,000 particles, 128 × 128, 3.77 Å/pix). Pose error is quantified by the mean out-of-plane angular error (mean and standard deviation over 5 runs). **d)** Pose estimation switches to SGD once at least 500,000 images have been processed by HPS. **e)** Gradient descent is 20x faster and leads to more accurate poses since it is not limited by the resolution of the search grid.

structures. This simplified model of heterogeneity often does not align with the true distribution of structures and, despite time-consuming iterative processing pipelines, still results in missed conformational states.

In this work, we show that cryo-EM and cryo-ET imaging data can be processed in a single run to reveal structural variability in biomolecular complexes. Our method, DRGN-AI (Deep Reconstructing Generative Networks - *Ab Initio*), leverages the expressivity of implicit neural representations to model complex distributions of structures that can be learned in an *ab initio* setting from either single particle or subtomogram tilt series images. To combine speed and robustness to noise, DRGN-AI relies on a new pose estimation strategy that switches from am exhaustive search avoiding local minima to stochastic gradient descent. We show that DRGN-AI can reconstruct the conformational and compositional landscape of the assembling bacterial ribosome, the pre-catalytic spliceosome, and the SARS CoV-2 spike protein in a single round of training without any manual intervention or prior pose estimation procedure. We show that DRGN-AI is robust to the presence of junk or outlier images and better resolves the structure of the DSL1/SNARE complex compared to existing processing workflows. By handling large and challenging heterogeneous datasets (i.e., containing highly dynamic complexes, junk particles or rare states), DRGN-AI reveals states that were missed in earlier processing approaches in a dataset of the V-ATPase complex [60] and of the human erythrocyte ankyrin-1 complex [66]. Unifying single particle analysis and subtomogram averaging under a common framework, DRGN-AI can process 2D tilt series collected by cryo-ET and reconstructs three states of the *M. pneumonia* 70S ribosome during its elongation cycle [64] in a single *ab initio* reconstruction step. DRGN-AI is released as an open-source software package at https://drgnai.cs.princeton.edu/.

2 RESULTS

2.1 Efficient pose estimation with a two-stage strategy

In single particle cryo-EM, each particle can be associated with a set of image-specific unknown parameters, called *latent variables*. In the case of *ab initio* heterogeneous reconstruction, this set includes the *pose* ϕ_i – the orientation of the particle with respect to the microscope and 2D translation to compensate for the imperfect centering of the particle in the image – and the *conformation* z_i – an abstract low-dimensional vector characterizing the shape of the particle among all the low-energy states of the molecule. Due to the binding-unbinding of monomers and due to their inherent flexibility, this space – referred to as the *conformational space* or *latent space* – can be seen as a finite union of smooth low-dimensional manifolds. In DRGN-AI, this space is represented by $\mathcal{V}_\theta$, an element of a parametric family $\Omega = \{\mathcal{V}_\theta, \theta \in \mathbb{R}^D\}$. Given a conformation $z_i \in \mathbb{R}^d$, $\mathcal{V}_\theta[z_i] : \mathbb{R}^3 \to \mathbb{R}$ represents a single electron scattering potential (also called density map). We parameterize the family Ω with a neural network (**Section 5.1.5**).

DRGN-AI relies on a new strategy for inferring these latent variables (**Figure 1b**). Unlike cryoDRGN [75] and cryoFIRE [40] which use an encoder neural network to map images to conformations, it stores the predicted conformation in a look-up table of real-valued vectors. The whole pipeline can therefore be seen as an autodecoder [4]. This design choice is motivated by the observation that, in the case of noisy datasets, a neural network tends to *memorize* the images and does not generalize to unseen images [14]. In **Figure S1**, we show that an encoder trained to predict poses tends to overfit on high-noise datasets (SNR ≤ -10 dB), preventing true amortization. We therefore reason that a look-up table serves as a simpler, faster and more memory-efficient object.

Pose estimation relies on a two-step strategy. First, poses are estimated by a hierarchical pose search step (HPS). During this step, poses are sampled on increasingly fine grids of $\mathrm{SO}(3) \times \mathbb{R}^2$, synthetic images are generated using the density map $\mathcal{V}_\theta[z_i]$ and compared to the observed particle image I_i. Although this approach is robust and often converges for a randomly initialized $\mathcal{V}_\theta$, querying the neural network is slow and the pose accuracy remains eventually limited by the resolution of the finest search grid. To overcome these limitations, after a set number of HPS steps, DRGN-AI initializes a look-up table to store the last pose estimated by HPS on each image and the entries of the table are further refined with stochastic gradient-descent (SGD). As shown in **Figure 1e**, optimization is 20x faster with SGD than with HPS. The pose accuracy being no longer bounded by the resolution of a grid, **Figure 1c** shows that SGD leads to a decrease in the mean out-of-plane angular error, whose effect on resolution we quantify in **Figure S2**. Empirically, we found that SGD reliably converges for a variety of real datasets if the following two conditions are satisfied: (1) each image in the dataset has been processed at least twice with HPS and (2) the algorithm has processed at least 500,000 images with HPS. We evaluate this pose estimation strategy for homogeneous reconstruction in **Figure S3** on a synthetic dataset of 100,000 images of the 80S ribosome.

2.2 Single shot *ab initio* heterogeneous reconstruction on benchmark datasets

To validate DRGN-AI's ability for *ab initio* heterogeneous reconstruction, we run DRGN-AI on three experimental benchmark datasets: the pre-catalytic spliceosome (EMPIAR-10180 [52]), the assembling 50S large ribosomal subunit (EMPIAR-10076 [13]) and the SARS-CoV-2 spike protein [68]. While previous analyses of these datasets typically operate on filtered versions (with less junk images) and use consensus

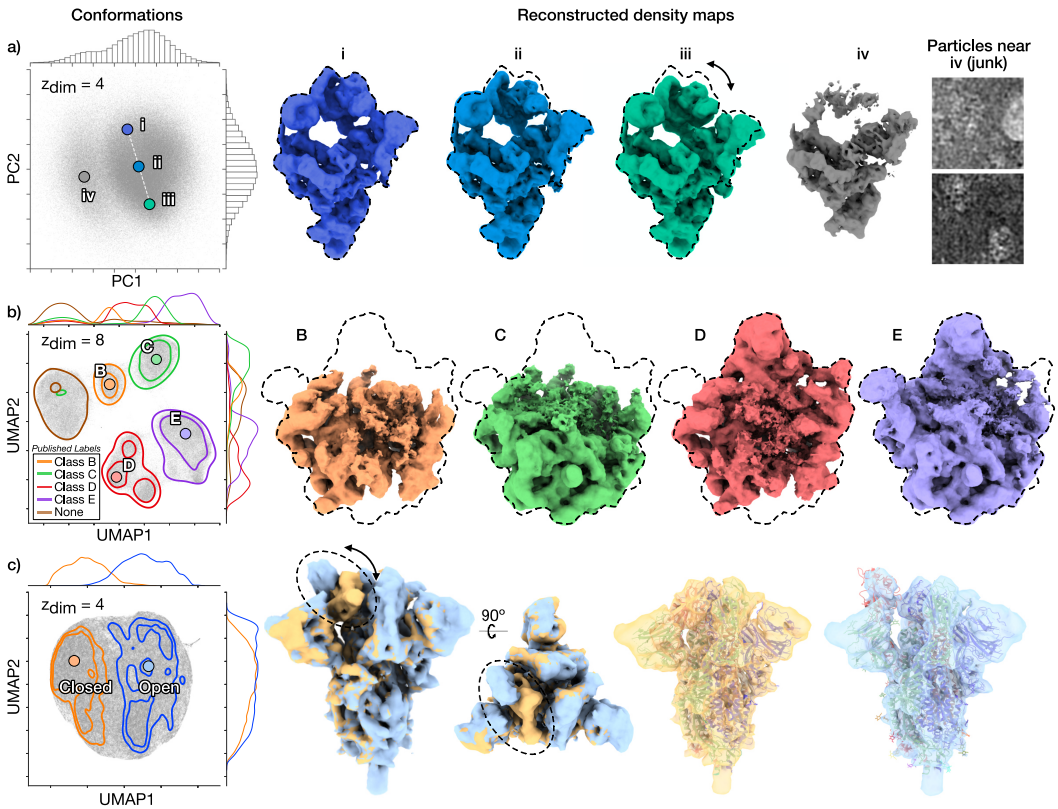


Fig. 2. **Single shot DRGN-AI *ab initio* heterogeneous reconstruction results on unfiltered benchmark datasets.** **a)** PCA and reconstructed maps for the pre-catalytic spliceosome dataset (EMPIAR-10180 [52], 2.6 MDa, 128x128, 4.25 Å/pix., 327,490 particles). Points i, iii and iv are obtained by k-means clustering ($k = 5$) and ii is a linear interpolation between i and iii. Density maps i-iii show a continuous bending motion of the molecule. Cluster iv contains outlier (junk) particles and is associated with a broken structure. Dashed lines indicate outlines of the extended spliceosome. **b)** UMAP and reconstructed maps for the assembling bacterial large ribosomal subunit dataset (EMPIAR-10076 [13], approx. 2.1 to 3.3 MDa, 256x256, 1.64 Å/pix., 131,899 particles). The distribution of the published labels is shown in the UMAP plot with a Gaussian kernel density estimation (KDE). Dashed lines indicate outlines of the fully mature 50S ribosome. DRGN-AI recovers the compositional variability of all major assembly states from the full unfiltered dataset (4 centroids obtained with k-means clustering, $k = 6$). **c)** UMAP and reconstructed maps for the SARS-CoV-2 spike protein dataset [68] (438 kDa, 128x128, 3.28 Å/pix., 369,429 particles). DRGN-AI reveals the closed and open states of the receptor binding domain (k-means clustering, $k = 2$). Atomic models are docked (PDB:6VXX, PDB:6VYB).

poses [9, 19, 28, 34, 42, 46, 53, 75, 78], we perform single shot reconstruction, i.e., a single round of training, on the deposited particle stack, fully *ab initio*.

In **Figure 2**, a visualization of the conformational latent space inferred by DRGN-AI (reduced to 2D via Principal Component Analysis (PCA) or Uniform Manifold Approximation and Projection (UMAP) [44]) is shown on the left, and representative density maps are shown on the right. On the EMPIAR-10180 dataset [52], DRGN-AI reconstructs the continuous bending motion of the pre-catalytic spliceosome while clustering a set of particle images corresponding to broken or denatured density maps in the latent space (**Figure 2a**, **S4**). DRGN-AI recovers the compositional heterogeneity in the assembly states of the bacterial large ribosomal subunit in the EMPIAR-10076 [13] dataset, while also grouping previously unclassified and discarded particles [13] in

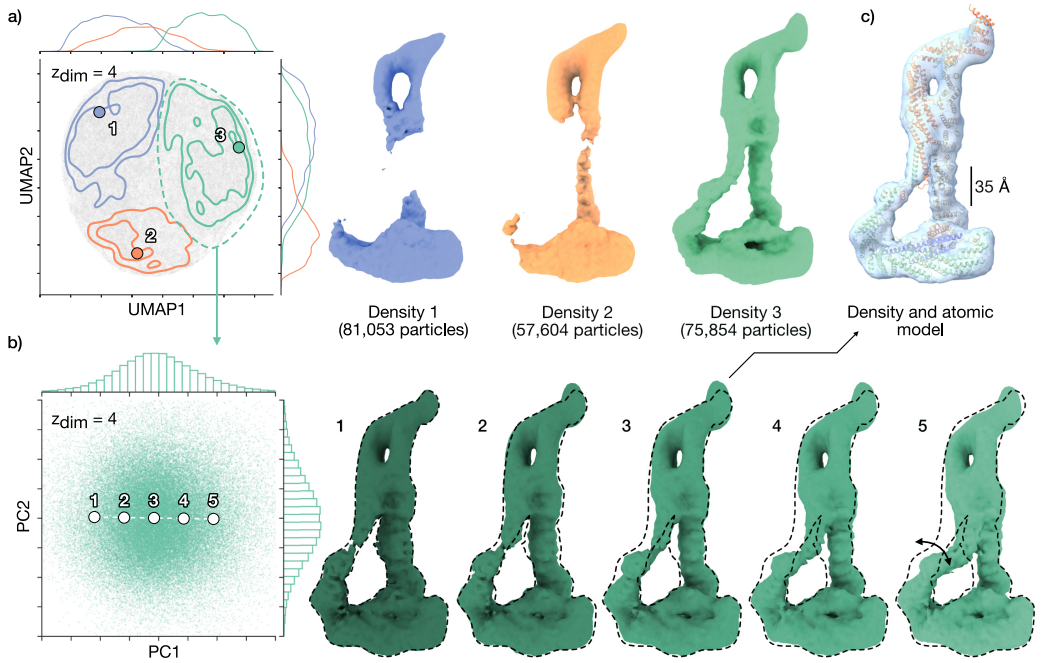


Fig. 3. **DRGN-AI *ab initio* reconstruction of the DSL1/SNARE complex [11], despite the presence of junk particles. a)** UMAP and reconstructed maps for the full dataset (214,511 particles, 128x128, 3.47 Å/pix.) with DRGN-AI. The conformations are clustered with k-means ($k = 3$). Particles associated with the green cluster are selected. **b)** PCA and reconstructed maps for the second DRGN-AI run on selected particles (75,854). Moving along the first principal component reveals the flexibility of the complex. Additional density maps are shown in **Supplementary Video 1**. **c)** Density map from **b** with docked atomic model (PDB: 8EKI).

the latent space (**Figure 2b**). Finally, on a much smaller protein complex of the SARS-CoV-2 spike protein (<500 kDa) [68], we show that DRGN-AI can reconstruct both the closed and open states of the spike protein fully *ab initio* and resolve the conformational change of the receptor binding domain (**Figure 2c**). In **Figure S11**, we provide additional results on a synthetic dataset with strong conformational heterogeneity for which DRGN-AI provides a better reconstruction than cryoSPARC's multi-class *ab initio* workflow.

