Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2025.12.30.000075. This version posted November 3, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

CryoDECO: disentangling compositional and conformational heterogeneity in cryo-EM with foundation-model priors

Yang Yan[1,2,6], Yanwanyu Xi[2,6], Shiqi Fan[1,2], Yifei Wang[3,4,5], Ziyun Tang[3,4,5], Fajie Yuan[2,*], and Huaizong Shen[3,4,5,*,†]

Submitted as a Brief Communication

[1]Zhejiang University, Hangzhou, Zhejiang, China

[2]School of Engineering, Westlake University, Hangzhou, Zhejiang, China

[3]Zhejiang Key Laboratory of Structural Biology, School of Life Sciences, Westlake University; Hangzhou, Zhejiang, China

[4]Westlake Laboratory of Life Sciences and Biomedicine, Hangzhou, Zhejiang, China

[5]Institute of Biology, Westlake Institute for Advanced Study, Hangzhou, Zhejiang, China

[6]These authors contributed equally to this work.

[*]Correspondence: F. Yuan (yuanfajie@westlake.edu.cn), H. Shen (shenhuaizong@smart.org.cn).

[†]Present address: Institute of Bio-Architecture and Bio-Interactions (IBABI), Shenzhen Medical Academy of Research and Translation (SMART), Shenzhen 518107, Guangdong, China.

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2025.12.30.000075. This version posted November 3, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

Abstract

Cryo-electron microscopy has transformed macromolecular structure determination, but heterogeneous biological samples contain multiple complexes and conformational states that remain difficult to resolve. This difficulty reflects a circular dependency among particle separation, pose estimation and structure reconstruction. We developed CryoDECO, a prior-guided framework that enters this coupled inference problem from structural representations learned from large-scale cryo-EM data rather than from random initialization, and that reduces manual tuning. CryoDECO separates simulated mixtures of up to 100 structures, resolves experimental and native-derived mixtures into populations suitable for high-resolution refinement, and captures continuous motion together with local density differences within a shared trajectory.

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2025.12.30.000075. This version posted November 3, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

Cryo-electron microscopy (cryo-EM) has transformed structural biology, especially since the resolution revolution[1–4]. The standard single-particle analysis (SPA) workflow reconstructs three-dimensional density maps by averaging millions of two-dimensional projection images according to the central slice theorem[5], assuming that all projections are different views of a structurally identical or near-identical object[2,6,7]. That assumption holds for rigid, highly purified samples, but biological reality is defined by structural heterogeneity. Heterogeneity takes two forms that are equally consequential and technically distinct: compositional complexity, in which a sample contains a mixture of different macromolecules, and conformational dynamics, the intrinsic flexibility of a single molecular machine[8,9].

Conventional pipelines treat such heterogeneity as a nuisance variable, and distinct structures or states are averaged into a blurred consensus[9]. For emerging frontiers such as native cell extracts[10–14] and time-resolved studies[15], however, heterogeneity is no longer merely a nuisance to be removed; it is also the biological signal to be resolved. Mapping the dynamic landscapes accessible to a complex[16–22] moves cryo-EM beyond the static snapshots of classical X-ray crystallography, toward resolving both discrete components and a continuum of states. The objective shifts: instead of one structure from an exhaustively purified sample, the aim is the simultaneous *in silico* reconstruction of many macromolecules directly from a crude mixture, whether a cell lysate or an environmental sample.

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2025.12.30.000075. This version posted November 3, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

Realizing this potential is hindered by a circular dependency in *ab initio* heterogeneous reconstruction[23–25]. Accurate classification requires reliable three-dimensional templates, yet template reconstruction depends on accurate classification, so class identity, structure and particle pose must be inferred together. Existing *ab initio* methods attack this coupled problem from random or weak initialization[18–21,26,27], which leaves the optimization vulnerable to local minima. Failure becomes likely when the number of states is large, when classes are weakly separated, or when particle orientations are uncertain[28], and it is amplified in high-cardinality mixtures and native samples, where an incorrect early assignment propagates through every subsequent reconstruction.

Here we introduce CryoDECO, which incorporates structural priors from a pretrained cryo-EM foundation model to mitigate this circular dependency and automates the dataset-specific decisions that would otherwise remain manual (Fig. 1a). Its image encoder, DECO-IEF, was pretrained on 425 datasets comprising approximately 134 million particle images using a Cryo-IEF-style contrastive objective[29] supplemented with structure-level supervision (Methods and Extended Data Table 1). The resulting particle representations condition neural-field density generation throughout optimization, providing structural guidance for the coupled classification–reconstruction problem. At the start of optimization, each particle representation is determined by its image and the pretrained encoder alone, independent of any current three-dimensional reconstruction. Particles from the same structure are therefore already adjacent in the learned space before a model exists, so the coupled inference is entered from a structurally organized partition rather than an arbitrary one.

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2025.12.30.000075. This version posted November 3, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

Because datasets differ in structural complexity, a fixed latent dimensionality is difficult to choose a *priori*. CryoDECO therefore introduces DimAda, an adaptive structural-*z* gate that learns which latent dimensions contribute to decoder conditioning within a specified nominal capacity budget (Fig. 1a,b). This allows decoder-visible capacity to adapt during reconstruction for both compositional and conformational analysis. For discrete mixtures, the appropriate partition count is also generally unknown, requiring manual specification or exploration in GMM-based workflows. After reconstruction, a data-driven *k*-estimator addresses this choice by combining global model-selection evidence with conditional hierarchical checks for unresolved substructure, without requiring a prespecified species count (Fig. 1c and Methods). The estimated count describes partitioning granularity in the learned representation and need not correspond one-to-one with the number of biochemical species. We show that CryoDECO separates mixtures containing up to 100 structures, resolves experimental and native-derived populations suitable for high-resolution refinement, and captures continuous conformational landscapes.

We first asked whether CryoDECO could separate mixtures of distinct structures with differing capacity requirements. We used two CryoBench benchmarks[28], Ribosembly and Tomotwin-100, which contain simulated particles from 16 and 100 different structures, respectively (Fig. 2a,d). They probe complementary forms of compositional complexity: Ribosembly tests separation of related assembly intermediates that share a common structural core, whereas Tomotwin-100 tests high-cardinality separation of structurally unrelated macromolecules. We began by examining how representation quality depended

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2025.12.30.000075. This version posted November 3, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

on fixed latent dimensionality (Extended Data Fig. 1 and Extended Data Table 2). On Ribosembly, CryoDECO reached its highest top-1 *k*-nearest-neighbor (*k*-NN) accuracy of 97.55% at $z = 32$ and still achieved 95.85% at $z = 4$. Tomotwin-100 was far more sensitive: top-1 *k*-NN accuracy rose from 43.20% at $z = 4$ to 97.95% at $z = 128$. These differing requirements motivate dataset-adaptive latent capacity.

With a nominal budget of $D_{cap} = 64$, the gate retained all $D_{act} = 64$ decoder-visible dimensions for Tomotwin-100 but only $D_{act} = 17$ for Ribosembly, consistent with the greater compositional complexity of the 100-structure mixture. Adaptive CryoDECO reached a top-1 *k*-NN accuracy of 97.23% on Ribosembly, against 86.5% for CryoDRGN-AI, and 96.21% on Tomotwin-100, against 44.0% for CryoDRGN-AI[27] (Fig. 2b,c,e,f and Extended Data Table 2). On Tomotwin-100 the gate had exhausted the nominal budget, so the budget rather than the gate was limiting: the adaptive result remained below the 97.95% reached at z = 128 in the fixed sweep. Across the two benchmarks, the adaptive setting thus came within two percentage points of the best fixed dimensionality while that dimensionality was not supplied in advance. In latent space, adaptive CryoDECO organized the 100 Tomotwin-100 structures into well-separated clusters, whereas CryoDRGN-AI produced a substantially less resolved organization (Fig. 2e). Separable representations also emerged earlier during training with CryoDECO on both datasets (Fig. 2c,f). We next evaluated whether these representations supported accurate partitioning, using the benchmark ground-truth labels for evaluation. On Tomotwin-100, in the standard setting with poses estimated rather than supplied, CryoDECO reached an adjusted Rand index (ARI) of 0.647 and an adjusted mutual information (AMI) of 0.871, compared with 0.086

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2025.12.30.000075. This version posted November 3, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

and 0.275 for CryoDRGN-AI and 0.116 and 0.374 for CryoDRGN2 (Extended Data Table 3). When ground-truth poses were supplied, which removes orientation estimation from the problem, CryoDECO reached ARI and AMI of 0.999, indicating near-complete separation of the 100 populations. These results demonstrated improved separation under uncertain orientations, while the near-complete partitioning with supplied poses highlighted the remaining challenge of pose estimation.

We next tested whether the separation observed in simulations translated to experimentally acquired particles. We applied CryoDECO to the "EM ladder" dataset (EMPIAR-11693)[30], a mixture of four purified macromolecular complexes: apoferritin (ApoF), β-galactosidase, tobacco mosaic virus (TMV) and PP7 virus-like particles. The adaptive gate retained $D_{act} = 57$ decoder-visible dimensions. Without a prespecified class number, the data-driven estimator selected an effective embedding-space partition count of four, and Gaussian mixture model (GMM)[31] partitioning recovered populations corresponding to the four complexes (Fig. 2g). The decisive case is PP7, the least abundant of the four and roughly 2% of the particles in the original processing: CryoDRGN-AI showed substantial overlap among the complexes in latent space and failed to isolate PP7, whereas CryoDECO produced a distinct PP7-associated cluster (Extended Data Fig. 2). The CryoDECO-separated populations gave homogeneous two-dimensional class averages (Extended Data Fig. 3a), and separate three-dimensional refinements in cryoSPARC produced maps at 2.49 Å for ApoF, 2.83 Å for β-galactosidase, 2.62 Å for TMV and 3.37 Å for PP7 (Fig. 2h), comparable to or better than the 2.52, 3.09, 2.85 and 3.40 Å reported for the same complexes in the original study. The informative outcome is not the change in

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2025.12.30.000075. This version posted November 3, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

resolution but that every population, including the minor species, was homogeneous enough for independent refinement with no species lost. CryoDECO thus acted as an in silico purification step.

Having resolved a defined mixture of purified complexes, we next asked whether CryoDECO could separate macromolecular populations from a more complex endogenous sample. We applied it to a preprocessed particle set from a native *Chaetomium thermophilum* lysate fraction[12]. The adaptive gate retained $D_{act} = 61$ decoder-visible dimensions, and the estimator returned a partition count of four without manually specifying *k*. GMM partitioning produced discrete particle populations in the projected embedding (Fig. 2i). Two-dimensional class averages of these populations are shown in Extended Data Fig. 3b. Downstream cryoSPARC refinement recovered maps for four assigned targets (Fig. 2j): the pyruvate dehydrogenase complex core (PDHc, 3.50 Å), fatty acid synthase (FAS, 3.96 Å), the oxoglutarate dehydrogenase complex E2 core (OGDHc, 3.83 Å) and the pre-60S ribosomal subunit (3.98 Å). After structural assignment of the four populations, the CryoDECO-assisted workflow improved on the nominal resolution reported in the original study for each target (3.84 Å, 4.47 Å, 4.38 Å and 4.52 Å, respectively). CryoDRGN-AI, by contrast, yielded mixed clusters and did not recover separate OGDHc and PDHc reconstructions in this analysis (Extended Data Fig. 2). These results extended CryoDECO separation from defined mixtures to a native lysate fraction, yielding multiple populations suitable for independent high-resolution refinement.

