CryoDECO: Deconstructing Extreme Compositional and Conformational Heterogeneity in Cryo-EM via 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,*]

[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@westlake.edu.cn).

Abstract

Resolving compositional and conformational heterogeneity remains the fundamental bottleneck in cryo-electron microscopy (cryo-EM). This challenge involves a circular dependency: accurate particle classification requires reliable 3D structural templates, yet template reconstruction demands high-fidelity classification. Current *ab initio* methods typically approach this as a joint optimization problem from a *tabula rasa* initialization, which frequently leads to optimization collapse when samples exhibit extreme complexity. Here, we present CryoDECO, an autoencoder framework that breaks this deadlock by integrating representation priors from pretrained cryo-EM foundation models. By projecting particle images onto a semantically structured manifold, CryoDECO effectively disentangles particle classification from structural reconstruction. We demonstrate that this prior-informed strategy robustly resolves extreme compositional heterogeneity, successfully classifying 100 distinct structures from a simulated mixture, and maps complex, continuous conformational landscapes. Applied to real-world datasets and unpurified native cell extracts, CryoDECO enables "Panoramic Structural Biology"—a high-throughput paradigm where *in silico* purification replaces laborious biochemical stabilization, allowing the simultaneous determination of diverse molecular machineries and their dynamic states.

Main

Cryo-electron microscopy (cryo-EM) has fundamentally transformed structural biology, especially since the resolution revolution[1–4]. The standard Single-Particle Analysis (SPA) workflow reconstructs 3D density maps by averaging millions of 2D projection images based on the Central Slice Theorem[5], which assumes that all 2D projection images derive from different views of a structurally identical or highly similar 3D object[2,6,7]. While this assumption holds for rigid, highly purified samples, biological reality is defined by ubiquitous structural heterogeneity, which manifests in two distinct but equally challenging forms: compositional complexity, arising from mixtures of different macromolecules, and conformational dynamics, driven by the intrinsic flexibility of biological machinery[8,9].

Conventional pipelines frequently treat these variations as noise, conflating distinct structures and states into blurred consensus maps[9]. However, for emerging frontiers such as the analysis of native cell extracts[10–13] and time-resolved studies[14], heterogeneity is increasingly recognized as a rich repository of biological insight. By offering an opportunity to explore the dynamic landscapes within complex biological samples[15–21], cryo-EM surpasses the limitations of classical X-ray crystallography, moving beyond static snapshots to capture both discrete components and a continuum of structural states. In these scenarios, the objective shifts from static structural determination to the analysis of dynamic, native landscapes. Here, we define "Panoramic Structural Biology" as a high-throughput paradigm: rather than resolving a single static structure following rigorous biochemical purification, this approach leverages *in silico* classification and reconstruction

to simultaneously identify and reconstruct myriad macromolecules and their dynamic states directly from crude mixtures, ranging from cell lysates to environmental samples.

Realizing this potential presents a formidable optimization challenge known as the *ab initio* circular dependency[22–24]. Accurate particle classification requires reliable 3D structural templates, but generating those templates first requires the correct classification of mixed particles. Current *ab initio* approaches generally attempt to resolve particle heterogeneity and reconstruct 3D templates simultaneously from initialization. Operating as *tabula rasa* learners without prior knowledge of the sample[17–20,25,26], these methods conduct a "blind" search in a high-dimensional landscape. Consequently, they frequently become trapped in local minima and fail to disentangle complex mixtures[27]. Compounding this difficulty, these methods must simultaneously search for correct particle orientations, further increasing the probability of task failure.

To overcome this bottleneck, we introduce CryoDECO (DEconstructing Extreme COmpositional and COnformational heterogeneity). Inspired by the transformative success of foundation models in capturing deep biological priors[28–30], this framework fundamentally alters the reconstruction paradigm by leveraging representation priors from pretrained cryo-EM foundation models, such as Cryo-IEF[31]. By exploiting the foundation model's demonstrated proficiency in classifying particle images by their distinct structures and states, CryoDECO effectively disentangles particle classification from 3D structural reconstruction. Treating the model as a semantic interpreter of 2D projections allows the framework to resolve both compositional and conformational heterogeneities before the

reconstruction loop begins. These learned priors provide a robust starting point that circumvents the "blind" search phase, effectively breaking the circular dependency and guiding optimization toward biologically valid solutions.

