Revealing biomolecular structure and motion with neural *ab initio* cryo-EM reconstruction

Axel Levy[1,2], Michal Grzadkowski[3], Frédéric Poitevin[2], Francesca Vallese[4,5,6], Oliver Biggs Clarke[4,5,6], 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) can access the intrinsic heterogeneity of these complexes and is therefore a key tool 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 a method, DRGN-AI, for *ab initio* heterogeneous cryo-EM reconstruction. With a two-step hybrid approach combining search and gradient-based optimization, DRGN-AI can reconstruct dynamic protein complexes from scratch without input poses or initial models. Using DRGN-AI, we reconstruct the compositional and conformational variability contained in a variety of benchmark datasets, process an unfiltered dataset of the DSL1/SNARE complex fully *ab initio*, and reveal a new "supercomplex" state of the human erythrocyte ankyrin-1 complex. With this expressive and scalable model for structure determination, we hope to unlock the full potential of cryo-EM as a high-throughput tool for structural biology and discovery.

1 INTRODUCTION

Single particle cryo-electron microscopy (cryo-EM) has transformed structural biology with its ability to visualize biomolecular structure at atomic resolution [38, 58] and uncover the conformations of dynamic molecular machines [4, 5, 11, 37, 44, 59]. 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, 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 [4, 44, 45, 59, 60] (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.

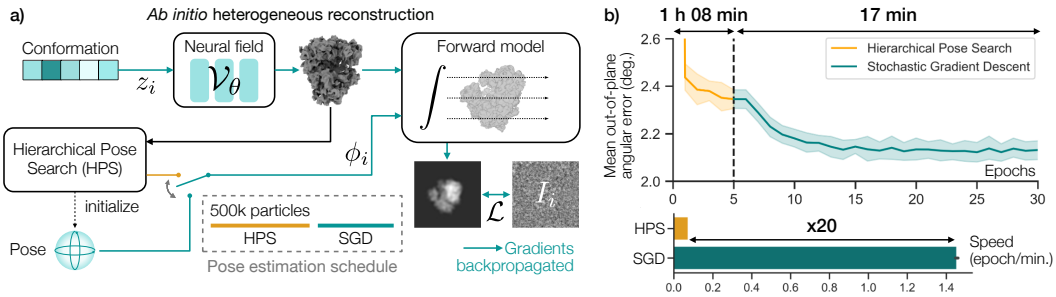


Fig. 1. **The DRGN-AI method for *ab initio* heterogeneous reconstruction. a)** Architecture overview. Objects in blue are optimized to minimize the loss $\mathcal{L}$. For each image, a synthetic image is generated by a differentiable forward model. Poses ϕ_i are first estimated using hierarchical pose search (HPS) and then refined by stochastic gradient descent (SGD). Conformations z_i are independently optimized by SGD. **b)** 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). HPS is used during the first five epochs, to provide a good initialization for SGD. Gradient descent is 20x faster and leads to more accurate poses since it is not limited by the resolution of the search grid.

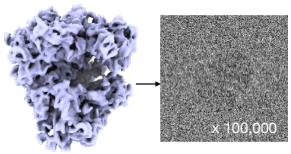
structures. This simplified model of heterogeneity often does not align with the true distribution of structures and, despite time-consuming iterative processing pipelines, still results in missed conformational states.

In this work, we show that cryo-EM imaging data can be processed end-to-end to reveal heterogeneous biomolecular structures in a single shot. 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 single particle images. DRGN-AI relies on a new pose estimation strategy that switches from a slow search that avoids local minima to stochastic gradient descent. We show that DRGN-AI can reconstruct the conformational and compositional landscape of the assembling bacterial ribosome, the precatalytic spliceosome, and the SARS CoV-2 spike protein in a single round of training without any manual intervention or prior pose estimation procedure. DRGN-AI is designed to scale to large, heterogeneous datasets, and we reconstruct a previously hypothesized "supercomplex" state of the human erythrocyte ankyrin-1 complex [52] that was missed in earlier processing approaches. Unlike existing state-of-the-art methods, we show that DRGN-AI is robust to the presence of junk or outlier images and can resolve the structure of the DSL1/SNARE complex. 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-step strategy

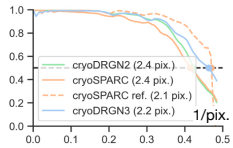
In single particle cryo-EM, each particle can be associated with a set of image-specific unknown parameters, called *latent variables*. In the case of *ab initio* heterogeneous reconstruction, this set includes the *pose* ϕ_i – the orientation of the particle with respect to the microscope and 2D translation 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



a) Ground truth b) Synthetic particle

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

c) Accuracy comparison



d) Fourier Shell Correlation

Fig. 2. **DRGN-AI *ab initio* homogeneous reconstruction of synthetic datasets. a,b)** Ground truth density map of the 80S ribosome and example particle image (100,000 particles, 128 × 128, 3.77 Å/pix). **c)** Pose error for DRGN-AI, DRGN-AI using HPS only and cryoSPARC [46]. **d)** FSC curves between the reconstructed density map and the ground truth density map and the corresponding resolution (FSC 0.5 criterion).