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2025.12.30.000075. This version posted November 3, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

Beyond compositional sorting, CryoDECO also captures continuous structural variation. The IgG-1D and IgG-RL CryoBench datasets[28] provide complementary tests of simple and complex conformational landscapes (Fig. 3a,d). Fixed-dimensional analyses showed that a compact latent space ($z = 4$) recovered the simple one-dimensional periodic motion of IgG-1D most effectively, whereas the highly flexible, randomized linker of IgG-RL benefited from an expanded space ($z = 64$) (Extended Data Fig. 1 and Extended Data Table 2). CryoDECO adapted its latent capacity to both landscapes within the same nominal budget of $D_{cap} = 64$. For IgG-1D, the gate retained $D_{act} = 4$ decoder-visible dimensions, yielding a well-organized circular manifold and reaching top-1 and top-10 *k*-NN accuracies of 45.1% and 92.3% (Fig. 3b,c and Extended Data Table 2). These results approached the best fixed $z = 4$ result and improved on the overparameterized $z = 64$ setting. For IgG-RL, the gate retained $D_{act} = 30$ dimensions and reached top-1 and top-10 accuracies of 72.6% and 96.9%, close to the best fixed $z = 64$ result (Fig. 3e,f and Extended Data Table 2). The same adaptive mechanism accommodated both simple and complex motion while approaching the best fixed-dimensional results, reducing the need for dataset-specific dimensionality selection.

We next tested conformational analysis on experimental αVβ8 integrin particles. The adaptive gate retained $D_{act} = 4$ decoder-visible dimensions, and traversal of the latent manifold recovered the continuous flexibility of the extended arm. The integrin analysis extended the conformational results beyond simulations, recovering continuous arm flexibility from experimentally acquired particles (Extended Data Fig. 4a,b and Supplementary Movie S1).

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2025.12.30.000075. This version posted November 3, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

Continuous motion can coexist with local differences between particle populations. We therefore asked whether CryoDECO could characterize local density differences within a shared conformational landscape, using the yeast U4/U6.U5 tri-snRNP complex dataset (EMPIAR-10073)[32], whose large-scale head-rotation dynamics have been described previously[20,22,33]. CryoDECO retained D_{act} = 5 decoder-visible dimensions. Using principal component analysis of the learned representations, we identified two regions of the embedding and designated them State 1 and State 2 for state-conditioned analysis (Fig. 3g,i). Decoding three representative positions sampled within each region yielded conformations showing progressive rotation of the head in both State 1 and State 2 (Fig. 3h,j and Supplementary Movies S2 and S3, respectively). In State 1, a peripheral density lobe adjacent to the head remained visible throughout the sampled trajectory. In State 2, this feature was locally depleted (the "missing blob" marked in Fig. 3j), and the depletion persisted across all three sampled conformations even though the overall head-rotation trajectory resembled that of State 1. Previous analyses of this dataset characterized the large-scale continuous motion but did not, to our knowledge, describe a head-adjacent density-depleted state[20,22,33]. CryoDECO therefore resolved a persistent local density difference between two embedding regions that followed similar global motions, extending the analysis beyond recovery of a shared trajectory.

These results support a broader role for cryo-EM in interpreting heterogeneous and native-derived specimens alongside high-resolution structure determination. CryoDECO combines pretrained structural representations with joint pose optimization and

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2025.12.30.000075. This version posted November 3, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

capacity-adaptive reconstruction to mitigate the coupling between classification and reconstruction. Adaptive latent capacity and data-driven partition estimation reduce two dataset-dependent choices within this framework. For the preprocessed *C. thermophilum* particles, computational sorting recovered populations suitable for independent refinement. Such *in silico* separation can complement biochemical sample preparation when heterogeneous populations remain after particle selection.

The tri-snRNP analysis further illustrates how local density differences can be examined alongside large-scale motion. Decoded conformations from two embedding regions showed similar head-rotation trajectories but differed in head-adjacent density. The two regions were selected from PCA representations, and their separation alone does not establish distinct biological states. The origin, structural identity and functional relevance of the density difference remain unresolved.

Performance depends on particle quality, pretraining coverage and embedding geometry, while adaptive capacity remains bounded by the nominal latent budget. For discrete mixtures, dimensionality reduction and GMM assumptions may merge structurally similar populations or subdivide a single species. The estimated k therefore describes embedding-space partitioning granularity and need not equal the number of biochemical species. The resulting populations still require independent image-based or reconstruction-based validation. In high-cardinality mixtures, pose estimation rather than representation quality remained the dominant source of error, as the near-complete partitioning obtained with supplied ground-truth poses indicates. Further improvements in pose-search efficiency

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2025.12.30.000075. This version posted November 3, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

and decoder evaluation will be needed to support larger particle datasets. Extension to cryo-electron tomography will require validation in crowded cellular volumes. By reducing the circular dependency between classification and reconstruction, CryoDECO provides a step toward scalable cryo-EM workflows that recover multiple structures and their conformational variability from increasingly complex heterogeneous and native-derived samples.

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2025.12.30.000075. This version posted November 3, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

Methods

Prior-Guided Heterogeneous Reconstruction

CryoDECO uses a prior-guided encoder–decoder architecture for heterogeneous cryo-EM reconstruction. The image encoder is initialized with the pretrained DECO-IEF model, which was trained on approximately 134 million cryo-EM particle images. For each input particle image, the encoder predicts a latent representation that is passed through the structural-z interface, with adaptive gating to produce the decoder-visible latent vector. This latent vector conditions a coordinate-based neural-field decoder that predicts Hartley-domain density values at queried 3D frequency coordinates. Particle orientations and in-plane translations are estimated through the reconstruction objective and are refined jointly with the network parameters during training. In this formulation, the pretrained encoder provides a structured representation prior, whereas the decoder and pose variables are optimized for dataset-specific reconstruction.

Pretraining of the DECO-IEF

To construct a stronger structural prior for CryoDECO, we pretrained a Cryo-IEF-style foundation encoder, DECO-IEF, on an expanded corpus of approximately 134 million cryo-EM particle images. Compared with the original Cryo-IEF model, which used a ViT-Base backbone[34], DECO-IEF uses a Vision Transformer Small (ViT-Small) backbone (12 layers, 384 embedding dimensions, 12 heads) by default.

DECO-IEF was trained from scratch using the Momentum Contrast v3 (MoCo v3) self-supervised learning framework[35], following the hyperparameter configuration of the

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2025.12.30.000075. This version posted November 3, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

original Cryo-IEF study[29]. We used the AdamW optimizer with a batch size of 2,048. The learning rate followed a cosine decay schedule with a base learning rate of 1.5×10^{-4} (scaled linearly with batch size: lr $\times$ BatchSize / 256) and a 5-epoch linear warmup. The contrastive loss temperature τ was 0.5, and the momentum update parameter for the key encoder was set to 0.99.

Preprocessing and data augmentation matched the original protocol. Input particle images were resized to 224 $\times$ 224 pixels and normalized. During training, we generated views via random cropping, colour jittering, and Gaussian blurring, encouraging representations that are robust to noise and common image-level perturbations. To improve structure-level semantic organization, we added a supervised contrastive regularization term based on structural identity, weighted by $\lambda_{\text{protein-supcon}} = 0.1$. This retraining ensures that the encoder learns a generalized latent manifold capable of separating complex structural states before fine-tuning for specific reconstruction tasks.

Model Architecture

Encoder as a Structural Compass. The encoder is a ViT-Small initialized with DECO-IEF weights. It processes 224 × 224 particle images using 14 × 14 patches and outputs an image-level latent code, which is projected into the structural-z interface before decoder conditioning.

Decoder as an AdaLN-conditioned neural field. The decoder is a residual coordinate-based neural field that predicts Hartley-domain density values. To synthesize a 2D

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2025.12.30.000075. This version posted November 3, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

projection under the Central Slice Theorem, CryoDECO first defines a 2D grid of continuous frequency coordinates, $\mathbf{x}$, on the central plane (k_z=0). For each particle, the orientation matrix R rotates these coordinates into 3D frequency space, so the rotated coordinates specify the central slice queried from the neural field. The rotated coordinates are then encoded using Gaussian random Fourier features, $\gamma(R\mathbf{x})$, by projecting them onto fixed Gaussian frequency vectors and concatenating the resulting sine and cosine features.

The decoder-visible structural latent vector, z, is obtained from the encoder latent code after structural-z gating and optional bottleneck processing. CryoDECO uses z in two complementary ways. First, z is concatenated with the Fourier-encoded coordinates as part of the coordinate query input, providing particle-specific structural context to the decoder. Second, z serves as the conditioning vector for adaptive layer normalization (AdaLN) inside the residual MLP blocks. In each AdaLN layer, a zero-initialized linear modulator maps z to feature-wise modulation terms $\Delta\alpha(z)$ and $\Delta\beta(z)$. For hidden activation h, the AdaLN-modulated activation is computed as $h' = \mathrm{LN}(h) \cdot (1 + s\Delta\alpha) + s\Delta\beta$, where s is a runtime AdaLN modulation scale. The zero initialization keeps the AdaLN branch neutral at the start, allowing the structural latent state to progressively regulate internal feature normalization during training. Thus, z controls both the coordinate-level decoder input and the hidden feature modulation, whereas the viewing direction is specified by R. In-plane translation t is handled separately by phase-shifting the Hartley-domain target image during loss computation.

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2025.12.30.000075. This version posted November 3, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

The structural-z gate determines which latent dimensions enter both decoder-conditioning pathways. Its mathematical formulation, training schedule, regularization terms and hard-thresholding procedure are described below in "Adaptive structural latent gating" and "Adaptive structural-z gate training."

Optimization Protocol

CryoDECO's optimization combines a three-stage reconstruction protocol[27] with an adaptive structural-z gating schedule. The reconstruction protocol adapts the pretrained DECO-IEF representations to each dataset, whereas the gating schedule controls the decoder-visible latent capacity during training; the corresponding reconstruction objective, gated latent formulation and gate regularization terms are defined in Eqs. (4) and (5). Unless otherwise stated, learning rates below refer to configured base values before launcher-level scaling. Per-GPU batch sizes are reported, and global batch size scales with the number of GPU processes.