2.3 DRGN-AI reconstructs the DSL1/SNARE complex despite the presence of junk particles

We next sought to characterize DRGN-AI on a dataset where a significant fraction of particles are outliers that do not contribute to the final structure. We ran our method on a dataset of the DSL1/SNARE complex (EMPIAR-11846) [11], a 255-kDa complex involved in the membrane fusion reaction of eukaryotic cells. In DAmico et al., a total of 214,511 particles were obtained after particle picking and an initial filtering via 2D classification. Reconstruction on several subsets of the dataset were explored until a low resolution density map was identified. The dataset was subsequently processed in cryoSPARC using three iterations of 3D classification to yield a final 6.2 Å consensus reconstruction from 49,947 particles.

We run DRGN-AI *ab initio* heterogeneous reconstruction on this dataset and observe a distribution of densities containing the fully intact complex as well as broken particles. We classify the images among three groups by k-means clustering in the latent space (**Figure 3a**). By visual inspection, we select the cluster

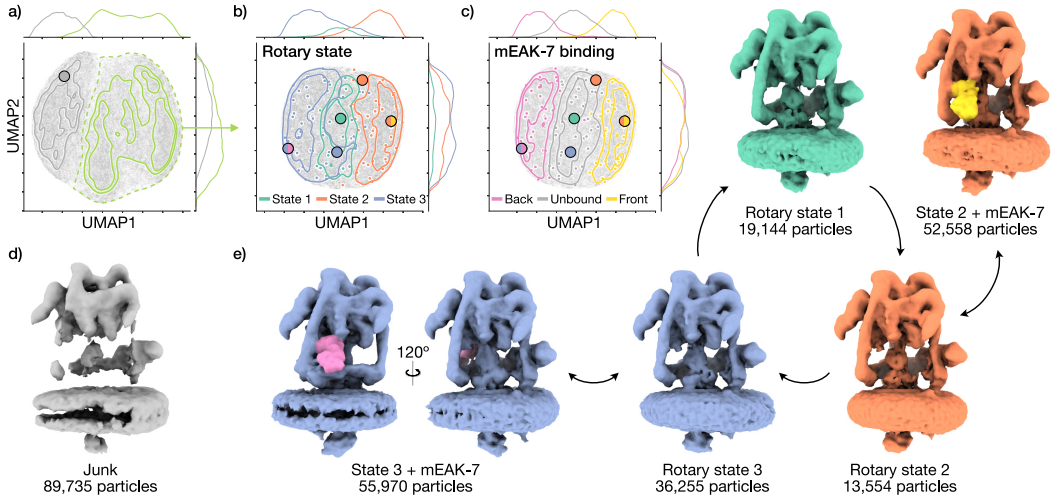


Fig. 4. **DRGN-AI *ab initio* reconstruction of the V-ATPase complex**. **a)** UMAP visualization of DRGN-AI latent embeddings of the full dataset (EMPIAR-10874 [60], 267,216 particles, 128x128, 3.97 Å/pix.). Selected particles are in green. **b, c)** UMAP visualization of DRGN-AI latent embeddings on the filtered dataset (177,481 particles) with three rotary states (**b**) or mEAK-7 binding location (**c**). Gaussian kernel density estimates are overlaid along with visually classified k-means centroids ($k = 100$) as small circles, and points corresponding to the maps in panel (**e**) as large circles. **d)** Sampled density map showing a broken complex from the unfiltered dataset. **e)** Sampled density maps of the three rotary states and mEAK-7 binding.

associated with the most resolved structure and run a second *ab initio* heterogeneous reconstruction on this subset of particles. Density maps sampled DRGN-AI from the filtered *ab initio* reconstruction reveal a continuous hinging motion of the DSL1/SNARE complex (**Figure 3b**). Unlike previous processing workflows, DRGN-AI is able to filter spurious particles and reveal the conformational heterogeneity of the protein complex. We provide a comparison to cryoSPARC-based workflows (multi-class *ab initio* and 2D classification) in **Figure S5**.

2.4 DRGN-AI captures conformational and compositional heterogeneity in a V-ATPase complex

We next evaluate DRGN-AI on a dataset consisting of significant compositional and conformational heterogeneity, in addition to containing a large fraction of junk particles. We process this dataset of the porcine kidney V-ATPase complex (EMPIAR-10874 [60]), a transmembrane proton pump involved in multiple signaling pathways. Through initial particle picking and 2D classification on the deposited micrographs, 267,216 particles were obtained.

Running DRGN-AI *ab initio*, we observe a distinct cluster in the latent space corresponding to broken particles (**Figure 4a, d**). We filter out these particles using DRGN-AI's interactive lasso tool and train a new *ab initio* heterogeneous model on the remaining 177,481 particles. Sampled maps show three conformational states of the complex, characterized by different orientations of the central rotor domains (**Figure 4b, e**). DRGN-AI also resolves compositional heterogeneity in rotary state 2, showing that the signaling protein mEAK-7 can bind to the front of the complex. Additionally, we reveal a previously unseen state corresponding to mEAK-7 bound to the back of rotary state 3. We observe that the latent space clusters by the binding location (or non-binding) of mEAK-7 (**Figure 4c**). The unbound cluster can be further decomposed into three clusters,

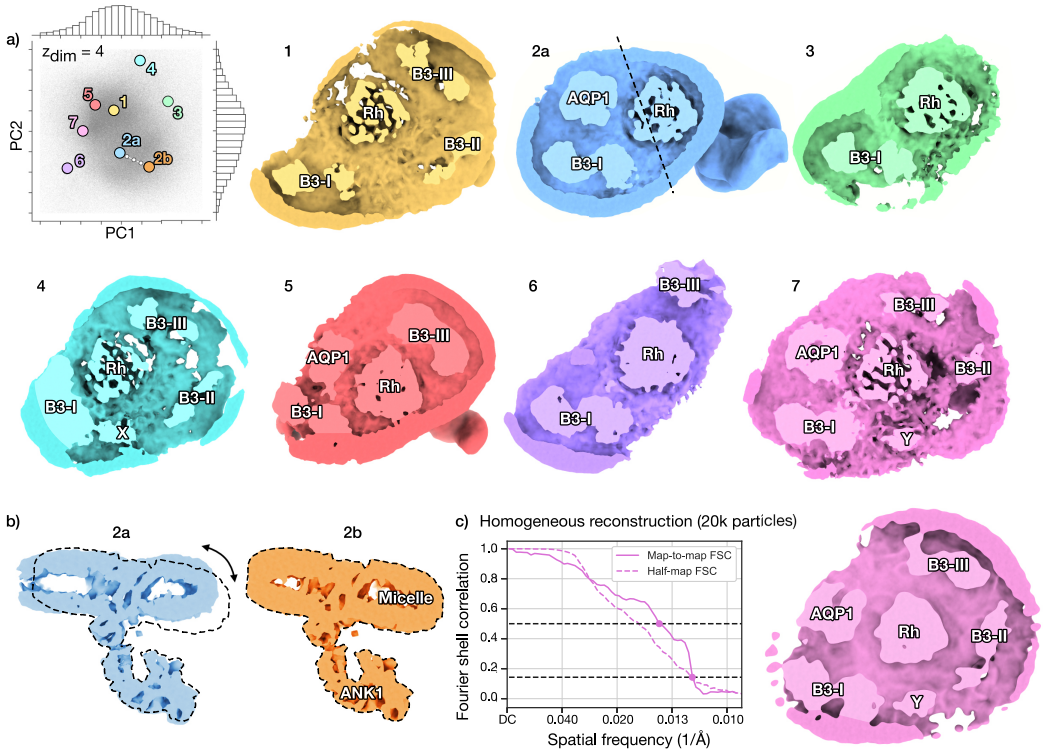


Fig. 5. **Single shot DRGN-AI *ab initio* reconstruction of the human erythrocyte ankyrin-1 complex and identification of a new "supercomplex" state. a)** PCA and reconstructed maps for the full dataset (EMPIAR-11043 [66], 710,437 particles, 128x128, 2.92 Å/pix.). Conformations can be distinguished according to micelle composition. The six classes from Vallese et al. [66] (1 to 6) are recovered. DRGN-AI also reveals the presence of a new "supercomplex" state (7), which simultaneously contains the rhesus heterotrimer (Rh), the aquaporin (AQP1), the band 3-I (B3-I), the band 3-II (B3-II), the band 3-III (B3-III) dimers and an unknown protein Y in the micelle. Y could correspond to the unknown protein reported in [66]. Additional density maps shown in **Supplementary Video 2**. **b)** Linear trajectory in the populated region between 2a and 2b reveals a continuous rotation of the ankyrin (ANK1) relative to the micelle. Plane of view shown as dotted line in 2a in panel **a**. **c)** Validation of the supercomplex structure. A homogeneous reconstruction in cryoSPARC [55] on 20k particles (20k closest maps from F) with poses from DRGN-AI. Map-to-map and half-map FSC curves.

associated with the three rotary states (**Figure S6**). We validate each state with a Fourier backprojection of the selected particles and compute a half-map Fourier shell correlation (**Figure S6**).

2.5 DRGN-AI reveals a new state of the human erythrocyte ankyrin-1 complex

An important aspect of DRGN-AI is its ability to process large cryo-EM datasets in an *ab initio* setting where rare states appear in sufficient quantity for detection and 3D reconstruction. We process a dataset of the human erythrocyte ankyrin-1 complex (EMPIAR-11043 [66]), a large membrane-embedded complex important for the shape and stability of red blood cells. The deposited dataset contains more than 700,000 picked particles and was shown to contain six different conformationally and compositionally varying states [66].

Figure 5 shows the results of *ab initio* heterogeneous reconstruction with DRGN-AI on the full dataset. Different forms of the complex can be identified by sampling conformations in the latent space. We run k-means clustering ($k = 6$) on the latent embeddings. By evaluating $\mathcal{V}_\theta$ on the k-means centroids, we reconstruct

the six previously discovered classes (**Figure 5a**). Additional conformational heterogeneity can be visualized in the same trained DRGN-AI model by sampling a series of conformations linearly interpolating the populated area between points 2a and 2b, corresponding to a rotation of ankyrin with respect to the micelle (**Figure 5b**). Increasing the number of sampled density maps to 100 (via k-means clustering) revealed a structure of the "supercomplex" state (**Figure 5a**), which was hypothesized but not shown in Vallese et al. [66]. This rare state simultaneously contains the Rhesus heterotrimer, aquaporin-1, the three band 3 I-II-III dimers, and an unknown protein Y, that could correspond to the unknown protein X revealed in class 4 by Vallese et al. [66]. We validated the presence of this new structure by running an independent homogeneous reconstruction on a subset of 20,000 particles (**Figure 5c**, **Section 5.4.4**).

Recent versions of the cryoSPARC software allow 3D classification with large numbers of classes (e.g., an order of magnitude greater than previous standard settings). As DRGN-AI estimates poses, we performed 3D classification with DRGN-AI poses and identified again the supercomplex (**Figure S7**), providing additional validation for this new structure.

2.6 DRGN-AI for *ab initio* subtomogram reconstruction

Finally, we recover detailed compositional and conformational variability of elongation states of the 70S ribosome *in situ* by performing *ab initio* subtomogram averaging (STA) of tilt series images obtained from cryo-ET. In STA, the relative orientations between subtomograms are constrained by the tilting scheme during hierarchical pose search but this constraint is relaxed when switching to stochastic gradient descent to compensate for potential jitter during the imaging process. **Figure 6** shows the results of an *ab initio* heterogeneous subtomogram reconstruction on a dataset of the *Mycoplasma pneumoniae* 70S ribosome (EMPIAR-10499 [64]), using 11 tilts per particle. The influence of the number of tilts is analyzed in **Figure S8**. By visual inspection of 100 density maps sampled by k-means clustering, we reconstruct three known representative states of the ribosome during the translation elongation cycle: the "P state" with a tRNA in the P site, the "EF-Tu, P state" with a tRNA attached to an elongation factor and the "A, P state" where the tRNA has moved to the A site (**Figure 6a, b**). We subsequently run a Fourier backprojection with the particles associated with the "A, P state" and show the resulting density maps in **Figure S9** and **Supplementary Video 3**. We show in **Figure S10** that the set of "junk" particles is consistently classified as outliers in replicas of the experiment. Additionally, DRGN-AI resolves heterogeneity in the cellular context of each particle, for example producing density maps containing neighboring ribosomes within polysomes (**Figure 6d**). The distribution of the translation states is visualized within a representative tomogram in **Figure 6c**.