Foundation Model Priors Break the Optimization Deadlock

The core strategy of CryoDECO involves injecting a strong inductive bias to accurately pre-classify particle projections. We achieve this using an autoencoder framework: the encoder projects images onto a semantically structured manifold derived from the Cryo-IEF foundation model, while the decoder reconstructs the structures and searches for correct particle orientations (Fig. 1a and Methods). We implemented several optimization strategies that significantly elevate CryoDECO's performance over its preliminary iteration, CryoSolver[31], achieving state-of-the-art heterogeneity disentanglement.

First, to ensure a truly universal representation of biological heterogeneity, we expanded the foundation model's training corpus by more than twofold—scaling from 65 million to approximately 134 million particle images. This unprecedented scale equips the encoder with generalized priors that break the optimization deadlock from the outset. Consequently, on the benchmark Ribosembly dataset (a resampled version comprising four particle species)[31], CryoDECO converges substantially faster and resolves clearer clusters than CryoDRGN-AI, a leading *ab initio* method (Fig. 1b-d and Extended Data Fig. 1). Notably, even with a frozen encoder (CryoDECO-Fixed), the framework demonstrated robust accuracy, underscoring the inherent discriminative power of the foundation model priors acquired during large-scale pretraining (Fig. 1b,e). Ablation studies on fine-tuning

strategies confirmed that fine-tuning the entire encoder consistently yields optimal performance (Extended Data Fig. 2), as it allows the encoder to move beyond universal generalized features and adapt deeper attention layers to the specific signal-to-noise profiles and structural nuances of the target dataset.

Second, to operationalize this massive prior without incurring prohibitive computational costs, we transitioned the encoder backbone from a ViT-Base to a ViT-Small architecture. Empirical benchmarking confirms this lightweight design significantly accelerates training throughput while retaining feature extraction fidelity (Extended Data Fig. 3a). Complementing the encoder, the decoder functions as a generative neural field for 3D density map reconstruction. We optimized this network by inserting Layer Normalization (LayerNorm) within the Multi-Layer Perceptron (MLP) blocks (Fig. 1a), which stabilized gradient flow, reduced training loss, and accelerated convergence (Extended Data Fig. 3b).

Finally, the interface between the encoder and decoder—the latent particle feature vector (z)—serves as a critical topological hyperparameter that must align with the complexity of the biological target (Fig. 1a and Extended Data Fig. 4). As will be discussed in the following sections, we observe that discrete compositional heterogeneity demands a high-dimensional interface (e.g., $z = 128$) to ensure orthogonality between disjoint structures. Conversely, simple conformational dynamics necessitate a tight bottleneck (e.g., $z = 4$) to enforce manifold smoothness, while complex conformational heterogeneity

benefits from an intermediate capacity (e.g., $z = 64$) to resolve high-dimensional non-linear motions without overfitting.

Resolving Extreme Compositional Heterogeneity

We rigorously evaluated the capacity of CryoDECO to handle extreme compositional heterogeneity using the Ribosembly and Tomotwin-100 datasets[27], which contain simulated particles from 16 and 100 different structures, respectively (Fig. 2a,b).