Warm-up adaptation. The network first undergoes a brief warm-up phase using a subset of 10,000 images. During this pretraining epoch, poses are randomly sampled and translations are set to zero. The frozen pretrained encoder provides structural latent features, while the hypervolume decoder is adapted by minimizing the reconstruction loss defined in Eq. 4 before pose optimization begins. In the reported configuration, the encoder learning rate was configured as 1×10^{-5} for later encoder-updating stages, and the decoder learning rate was initialized at 1×10^{-4} during reconstruction training. The per-GPU batch size was 64.

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2025.12.30.000075. This version posted November 3, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

Global pose discovery (Hierarchical search). Pose estimation in *ab initio* reconstruction is challenging because it is coupled with unknown structural heterogeneity. A coarse-to-fine hierarchical search strategy is employed in this stage, parameterizing the search space on the $SO(3)$ group using the Hopf fibration and discretizing it into a multi-resolution grid (HEALPix). The hierarchical pose search is performed on the dataset to determine the optimal discrete orientation for each particle. The structural priors from the pretrained DECO-IEF are expected to smooth the optimization landscape, which in our runs allowed this iterative search to converge more reliably toward the global optimum. In this stage, the per-GPU batch size was reduced to 22 to accommodate memory constraints during hierarchical pose search.

Continuous autodecoding refinement. Finally, CryoDECO transitions from discrete grid search to continuous optimization. For each particle i, the pose parameters $\phi_i = (R_i, t_i)$ are instantiated as learnable lookup-table variables initialized from the best poses obtained by hierarchical search. The reconstruction model and pose table are then jointly optimized with AdamW under the reconstruction objective in Eq. 4. Backpropagation flows directly into the pose lookup table, allowing the model to refine orientations and translations continuously beyond the resolution limits of the initial search grid. This stage runs for 100 epochs with a per-GPU batch size of 192. The pose-table learning rate is 1×10^{-4}, and the global batch size and learning rate are scaled with the number of GPU processes.

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2025.12.30.000075. This version posted November 3, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

Adaptive structural-z gate training. The structural-z gate is trained according to a four-step schedule: warm-up initialization, soft-gate learning, high-gate locking and low-gate pruning (Fig. 1b). First, the gate is initialized in a near-open state, allowing early reconstruction to proceed with almost the full nominal latent-capacity budget D_{cap}. After the five-epoch structural-z warm-up used in our experiments, CryoDECO activates the trainable per-dimension soft gate and optimizes the gate jointly with the reconstruction model. The gate parameters are placed in a separate optimizer group, allowing the gate to use its own learning rate and weight decay, while the decoder, encoder, latent codes and pose variables follow the standard reconstruction schedule.

During soft-gate learning, the reconstruction loss remains the physics-based generative objective described in "Generative reconstruction." Gate-specific regularization is applied only while the gate remains trainable and is defined formally in "Adaptive structural latent gating" (Eq. 5). After mask locking, gate regularization is disabled and the reconstruction model continues optimization with the fixed decoder-visible subspace.

After the soft gate has adapted, CryoDECO converts the learned gate profile into a hard decoder-visible mask. In the reported adaptive configuration, mask locking begins at epoch 50 with a gate threshold of 0.50, followed by a five-epoch annealing window that gradually suppresses inactive dimensions before freezing the final mask. The final retained subspace is then used for the remaining reconstruction refinement, with a minimum-active-dimension safeguard preventing collapse of the decoder-visible interface below the configured lower bound.

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2025.12.30.000075. This version posted November 3, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

Objective Functions

DECO-IEF pre-training. The DECO-IEF encoder backbone is pre-trained using an InfoNCE contrastive loss [36] with a temperature parameter $\tau = 0.5$, together with a structure-level supervised contrastive regularization term weighted by $\lambda_{\text{protein-supcon}} = 0.1$. This objective maximizes the similarity between different augmented views of the same particle (positive pairs) while minimizing similarity with other particles (negative pairs), thereby driving the encoder to learn representations that are robust to noise and orientation. The loss for a query representation $\mathbf{q}$ is defined as:

$$\mathcal{L}_{contrastive} = -\log \frac{\exp(\mathbf{q} \cdot \mathbf{k}^{+}/\tau)}{\exp(\mathbf{q} \cdot \mathbf{k}^{+}/\tau) + \sum_{\mathbf{k}^{-}} \exp(\mathbf{q} \cdot \mathbf{k}^{-}/\tau)} \quad (1)$$

where $\boldsymbol{k}^{+}$ represents the positive key (an augmented view of the same particle) and $\boldsymbol{k}^{-}$ represents the set of negative keys (views of other particles in the batch).

To further encourage particles from the same structure to form coherent representations, we additionally introduce a structure-level supervised contrastive loss:

$$\mathcal{L}_{\text{protein-supcon}} = -\frac{1}{N}\sum_{i=1}^{N}\frac{1}{|\mathcal{P}(i)|}\sum_{p\in\mathcal{P}(i)}\log\frac{\exp(\mathbf{q}_i \cdot \mathbf{k}_p/\tau)}{\sum_{a=1}^{M}\exp(\mathbf{q}_i \cdot \mathbf{k}_a/\tau)} \quad (2)$$

where $\mathcal{P}(i) = \{p : y_p = y_i\}$ denotes the set of keys whose protein identity is the same as that of query i, N is the number of query samples, and M is the number of gathered key samples. The final pre-training objective is:

$$\mathcal{L}_{\text{pretrain}} = \mathcal{L}_{\text{contrastive}} + 0.1\,\mathcal{L}_{\text{protein-supcon}} \quad (3)$$

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2025.12.30.000075. This version posted November 3, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

Generative reconstruction. To bridge the gap between latent feature extraction and 3D structural determination, the full CryoDECO framework (encoder and decoder) is fine-tuned by minimizing a physics-based reconstruction loss. This loss quantifies the discrepancy between the experimental particle images and the theoretical projections generated by the network. The specific objective function is given by:

$$\mathcal{L}(\theta, \{\phi_i\}, \{z_i\}) = \sum_{i=1}^{N} \| \mathcal{H}[I_i] - \widehat{\boldsymbol{C}}_i \odot T_{t_i} S_{R_i} \mathcal{V}_\theta[\tilde{z}_i] \|_2^2 \quad (4)$$

Here, θ denotes network parameters, and $\phi_i = (\boldsymbol{R}_i, \boldsymbol{t}_i)$ is the pose of the i-th particle, updated via pose search and then refined with SGD. $\mathcal{H}[\cdot]$ represents the Hartley transform. The decoder D_θ is implemented as an AdaLN-conditioned coordinate neural field that maps Fourier-encoded 3D frequency coordinates and the decoder-visible latent vector $\tilde{z}_i$ to Hartley-domain density values. The operator $S_{\boldsymbol{R}_i}$ extracts a central slice perpendicular to viewing direction $\boldsymbol{R}_i$. The phase-shift operator $\boldsymbol{T}_{\boldsymbol{t}_i}$ corrects for in-plane translation $\boldsymbol{t}_i$, and $\widehat{\boldsymbol{C}}_i$ is the Contrast Transfer Function (CTF). Minimizing this objective jointly optimizes the network for accurate density generation and infers the latent conformation $\boldsymbol{z}_i$ and pose ϕ_i for each particle. In adaptive latent-capacity runs, $\tilde{z}_i$ denotes the gated decoder-visible vector defined in Adaptive Structural Latent Gating; it is both concatenated with the encoded coordinates and used as the AdaLN conditioning vector. In fixed-dimensional reference runs, $\tilde{z}_i = z_i$; in adaptive runs, $\tilde{z}_i = z_i \odot g_{\text{eff}}$.

Adaptive structural latent gating. For adaptive-capacity runs, the encoder output for particle *i* is first projected to a nominal decoder-visible latent vector $z_i \in \mathbb{R}^{D_{\text{cap}}}$, where D_{cap} denotes the maximum latent-capacity budget available to the decoder. CryoDECO

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2025.12.30.000075. This version posted November 3, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

learns a global per-dimension structural gate $g \in [0,1]^{D_{\mathrm{cap}}}$ that controls the decoder visibility of each latent dimension. The gated decoder input is defined as $\tilde{z}_i = z_i \odot g_{\mathrm{eff}}$, where $g_{\mathrm{eff},d} = g_d[m_d + (1-\alpha)(1-m_d)]$, $g_d = \sigma(a_d)$, $m_d = \mathbf{1}(g_d \geq \tau)$, for $d = 1, \ldots, D_{\mathrm{cap}}$. Here a_d is the learned gate logit, τ is the hard-gate threshold, m_d is the thresholded binary mask and α is an annealing coefficient that increases from 0 to 1 after the hard-gate start epoch. The active decoder-visible dimensionality is reported as $D_{\mathrm{act}} = \sum_{d=1}^{D_{\mathrm{cap}}} m_d$.

Gate regularization is added as an auxiliary loss only while the soft gate remains trainable; the reconstruction loss term remains unchanged. The auxiliary gate loss is

$$\mathcal{L}_{\mathrm{gate}} = \lambda_{\mathrm{budget}} \max(\mathrm{mean}(g) - \rho, 0) + \lambda_{\mathrm{polar}} \mathrm{mean}\big(g \odot (1-g)\big), \quad (5)$$

where the hinge budget term penalizes the mean gate value only when it exceeds the target budget ρ, and the polarization term discourages ambiguous mid-range gate values. In the adaptive simulated runs, we used $D_{\mathrm{cap}} = 64$, $\tau = 0.50$, ρ=0.05 and $\lambda_{\mathrm{polar}} = \lambda_{\mathrm{budget}} = 0.2$.

Data-driven estimation of effective embedding-space partition count

After CryoDECO training, discrete compositional analysis requires choosing the number *k* of GMM components used to partition the learned particle representation. Rather than specifying *k* manually, we estimate the partitioning granularity supported by the embedding itself. Importantly, *k* is defined as an effective embedding-space partition count and is not assumed to correspond one-to-one with the number of biochemical species. The estimator

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2025.12.30.000075. This version posted November 3, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

combines GMM-based Bayesian information criterion (BIC)[37], for which lower values indicate a better fit–complexity trade-off, with the L-method for detecting the point of diminishing improvement in the BIC profile[38]. The procedure consists of three stages: adaptive coarse search, integer-resolution local refinement and, when the coarse and refined estimates disagree, a conditional hierarchical split check (Fig. 1c and Extended Data Fig. 5).

In our experiments, the learned representations were projected to a 16-dimensional space using Uniform Manifold Approximation and Projection (UMAP)[39], followed by random selection of up to 20,000 projected particles. For all BIC evaluations, candidate partition counts were assessed using full-covariance GMMs with five random initializations. BIC values were averaged over two independent 80% subsamples of the projected embeddings.

Adaptive coarse search.