$z_i \in \mathbb{R}^d$, $\mathcal{V}_\theta[z_i] : \mathbb{R}^3 \to \mathbb{R}$ represents a single electron scattering potential (also called density map). We parameterize the family Ω with a neural network (Section 5).

DRGN-AI relies on a new strategy for inferring these latent variables (**Fig. 1a**). Unlike cryoDRGN [59] and cryoFIRE [31] 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 [2]. 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 ([9] and **Fig. S2**). 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 1b**, optimization is 20x faster with SGD than with HPS. The pose accuracy being no longer bounded by the resolution of a grid, **Figure 1b** shows that SGD leads to a decrease of the mean out-of-plane angular error, whose effect on resolution we quantify in **Fig S1**. 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 first evaluate this pose estimation strategy for homogeneous reconstruction in **Figure 2**. We use a synthetic dataset of 100,000 images of the 80S ribosome (example particles shown in **Fig. S3**), where ground truth poses are known, and perform *ab initio* homogeneous reconstructions with cryoSPARC [46] and DRGN-AI. DRGN-AI leads to pose error that is on par with cryoSPARC (*ab initio* + homogeneous refinement). Switching from HPS to SGD leads to a 10 to 30% decrease in pose error.

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 [43]), the assembling 50S large ribosomal subunit (EMPIAR-10076 [8]) and the SARS-CoV-2 spike protein [54]. While previous analyses of these datasets typically operate on filtered versions (with less junk images) and use consensus