3 RELATED WORK

Modeling conformational heterogeneity in cryo-EM. Recent methods for heterogeneous cryo-EM reconstruction have explored the possibility of using linear [18, 53] and nonlinear [12, 16, 22, 38, 43, 45] mappings between latents and 3D structures. Nonlinear methods can be further categorized depending on the support they use for representing density maps: neural networks [75, 76], 3D Zernike polynomials [27], Gaussian mixture models [9, 77] or flow fields [54]. Normal mode-based approaches [23, 24, 32, 48] introduce a linear and physically interpretable mapping between latents and atomic coordinates. However, these methods need a coarse initialization of the density map or the poses to be provided by a pre-processing step of homogeneous reconstruction. In practice, this *ab initio* step is error-prone due to (1) the presence of junk particles and (2)

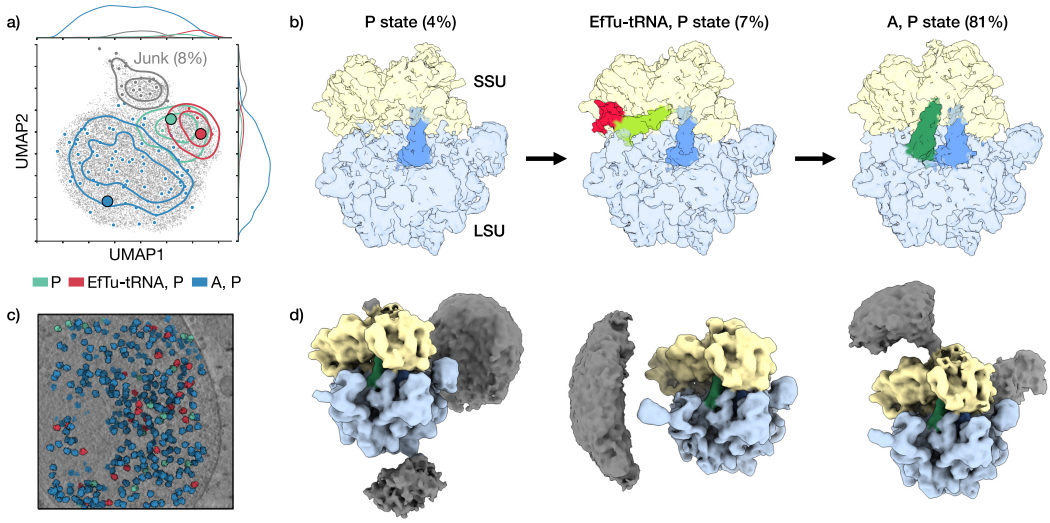


Fig. 6. **Single shot heterogeneous *ab initio* subtomogram reconstruction with DRGN-AI. a)** UMAP visualization of DRGN-AI latent embeddings of a dataset of the *M. pneumoniae* 70S ribosome (EMPIAR-10499 [64]). The 100 maps obtained by k-means clustering are classified by visual inspection according to the content of the tRNA sites, between the small sub-unit (SSU) and the large sub-unit (LSU). On top of a 2D scatterplot of the predicted conformations, the classified centroids and a Gaussian KDE of the distribution of states are shown. Large circles indicate the conformations of maps shown in panel **b**. **b)** Three intermediate states of the ribosome during the translation elongation cycle. The first state (P state) only contains a tRNA in the P site. The second state (EF-Tu, P state) contains a tRNA attached to an elongation factor. In the third state (A, P state), the tRNA has moved to the A site. **c)** Visualization of a representative *M. pneumoniae* tomogram with ribosomes colored by their class assignments from panel **a**. **d)** Sampled maps displaying density for neighboring ribosomes in polysomes.

large conformational or strong compositional heterogeneity. DRGN-AI runs directly on unposed particle images and is therefore not subject to the above limitations.

Ab initio single particle analysis. While early algorithms for heterogeneous reconstruction were based on expectation-maximization [15, 57, 58], stochastic gradient-based optimization techniques were later introduced to allow *ab initio* reconstruction and released as part of the cryoSPARC software [5, 55]. These methods, termed "3D classification", approximate the conformational landscape with a small set of independent voxel grids. While these methods can be used in practice to recover diverse structures from complex samples [30, 59], they are fundamentally limited to modeling discrete heterogeneity over a user-defined number of classes. CryoDRGN2 [78] and cryoFIRE [40] introduced preliminary neural *ab initio* methods for heterogeneous reconstruction. However, CryoDRGN2 is limited in speed and scales poorly for large datasets due to its pose search strategy. CryoFIRE uses an autoencoder to estimate poses but amortized inference is not well-suited for highly noisy datasets (e.g., for smaller complexes), where the encoder may memorize the data [14] (**Figure S1**). We propose a two-step pose estimation strategy and an autodecoding architecture to explicitly address the limitations of previous neural approaches, and in doing so, reveal both the compositional and conformational heterogeneity of large, unfiltered datasets.

Subtomogram averaging. Subtomogram averaging traditionally is done on "3D subtomograms", i.e. subvolumes extracted from the 3D tomogram that is obtained from a backprojection of 2D tilt series images [3, 7, 24, 25, 62]. As samples cannot be imaged at high tilt angles, a "missing wedge" in Fourier space leads to degraded high-frequency information in the 3D subtomograms. Methods using per-particle and per-tilt corrections [8, 29] or unifying single particle imaging and subtomogram averaging under a common framework [63, 65] improved the accuracy of subtomogram averaging but do not handle *ab initio* heterogeneous reconstruction. Rangan et. al. [56] recently showed that continuous conformational landscapes could be resolved directly from particle tilt images but required known poses. DRGN-AI is the first method that directly processes 2D data without building intermediate 3D subtomograms, enabling faster and single shot processing of electron-tomography data while resolving structural heterogeneity.

4 DISCUSSION

In this work, we demonstrate new capabilities for recovering 3D structure and variability from diverse, challenging single-particle cryo-EM and cryo-ET subtomogram datasets. Unlike most existing methods for heterogeneity analysis, DRGN-AI can be trained in an *ab initio* setting, and we leverage an expressive implicit neural representation to reconstruct compositionally and conformationally diverse structures. We find that our hybrid pose search and autodecoding algorithm leads to successful optimization of the model in challenging settings, in particular, in the regime of large, unfiltered cryo-EM datasets, demonstrating the utility of advanced heterogeneity analysis tools in an earlier stage of image processing. For example, we identified a new state of the human erythrocyte ankyrin-1 complex and of the v-ATPase complex. We additionally recovered the structure and dynamics of the DSL1/SNARE complex from a dataset containing a significant fraction of junk particles, a common occurrence in cryo-EM datasets that typically leads to complex and ad hoc processing pipelines [80].

How should the learned distribution of structures be interpreted and analyzed? Although we do not use any explicit regularization scheme as in variational approaches to generative modeling [36], we empirically observe that similar density maps tend to be close to each other in latent space. As a consequence, we found that compositional heterogeneity could effectively be revealed using k-means clustering, while conformational heterogeneity could be visualized through linear trajectories along the principal components of the latent space. We note that the distribution of latent vectors cannot be readily interpreted as a free energy landscape since the mapping between the latent space and the space of atomic models is unknown (and, in general, not an isometry) [37]. Imbuing a thermodynamic interpretation to distances in latent space is a remaining milestones.

Over the past decade, cryo-EM has become a dominant approach for experimental biomolecular structure determination. While recent machine learning breakthroughs have transformed our ability to computationally predict protein structure from sequence [1, 33] and design protein structures *de novo* [31, 69], these tools propose structures purely *in silico*, and therefore require experimental validation for many downstream applications or scientific study. With DRGN-AI, we seek to contribute to automated, reproducible pipelines that remove the manual, subjective decisions in cryo-EM data processing, thus increasing the throughput of experimental structure determination. In addition, we hope to expand the capabilities of cryo-EM reconstruction to more complex samples where the current paradigm of assigning poses separately before heterogeneity analysis is insufficient. Lastly, by offering new computational capabilities in protein structure determination,

we hope to enable the creation of larger and information-richer training datasets as well as more reliable experimental protocols for validation, thereby contributing to other machine learning methods in biology.

5 METHODS

5.1 Forward Model

5.1.1 Single-particle analysis. In single particle analysis, an image I_i is considered to be a random realization following [57, 67]:

$$I_i = C_i * P_{\phi_i} \mathcal{V}[z_i] + \eta_i, \tag{1}$$

where C_i is the (discretized) Contrast Transfer Function (CTF), $*$ the convolution operator and η_i additive white Gaussian noise. $\mathcal{V}$ maps conformations ($z \in \mathbb{R}^d$) to 3D density maps ($V[z] : \mathbb{R}^3 \to \mathbb{R}$). The operator P_ϕ represents an orthographic projection of the map $\mathcal{V}[z_i]$ from the pose $\phi = (R, \mathbf{t}) \in \mathrm{SO}(3) \times \mathbb{R}^2$, followed with a discretization of the projected image. For $V : \mathbb{R}^3 \to \mathbb{R}$,

$$P_{(R,\mathbf{t})} V = \mathcal{D}\left[(x, y) \mapsto \int_s V(R \cdot [x - t_x, y - t_y, s]^T) ds\right], \tag{2}$$

where, for $I : \mathbb{R}^2 \to \mathbb{R}$ and a D^2 grid $\{(x_m, y_n)\}_{m,n \in \{1,\ldots,D\}}$,

$$\mathcal{D}I : (m, n) \in \{1, \ldots, D\}^2 \mapsto I(x_m, y_n). \tag{3}$$

5.1.2 Subtomogram averaging. Similar to (1), the image formation model of *in situ* tomography follows

$$I_i^j = C_i^j * P_{\phi_i^j} \mathcal{V}[z_i] + \eta_i^j, \tag{4}$$

where the CTFs, poses and noises are tilt-dependent while the conformations are only particle-dependent.

5.1.3 Contrast transfer function. In all the experiments, the contrast transfer function is assumed to be known, given by an upstream software.

5.1.4 Dose-exposure correction. To account for accumulated radiation damage in tomography, we extend the CTF to account for lower signal-to-noise ratio (SNR) in tilts collected at later time-points and higher angles. We include a dose exposure correction to account for frequency-dependent signal attenuation in later tilt images, as described in [21]. Additionally, since sample thickness effectively increases at higher tilt angles leading to decreasing SNR for these tilts, we further multiply the CTF by the cosine of the tilt angle [2].

5.1.5 Implicit neural representation. The "hypervolume" [38] $\mathcal{V}_\theta : \mathbb{R}^{d \times 3} \to \mathbb{R}$ is stored in a neural network. Conditioned on a conformation $z_i \in \mathbb{R}^d$, $\mathcal{V}_\theta[z_i] : (\mathbf{k} \in)\mathbb{R}^3 \to \mathbb{R}$ represents the Hartley transform of the 3D electron scattering potential of a single particle. The frequency coordinate $\mathbf{k} \in [-0.5, 0.5]^3$ is expanded in a sinusoidal basis using Fourier features [61]. We use 64 base frequencies randomly sampled from a 3D Gaussian distribution of standard deviation 0.5. The neural network follows the same architecture as in cryoDRGN [75]: the concatenation of $\mathbf{z}_i$ with the positionally-encoded frequency is passed through 3 hidden residual layers [26] of size 256 with ReLU non-linearities. The last layer is linear and provides a one-dimensional output.

5.1.6 Hartley space. In order to avoid the computation of integrals, we simulate the image formation model in Hartley space, thereby allowing the use of the Fourier Slice Theorem. A synthetic image $\hat{I}_i$ is generated in Hartley space following

$$\hat{I}_i = \hat{C}_i \odot T_{\mathbf{t}_i} S_{R_i} \mathcal{V}_\theta[z_i], \tag{5}$$

where $\hat{C}_i$ is the Hartley transform of C_i and $\odot$ the element-wise multiplication. For a pose $\phi = (R, \mathbf{t})$ and $V : \mathbb{R}^3 \to \mathbb{R}$, S_R corresponds to a "slicing" operation:

$$S_R V : (k_x, k_y) \mapsto V(R \cdot [k_x, k_y, 0]^T). \tag{6}$$

$T_{\mathbf{t}}$ corresponds to a translation in Hartley space. Given $H : \mathbb{R}^2 \to \mathbb{R}$ and a D^2 grid $\{(k_m^x, k_n^y)\}_{m,n \in \{1,\dots,D\}}$, it is defined by

$$T_{\mathbf{t}} H : (m, n) \mapsto \cos(2\pi \mathbf{t} \cdot \mathbf{k}_{m,n}) H(\mathbf{k}_{m,n}) + \sin(2\pi \mathbf{t} \cdot \mathbf{k}_{m,n}) H(-\mathbf{k}_{m,n}), \tag{7}$$

where $\mathbf{k}_{m,n} = [k_m^x, k_n^y]$. In practice, we use a regular grid centered around the origin.