First, the approximate range of plausible partition counts is identified using a sparsely sampled grid of candidate k values. For each candidate k, a GMM was fitted to repeated random subsamples of the projected particle representation, and the resulting BIC values were averaged to reduce sensitivity to sampling and model initialization. Because lower BIC indicates stronger model support, the resulting BIC profile typically decreases rapidly while additional components capture major population structure and then flattens as further partitioning provides diminishing benefit. We therefore apply the L-method to the sampled BIC profile to obtain a provisional elbow k_{elbow}.

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2025.12.30.000075. This version posted November 3, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

The elbow is used primarily to determine whether the currently explored range is sufficiently large rather than being accepted immediately as the final partition count. We quantify the residual improvement beyond the elbow using

$$\mathcal{R}_{post} = \frac{[\overline{BIC}(k_{elbow}) - \overline{BIC}(k_{min})]/(k_{min} - k_{elbow})}{[\overline{BIC}(1) - \overline{BIC}(k_{elbow})]/(k_{elbow} - 1)}, \quad (6)$$

where k_{min} is the candidate with the lowest BIC within the current search range. Thus, $\mathcal{R}_{post}$ measures the average post-elbow BIC improvement relative to that observed before the elbow. A large value indicates that substantial improvement persists beyond k_{elbow}, suggesting that the current search range is insufficient; in this case, the upper bound k_{max} is expanded and the coarse search is repeated. When the BIC profile instead exhibits a sustained rebound after its minimum, the minimum has been bracketed and its location is used to define k_{coarse}. In our implementation, the coarse search was initialized with $k_{max} = 20$ and a step size of 4. Given the initial candidate grid $k = \{1,4,8,12,16,20\}$, the search range was expanded when $\mathcal{R}_{post} > 8\%$. A rebound was defined when more than 60% of consecutive BIC changes in the terminal 20% of the sampled profile were increases. k_{min} was then taken as k_{coarse}.

Local refinement.

The coarse stage is designed to identify the relevant range efficiently and therefore evaluates only a sparse subset of candidate partition counts. We next evaluate every integer-valued k within a local window centred on k_{coarse}. The refinement window was defined as $[k_{coarse} - 4, k_{coarse} + 4]$, corresponding to one coarse-search step on each

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2025.12.30.000075. This version posted November 3, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

side. The L-method is reapplied to this densely sampled BIC profile to obtain the integer-resolution estimate k_{refine}.

Conditional hierarchical split check. Disagreement between k_{coarse} and k_{refine} indicates that the inferred partition count remains sensitive to search resolution. Only in this case do we perform an additional hierarchical validation step. We first fit the k_{refine}-component GMM and treat the resulting components as macro-partitions. Each macro-partition is then independently tested for reproducible evidence of further subdivision across repeated subsamples. A proposed local split is retained only when (i) it is supported in at least 9 of 10 subsamples, with a median relative BIC improvement of at least 5%, and (ii) incorporating the split yields a normalized global BIC gain of at least 0.5%. Accepted subdivisions are added to obtain k_{final}. Stage-wise estimates across the evaluated datasets are summarized in Extended Data Table 4.

Latent Space Clustering. In this step, particles were assigned to k_{final} discrete groups by fitting a k_{final}-component GMM to the 16-dimensional UMAP embedding[39]. The GMM parameters were optimized using the expectation–maximization algorithm, and each particle was assigned to the component with the highest posterior probability. The resulting particle groups were used for downstream homogeneous three-dimensional refinement. We reused the UMAP embedding generated during partition-count estimation to avoid an additional dimensionality-reduction step.

Data

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2025.12.30.000075. This version posted November 3, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

Dataset for pre-training DECO-IEF. On the basis of the 117 datasets (~65 million particles) used in previous work[29], we further expanded the pre-training dataset for DECO-IEF by incorporating an additional 308 publicly available cryo-EM datasets from EMPIAR[40], as summarized in Extended Data Table 1. The detailed processing steps for each dataset are as described previously[29]. The expanded dataset comprises approximately 134 million particle images, significantly enhancing the diversity and representativeness of the training data.

CryoBench datasets[28]. The Ribosembly dataset contains 16 distinct structures of the ribosome assembly intermediate. It presents a discrete heterogeneity problem where structures share a common core that progressively acquires additional proteins and ribosomal RNA. The particle distribution across states is non-uniform, totaling 335,240 particles, mimicking realistic experimental conditions. The data were simulated at a signal-to-noise ratio (SNR) of 0.01, testing the method's ability to resolve and classify discrete compositional states under typical noise levels. In addition to the full 16-state dataset, a resampled sub-dataset comprising four distinct particle species was generated to evaluate early-stage clustering dynamics (Fig. 1c).

The Tomotwin-100 dataset represents an extreme case of compositional heterogeneity[28]. This dataset contains a mixture of 100 different macromolecular complexes from a curated cellular catalogue, with molecular weights ranging from approximately 200 kDa to 4 MDa. It comprises 100,000 particle images at an SNR of 0.01. It challenges the method's capacity to perform ab initio reconstruction and classification in a highly complex

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2025.12.30.000075. This version posted November 3, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

mixture without a common structural core, simulating a scenario akin to analyzing cellular lysates.

The IgG-1D dataset features continuous conformational heterogeneity simulated from an IgG antibody (PDB: 1HZH). The heterogeneity is generated by a simple, one-dimensional rotation of a Fab domain around a dihedral angle, creating a continuous circular motion across 100 conformations. The dataset contains 100,000 particles simulated at an SNR of 0.01. It serves as a diagnostic benchmark for a method's ability to recover a simple, continuous low-dimensional manifold of conformational change.

The IgG-RL dataset presents a more challenging and realistic form of conformational heterogeneity. It uses the same IgG structure but introduces flexibility by randomizing the conformation of a 5-residue peptide linker, causing one Fab domain to sample a wide range of random orientations. This dataset also contains 100,000 particles at an SNR of 0.01, testing the method's robustness in handling complex, non-linear motions.

EM Ladder dataset. Micrographs for the "EM ladder" dataset were downloaded from EMPIAR-11693[30] and imported into cryoSPARC (v4.6.0)[25]. We performed CTF estimation (CTFFIND4[41]), particle picking (Blob Picker), particle extraction, and 2D classification. This process yielded 117,530 particle images with a box size of 400 pixels, which were then downsampled to a 128-pixel box size using the Fourier crop tool in cryoSPARC for classification in CryoDECO.

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2025.12.30.000075. This version posted November 3, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

Without a pre-specified class number, the *k*-estimator selected an effective embedding-space partition count of four, and GMM partitioning produced four particle populations corresponding to ApoF, β-galactosidase, TMV and PP7.

Following classification, particles from each final population were mapped back to their original 400-pixel box size for high-resolution reconstruction in cryoSPARC. ApoF, β-galactosidase and PP7 were initially refined using Homogeneous Refinement with O, D2 and I symmetry, respectively. TMV was processed using Helical Refinement with a rise of 1.408 Å and a twist of 22.04°. The resulting particle populations were then subjected sequentially to Global CTF Refinement and Local CTF Refinement. A final round of Non-Uniform Refinement was performed to generate the reported reconstructions. The original study[30] reported reconstructions of ApoF at 2.52 Å using 101,354 particles, β-galactosidase at 3.09 Å using 5,696 particles, PP7 at 3.40 Å using 2,351 particles and TMV at 2.85 Å using 4,972 particles.

Cryo-EM data from a *Chaetomium thermophilum* native cell extract. This dataset (EMPIAR-10892)[12] comprises 2,808 movies. The movies were processed in cryoSPARC[25] using Patch Motion Correction and Patch CTF Estimation. A total of 1,846,404 particles were picked using Blob Picker and extracted from all micrographs with a box size of 320 pixels. Several rounds of 2D classification and template-based picking yielded 77,316 particles for heterogeneous reconstruction. These particles were Fourier-cropped to a box size of 128 pixels for classification with CryoDECO. Following classification, particles from each recovered population were mapped back to their original 320-pixel box size for

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2025.12.30.000075. This version posted November 3, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

high-resolution reconstruction in cryoSPARC. An initial round of Non-Uniform Refinement was followed by Reference-Based Motion Correction and a second round of Non-Uniform Refinement. Global CTF Refinement and Local CTF Refinement were then performed sequentially, followed by a final round of Non-Uniform Refinement to generate the reported reconstructions.

Cryo-EM data of tri-snRNP. This dataset (EMPIAR-10073)[32] consists of 138,899 particles of the U4/U6.U5 tri-snRNP (tri-small nuclear ribonucleoprotein) complex from Saccharomyces cerevisiae. We analyzed the 138,899-particle cleaned subset derived from EMPIAR-10073. Particles were downsampled to a 128-pixel box size before analysis.

Cryo-EM data of $\alpha V\beta 8$ integrin. This dataset (EMPIAR-10345)[42] consists of 84,266 particles of the $\alpha V\beta 8$ integrin complex. These particles, selected after 2D classification, were downsampled to a 128-pixel box size.

CryoDRGN-AI Analysis.

CryoDRGN-AI[27] was used as a benchmark method for comparison with CryoDECO. Instead of using traditional encoder for feature extraction, CryoDRGN-AI employs a look-up table to directly learn a latent vector for each particle image. In our experiments, we followed the official guidelines provided in the CryoDRGN-AI repository (https://github.com/ml-struct-bio/drgnai) to set up and run the experiments. All hyperparameters were set to their default values unless specified otherwise.

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2025.12.30.000075. This version posted November 3, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

Computational Resources.

All particles were downsampled to a 128-pixel box size prior to training. Experiments were conducted using NVIDIA A40 (48 GB memory) and NVIDIA A100 (80 GB memory) GPUs. Approximated training times for CryoDECO for each dataset were as follows: CryoBench (100,000 particles; IgG-1D, IgG-RL, Tomotwin-100): ~13 hours on 2x NVIDIA A40 GPUs. CryoBench (335,240 particles; Ribosembly): ~39 hours on 2x NVIDIA A40 GPUs. EM Ladder (117,530 particles; EMPIAR-11693): ~14 hours on 2x NVIDIA A40 GPUs. C. thermophilum (77,316 particles; EMPIAR-10892): ~9 hours on 2x NVIDIA A40 GPUs.

Data availability.

The raw micrographs are available in the Electron Microscopy Public Image Archive https://www.ebi.ac.uk/empiar/ under accession code EMPIAR-11693, EMPIAR-10892, EMPIAR-10073 and EMPIAR-10345. The IgG-1D and IgG-RL datasets from CryoBench are available at https://zenodo.org/records/11629428, and the Ribosembly and Tomotwin-100 datasets from CryoBench are available at https://zenodo.org/records/12528292.

Code availability.

The CryoDECO source code and pretrained DECO-IEF weights are available at https://github.com/yanyang1998/CryoDECO.