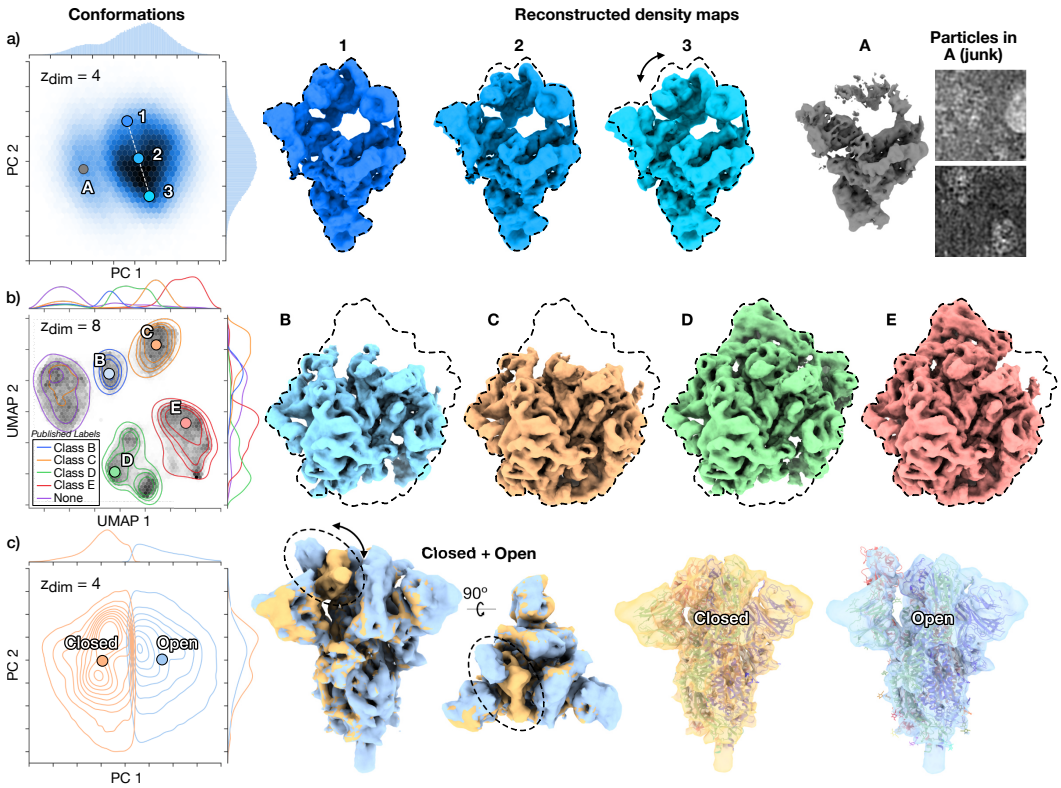


Fig. 3. **Single shot DRGN-AI *ab initio* heterogeneous reconstruction results on unfiltered benchmark datasets. a)** Pre-catalytic spliceosome dataset (EMPIAR-10180 [43], 2.6 MDa, 128x128, 4.25 Å/pix., 327,490 particles). A histogram of the two first principal components of the predicted conformations is shown. Points A, 1 and 3 are obtained by k-means clustering ($k = 5$) and 2 is a linear interpolation between 1 and 3. Density maps 1-3 show a continuous bending motion of the molecule. An outlier cluster resembling broken or junk particles and two randomly selected particles images from that cluster. Dashed lines indicate outlines of the extended spliceosome. **b)** Assembling bacterial large ribosomal subunit dataset (EMPIAR-10076 [8], approx. 2.1 to 3.3 MDa, 256x256, 1.64 Å/pix., 131,899 particles). Conformations are embedded in a two-dimensional space with UMAP. The distribution of the published labels is shown in this space with a Gaussian kernel density estimation (KDE) and four of the maps obtained by k-means clustering ($k = 6$). 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. **c)** SARS-CoV-2 spike protein dataset [54] (438 kDa, 128x128, 3.28 Å/pix., 369,429 particles). Predicted conformations are clustered with k-means and the distribution of two first principal components of each cluster is approximated with Gaussian KDE. Two sampled density maps correspond to the closed and open states of the receptor binding domain (RBD) with docked atomic models (PDB:6VXX, PDB:6VYB).

poses [4, 13, 20, 25, 33, 37, 44, 59, 62], we perform single shot reconstruction, i.e., a single round of training, on the unfiltered deposited dataset, fully *ab initio*.

In **Figure 3**, 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) [35]) is shown on the left, and representative density maps are shown on the right. On the EMPIAR-10180 dataset [43], 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 (**Fig. 3a**). DRGN-AI recovers the compositional heterogeneity in the assembly states of the bacterial large ribosomal subunit in the