5.2 Optimization

5.2.1 Objective function. DRGN-AI aims at reproducing the image formation model and minimizing the reconstruction error in Hartley space (negative log-likelihood of the observations):

$$\mathcal{L}(\theta, \{\phi_i = (\mathbf{t}_i, R_i)\}, \{z_i\}) = \sum_{i=1}^{N} \|\mathcal{H}[I_i] - \hat{C}_i \odot T_{\mathbf{t}_i} S_{R_i} \mathcal{V}_\theta[z_i]\|_2^2, \tag{8}$$

where $\mathcal{H}$ is the discrete Hartley transform.

5.2.2 Optimization parameters. All the optimizable parameters are optimized with the Adam optimizer [35]. The learning rates are 0.0001 for θ, 0.001 for the poses $\{\phi_i\}$ and 0.01 for the conformations $\{z_i\}$. The batch size is set to 8 during HPS and 256 for the rest of the training. We initially tested the second-order method L-BFGS [41] instead of the Adam optimizer for poses, but this produced unstable results in our early experimentation.

5.2.3 Initialization. The weights of the neural network are randomly initialized with the default initialization scheme for linear layers in PyTorch [50]. The neural network is then trained with 10,000 images using fixed random poses and conformations ("pretraining phase"). The conformation table is initialized randomly from a d-dimensional Gaussian distribution of standard deviation 0.1. When switching from HPS to SGD, the pose table is initialized with the last estimated poses, once 500,000 images have been processed (with a minimum of 2 epochs of pose search).

5.2.4 Hardware. All the experiments are ran on 4 A100s NVIDIA GPUs using data parallelization.

5.3 Poses and Conformations

5.3.1 Hierarchical Pose Search. The pose search strategy is similar to [78]. The "reprojection error" is defined as

$$\|I_i - C_i * P_{\phi_i} \mathcal{V}_\theta[z_i]\|_2^2 \tag{9}$$

in the single-particle analysis setup and

$$\sum_{j=1}^{J} \|I_i^j - C_i^j * P_{\phi_i^j} \mathcal{V}_\theta[z_i]\|_2^2 \tag{10}$$

in the subtomogram averaging setup. Both are directly computed in Hartley space. In the latter case, a single pose is optimized per particle and the known tilting scheme constrains the relative orientations between $\{\phi_i^j\}_j$. The first search step is done on a predefined grid over $\mathrm{SO}(3) \times \mathbb{R}^2$ (4,608 rotations and 49 translations in [−10 pix., 10 pix.]). The first 8 poses minimizing the reprojection error are kept and refined with a local

search over 8 neighbors during 4 additional steps. The images are band-limited during pose search and the cutoff frequency increases linearly from $k_{\min}$ to $k_{\max}$ ($k_{\min} = 6$, $k_{\max} = 16$ in (image length)$^{-1}$). Grids are parameterized using the Hopf fibration [73], product of the Healpix [20] grid on the 2-sphere and a regular grid on the circle.

5.3.2 Conformational landscape analysis. For heterogeneous reconstruction experiments, the set of predicted conformations can be visually inspected in a 2D space after training using a dimensionality reduction method. We use PCA and UMAP [44]. We generally find it easier to reveal compositional heterogeneity with UMAP (as in **Figure 2b**, **c**) and continuous deformations with PCA (as in **Figure 2a**). To generate representative maps, we run k-means clustering in the latent space, find in each cluster the predicted conformation that is the closest to the cluster's centroid, and use this conformation to condition the hypervolume. The conditioned hypervolume is queried on a 3D grid of points and a 3D inverse Hartley transform is applied in order to obtain an explicit voxel-based representation of the electron scattering potential.

5.3.3 Pose-conformation entanglement. No measures are explicitly taken to disentangle poses and conformations (i.e., to guarantee that a change of conformation does not represent only a rigid transformation of the density map). The conformation table is randomly initialized (see **Section 5.2.3**) and is only frozen during the pretraining phase. We refer to Klindt et al. [37] for an in-depth discussion on ways to evaluate and potentially correct for pose-conformation entanglement.

5.4 Metrics

5.4.1 Per-image FSC. For synthetic heterogeneous datasets (on which we have access to ground truth density maps for each particle image) we use the "per-image FSC" to quantify the accuracy of a reconstruction. It corresponds to the Fourier Shell Correlation between the ground truth density map and the predicted 3D map associated with each image, averaged across the dataset. For DRGN-AI, the predicted map is obtained by feeding the predicted conformation of a given image to the decoder. In cryoSPARC, each image is classified among a finite number of possible classes and each class is associated with a predicted density map.

5.4.2 Rotation accuracy. We report the rotation accuracy using the "out-of-plane" an "in-plane" angular error. The estimated rotations $\mathbf{r}_i$ are first aligned to the reference rotations $\mathbf{R}_i$ using 100 candidates ($\mathbf{R} \in \{\mathbf{r}_i^{-1} \cdot \mathbf{R}_i, i \in \{0, \ldots, 99\}\}$) and keeping the minimizer of the mean Frobenius norm of the difference between the reference rotation matrices and the estimated aligned rotation matrices $\mathbf{r}_i \cdot \mathbf{R}$. The rotation matrices are then decomposed into viewing directions and viewing angles using the Euler decomposition in the canonical reference frame. The "out-of-plane" error is then defined as the angular error between the estimated (aligned) viewing directions and the reference viewing directions. The "in-plane" error is defined as the angular error between the viewing angles.

5.4.3 Translation accuracy. When computing the 2D translation error, we account for the fact that the origin of the 3D reference frame is unknown. Given a set of estimated rotation matrices $\mathbf{r}_i \in \mathbb{R}^{3\times 3}$, a set of estimated 2D translations $\mathbf{t}_i \in \mathbb{R}^2$ and a set of reference 2D translations $\mathbf{T}_i \in \mathbb{R}^2$, the "optimal 3D shift" $\mathbf{u} \in \mathbb{R}^3$ is defined as the minimizer of the mean square error of the corrected translations

$$\mathbf{u}^* = \operatorname{argmin}_{\mathbf{u}} \left\| \mathbf{T}_i - \left(\mathbf{t}_i - \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \cdot \mathbf{r}_i \cdot \mathbf{u} \right) \right\|_2^2 . \tag{11}$$

The minimizer is obtained with a numpy least-square solver and the corrected translations are defined as

$$\mathbf{t}_i' = \mathbf{t}_i - \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \cdot \mathbf{r}_i \cdot \mathbf{u}^*. \tag{12}$$

5.4.4 Validation in cryoSPARC. We used cryoSPARC version 4.4 [55] to validate the structure of the new "supercomplex" state in the ankyrin-1 complex dataset. We computed the L2 distance between the 710,437 reconstructed maps and kept the images associated with the 20,000 maps that were the closest from the identified "supercomplex" state (in L2 distance). Validation was obtained running a homogeneous reconstruction job in cryoSPARC, based on the poses predicted by DRGN-AI. We report two FSC curves: half-map and map-to-map (i.e., we compare cryoSPARC's density map to DRGN-AI's density map). We run cryoSPARC 3D classification using 80 classes, with poses provided by DRGN-AI, at a target resolution of 12 Å, for 40 O-EM epochs, with initial low pass filtering to 60 Å, class similarity of 0.1, auto-tuning of initial class similarity as True, and all other parameters set to their default values. The resulting density map of the supercomplex was identified by visual inspection of classes. A comparison to DRGN-AI's conformations is provided in **Figure S7**.

5.5 Analysis of Pose Encoder

We perform an additional experiment to evaluate the capability of a neural-based encoder to predict latent variables (poses in this case) purely from images. A neural network mapping particle images to poses is trained supervisedly on synthetic cryo-EM datasets of the 80S ribosome with different noise levels (generation protocol in **Section 5.7.1**). Each dataset is split into a training set (500k particles) and a test set (10k particles). The training poses are obtained by running DRGN-AI (HPS only) on the training datasets. The neural network is built with a stack of 2D convolutional layers (kernel size size of 3, "reflection" mode for padding) interlaced with ReLU activation functions at every layer and a group normalization layer (32 groups) followed by an average pooling layer every other layer. The sequential CHW sizes are: $1 \times 128 \times 128$, $32 \times 128 \times 128$, $32 \times 64 \times 64$, $64 \times 64 \times 64$, $64 \times 32 \times 32$, $128 \times 32 \times 32$, $128 \times 16 \times 16$, $256 \times 16 \times 16$, $256 \times 8 \times 8$, $512 \times 8 \times 8$. The last layer is a tanh activation function followed by a linear layer with 6 output dimensions. The output is interpreted as a rotation using the S2S2 parameterization [79] and converted into a quaternion. The neural network is optimized with the Adam optimizer, a learning rate of 0.0001, a batch size of 32 for 100 epochs. The loss for a pair of poses $(\mathbf{q}, \mathbf{q}')$, represented as quaternions, is defined by

$$\mathcal{L} = 1 - (\mathbf{q} \cdot \mathbf{q}')^2. \tag{13}$$

In **Figure S1**, we show that the error on the test set gets significant ($\geq$ 10 deg. of mean out-of-plane error) on high-noise datasets (SNR ≤ -10 dB), preventing true amortization.

5.6 Visualization

Molecular graphics and analyses performed with UCSF ChimeraX [51], developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from National Institutes of Health R01-GM129325 and the Office of Cyber Infrastructure and Computational Biology, National Institute of Allergy and Infectious Diseases. In all the figures, we show a single isosurface of the potential.

5.7 Data

A summary of all the datasets analyzed in this study and DRGN-AI dataset-specific training settings are found in Table S1. We also provide a summary of the parameters parameters of the deposited particle stacks in Table S2, for the dataset that were processed from picked particles.

5.7.1 Synthetic 80S Ribosome. We generate synthetic datasets of the 80S ribosome using the same procedure as Levy et al. [39] by following the image formation model (Equation 1). The electron scattering potential was simulated in ChimeraX [51] at 6.0 Å resolution from a combination of two atomic models: the small subunit (PDB:3J7A) and the large subunit (PDB:3J79) [70]. The dataset contains 100,000 particles with an image size of 128×128 and a pixel size of 3.77 Å. Rotations are uniformly sampled over SO(3) and images are centered (no translations). CTF defocus values are randomly sampled using log-normal distributions (details in the SI of Levy et al. [39]). Zero-mean white Gaussian noise is added with a standard deviation of 0.5. Example images are shown in Fig. S3.

5.7.2 Pre-catalytic spliceosome. The EMPIAR-10180 [52] dataset was downsampled to 128×128 from the original image size of 320×320, giving a pixel size of 4.25 Å/pix. We process the full dataset of deposited particles (327,490 particles).

5.7.3 Assembling 50S ribosome. The EMPIAR-10076 [13] dataset was downsampled to 256×256 from the original image size of 320×320, giving a pixel size of 1.64 Å/pix. We process the full dataset of deposited particles (131,899 particles).

5.7.4 SARS-CoV-2 spike protein. The dataset from Walls et al. [68] was provided courtesy of the authors. Particle images were downsampled to 128×128 from the original resolution 400×400, giving a pixel size of 3.28 Å/pix. We process the full dataset of 369,429 particles.

5.7.5 DSL1/SNARE complex. The EMPIAR-11846 [11] dataset was downsampled to 128×128 from the original image size of 400×400, giving a pixel size of 3.47 Å/pix. The original dataset contains 286,801 picked particles and was filtered down to 214,511 particles by visual inspection of the latent space obtained with cryoDRGN [75].

5.7.6 V-ATPase complex. V-ATPase particles were obtained from reanalysis of EMPIAR-10874 [60] dataset entry #10 (unaligned multiframe movies of Pig Kidney V-ATPase bound to mEAK-7 with ATP collected using Titan Krios and Falcon4). Initial preprocessing prior to DRGN-AI analysis was performed in RELION5 [6]. Briefly, 5,498 movies were aligned and motion corrected using the RELION implementation of MotionCor2. 1,517,258 particles were picked using 2D class average templates generated from classification of an initial Laplacian-of-Gaussian blob pick on a subset of images. Iterative 2D classification was used to remove classes that appeared to contain false positive picks and empty micelles resulting in a particle stack containing 267,216 particle images downsampled to 128×128 (3.97 Å/pix).

5.7.7 Ankyrin-1 complex. The EMPIAR-11043 [66] dataset was downsampled to 128×128 from the original image size of 450×450, giving a pixel size of 2.92 Å/pix. We process the full dataset of ankyrin complex particles (710,437 particles).