Acknowledgements. We thank the HPC Center of Westlake University for providing computational facility support and technical assistance. This work was supported by the

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2025.12.30.000075. This version posted November 3, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

Ministry of Science and Technology (MOST) of the People's Republic of China (2024YFA0916903 to H.S.), the National Natural Science Foundation of China (62576286 to F.Y.), the Zhejiang Provincial Natural Science Foundation (DQ24C050001 to H.S.), the Research Center for Industries of the Future (RCIF), Westlake University, and the Westlake Education Foundation (to H.S. and F.Y.). We acknowledge the use of data from EMDB and EMPIAR for training our models.

Author information. These authors contributed equally: Yang Yan and Yanwanyu Xi. Correspondence to Fajie Yuan or Huaizong Shen.

Author contributions. The project was conceived and supervised by F.Y. and H.S. Y.Y. and Y.X. developed the model and performed related tests. Y.Y. and S.F. handled the preparation and processing of Cryo-EM data for DECO-IEF training. Y.W. and Z.T. performed wet-lab experiments for in-house data collection. The initial draft of the manuscript was written by Y.Y. and subsequently revised and finalized by F.Y. and H.S. All authors reviewed and provided feedback on the manuscript.

Competing interests. The authors declare no competing interests.

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2025.12.30.000075. This version posted November 3, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

1. Kühlbrandt, W. The Resolution Revolution. *Science* **343**, 1443–1444 (2014).
2. Cheng, Y. Single-Particle Cryo-EM at Crystallographic Resolution. *Cell* **161**, 450–457 (2015).
3. Liao, M., Cao, E., Julius, D. & Cheng, Y. Structure of the TRPV1 ion channel determined by electron cryo-microscopy. *Nature* **504**, 107–112 (2013).
4. Amunts, A. *et al.* Structure of the Yeast Mitochondrial Large Ribosomal Subunit. *Science* **343**, 1485–1489 (2014).
5. De Rosier, D. J. & Klug, A. Reconstruction of Three Dimensional Structures from Electron Micrographs. *Nature* **217**, 130–134 (1968).
6. Frank, J., Goldfarb, W., Eisenberg, D. & Baker, T. S. Reconstruction of glutamine synthetase using computer averaging. *Ultramicroscopy* **3**, 283–290 (1978).
7. Cheng, Y. Single-particle cryo-EM—How did it get here and where will it go. *Science* **361**, 876–880 (2018).
8. Nogales, E. & Scheres, S. H. W. Cryo-EM: A Unique Tool for the Visualization of Macromolecular Complexity. *Mol. Cell* **58**, 677–689 (2015).
9. Scheres, S. H. W. Processing of structurally heterogeneous cryo-EM data in RELION. in *Methods in Enzymology* vol. **579**, 125–157 (Academic Press, 2016).
10. Kastritis, P. L. & Gavin, A.-C. Enzymatic complexes across scales. *Essays Biochem.* **62**, 501–514 (2018).
11. Tüting, C. *et al.* Cryo-EM snapshots of a native lysate provide structural insights into a metabolon-embedded transacetylase reaction. *Nat. Commun.* **12**, 6933 (2021).
12. Skalidis, I. *et al.* Cryo-EM and artificial intelligence visualize endogenous protein community members. *Structure* **30**, 575-589.e6 (2022).
13. Semchonok, D. A., Kyrilis, F. L., Hamdi, F. & Kastritis, P. L. Cryo-EM of a heterogeneous biochemical fraction elucidates multiple protein complexes from a multicellular thermophilic eukaryote. *J. Struct. Biol. X* **8**, 100094 (2023).
14. Ho, C.-M. *et al.* Bottom-up structural proteomics: cryoEM of protein complexes enriched from the cellular milieu. *Nat. Methods* **17**, 79–85 (2020).
15. Torino, S., Dhurandhar, M., Stroobants, A., Claessens, R. & Efremov, R. G. Time-resolved cryo-EM using a combination of droplet microfluidics with on-demand jetting. *Nat. Methods* **20**, 1400–1408 (2023).
16. Nakane, T., Kimanius, D., Lindahl, E. & Scheres, S. H. Characterisation of molecular motions in cryo-EM single-particle data by multi-body refinement in RELION. *eLife* **7**, e36861 (2018).
17. Punjani, A. & Fleet, D. J. 3D variability analysis: Resolving continuous flexibility and discrete heterogeneity from single particle cryo-EM. *J. Struct. Biol.* **213**, 107702 (2021).
18. Zhong, E. D., Bepler, T., Berger, B. & Davis, J. H. CryoDRGN: reconstruction of heterogeneous cryo-EM structures using neural networks. *Nat. Methods* **18**, 176–185 (2021).
19. Chen, M. & Ludtke, S. J. Deep learning-based mixed-dimensional Gaussian mixture model for characterizing variability in cryo-EM. *Nat. Methods* **18**, 930–936 (2021).
20. Punjani, A. & Fleet, D. J. 3DFlex: determining structure and motion of flexible proteins from cryo-EM. *Nat. Methods* **20**, 860–870 (2023).
21. Luo, Z., Ni, F., Wang, Q. & Ma, J. OPUS-DSD: deep structural disentanglement for cryo-EM single-particle analysis. *Nat. Methods* **20**, 1729–1738 (2023).

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2025.12.30.000075. This version posted November 3, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

22. Gilles, M. A. & Singer, A. Cryo-EM heterogeneity analysis using regularized covariance estimation and kernel regression. *Proc. Natl. Acad. Sci.* **122**, e2419140122 (2025).
23. Sigworth, F. J. A Maximum-Likelihood Approach to Single-Particle Image Refinement. *J. Struct. Biol.* **122**, 328–339 (1998).
24. Scheres, S. H. W. RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
25. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
26. Zhong, E. D., Lerer, A., Davis, J. H. & Berger, B. CryoDRGN2: ab initio neural reconstruction of 3D protein structures from real cryo-EM images. in *Proc. IEEE/CVF International Conference on Computer Vision (ICCV)* 4066–4075 (2021).
27. Levy, A. *et al.* CryoDRGN-AI: neural ab initio reconstruction of challenging cryo-EM and cryo-ET datasets. *Nat. Methods* **22**, 1486–1494 (2025).
28. Jeon, M. et al. CryoBench: diverse and challenging datasets for the heterogeneity problem in cryo-EM. in *Advances in Neural Information Processing Systems* vol. **37**, 89468–89512 (2024).
29. Yan, Y., Fan, S., Yuan, F. & Shen, H. A comprehensive foundation model for cryo-EM image processing. *Nat. Methods* 23, 88–95 (2026) doi:10.1038/s41592-025-02916-8.
30. Bobe, D., Kopylov, M., Miller, J., Owji, A. P. & Eng, E. T. Multi-species cryoEM calibration and workflow verification standard. *Acta Crystallogr. Sect. F* **80**, 320–327 (2024).
31. Reynolds, D. Gaussian Mixture Models. in *Encyclopedia of Biometrics* 659–663 (Springer, Boston, MA, 2009). doi:10.1007/978-0-387-73003-5_196.
32. Nguyen, T. H. D. *et al.* Cryo-EM structure of the yeast U4/U6.U5 tri-snRNP at 3.7 Å resolution. *Nature* **530**, 298–302 (2016).
33. Li, Y., Zhou, Y., Yuan, J., Ye, F. & Gu, Q. CryoSTAR: leveraging structural priors and constraints for cryo-EM heterogeneous reconstruction. *Nat. Methods* **21**, 2318–2326 (2024).
34. Dosovitskiy, A. et al. An image is worth 16x16 words: transformers for image recognition at scale. in *Proc. International Conference on Learning Representations (ICLR)* (2021).
35. Chen, X., Xie, S. & He, K. An Empirical Study of Training Self-Supervised Vision Transformers. in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)* 9620–9629 (2021). doi:10.1109/ICCV48922.2021.00950.
36. Oord, A. van den, Li, Y. & Vinyals, O. Representation Learning with Contrastive Predictive Coding. Preprint at https://doi.org/10.48550/arXiv.1807.03748 (2019).
37. Fraley, C. & Raftery, A. E. Model-Based Clustering, Discriminant Analysis, and Density Estimation. *J. Am. Stat. Assoc.* **97**, 611–631 (2002).
38. Salvador, S. & Chan, P. Determining the number of clusters/segments in hierarchical clustering/segmentation algorithms. in *16th IEEE International Conference on Tools with Artificial Intelligence* 576–584 (IEEE Comput. Soc, Boca Raton, FL, USA, 2004). doi:10.1109/ICTAI.2004.50.
39. McInnes, L., Healy, J., Saul, N. & Großberger, L. UMAP: Uniform Manifold Approximation and Projection. *J. Open Source Softw.* **3**, 861 (2018).

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2025.12.30.000075. This version posted November 3, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

748 40. Iudin, A., Korir, P. K., Salavert-Torres, J., Kleywegt, G. J. & Patwardhan, A.
749 EMPIAR: a public archive for raw electron microscopy image data. *Nat. Methods* **13**, 387–
750 388 (2016).
751 41. Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from
752 electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).
753 42. Campbell, M. G. *et al.* Cryo-EM Reveals Integrin-Mediated TGF-β Activation
754 without Release from Latent TGF-β. *Cell* **180**, 490-501.e16 (2020).
755 43. Pettersen, E. F. *et al.* UCSF ChimeraX: Structure visualization for researchers,
756 educators, and developers. *Protein Sci.* **30**, 70–82 (2021).
757

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2025.12.30.000075. This version posted November 3, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

Figures

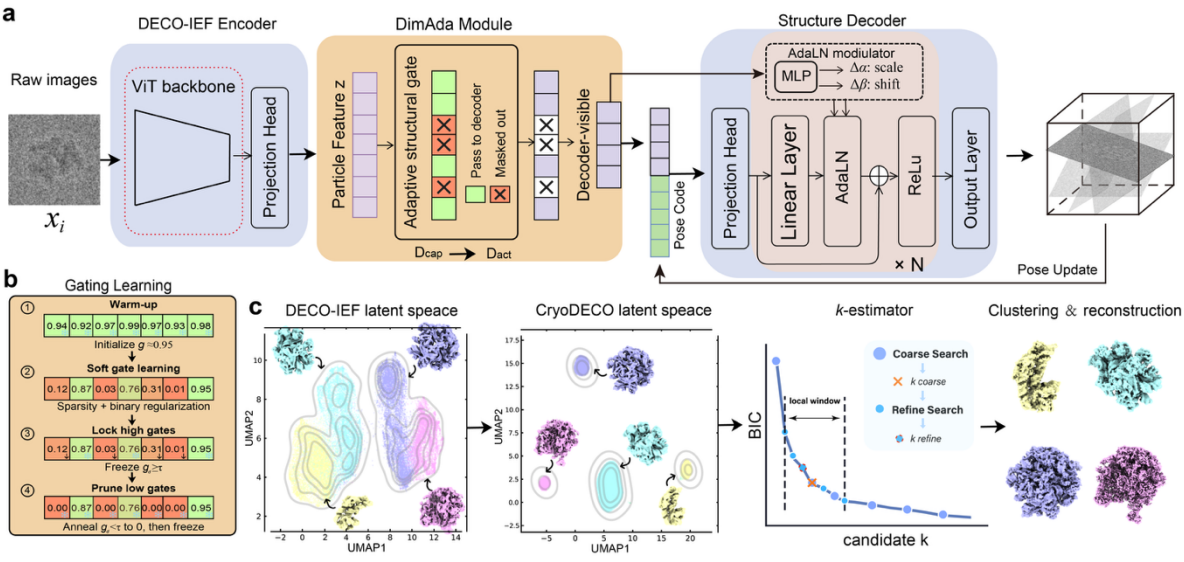


Fig. 1 | Overview of the CryoDECO framework and automated heterogeneity decomposition. (a) Prior-guided encoder–decoder architecture. DECO-IEF extracts particle features, the adaptive structural-*z* gate controls decoder-visible latent capacity, and the AdaLN-conditioned neural-field decoder reconstructs Hartley-domain structures while particle poses are optimized. (b) Adaptive structural-*z* gate training proceeds through warm-up initialization, soft-gate learning, high-gate locking and low-gate pruning, yielding a dataset-adaptive active capacity D_{act} within the nominal budget D_{cap}. (c) Workflow for compositional decomposition, illustrated using the resampled four-structure Ribosembly subset described in Methods. DECO-IEF provides an initially structured representation that CryoDECO reorganizes into particle populations. The *k*-estimator selects an effective embedding-space partition count through coarse-to-fine GMM–BIC analysis, followed by clustering and population-specific reconstruction.