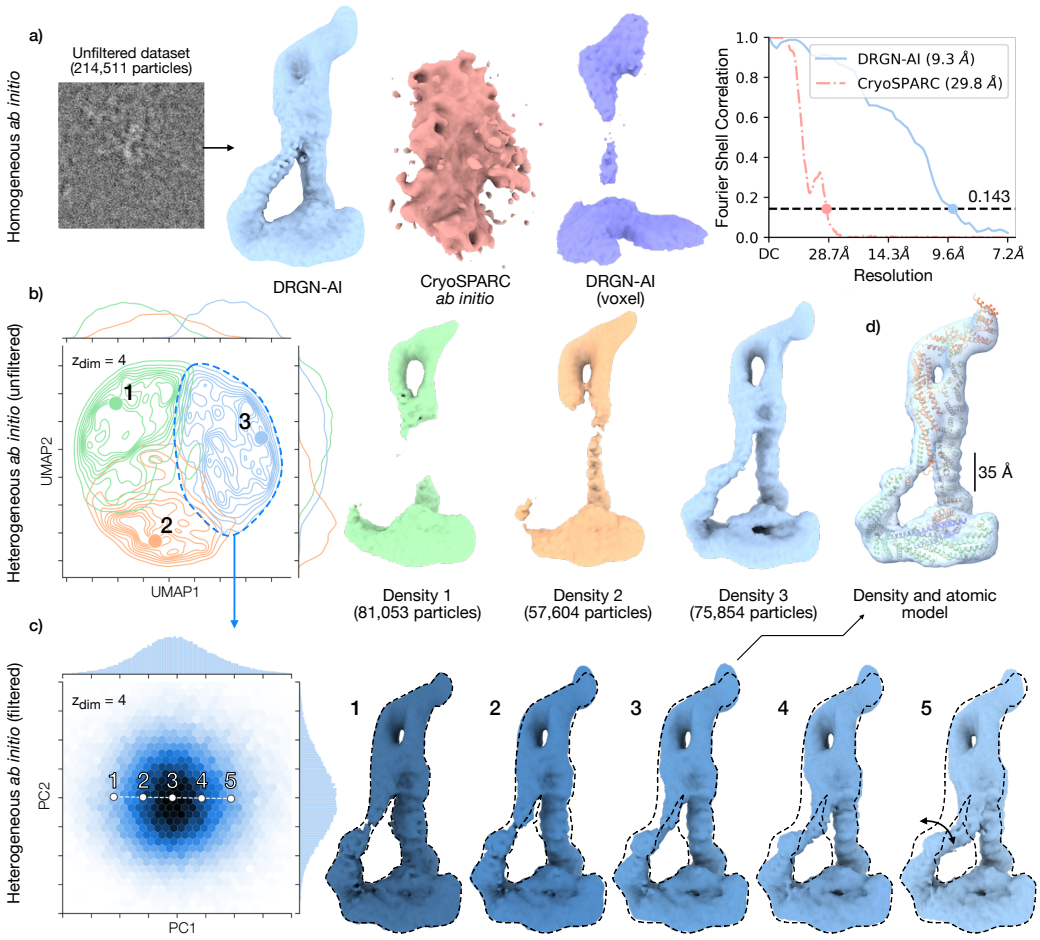


Fig. 4. **DRGN-AI *ab initio* reconstruction of a highly heterogeneous dataset of the DSL1/SNARE complex [6].** **a)** Homogeneous *ab initio* reconstruction of the full dataset (214,511 particles, 128x128, 3.47 Å/pix.) and resulting density map from DRGN-AI, cryoSPARC [46] and DRGN-AI after replacing the neural representation with an explicit voxel-based representation. DRGN-AI's neural representation converges to the expected structure despite the presence of junk particles. **b)** Heterogeneous *ab initio* reconstruction with DRGN-AI. The conformations are clustered with k-means ($k = 3$) and visualized with UMAP [35]. Particles associated with the blue cluster are selected. **c)** Heterogeneous *ab initio* reconstruction of the selected images (blue cluster) and PCA of the learned conformations. Moving along the first principal component reveals the flexibility of the complex. Additional density maps are shown in **Supplementary Video 1**. **d)** Density map from (**c**) with docked atomic model (PDB: 8EKI).

EMPIAR-10076 dataset [8], while also grouping previously unclassified and discarded particles in the latent space (**Fig. 3b**). Finally, on a much smaller protein complex of the SARS-CoV-2 spike protein (< 500 kDa) [54], 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 (**Fig. 3c**).

2.3 DRGN-AI reconstructs the DSL1/SNARE complex and its flexibility *ab initio*

We next sought to characterize DRGN-AI on a dataset where a majority fraction of particles are outliers that do not contribute to the final structure. We run our method on a dataset of the DSL1/SNARE complex (EMPIAR-11846) [6], 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 structure from 49,947 particles.

On the full dataset, we find that DRGN-AI converges fully *ab initio* to a low resolution structure of the DLS1/SNARE complex on the full dataset. Similar to the original processing, cryoSPARC's *ab initio* homogeneous reconstruction does not converge to a meaningful structure (**Fig. 4a**). To test the importance of the neural representation, we perform an ablation study replacing $\mathcal{V}_\theta$ with an explicit representation (a voxel array, Section 5). Our explicit baseline fails to converge and we conclude that, without any frequency marching strategies, the neural representation is more suited for 3D reconstruction with unknown poses. As hypothesised in Levy et al. [31], this could come from inductive biases of neural representation, which tend to fit low-frequency information before high-frequency details. We acknowledge that a scheduled coarse-to-fine optimization strategy could enable the explicit representation to converge, but a single voxel array would remain irrelevant for heterogeneous reconstruction.

We next run DRGN-AI *ab initio* heterogeneous reconstruction on the full 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 (**Fig. 4b**). By visual inspection, we select the cluster 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 (**Fig. 4c**).

2.4 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 next process a dataset of the human erythrocyte ankyrin-1 complex (EMPIAR-11043 [52]), 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 [52].

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 (**Fig. 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 (**Fig. 5b**). Increasing the number of sampled density maps to 100 (via k-means centroids) revealed a structure of the "supercomplex" state (**Fig. 5a**), which was hypothesized but not shown in Vallese et al. [52]. 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. [52]. We validated