5.7.8 Mycoplasma pneumoniae 70S ribosome. Subtomograms were picked and processed from EMPIAR-10499 [64] with the same protocol as performed in Rangan et al. [56]. The dataset contains 18,466 particles

with 41 tilts each (757,106 subtilt images). We keep the first 11 tilts (3 degrees between consecutive tilts, total coverage of ±15 degrees) at a resolution of 128 × 128 (3.9 Å/pix.). The DRGN-AI model was trained with an architecture dimension of 1024, in addition to the settings described in Table S1. The backprojection of the selected "A, P" state particles in Figure S9 was done at a resolution of 294 × 294 (1.7 Å/pix.).

5.7.9 Synthetic 1D rotation. We used the synthetic dataset introduced in Zhong et al. [75] to simulate strong conformational heterogeneity. The dataset contains 50,000 particles. 50 density maps were generated along a 1D reaction coordinate defined by rotating a dihedral angle in the atomic model of a hypothetical protein complex (0.03 rad. increment, **Figure S11a**). Density maps were generated using the molmap command in ChimeraX [51] at a grid spacing of 6 Å per pixel, resolution of 12 Å and a box size of 128. Each density map is projected 1,000 times using random poses (translations of ±10 pixels). CTF and noise at a signal-to-noise ratio of 0.1 were added to the dataset. The CTF defocus values were randomly sampled from EMPIAR-10028 [71].

6 CODE AVAILABILITY

DRGN-AI is available as open-source software at https://drgnai.cs.princeton.edu.

7 DATA AVAILABILITY

The following datasets can be accessed on EMPIAR (https://www.ebi.ac.uk/empiar/): pre-catalytic spliceosome (EMPIAR-10180), assembling 50S ribosome (EMPIAR-10076), DLS1/SNARE complex (EMPIAR-11846), V-ATPase complex (EMPIAR-10874), ankyrin-1 complex (EMPIAR-11043), *Mycoplasma pneumoniae* 70S ribosome (EMPIAR-10499). Additional datasets of the covid spike protein are available upon request. DRGN-AI model weights and any associated files needed to reproduce this analysis will be deposited to Zenodo. Atomic models used from previous studies were obtained from the PDB (3J7A, 3J79, 6VXX, 6VYB, 8EKI).

8 ACKNOWLEDGEMENTS

We thank Mike Dunne, Ramya Rangan, and members of the Zhong lab for assistance on the manuscript. We thank Kevin DAmico and Fred Hughson for guidance on structure interpretation. We thank Bridget Carragher, Clint Potter, and members of the Chan Zuckerberg Imaging Institute for feedback on the cryo-ET experiments. We thank Vineet Bansal for software engineering support. We acknowledge the use of computational resources from the SLAC Shared Scientific Data Facility and from Princeton Research Computing, which is consortium of groups led by the Princeton Institute for Computational Science and Engineering (PICSciE) and Office of Information Technology's Research Computing. Some of this work was performed at the Simons Electron Microscopy Center at the New York Structural Biology Center, with support from the Simons Foundation (SF349247), NIH (U24 GM129539), and NY State Assembly. This work was partly supported by the U.S. Department of Energy, under DOE Contract No. DE-AC02-76SF00515 and by the SLAC LDRD project "AtomicSPI: Learning atomic scale biomolecular dynamics from single-particle imaging data". The Zhong lab gratefully acknowledges support from the Princeton Catalysis Initiative, Princeton School of Engineering and Applied Sciences, and Generate Biomedicines. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

9 AUTHOR CONTRIBUTIONS

A.L., F.P., G.W., and E.Z. conceived of the work; A.L., R.R., J.R.F. and E.Z implemented the methods and performed the computational experiments; A.L., R.R., M.G., and E.Z. developed the software package; J.J. preprocessed subtomogram data; F.V. and O.B.C. analyzed the results on the ankyrin dataset; A.L., R.R. and E.Z. wrote the manuscript with feedback from all the authors; G.W. and E.Z. supervised the project.

10 COMPETING INTERESTS

The authors declare no competing interests.

REFERENCES

[1] Josh Abramson, Jonas Adler, Jack Dunger, Richard Evans, Tim Green, Alexander Pritzel, Olaf Ronneberger, Lindsay Willmore, Andrew J Ballard, Joshua Bambrick, Sebastian W Bodenstein, David A Evans, Chia-Chun Hung, Michael O'Neill, David Reiman, Kathryn Tunyasuvunakool, Zachary Wu, Akvilė Žemgulytė, Eirini Arvaniti, Charles Beattie, Ottavia Bertolli, Alex Bridgland, Alexey Cherepanov, Miles Congreve, Alexander I Cowen-Rivers, Andrew Cowie, Michael Figurnov, Fabian B Fuchs, Hannah Gladman, Rishub Jain, Yousuf A Khan, Caroline M R Low, Kuba Perlin, Anna Potapenko, Pascal Savy, Sukhdeep Singh, Adrian Stecula, Ashok Thillaisundaram, Catherine Tong, Sergei Yakneen, Ellen D Zhong, Michal Zielinski, Augustin Žídek, Victor Bapst, Pushmeet Kohli, Max Jaderberg, Demis Hassabis, and John M Jumper. 2024. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* (May 2024).

[2] Tanmay AM Bharat, Christopher J Russo, Jan Löwe, Lori A Passmore, and Sjors HW Scheres. 2015. Advances in single-particle electron cryomicroscopy structure determination applied to sub-tomogram averaging. *Structure* 23, 9 (2015), 1743–1753.

[3] Tanmay AM Bharat and Sjors HW Scheres. 2016. Resolving macromolecular structures from electron cryo-tomography data using subtomogram averaging in RELION. *Nature protocols* 11, 11 (2016), 2054–2065.

[4] Piotr Bojanowski, Armand Joulin, David Lopez-Paz, and Arthur Szlam. 2017. Optimizing the latent space of generative networks. *arXiv preprint arXiv:1707.05776* (2017).

[5] Marcus A Brubaker, Ali Punjani, and David J Fleet. 2015. Building Proteins in a Day: Efficient 3D Molecular Reconstruction. (April 2015). arXiv:1504.03573 [cs.CV]

[6] Alister Burt, Bogdan Toader, Rangana Warshamanage, Andriko von Kügelgen, Euan Pyle, Jasenko Zivanov, Dari Kimanius, Tanmay AM Bharat, and Sjors HW Scheres. 2024. An image processing pipeline for electron cryo-tomography in RELION-5. *FEBS Open Bio* (2024).

[7] Daniel Castaño-Díez, Mikhail Kudryashev, Marcel Arheit, and Henning Stahlberg. 2012. Dynamo: a flexible, user-friendly development tool for subtomogram averaging of cryo-EM data in high-performance computing environments. *Journal of structural biology* 178, 2 (2012), 139–151.

[8] Muyuan Chen, James M Bell, Xiaodong Shi, Stella Y Sun, Zhao Wang, and Steven J Ludtke. 2019. A complete data processing workflow for cryo-ET and subtomogram averaging. *Nature methods* 16, 11 (2019), 1161–1168.

[9] Muyuan Chen and Steven J Ludtke. 2021. Deep learning-based mixed-dimensional Gaussian mixture model for characterizing variability in cryo-EM. *Nature methods* 18, 8 (2021), 930–936.

[10] Eugene Y D Chua, Joshua H Mendez, Micah Rapp, Serban L Ilca, Yong Zi Tan, Kashyap Maruthi, Huihui Kuang, Christina M Zimanyi, Anchi Cheng, Edward T Eng, Alex J Noble, Clinton S Potter, and Bridget Carragher. 2022. Better, Faster, Cheaper: Recent Advances in Cryo-Electron Microscopy. *Annu. Rev. Biochem.* 91 (June 2022), 1–32.

[11] Kevin A DAmico, Abigail E Stanton, Jaden D Shirkey, Sophie M Travis, Philip D Jeffrey, and Frederick M Hughson. 2023. Structure of a Membrane Tethering Complex Incorporating Multiple SNAREs. *bioRxiv* (2023), 2023–01.

[12] Ali Dashti, Ghoncheh Mashayekhi, Mrinal Shekhar, Danya Ben Hail, Salah Salah, Peter Schwander, Amedee des Georges, Abhishek Singharoy, Joachim Frank, and Abbas Ourmazd. 2020. Retrieving functional pathways of biomolecules from single-particle snapshots. *Nature communications* 11, 1 (2020), 4734.

[13] Joseph H Davis, Yong Zi Tan, Bridget Carragher, Clinton S Potter, Dmitry Lyumkis, and James R Williamson. 2016. Modular assembly of the bacterial large ribosomal subunit. *Cell* 167, 6 (2016), 1610–1622.

[14] Daniel G Edelberg and Roy R Lederman. 2023. Using VAEs to Learn Latent Variables: Observations on Applications in cryo-EM. *arXiv preprint arXiv:2303.07487* (2023).

[15] Dominika Elmlund and Hans Elmlund. 2012. SIMPLE: Software for ab initio reconstruction of heterogeneous single-particles. *Journal of structural biology* 180, 3 (2012), 420–427.

[16] Joachim Frank and Abbas Ourmazd. 2016. Continuous changes in structure mapped by manifold embedding of single-particle data in cryo-EM. *Methods* 100 (2016), 61–67.

[17] Jesus G Galaz-Montoya and Steven J Ludtke. 2017. The advent of structural biology in situ by single particle cryo-electron tomography. *Biophysics reports* 3 (2017), 17–35.

[18] Marc Aurèle Gilles and Amit Singer. 2023. A Bayesian Framework for Cryo-EM Heterogeneity Analysis using Regularized Covariance Estimation. *bioRxiv* (Nov. 2023).

[19] Marc Aurele T Gilles and Amit Singer. 2023. A Bayesian framework for cryo-EM heterogeneity analysis using regularized covariance estimation. *bioRxiv* (2023), 2023–10.

[20] Krzysztof M Gorski, Eric Hivon, Anthony J Banday, Benjamin D Wandelt, Frode K Hansen, Mstvos Reinecke, and Matthia Bartelmann. 2005. HEALPix: A framework for high-resolution discretization and fast analysis of data distributed on the sphere. *The Astrophysical Journal* 622, 2 (2005), 759.

[21] Timothy Grant and Nikolaus Grigorieff. 2015. Measuring the optimal exposure for single particle cryo-EM using a 2.6 Å reconstruction of rotavirus VP6. *elife* 4 (2015), e06980.

[22] Harshit Gupta, Thong H Phan, Jaejun Yoo, and Michael Unser. 2020. Multi-CryoGAN: Reconstruction of continuous conformations in Cryo-EM using generative adversarial networks. In *European Conference on Computer Vision*. Springer, 429–444.

[23] Ilyes Hamitouche and Slavica Jonic. 2022. DeepHEMNMA: ResNet-based hybrid analysis of continuous conformational heterogeneity in cryo-EM single particle images. *Frontiers in Molecular Biosciences* 9 (2022), 965645.

[24] Mohamad Harastani, Mikhail Eltsov, Amélie Leforestier, and Slavica Jonic. 2021. Hemnma-3d: Cryo electron tomography method based on normal mode analysis to study continuous conformational variability of macromolecular complexes. *Frontiers in molecular biosciences* 8 (2021), 663121.

[25] Mohamad Harastani, Mikhail Eltsov, Amélie Leforestier, and Slavica Jonic. 2022. TomoFlow: analysis of continuous conformational variability of macromolecules in cryogenic subtomograms based on 3D dense optical flow. *Journal of molecular biology* 434, 2 (2022), 167381.

[26] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.

[27] David Herreros, Roy R Lederman, James Krieger, Amaya Jiménez-Moreno, Marta Martínez, David Myška, David Strelak, Jiri Filipovic, Ivet Bahar, Jose Maria Carazo, et al. 2021. Approximating deformation fields for the analysis of continuous heterogeneity of biological macromolecules by 3D Zernike polynomials. *IUCrJ* 8, 6 (2021), 992–1005.

[28] David Herreros, Roy R Lederman, James M Krieger, Amaya Jiménez-Moreno, Marta Martínez, David Myška, D Strelak, J Filipovic, Carlos OS Sorzano, and José M Carazo. 2023. Estimating conformational landscapes from Cryo-EM particles by 3D Zernike polynomials. *Nature Communications* 14, 1 (2023), 154.

[29] Benjamin A Himes and Peijun Zhang. 2018. emClarity: software for high-resolution cryo-electron tomography and subtomogram averaging. *Nature methods* 15, 11 (2018), 955–961.

[30] Chi-Min Ho, Xiaorun Li, Mason Lai, Thomas C Terwilliger, Josh R Beck, James Wohlschlegel, Daniel E Goldberg, Anthony W P Fitzpatrick, and Z Hong Zhou. 2020. Bottom-up structural proteomics: cryoEM of protein complexes enriched from the cellular milieu. *Nat. Methods* 17, 1 (Jan. 2020), 79–85.