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2025.12.30.000075. This version posted November 3, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

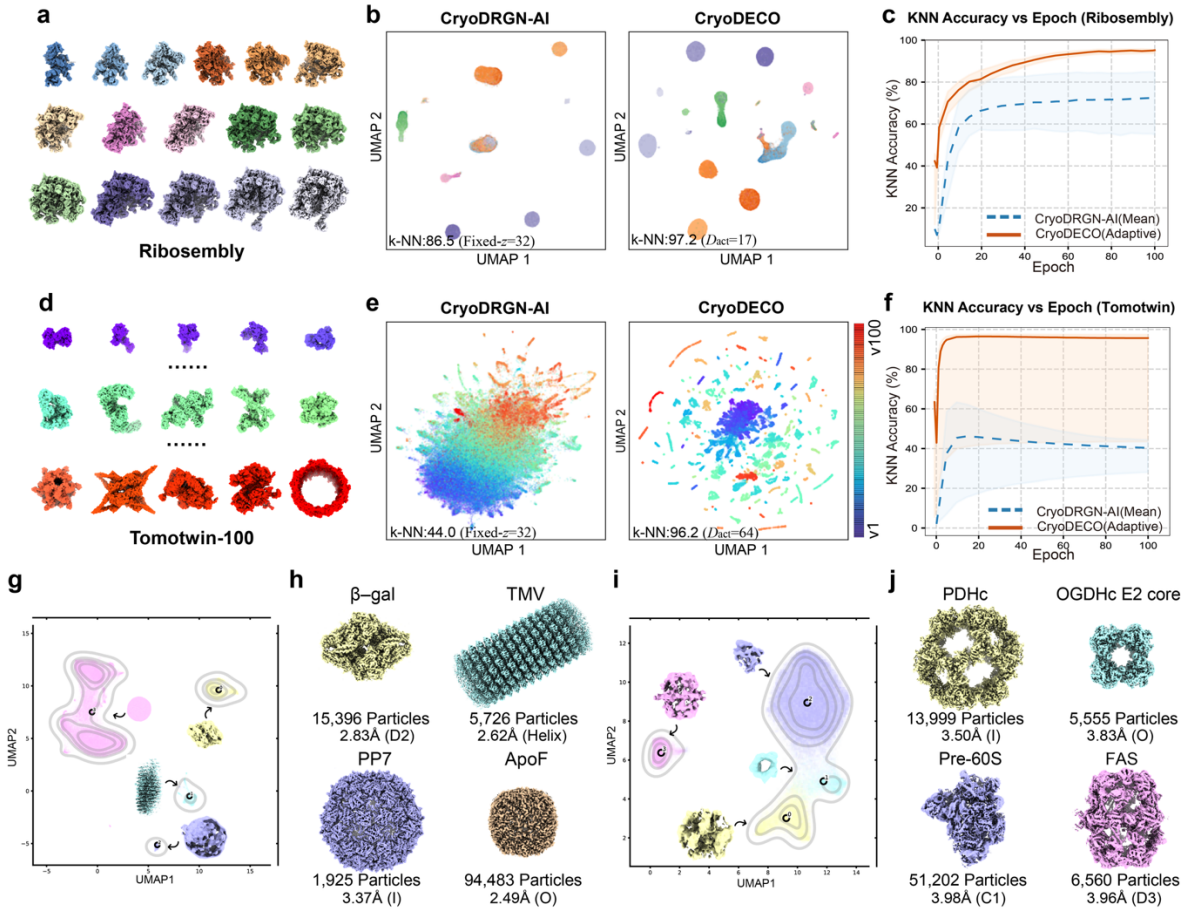
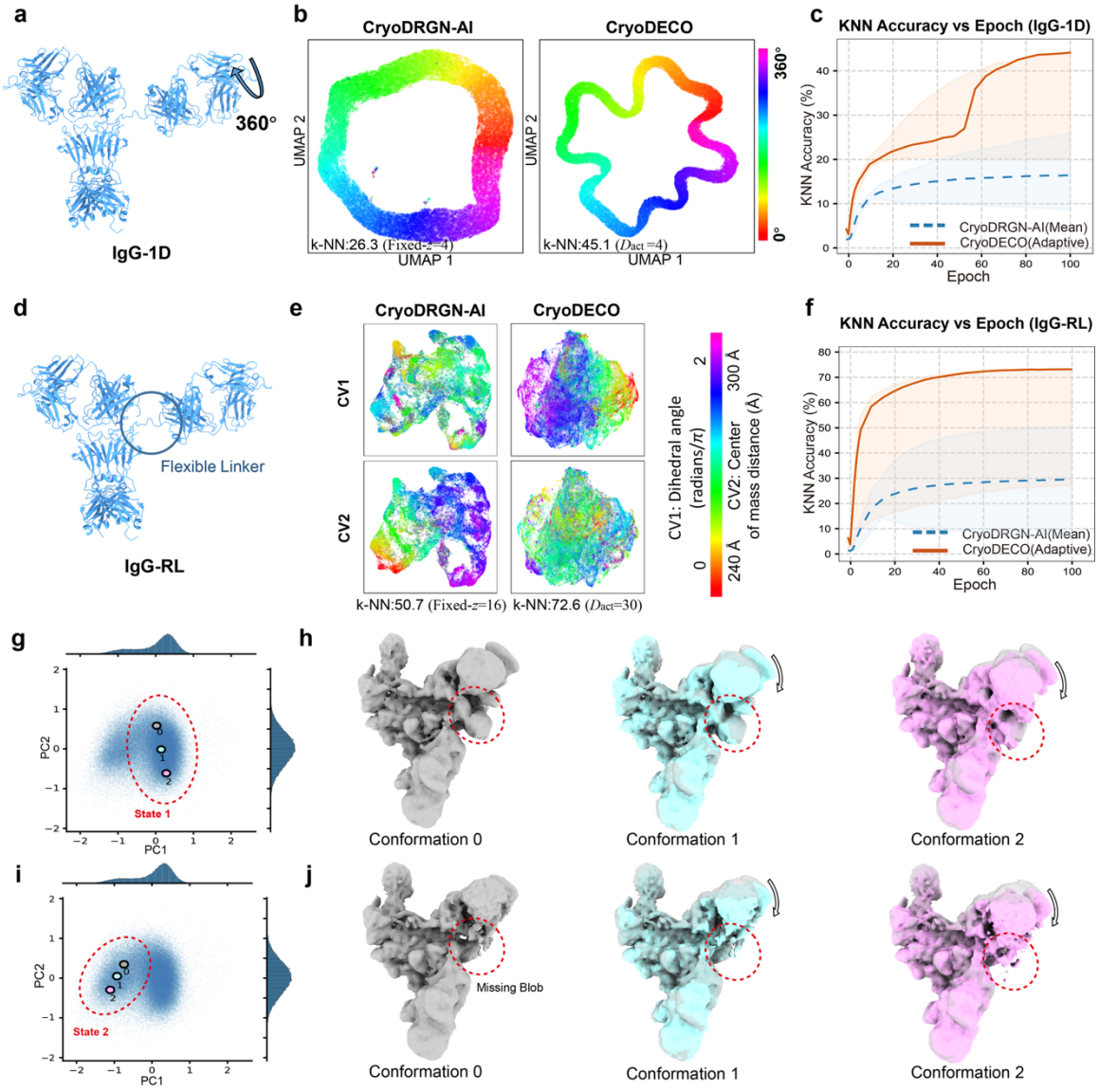


Fig. 2 | CryoDECO separates simulated and experimental compositional mixtures. (a) The 16 ground-truth structures in Ribosembly. (b) UMAP projections of Ribosembly particle representations learned by CryoDRGN-AI and CryoDECO, colored by ground-truth structure. (c) Top-1 *k*-NN accuracy during training on Ribosembly. (d) Representative structures from Tomotwin-100. (e) UMAP projections for Tomotwin-100, colored by ground-truth structure. (f) Top-1 k-NN accuracy during training on Tomotwin-100. In b and e, CryoDRGN-AI uses fixed z = 32; adaptive CryoDECO retains D_{act} = 17 and 64 for Ribosembly and Tomotwin-100, respectively. In c and f, orange curves show adaptive CryoDECO and blue dashed curves summarize CryoDRGN-AI across fixed *z* values of 4, 16, 32, 64 and 128. (g,i) UMAP projections of CryoDECO representations for the EM

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2025.12.30.000075. This version posted November 3, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

ladder mixture (g) and C. thermophilum native extract (i). Colors identify GMM-assigned populations, and inset structures are decoder-generated maps at representative cluster positions. (h,j) Final cryoSPARC reconstructions from the corresponding populations in g and i. Particle counts, nominal resolutions and refinement symmetries are indicated in the panels. Two-dimensional class averages are shown in Extended Data Fig. 3, and CryoDRGN-AI comparisons in Extended Data Fig. 2. I, icosahedral; O, octahedral; D2 and D3, dihedral; C1, no imposed symmetry; Helix, helical refinement.