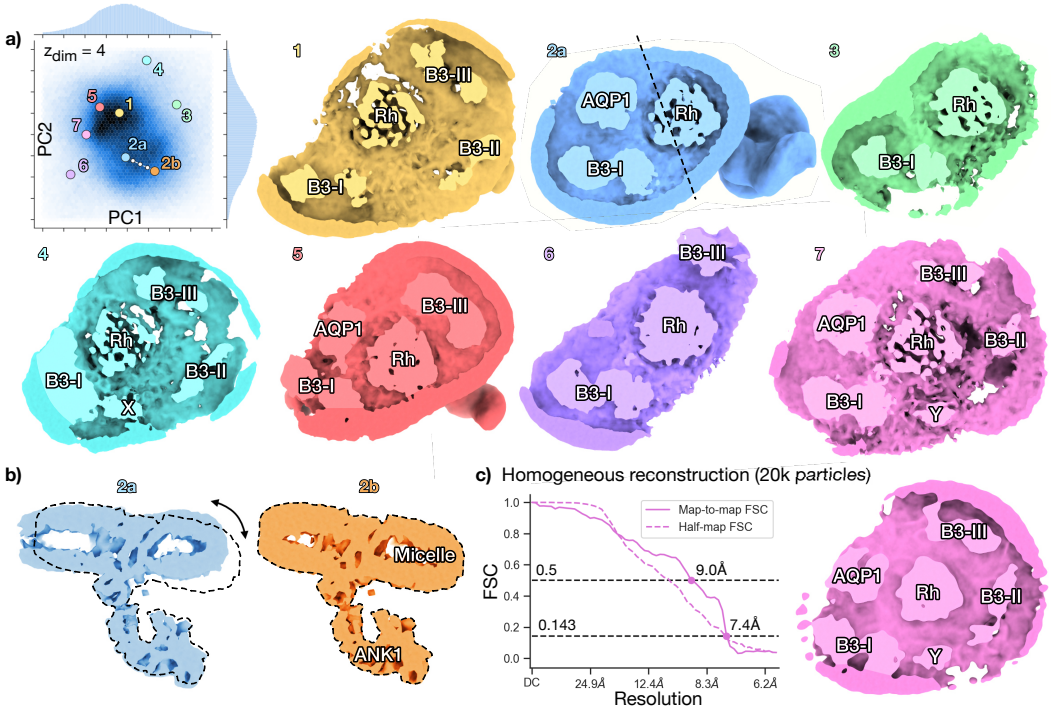


Fig. 5. **Single shot DRGN-AI *ab initio* reconstruction of the human erythrocyte ankyrin-1 complex and identification of a new "supercomplex" state. a)** Heterogeneous *ab initio* reconstruction of the full dataset (EMPIAR-11043 [52], 710,437 particles, 128x128, 2.92 Å/pix.). PCA on the latent space of predicted conformations and sampled density maps from their corresponding location in latent space. Conformations can be distinguished according to micelle composition. The six classes from Vallese et al. [52] (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 [52]. Additional density maps shown in **Supplementary Video 2**. **b)** In addition to compositional heterogeneity, DRGN-AI reveals conformational heterogeneity. Sampling maps 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 [46] on 20k particles (20k closest maps from F) with poses from DRGN-AI. Map-to-map and half-map FSC curves.

the presence of this new structure by running an independent, homogeneous reconstruction on a subset of 20,000 particles (**Fig. 5c**, Section 5).

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). In parallel, we were able to identify the supercomplex by 3D classification without alignments using 80 classes after extensive tuning of classification parameters (Section 5). As DRGN-AI estimates poses, we performed 3D classification with DRGN-AI poses and identified again the supercomplex (**Fig. S4**), providing additional validation for this new structure.

3 RELATED WORK

Modeling continuous motion in cryo-EM. Recent methods for heterogeneous cryo-EM reconstruction have explored the possibility of using linear [12, 44] and nonlinear [7, 11, 15, 29, 34, 36] mappings between

latents and 3D structures. Nonlinear methods can be further categorized depending on the support they use for representing density maps: neural networks [59, 60], 3D Zernike polynomials [19], Gaussian mixture models [4, 61] or flow fields [45]. Normal mode-based approaches [16, 17, 23, 39] 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) large conformational motions 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 [10, 48, 49], stochastic gradient-based optimization techniques were later introduced to allow *ab initio* reconstruction and released as part of the cryoSPARC software [3, 46]. 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 [21, 50], they are fundamentally limited to modeling discrete heterogeneity. CryoDRGN2 [62] and cryoFIRE [31] 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, such as for smaller complexes, where the encoder may memorize the data [9] (**Fig. S2**). 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 compositional heterogeneity and conformational motions of large, unfiltered datasets.

4 DISCUSSION

In this work, we demonstrate new capabilities for recovering biomolecular structural variability from diverse, challenging single-particle cryo-EM 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. For example, we identified a new "supercomplex" state of the human erythrocyte ankyrin-1 complex. We additionally recovered the structure and motion 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 [63].

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 [27], we empirically observe that similar density maps tend to be close to each other in latent space. As the latent space cannot be meaningfully interpreted in terms of energy landscape and conformational variation [28], we found that visual inspection of the density maps sampled from k-means centroids or linear trajectories in latent space remains an effective way of analyzing the structural heterogeneity. We also propose 3D classification with poses from DRGN-AI as a complementary approach for exploring the learned distribution and assigning states. Going forward, imbuing a thermodynamic interpretation to distances in latent space is one of the 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, 24] and design protein structures *de novo* [22, 55], 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 to increasing the throughput of cryo-EM, 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. In particular, we envision building upon recent deep learning advances for heterogeneity analysis *in situ* [47] to address outstanding challenges in *ab initio* reconstruction of biomolecules from the cellular millieu [40]. Lastly, new machine learning-based reconstruction methods that enable the study of increasingly complex samples can be further integrated with other AI methods in structural biology to provide major advances in our understanding and engineering of biological molecules.

5 METHODS

5.1 Computational Details

5.1.1 Implicit Representation. The "hypervolume" [29] $\mathcal{V}_\theta : \mathbb{R}^{3\times d} \to \mathbb{R}$ is stored in a neural network. Conditioned on a conformation $z_i \in \mathbb{R}^d$, $\mathcal{V}_\theta[z_i]$ 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 [51] (64 base frequencies are randomly sampled from a 3D Gaussian distribution of standard deviation 0.5). The neural network follows the same architecture as in cryoDRGN [59]: it contains 3 hidden residual layers [18] of size 256 with ReLU non-linearities, except at the last layer.