[31] John B Ingraham, Max Baranov, Zak Costello, Karl W Barber, Wujie Wang, Ahmed Ismail, Vincent Frappier, Dana M Lord, Christopher Ng-Thow-Hing, Erik R Van Vlack, et al. 2023. Illuminating protein space with a programmable generative model. *Nature* (2023), 1–9.

[32] Qiyu Jin, Carlos Oscar S Sorzano, José Miguel de La Rosa-Trevín, José Román Bilbao-Castro, Rafael Núñez-Ramírez, Oscar Llorca, Florence Tama, and Slavica Jonić. 2014. Iterative elastic 3D-to-2D alignment method using normal modes for studying structural dynamics of large macromolecular complexes. *Structure* 22, 3 (2014), 496–506.

[33] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, et al. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 7873 (2021), 583–589.

[34] Satinder Kaur, Josue Gomez-Blanco, Ahmad AZ Khalifa, Swathi Adinarayanan, Ruben Sanchez-Garcia, Daniel Wrapp, Jason S McLellan, Khanh Huy Bui, and Javier Vargas. 2021. Local computational methods to improve the interpretability and analysis of cryo-EM maps. *Nature communications* 12, 1 (2021), 1240.

[35] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

[36] Diederik P Kingma and Max Welling. 2013. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114* (2013).

[37] David Alexander Klindt, Aapo Hyvarinen, Axel Levy, Nina Miolane, and Frederic Poitevin. 2024. Towards Interpretable Cryo-EM: Disentangling Latent Spaces of Molecular Conformations. *bioRxiv* (2024), 2024–03.

[38] Roy R Lederman and Amit Singer. 2017. Continuously heterogeneous hyper-objects in cryo-EM and 3-D movies of many temporal dimensions. *arXiv preprint arXiv:1704.02899* (2017).

[39] Axel Levy, Frédéric Poitevin, Julien Martel, Youssef Nashed, Ariana Peck, Nina Miolane, Daniel Ratner, Mike Dunne, and Gordon Wetzstein. 2022. Cryoai: Amortized inference of poses for ab initio reconstruction of 3d molecular volumes from real cryo-em images. In *European Conference on Computer Vision*. Springer, 540–557.

[40] Axel Levy, Gordon Wetzstein, Julien NP Martel, Frederic Poitevin, and Ellen Zhong. 2022. Amortized Inference for Heterogeneous Reconstruction in Cryo-EM. *Advances in Neural Information Processing Systems* 35 (2022), 13038–13049.

[41] Dong C Liu and Jorge Nocedal. 1989. On the limited memory BFGS method for large scale optimization. *Mathematical programming* 45, 1 (1989), 503–528.

[42] Zhenwei Luo, Fengyun Ni, Qinghua Wang, and Jianpeng Ma. 2023. OPUS-DSD: deep structural disentanglement for cryo-EM single-particle analysis. *Nature Methods* 20, 11 (2023), 1729–1738.

[43] Suvrajit Maji, Hstau Liao, Ali Dashti, Ghoncheh Mashayekhi, Abbas Ourmazd, and Joachim Frank. 2020. Propagation of conformational coordinates across angular space in mapping the continuum of states from cryo-EM data by manifold embedding. *Journal of chemical information and modeling* 60, 5 (2020), 2484–2491.

[44] Leland McInnes, John Healy, and James Melville. 2018. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426* (2018).

[45] Amit Moscovich, Amit Halevi, Joakim Andén, and Amit Singer. 2020. Cryo-EM reconstruction of continuous heterogeneity by Laplacian spectral volumes. *Inverse Problems* 36, 2 (2020), 024003.

[46] Takanori Nakane, Dari Kimanius, Erik Lindahl, and Sjors HW Scheres. 2018. Characterisation of molecular motions in cryo-EM single-particle data by multi-body refinement in RELION. *elife* 7 (2018), e36861.

[47] Takanori Nakane, Abhay Kotecha, Andrija Sente, Greg McMullan, Simonas Masiulis, Patricia MGE Brown, Ioana T Grigoras, Lina Malinauskaite, Tomas Malinauskas, Jonas Miehling, et al. 2020. Single-particle cryo-EM at atomic resolution. *Nature* 587, 7832 (2020), 152–156.

[48] Youssef Nashed, Ariana Peck, Julien Martel, Axel Levy, Bongjin Koo, Gordon Wetzstein, Nina Miolane, Daniel Ratner, and Frédéric Poitevin. 2022. Heterogeneous reconstruction of deformable atomic models in Cryo-EM. *arXiv preprint arXiv:2209.15121* (2022).

[49] Catherine M Oikonomou and Grant J Jensen. 2017. Cellular electron cryotomography: toward structural biology in situ. *Annual review of biochemistry* 86 (2017), 873–896.

[50] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. 2019. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems* 32 (2019).

[51] Eric F Pettersen, Thomas D Goddard, Conrad C Huang, Elaine C Meng, Gregory S Couch, Tristan I Croll, John H Morris, and Thomas E Ferrin. 2021. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Science* 30, 1 (2021), 70–82.

[52] Clemens Plaschka, Pei-Chun Lin, and Kiyoshi Nagai. 2017. Structure of a pre-catalytic spliceosome. *Nature* 546, 7660 (2017), 617–621.

[53] Ali Punjani and David J Fleet. 2021. 3D variability analysis: Resolving continuous flexibility and discrete heterogeneity from single particle cryo-EM. *Journal of structural biology* 213, 2 (2021), 107702.

[54] Ali Punjani and David J Fleet. 2023. 3DFlex: determining structure and motion of flexible proteins from cryo-EM. *Nature Methods* (2023), 1–11.

[55] Ali Punjani, John L Rubinstein, David J Fleet, and Marcus A Brubaker. 2017. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nature methods* 14, 3 (2017), 290–296.

[56] Ramya Rangan, Sagar Khavnekar, Adam Lerer, Jake Johnston, Ron Kelley, Martin Obr, Abhay Kotecha, and Ellen D Zhong. 2023. Deep reconstructing generative networks for visualizing dynamic biomolecules inside cells. *bioRxiv* (2023), 2023–08.

[57] Sjors HW Scheres. 2012. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *Journal of structural biology* 180, 3 (2012), 519–530.

[58] Sjors HW Scheres, Haixiao Gao, Mikel Valle, Gabor T Herman, Paul PB Eggermont, Joachim Frank, and Jose-Maria Carazo. 2007. Disentangling conformational states of macromolecules in 3D-EM through likelihood optimization. *Nature methods* 4, 1 (2007), 27–29.

[59] Chih-Chia Su, Meinan Lyu, Christopher E Morgan, Jani Reddy Bolla, Carol V Robinson, and Edward W Yu. 2021. A 'Build and Retrieve' methodology to simultaneously solve cryo-EM structures of membrane proteins. *Nat. Methods* 18, 1 (Jan. 2021), 69–75.

[60] Yong Zi Tan, Yazan M Abbas, Jing Ze Wu, Di Wu, Kristine A Keon, Geoffrey G Hesketh, Stephanie A Bueler, Anne-Claude Gingras, Carol V Robinson, Sergio Grinstein, and John L Rubinstein. 2022. CryoEM of endogenous mammalian V-ATPase interacting with the TLDc protein mEAK-7. *Life Science Alliance* 5, 11 (2022), e202201527.

[61] Matthew Tancik, Pratul Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan Barron, and Ren Ng. 2020. Fourier features let networks learn high frequency functions in low dimensional domains. *Advances in Neural Information Processing Systems* 33 (2020), 7537–7547.

[62] Guang Tang, Liwei Peng, Philip R Baldwin, Deepinder S Mann, Wen Jiang, Ian Rees, and Steven J Ludtke. 2007. EMAN2: an extensible image processing suite for electron microscopy. *Journal of structural biology* 157, 1 (2007), 38–46.

[63] Dimitry Tegunov and Patrick Cramer. 2019. Real-time cryo-electron microscopy data preprocessing with Warp. *Nature methods* 16, 11 (2019), 1146–1152.

[64] Dimitry Tegunov, Liang Xue, Christian Dienemann, Patrick Cramer, and Julia Mahamid. 2021. Multi-particle cryo-EM refinement with M visualizes ribosome-antibiotic complex at 3.5 Å in cells. *Nature Methods* 18, 2 (2021), 186–193.

[65] Dimitry Tegunov, Liang Xue, Christian Dienemann, Patrick Cramer, and Julia Mahamid. 2021. Multi-particle cryo-EM refinement with M visualizes ribosome-antibiotic complex at 3.5 Å in cells. *Nature Methods* 18, 2 (2021), 186–193.

[66] Francesca Vallese, Kookjoo Kim, Laura Y Yen, Jake D Johnston, Alex J Noble, Tito Calì, and Oliver Biggs Clarke. 2022. Architecture of the human erythrocyte ankyrin-1 complex. *Nature Structural & Molecular Biology* 29, 7 (2022), 706–718.

[67] Miloš Vulović, Raimond BG Ravelli, Lucas J van Vliet, Abraham J Koster, Ivan Lazić, Uwe Lücken, Hans Rullgård, Ozan Öktem, and Bernd Rieger. 2013. Image formation modeling in cryo-electron microscopy. *Journal of structural biology* 183, 1 (2013), 19–32.

[68] Alexandra C Walls, Young-Jun Park, M Alejandra Tortorici, Abigail Wall, Andrew T McGuire, and David Veesler. 2020. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell* 181, 2 (2020), 281–292.

[69] Joseph L Watson, David Juergens, Nathaniel R Bennett, Brian L Trippe, Jason Yim, Helen E Eisenach, Woody Ahern, Andrew J Borst, Robert J Ragotte, Lukas F Milles, et al. 2023. De novo design of protein structure and function with RFdiffusion. *Nature* 620, 7976 (2023), 1089–1100.

[70] Wilson Wong, Xiao-chen Bai, Alan Brown, Israel S Fernandez, Eric Hanssen, Melanie Condron, Yan Hong Tan, Jake Baum, and Sjors HW Scheres. 2014. Cryo-EM structure of the Plasmodium falciparum 80S ribosome bound to the anti-protozoan drug emetine. *Elife* 3 (2014), e03080.

[71] Wilson Wong, Xiao-chen Bai, Alan Brown, Israel S Fernandez, Eric Hanssen, Melanie Condron, Yan Hong Tan, Jake Baum, and Sjors HW Scheres. 2014. Cryo-EM structure of the Plasmodium falciparum 80S ribosome bound to the anti-protozoan drug emetine. *Elife* 3 (2014), e03080.

[72] Liang Xue, Swantje Lenz, Maria Zimmermann-Kogadeeva, Dimitry Tegunov, Patrick Cramer, Peer Bork, Juri Rappsilber, and Julia Mahamid. 2022. Visualizing translation dynamics at atomic detail inside a bacterial cell. *Nature* 610, 7930 (2022), 205–211.

[73] Anna Yershova, Swati Jain, Steven M Lavalle, and Julie C Mitchell. 2010. Generating uniform incremental grids on SO (3) using the Hopf fibration. *The International journal of robotics research* 29, 7 (2010), 801–812.

[74] Ka Man Yip, Niels Fischer, Elham Paknia, Ashwin Chari, and Holger Stark. 2020. Atomic-resolution protein structure determination by cryo-EM. *Nature* 587, 7832 (2020), 157–161.

[75] Ellen D Zhong, Tristan Bepler, Bonnie Berger, and Joseph H Davis. 2021. CryoDRGN: reconstruction of heterogeneous cryo-EM structures using neural networks. *Nature methods* 18, 2 (2021), 176–185.

[76] Ellen D Zhong, Tristan Bepler, Joseph H Davis, and Bonnie Berger. 2020. Reconstructing continuous distributions of 3D protein structure from cryo-EM images.. In *International Conference on Learning Representations (ICLR)*.

[77] Ellen D Zhong, Adam Lerer, Joseph H Davis, and Bonnie Berger. 2020. Exploring generative atomic models in cryo-EM reconstruction, In NeurIPS Workshop on Machine Learning for Structural Biology (MLSB). *arXiv preprint arXiv:2107.01331*.

[78] Ellen D Zhong, Adam Lerer, Joseph H Davis, and Bonnie Berger. 2021. Cryodrgn2: Ab initio neural reconstruction of 3d protein structures from real cryo-em images. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 4066–4075.

[79] Yi Zhou, Connelly Barnes, Jingwan Lu, Jimei Yang, and Hao Li. 2019. On the continuity of rotation representations in neural networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 5745–5753.

[80] Jianying Zhu, Qi Zhang, Hui Zhang, Zuoqiang Shi, Mingxu Hu, and Chenglong Bao. 2023. A minority of final stacks yields superior amplitude in single-particle cryo-EM. *Nat. Commun.* 14, 1 (Dec. 2023), 7822.