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2025.12.30.000075. This version posted November 3, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

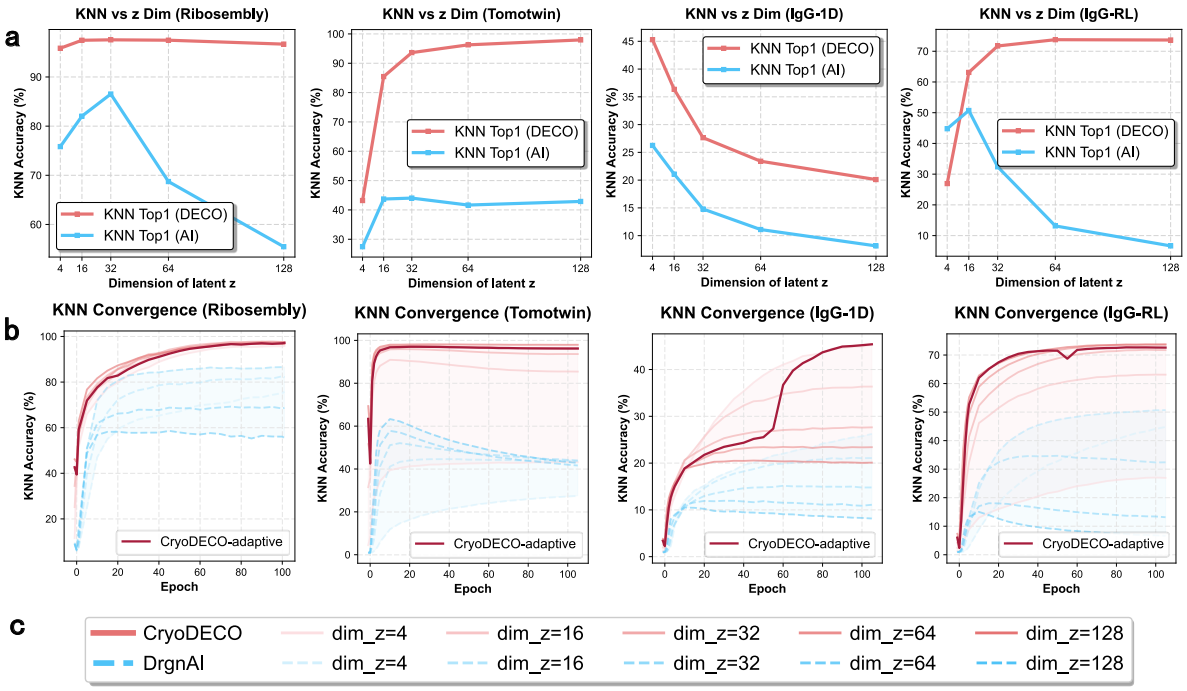


Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2025.12.30.000075. This version posted November 3, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

Fig. 3 | CryoDECO captures conformational variation and local density differences. (a,d) Conformational variability in IgG-1D (a), generated by rotation of a Fab-domain dihedral angle, and IgG-RL (d), generated by randomizing a linker region. Both CryoBench datasets contain 100 conformations with 1,000 particles per conformation at SNR = 0.01. (b) UMAP projections for IgG-1D, colored by rotation angle. (e) UMAP projections for IgG-RL, colored by the conformational variables indicated: dihedral angle (CV1) and center-of-mass distance (CV2). CryoDRGN-AI representations are shown on the left and CryoDECO representations on the right. CryoDRGN-AI uses fixed z = 4 for IgG-1D and *z* = 16 for IgG-RL; adaptive CryoDECO retains D_{act} = 4 and 30, respectively. (c,f) Top-1 *k*-NN accuracy during training for IgG-1D (c) and IgG-RL (f). Orange curves show adaptive CryoDECO; blue dashed curves summarize CryoDRGN-AI across fixed *z* values of 4, 16, 32, 64 and 128. (g,i) PCA projections of tri-snRNP representations. Red dashed ellipses mark regions designated State 1 (g) and State 2 (i). Marginal histograms show particle distributions along PC1 and PC2. Three representative positions per region are labeled 0–2. (h,j) Decoder-generated maps at those positions for State 1 (h) and State [2] (j), shown in gray, cyan and magenta. Maps 1 and 2 are superimposed on map 0 to show head rotation. Red dashed circles identify the head-adjacent density retained in State 1 and depleted in State 2. Three-dimensional visualizations were generated with ChimeraX[43].

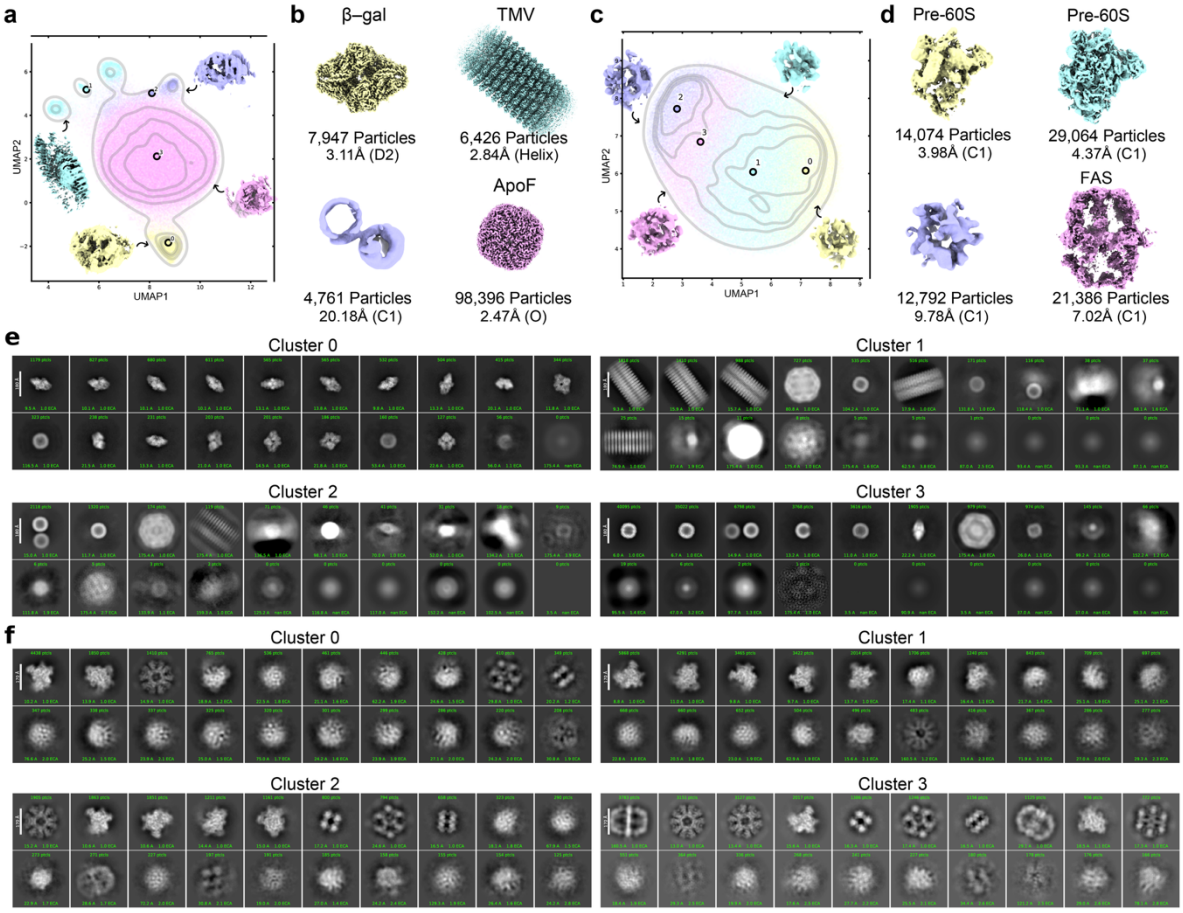
Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2025.12.30.000075. This version posted November 3, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

Extended Data Figures and Tables



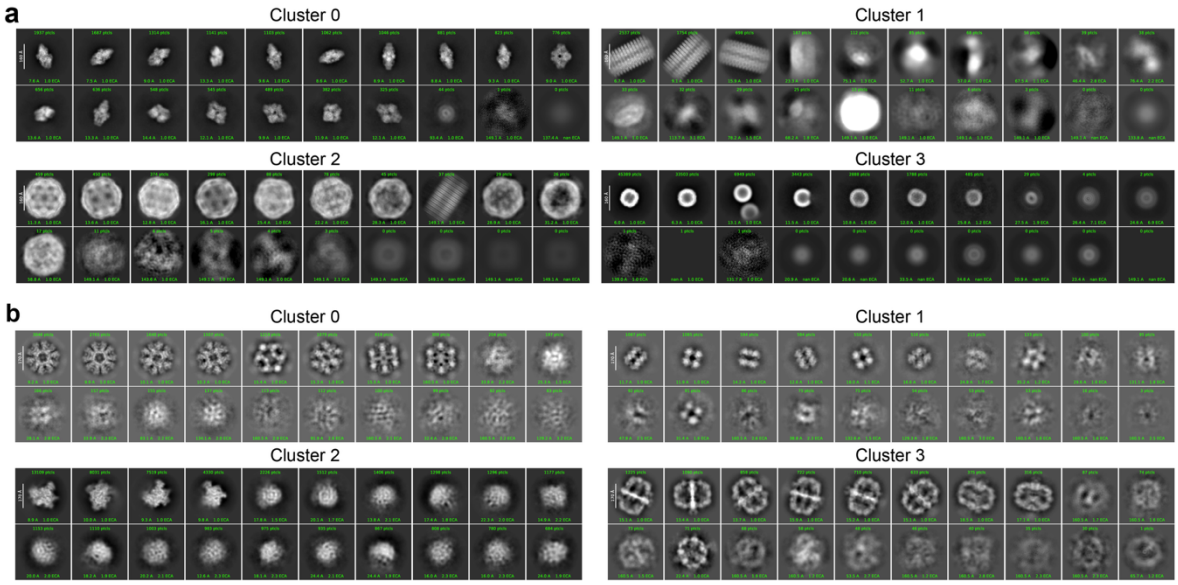
Extended Data Fig. 1 | Dependence of representation quality on latent capacity. CryoDECO and CryoDRGN-AI were evaluated at fixed particle-feature dimensions $z = 4$, 16, 32, 64 and 128 on Ribosembly, Tomotwin-100, IgG-1D and IgG-RL. (a) Top-1 k-NN accuracy as a function of fixed latent dimensionality. (b) Top-1 k-NN accuracy over training epochs for the displayed latent-capacity settings, including adaptive CryoDECO with a nominal latent budget of 64 (bold red curves). (c) Model, line-style and latent-dimensionality encodings used in a and b. Red denotes CryoDECO and blue denotes CryoDRGN-AI.