5.1.2 Image Formation Model. In single particle imaging, an image I_i is considered to be a random realization following [48, 53]:

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

where C_i is the (discretized) Point Spread Function, $*$ the convolution operator and η_i additive white Gaussian noise. The operator P_ϕ represents an orthographic projection of the map $\mathcal{V}_\theta[z_i]$ according to the pose $\phi = (R, \mathbf{t}) \in \mathrm{SO}(3) \times \mathbb{R}^2$ and 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 $\{(x_m, y_n)\}_{m,n\in\{1,\ldots,D\}}$,

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

5.1.3 Objective Function. DRGN-AI aims at reproducing this model and minimizing the reconstruction error (negative log-likelihood of the observations):

$$\mathcal{L}(\theta, \{\phi_i\}, \{z_i\}) = \sum_{i=1}^{N} \|I_i - C_i * P_{\phi_i} \mathcal{V}_\theta[z_i]\|_2^2 \tag{4}$$

5.1.4 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 $\{(k_m^x, k_n^y)\}_{m,n \in \{1,\ldots,D\}}$, 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. Finally, the loss $\mathcal{L}$ (4) is computed in Hartley space, which is equivalent to computing it in real space by Parseval's theorem.

5.1.5 Hierarchical Pose Search. The pose search strategy is similar to [62]. The "reprojection error" is defined as

$$\|I_i - C_i * P_{\phi_i} \mathcal{V}_\theta[z_i]\|_2^2. \tag{8}$$

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 [57], product of the Healpix [14] grid on the 2-sphere and a regular grid on the circle.

5.1.6 Initialization. The weights of the neural network are randomly initialized with the default initialization scheme for linear layers in PyTorch [41]. The neural network is then trained with 10,000 images by default using random poses and conformations ("pretraining phase"). The pose table is initialized with the last poses estimated by the hierarchical pose search, once 500,000 images by default have been processed (with a minimum of 2 epochs of pose search). The conformation table is initialized randomly from a d-dimensional Gaussian distribution of standard deviation 0.1.

5.1.7 Explicit Representation. We replace the neural representation with a voxel array to store the Hartley transform of the density map in Figure 4 (DRGN-AI + voxel). This "explicit" representation is implemented with a differentiable array of size $128 \times 128 \times 128$.

5.1.8 Optimization. All the optimizable parameters are optimized with the Adam optimizer [26]. 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 [32] instead of the Adam optimizer for poses, but this produced unstable results in our early experimentation.

5.1.9 Hardware. All the experiments are ran on 4 A100s Nvidia GPUs using data parallelization.

5.1.10 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 Principal Component Analysis (PCA) and UMAP [35]. 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.

5.1.11 Validation in cryoSPARC. We used cryoSPARC version 4.4 [46] 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 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, poses provided by DRGN-AI, a target resolution of 12 Å, 40 O-EM epochs, 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 is provided in **Figure S4)**.

5.1.12 Visualization. Molecular graphics and analyses performed with UCSF ChimeraX [42], 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.

5.2 Data

A summary of all the datasets analyzed in this study and DRGN-AI training settings are found in Table S1. All datasets are downloaded from EMPIAR as extracted particle stacks.

5.2.1 Synthetic 80S Ribosome. We generate a synthetic dataset of the 80S ribosome using the same procedure as Levy et al. [30] by following the image formation model (Equation 1). The electron scattering potential was simulated in ChimeraX [42] at 6.0 Å resolution from a combination of two atomic models: the small subunit (PDB:3J7A) and the large subunit (PDB:3J79) [56]. 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. [30]). Zero-mean white Gaussian noise is added with a standard deviation of 0.5. Example images are shown in Fig. S3.

5.2.2 Pre-catalytic spliceosome. The EMPIAR-10180 [43] 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).

5.2.3 Assembling 50S ribosome. The EMPIAR-10076 [8] 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).

5.2.4 SARS-CoV-2 spike protein. The dataset from Walls et al. [54] 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.

5.2.5 DSL1/SNARE complex. The EMPIAR-11846 [6]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 [59].

5.2.6 Ankyrin-1 complex. The EMPIAR-11043 [52] dataset was downsampled to 128×128 from theoriginal 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).

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), ankyrin-1 complex (EMPIAR-11043), DLS1/SNARE complex (EMPIAR-11846). 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, Ryan Feathers, 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 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. 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 authors acknowledge Princeton University startup funds for support of this work.

9 AUTHOR CONTRIBUTIONS

A.L., F.P., G.W., and E.Z. conceived of the work; A.L. and E.Z. implemented the methods and performed the computational experiments; A.L., M.G., and E.Z. developed the software package; F.V. and O.B.C. analyzed the results on the ankyrin dataset; A.L. 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] Piotr Bojanowski, Armand Joulin, David Lopez-Paz, and Arthur Szlam. 2017. Optimizing the latent space of generative networks. *arXiv preprint arXiv:1707.05776* (2017).