SUPPLEMENTARY INFORMATION

11 SUPPLEMENTARY FIGURES

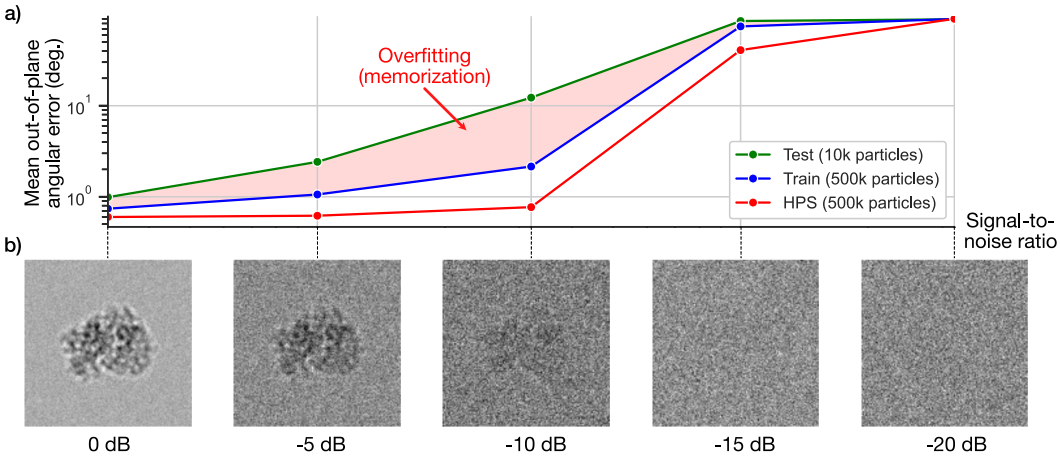


Fig. S1. **Limits of amortized inference in high-noise regimes.** See **Section 5.5** for experimental details. **a)** Training, testing and DRGN-AI (HPS) errors compared the "ground truth" poses, used for generation. **b)** Projection from the same pose for the different noise levels used in this experiment.

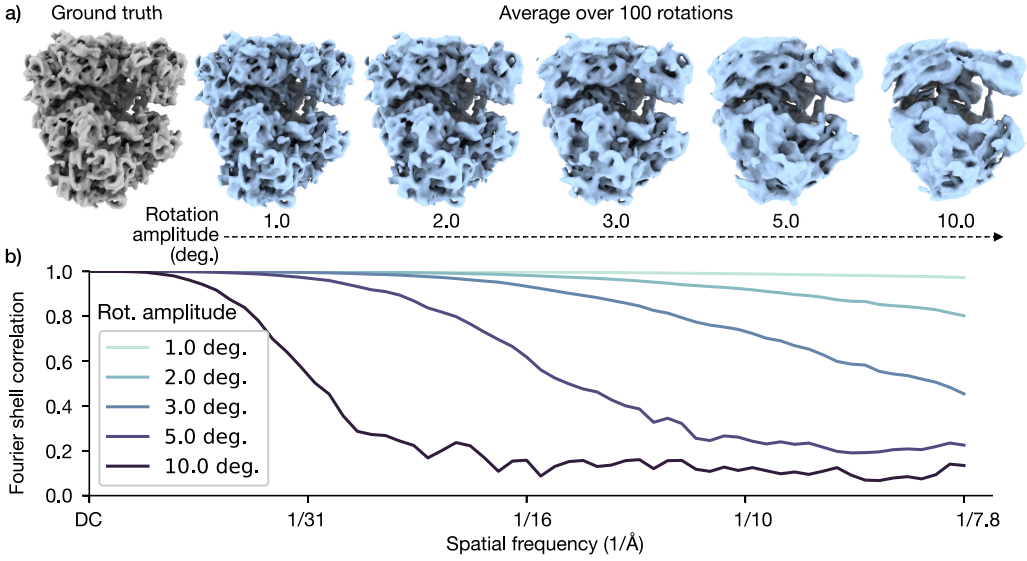
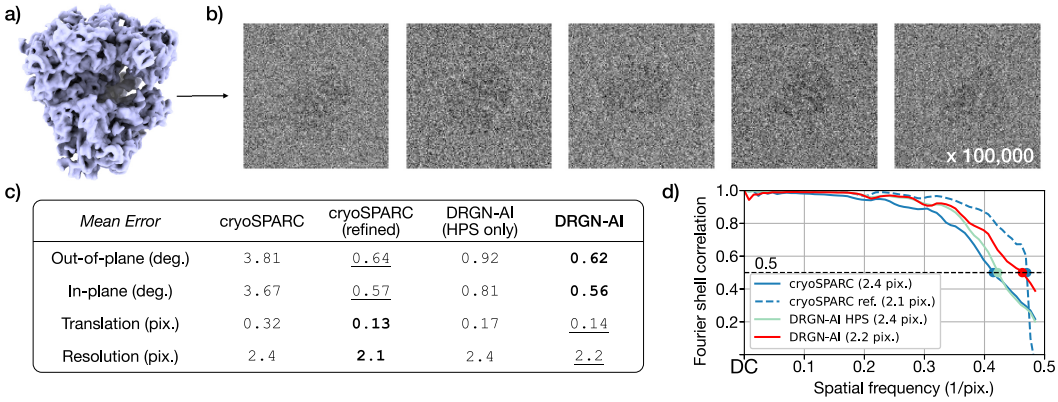


Fig. S2. **Influence of pose error on resolution.** To illustrate the impact of inaccurate poses on the reconstructed density map, 100 randomly rotated density maps of the 80S ribosome ($128 \times 128 \times 128$, 3.77 Å/pix.) are averaged, using a rotation amplitude between 1 and 10 degrees. **a)** Ground truth and averaged density maps for different rotation amplitudes. **b)** Fourier shell correlation for different rotation amplitudes.



Mean Error	cryoSPARC	cryoSPARC (refined)	DRGN-AI (HPS only)	**DRGN-AI**
Out-of-plane (deg.)	3.81	0.64	0.92	**0.62**
In-plane (deg.)	3.67	0.57	0.81	**0.56**
Translation (pix.)	0.32	**0.13**	0.17	0.14
Resolution (pix.)	2.4	**2.1**	2.4	2.2

Fig. S3. **DRGN-AI *ab initio* homogeneous reconstruction of synthetic datasets. a)** Ground truth density map of the 80S ribosome. **b)** Five example of particle images (100,000 particles, 128 × 128, 3.77 Å/pix). **c)** Pose error for DRGN-AI, DRGN-AI using HPS only and cryoSPARC [55]. **d)** FSC curves between the reconstructed density map and the ground truth density map (resolution at FSC 0.5 between parenthesis).

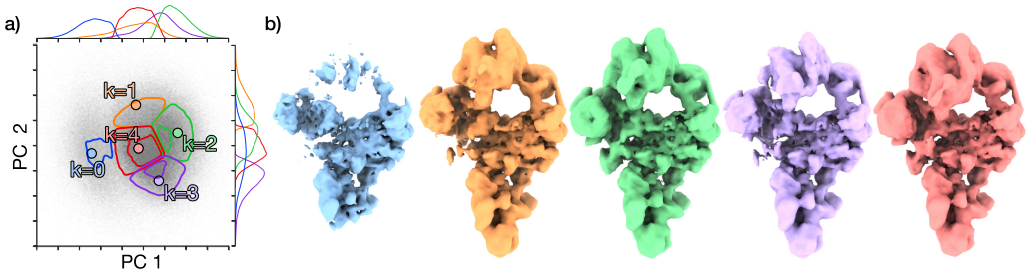


Fig. S4. **Additional visualizations of the latent space for the spliceosome dataset. a)** UMAP, 5 centroids obtained by k-means clustering and KDE plots of their regions of support. **b)** Associated density maps.

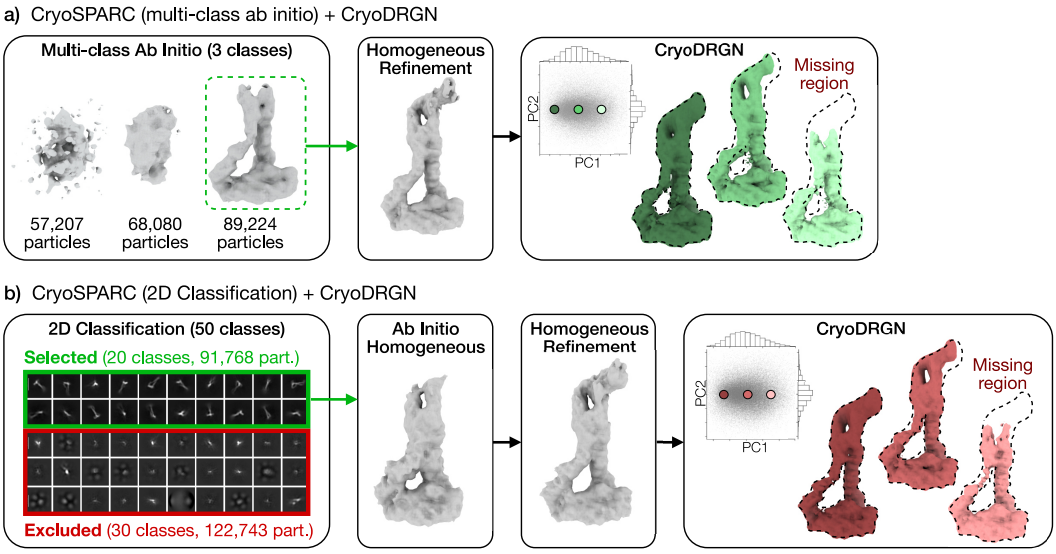


Fig. S5. **Comparison to other cryo-EM workflows for the DSL1 dataset, based on the cryoSPARC software [55].** **a)** Spurious particles are filtered using multi-class *ab initio* reconstruction with 3 classes. Poses are then refined with a step of homogeneous refinement. The filtered and posed particle stack is finally processed with cryoDRGN [75]. **b)** Filtering is done by 2D classification and followed by a step of *ab initio* reconstruction and a step of homogeneous refinement. The posed particle stack is then processed with cryoDRGN.

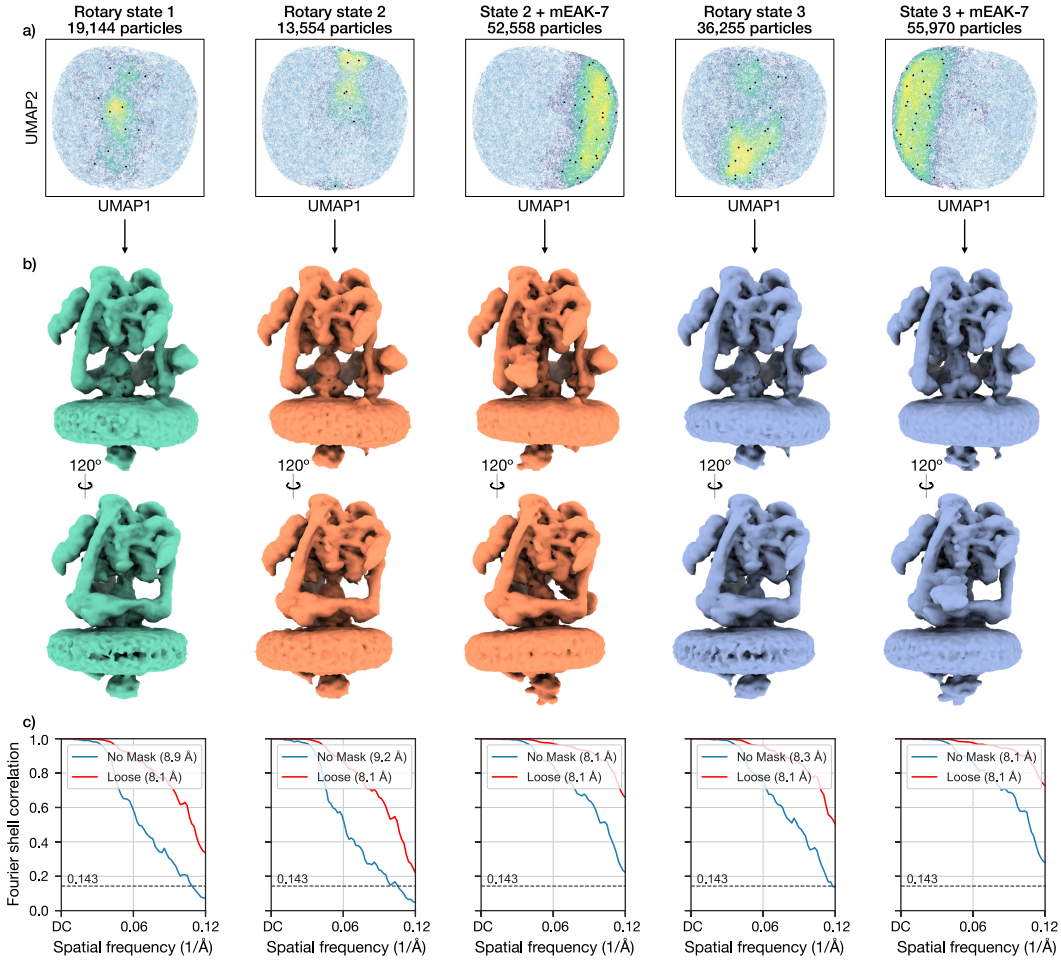


Fig. S6. **Backprojections of V-ATPase states. a)** UMAP visualizations of the latent embeddings, colored by the Gaussian kernel density estimate of the selected particles from each state. **b)** Two views of backprojected density maps from the selected particles of each state. **c)** Half-map FSC with soft mask at half of maximum value, 15 Å cosine edge width, and 25 Å dilation.