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2025.12.30.000075. This version posted November 3, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0



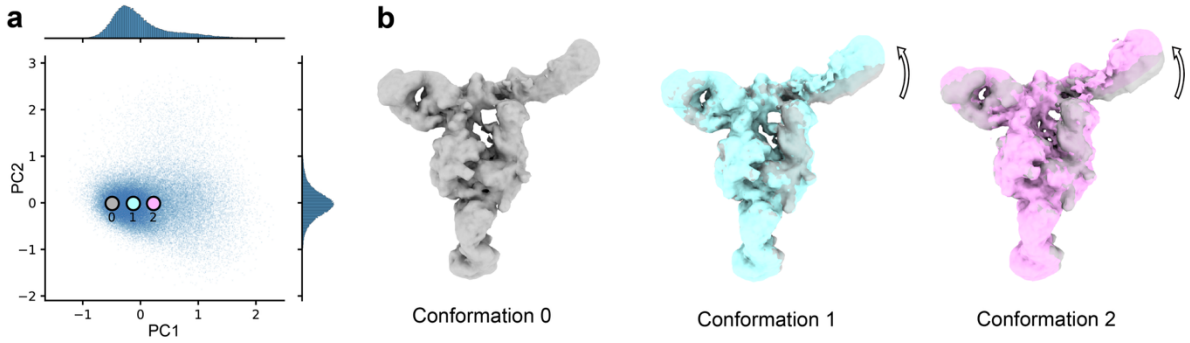
Extended Data Fig. 2 | CryoDRGN-AI analysis of experimental compositional mixtures. (a,c) UMAP projections of particle representations for the EM ladder mixture (a) and C. thermophilum native extract (c), with GMM-assigned populations shown in different colors. Insets show decoder-generated maps at representative cluster positions. (b,d) Reconstructions obtained from the corresponding particle populations after downstream processing in cryoSPARC. Particle counts, nominal resolutions and refinement symmetries are indicated in the panels. (e,f) Two-dimensional class averages for the EM ladder populations (e) and native-extract populations (f).

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2025.12.30.000075. This version posted November 3, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0



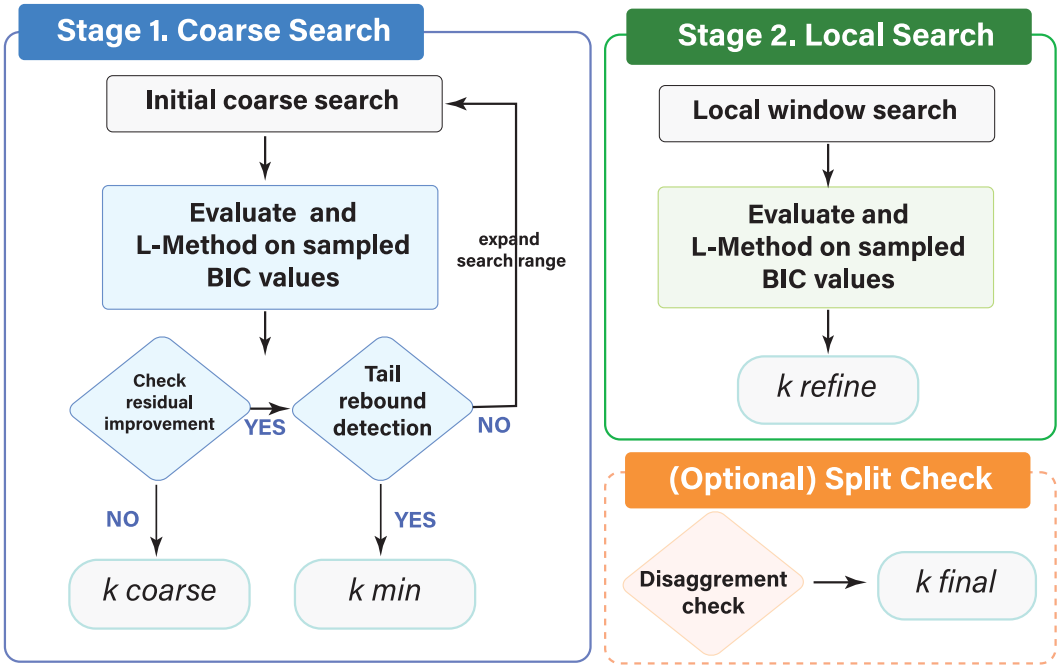
Extended Data Fig. 3 | Two-dimensional class averages of CryoDECO-separated particle populations. (a) Class averages for the four populations from the EM ladder mixture shown in Fig. 2. (b) Class averages for the four populations from the C. thermophilum native extract shown in Fig. 2.

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2025.12.30.000075. This version posted November 3, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0



Extended Data Fig. 4 | Analysis of the continuous conformational heterogeneity of αVβ8 integrin. (a) PCA visualization of the feature space for αVβ8 integrin. Three points were selected along PC1 to represent distinct conformations (gray, cyan and violet). (b) Reconstructions from the selected data points, color-coded by state. Superimposed structures illustrate the sequential conformational change. All three-dimensional visualizations were generated with ChimeraX.

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2025.12.30.000075. This version posted November 3, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0



Extended Data Fig. 5 | Adaptive coarse-to-fine estimation of the effective embedding-space partition count. The *k*-estimator first identifies a coarse partition count k_{coarse} by adaptively evaluating sampled candidate values with GMM–BIC scoring and L-method elbow detection. The search range is expanded when substantial post-elbow BIC improvement remains, whereas a sustained terminal rebound indicates that the BIC minimum has been passed. The resulting k_{coarse} is then refined within an integer-spaced local window to obtain k_{refine}. A conditional split check is applied when the coarse and refined estimates disagree, yielding the final estimate k_{final}.

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2025.12.30.000075. This version posted November 3, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

Extended Data Table 1 | Summary of Expanded Data Downloaded from EMPIAR for DECO-IEF Pre-training.

EMPIAR ID

10038	10153	10181	10193	10203	10249	10257	10258	10261	10264
10281	10285	10288	10290	10299	10314	10317	10332	10334	10335
10342	10344	10350	10352	10360	10362	10379	10380	10391	10399
10400	10407	10420	10422	10423	10425	10429	10433	10437	10438
10439	10443	10470	10477	10485	10486	10487	10495	10503	10507
10518	10521	10522	10530	10536	10540	10543	10544	10547	10549
10552	10557	10561	10573	10574	10575	10577	10580	10581	10582
10584	10594	10595	10596	10597	10599	10600	10604	10605	10611
10626	10628	10629	10630	10642	10652	10654	10656	10657	10661
10663	10666	10667	10682	10684	10695	10706	10707	10714	10716
10721	10726	10728	10752	10758	10759	10763	10777	10778	10793
10803	10808	10810	10811	10833	10834	10837	10842	10846	10850
10854	10855	10856	10858	10863	10864	10866	10877	10878	10880
10893	10894	10902	10909	10910	10915	10919	10927	10931	10932
10941	10942	10945	10951	10952	10964	10965	10966	10974	10975
10977	10978	10979	10984	10990	11006	11007	11008	11012	11021
11022	11028	11030	11031	11033	11036	11046	11047	11048	11049
11052	11053	11059	11071	11072	11073	11075	11076	11079	11080
11081	11091	11099	11110	11124	11127	11128	11132	11133	11134
11135	11137	11138	11143	11158	11171	11176	11177	11178	11179
11180	11182	11184	11185	11191	11192	11193	11194	11197	11218
11223	11224	11225	11227	11229	11231	11232	11234	11235	11241
11242	11251	11255	11256	11266	11268	11279	11295	11298	11304
11305	11307	11308	11313	11326	11328	11329	11330	11331	11333
11334	11335	11336	11337	11340	11341	11348	11352	11355	11364
11373	11374	11378	11382	11389	11390	11409	11441	11492	11512
11513	11524	11536	11559	11560	11568	11600	11604	11608	11624
11642	11645	11648	11654	11655	11656	11662	11663	11689	11697
11708	11713	11719	11725	11726	11727	11728	11760	11761	11780
11782	11786	11792	11797	11804	11810	11824	11859	11868	11887
11894	11900	11915	11916	12023	12024	12112	12146		

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2025.12.30.000075. This version posted November 3, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

Extended Data Table 2 | Effects of latent particle feature dimensionality on k-NN classification accuracy

Dataset		Latent dimension	Top-1 k-NN	Top-10 k-NN
Tomotwin-100	Fixed z	4	43.203	69.145
		16	84.274	96.087
		32	93.629	98.865
		64	96.276	99.456
		128	97.955	99.663
	Adaptive z*	D_{act} =64	96.213	99.459
Ribosembly	Fixed z	4	95.850	99.781
		16	97.469	99.844
		32	97.556	99.788
		64	97.481	99.819
		128	96.681	99.775
	Adaptive z*	$D_{\mathrm{act}} = 17$	97.231	99.788
IgG-1D	Fixed z	4	45.288	91.854
		16	36.352	92.954
		32	27.642	89.573
		64	23.400	85.182
		128	20.101	80.441
	Adaptive z*	D_{act} =4	45.149	92.321
IgG-RL	Fixed z	4	26.931	62.493
		16	63.114	93.856
		32	71.715	96.757
		64	73.761	97.409
		128	73.654	97.561
	Adaptive z*	D_{act} =30	72.604	96.860

* For fixed-dimensional runs, latent dimension denotes the predefined latent size z; for adaptive runs, it denotes the number of active dimensions D_{act} after gate locking. All adaptive runs were initialized with a nominal capacity $D_{cap} = 64$.

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2025.12.30.000075. This version posted November 3, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

Extended Data Table 3 | Clustering accuracy on compositional heterogeneity benchmarks

Method		Ribosembly		Tomotwin-100	
		ARI	**AMI**	**ARI**	**AMI**
With prior poses†	CryoDRGN[18]	0.873	0.935	0.956	0.983
	CryoDRGN-AI-fixed[27]	0.624	0.771	0.791	0.906
	Opus-DSD[21]	0.891	0.934	0.500	0.781
	3DVA[17]	0.666	0.823	0.058	0.335
	RECOVAR[22]	0.968	0.976	0.315	0.649
	CryoDECO	**0.984**	**0.979**	**0.999**	**0.999**
ab initio‡	CryoDRGN2[26]	0.529	0.618	0.116	0.374
	CryoDRGN-AI[27]	0.644	0.729	0.086	0.275
	CryoDECO	**0.928**	**0.935**	**0.647**	**0.871**

† Reconstruction with fixed (ground truth) poses.

‡ *ab initio* reconstruction where no input poses are provided.

* Results from CryoBench paper[28].

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2025.12.30.000075. This version posted November 3, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

Extended Data Table 4 | Stage-wise estimates of the effective-partition-count estimator (*k*-estimator)

Dataset	Nominal k	k_{coarse}	k_{refine}	k_{final}
EMPIAR-10892	4	4	3	4
EMPIAR-11693	4	4	4	4
Ribosembly	16	20	18	18
Tomotwin-100	100	116	114	114

* The effective embedding-space partition count was estimated under the default configuration, with a coarse step size 4 and an initial search upper bound 20. k_{coarse}: the estimate obtained from the adaptive coarse search. k_{refine}: the estimate after integer-spaced local refinement. k_{final}: the final estimate after the conditional hierarchical split check. The nominal partition count k is shown for reference.