[3] 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]

[4] 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.

[5] 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.

[6] Kevin A DAmico, Abigail E Stanton, Jaden D Shirkey, Sophie M Travis, Philip D Jeffrey, and Frederick M Hughson. 2024. Structure of a membrane tethering complex incorporating multiple SNAREs. *Nat. Struct. Mol. Biol.* 31, 2 (Feb. 2024), 246–254.

[7] 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.

[8] 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.

[9] 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).

[10] 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.

[11] 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.

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

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

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] 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.

[21] 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.

[22] 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.

[23] 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.

[24] 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.

[25] 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.

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

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

[28] 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.

[29] 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).

[30] 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.

[31] 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.

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

[33] 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.

[34] 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.

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

[36] 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.

[37] 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.

[38] 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.

[39] 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).

[40] Eva Nogales and Julia Mahamid. 2024. Bridging structural and cell biology with cryo-electron microscopy. *Nature* 628, 8006 (April 2024), 47–56.

[41] 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).

[42] 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.

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

[44] 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.

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

[46] 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.

[47] 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.

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

[49] 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.

[50] 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.

[51] 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.

[52] 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.

[53] 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.

[54] 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.

[55] 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.

[56] 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.

[57] 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.

[58] 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.

[59] 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.

[60] 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)*.

[61] 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*.

[62] 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.

[63] 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

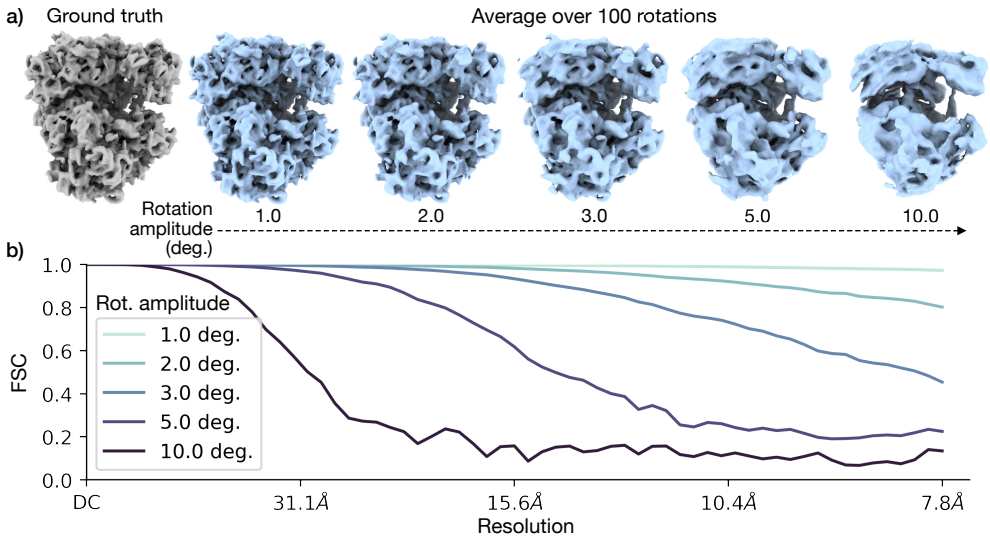


Fig. S1. **Influence of pose error on resolution.** To illustrate the impact of inaccurate poses on the reconstructed density map, 100 randomly rotated density maps of the 80S ribosome ($128 \times 128 \times 128$, 3.77 Å/pix.) are averaged, using a rotation amplitude between 1 and 10 degrees. **a)** Ground truth and averaged density maps for different rotation amplitudes. **b)** Fourier shell correlation for different rotation amplitudes.

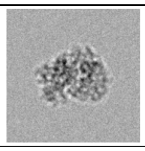
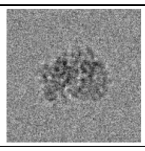
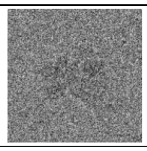
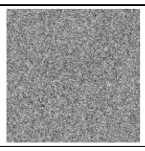
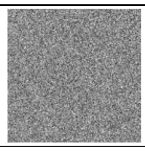
Particles							
Signal-to-noise ratio			0 dB	-5 dB	-10 dB	-15 dB	-20 dB
Mean out-of-plane angular error (deg.)	HPS	50k particles	0.60	0.62	0.77	41	90
		500k particles	0.48	0.51	0.78	41	90
	Train	50k particles	0.74	1.1	2.2	73	90
		500k particles	0.46	0.60	1.7	51	90
	Test	50k particles	0.99	2.42	12	85	90
		500k particles	0.51	1.2	6.2	81	90