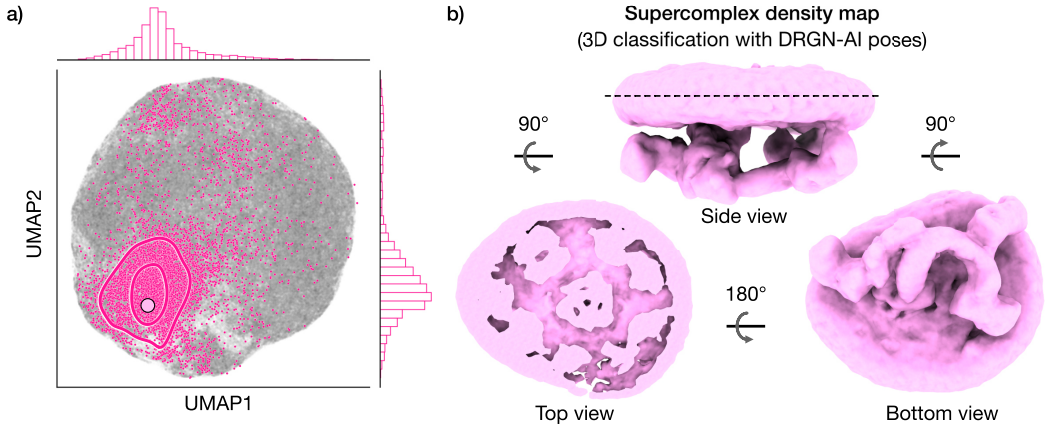


Fig. S7. **3D classification with cryoSPARC [55] on the ankyrin dataset**, using the poses estimated by DRGN-AI and 80 classes. See **Section 5.4.4** for experimental details. **a)** UMAP of the predicated conformations with DRGN-AI (grey). The large pink circle indicates the conformation of the supercomplex shown in **Figure 5**. The smaller dots indicate the conformations of the particles belonging to the supercomplex class, according to cryoSPARC (8,523 particles). These conformations are approximated with a Gaussian KDE shown with pink contours. **b)** Three views of the supercomplex state reconstructed with cryoSPARC (one class out of 80).

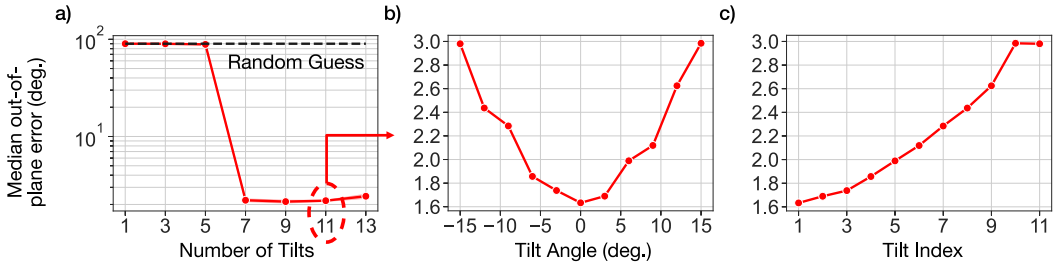


Fig. S8. **Pose accuracy of subtomogram reconstruction on the mycoplasma pneumoniae 70S ribosome** (EMPIAR-10499 [64]). Median out-of-plane angular error vs. number of tilts used for reconstruction (**a**), vs. tilt angle (**b**) and vs. tilt index (**c**). **b** and **c** were obtained using 11 tilts per particle. Experiments were run on the filtered subset of 16,655 particles from Rangan et al. [56]. This dataset was captured with a dose-symmetric tilting scheme.

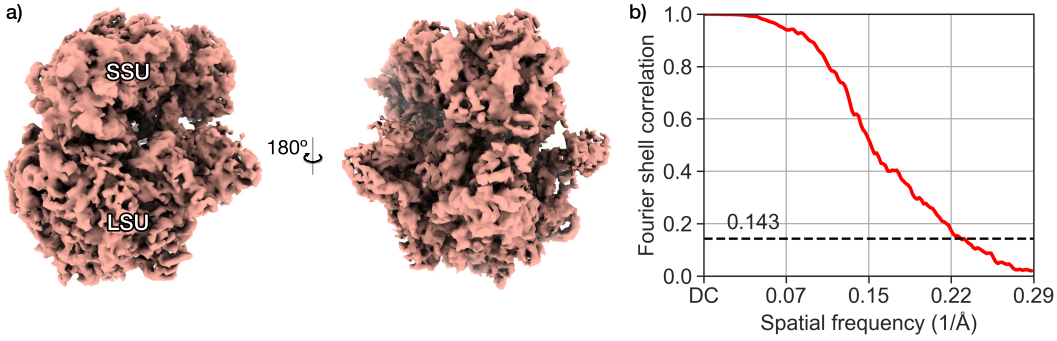


Fig. S9. **Homogeneous reconstruction of the "A, P state" of the mycoplasma pneumoniae 70S ribsome** (EMPIAR-10499 [64], 294x294, 1.70 Å/pix.) from 14,857 selected particles (11 tilts per particle, 163,427 subtilt images total). **a)** Reconstructed density map. **b)** Half-map FSC with soft mask at half of maximum value, 15 Å cosine edge width, and 25 Å dilation.

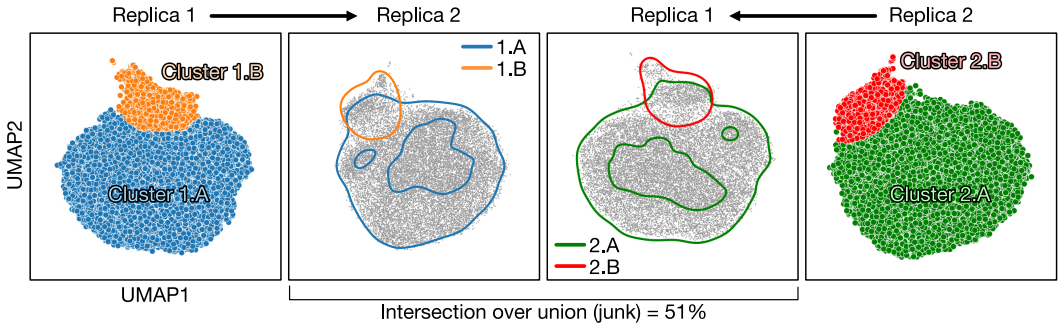


Fig. S10. **Consistent classification of junk in the mycoplasma pneumoniae 70S ribsome dataset.** We run two independent experiments (replica 1 and 2) and manually select an outlying cluster in the UMAP plots (1.B and 2.B). Each particle is labeled according to the cluster it appears in after reconstruction 1 (resp. 2), and their distributions are shown in the UMAP plot of reconstruction 2 (resp. 1). Intersection over union would be 4.8% if the two classifications were uncorrelated.

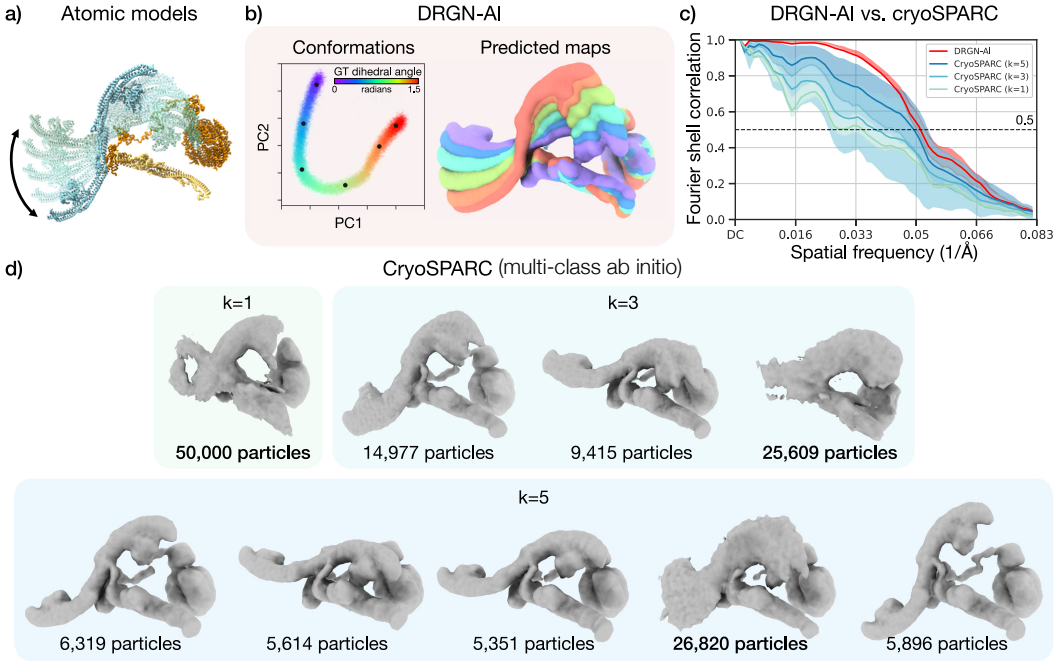


Fig. S11. **Results of DRGN-AI on a synthetic datatset with strong conformational heterogeneity and comparison to cryoSPARC (multi-class *ab initio*). a)** The dataset simulates strong conformational heterogeneity (rotation along a dihedral angle). We show 5 models separated by a 0.3 rad. increment each (full description in **Section 5.7.9**). **b)** Output of DRGN-AI. PCA on the predicted conformations on the left (the hue represents, for each image, the true dihedral angle) and associated density maps on the right (sampled on black dots). **c)** Comparison between DRGN-AI and cryoSPARC multiclass *ab initio*. We show the mean and inter-quartile range of the per-image FSC computed on 50 images uniformly sampling the range of motion (see **Section 5.4.1**). **d)** Results obtained with cryoSPARC multi-class *ab initio* (default parameters), using different numbers of classes. We indicate the number of particles associated to each class and bold the highest number.

12 SUPPLEMENTARY TABLES

Dataset	EMPIAR	SPA/ET	D	Å/pix.	Particles	Tilts	z_{dim}	Epochs	Time	GPUs
Pre-catalytic spliceosome [52]	10180	SPA	128	4.25	327,490	–	4	100	04:41 h	4xA100
Assembling ribosome [13]	10076	SPA	256	1.64	131,899	–	8	104	11:36 h	4xA100
Spike protein [68]	N/A	SPA	128	3.28	369,429	–	4	100	05:19 h	4xA100
DSL1/SNARE (heterogeneous) [11]	11846	SPA	128	3.47	214,511	–	4	100	03:43 h	4xA100
DSL1/SNARE (heterogeneous) [11]	11846	SPA	128	3.47	75,854	–	4	100	02:11 h	4xA100
V-ATPase [60]	10874	SPA	128	3.97	267,216	–	4	101	08:40 h	4xA100
V-ATPase [60]	10874	SPA	128	3.97	177,481	–	4	102	07:05 h	4xA100
Ankyrin [66]	11043	SPA	128	2.92	710,437	–	4	720	48:04 h	4xA100
70S Ribosome [64]	10499	ET	128	3.90	18,466	11	16	108	46:02 h	4xA100
Synthetic 80S ribosome	N/A	SPA	128	3.77	100,000	–	–	30	01:25 h	4xA100
Synthetic 1D rotation [75]	N/A	SPA	128	6.00	50,000	–	8	100	03:02 h	4xA100

Table S1. **Datasets and DRGN-AI training settings.** D refers to the box size in pixels.

Dataset	EMPIAR	Orig. box size	Orig. pixel size (Å)	Orig. number of particles	Best EMDB resolution (Å)
Pre-catalytic spliceosome [52]	10180	320	1.699	327,490	3.6
Assembling ribosome [13]	10076	320	1.31	131,899	3.7
DSL1/SNARE (homogeneous) [11]	11846	400	1.114	469,193	4.5
Ankyrin [66]	11043	450	0.83	710,437	2.4

Table S2. **Original dataset parameters.**

13 SUPPLEMENTARY VIDEOS

Supplementary Video 1: Additional DRGN-AI results from the filtered DSL1/SNARE dataset (**Figure 3b**). Density maps are reconstructed following a random trajectory in the latent conformation space, generated by simulating a Brownian motion in the PC1/2 space of the latent embeddings, biased towards the origin with a harmonic potential.

Supplementary Video 2: Additional DRGN-AI results from the ankyrin-1 dataset (**Figure 5**). Density maps are reconstructed along a trajectory generated by linear interpolation between the k-means cluster centroids (k=20) of the latent embeddings.

Supplementary Video 3: Density map of a homogeneous reconstruction of the AP state of the 70S ribosome shown in **Figure S9**.