Fig. S2. **Limits of amortized inference in high-noise regimes.** An encoder mapping particle images to poses is trained supervisedly on synthetic cryo-EM datasets of the 80S ribosome of different sizes with different noise levels. Each dataset is split into a training set (50k or 500k particles) and a test set (10k particles). To test whether an encoder can be used to amortize pose estimation in an *ab initio* reconstruction setup, the training poses are obtained by running DRGN-AI on the training datasets. We compare the training and testing errors to the results DRGN-AI, which uses hierarchical pose search (HPS). As the size of the training dataset increases, the testing error decreases but remains significantly larger than the HPS error for datasets with realistic noise levels (SNR ≤ -10 dB). In highly noisy datasets, the encoder tends to memorize images, preventing true amortization.

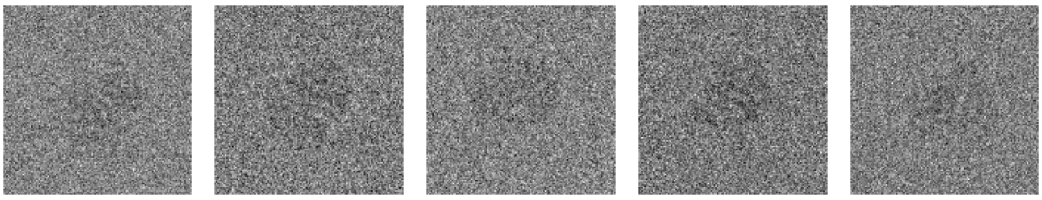


Fig. S3. **Synthetic 80S ribosome dataset.** Five images of our synthetic dataset of the 80S ribosome.

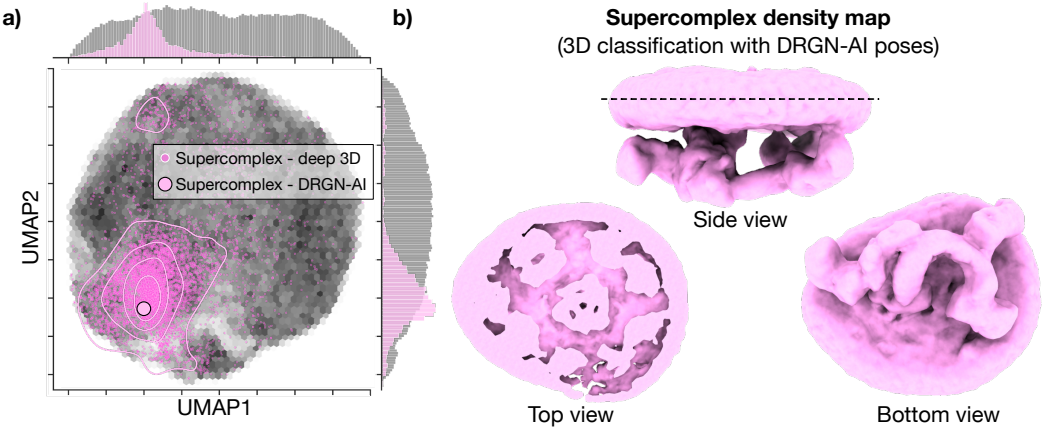


Fig. S4. **3D classification with cryoSPARC [46] on the ankyrin dataset.** 3D classification with cryoSPARC on the ankyrin dataset, using the poses estimated by DRGN-AI and 80 classes. **(a)** UMAP of the predicated conformations with DRGN-AI (grey). The large pink circle indicates the conformation of the supercomplex shown in Fig. 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).

Dataset	EMPIAR	D	Å/pix.	Particles	z_{dim}	Epochs	Time	GPUs
Pre-catalytic spliceosome[43]	10180	128	4.25	327,490	4	100	04:41 h	4xA100
Assembling ribosome[8]	10076	256	1.64	131,899	8	104	11:36 h	4xA100
Spike protein[54]	N/A	128	3.28	369,429	4	100	05:19 h	4xA100
DSL1/SNARE (homogeneous)[6]	11846	128	3.47	214,511	–	103	04:19 h	4xA100
DSL1/SNARE (heterogeneous)[6]	11846	128	3.47	214,511	4	100	03:43 h	4xA100
DSL1/SNARE (heterogeneous)[6]	11846	128	3.47	75,854	4	100	02:11 h	4xA100
Ankyrin[52]	11043	128	2.92	710,437	4	720	48:04 h	4xA100
Synthetic 80S ribosome	N/A	128	3.77	100,000	–	30	01:25 h	4xA100

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

11 SUPPLEMENTARY VIDEOS

Supplementary Video 1: Additional DRGN-AI results from the filtered DSL1/SNARE dataset (Fig. 4c). 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 (Fig. 5). Density maps are reconstructed along a trajectory generated by linear interpolation between the k-means cluster centroids (k=20) of the latent embeddings.