Benchmarking against a baseline method, CryoDRGN-AI[26], revealed the decisive advantage of our prior-informed strategy (Fig. 2c-f). Lacking strong inductive biases, CryoDRGN-AI underperformed on both the Ribosembly (top-1 k-NN score: 86.5% vs 93.1%) and Tomotwin-100 datasets (top-1 k-NN score: 44.0% vs 94.0%) (Fig. 2c,d). Moreover, the training dynamics shows that CryoDECO not only achieved higher accuracies but did so with significantly faster convergence rates (Fig. 2e,f), demonstrating that the learned priors effectively shortcut the initial "blind" search phase of classification (Fig. 2e,f). Notably, on the Tomotwin-100 dataset, CryoDECO successfully projected particles from 100 distinct species into sharp, well-separated clusters, a task where CryoDRGN-AI severely struggled (Fig. 2d). Quantitative clustering metrics further underscored this performance gap. On the challenging Tomotwin-100 dataset, CryoDECO achieved an Adjusted Rand Index (ARI) of 0.622 and an Adjusted Mutual Information (AMI) score of 0.853 (Table 1). In contrast, CryoDRGN-AI (ARI: 0.086; AMI: 0.275) and CryoDRGN2 (ARI: 0.116; AMI: 0.374) essentially failed to resolve the mixture's high cardinality. Furthermore, when utilizing prior poses, CryoDECO’s performance reached

158 near-perfection, yielding ARI and AMI scores of 0.999 and confirming its ability to resolve
159 this hyper-complex mixture with total accuracy (Table 1).
160
161 We further investigated how the latent dimensionality (z) influences classification
162 performance (Extended Data Fig. 4). While varying z yielded consistently high top-1 k-
163 NN scores on the Ribosembly benchmark, accuracy on the more challenging Tomotwin-
164 100 dataset scaled markedly with dimensionality—increasing from 41.3% at $z = 4$ to
165 94.0% at $z = 128$. This trend validates that a high-dimensional interface is essential for
166 maintaining the orthogonality required to resolve hyper-complex mixtures.
167
168 ***In Silico* Classification of Real-World Mixtures**
169 To demonstrate the framework's utility in realistic experimental scenarios, we applied
170 CryoDECO to the "EM ladder" dataset (EMPIAR-11693)[32], a mixture of four purified
171 macromolecular complexes: Apoferritin, β-galactosidase, Tobacco Mosaic Virus (TMV),
172 and PP7 virus-like particles.
173
174 While the baseline method, CryoDRGN-AI, exhibited significant overlap between
175 different complexes in its latent space, failing to isolate PP7 particles entirely (Extended
176 Data Fig. 5), CryoDECO successfully disentangled the mixture into four distinct, well-
177 separated latent clusters (Fig. 3a). 2D class averages confirmed the high purity of these
178 classified populations (Fig. 3b). Independent 3D refinements in CryoSPARC yielded high-
179 resolution maps for all targets: 3.00 Å for β-gal, 2.48 Å for ApoF, 2.83 Å for TMV, and
180 3.38 Å for PP7 (Fig. 3c). These resolutions outperform the original report (3.09 Å, 2.52 Å,

2.85 Å, and 3.40 Å, respectively), confirming that CryoDECO effectively acts as an "*in silico* purification" tool, extracting structural populations of sufficient purity to yield superior reconstructions.

Enabling Panoramic Structural Biology in Native Extracts

To validate the Panoramic Structural Biology paradigm, we applied CryoDECO to resolve protein communities directly from a native *Chaetomium thermophilum* lysate fraction[12]. This dataset presents a massive challenge, requiring the isolation of stable structures from a complex endogenous background without relying on 3D templates or extensive prior structural knowledge.

CryoDECO successfully deconvolved this unpurified mixture into four discrete clusters (Fig. 4a,b). Subsequent homogeneous reconstruction recovered high-resolution maps for the expected targets (Fig. 4c): the Pyruvate Dehydrogenase Complex core (PDHc, 3.65 Å), Fatty Acid Synthase (FAS, 4.02 Å), the Oxoglutarate Dehydrogenase Complex E2 core (OGDHc, 3.86 Å), and the pre-60S ribosomal subunit (4.01 Å). CryoDECO not only automated identification but also achieved higher resolutions than the original manual study (4.47 Å, 3.84 Å, 4.38 Å, and 4.52 Å, respectively). Conversely, CryoDRGN-AI failed to disentangle the feature space, resulting in contaminated clusters and the complete loss of the OGDHc and PDHc complexes (Extended Data Fig. 6).

Mapping Conformational Dynamics

Beyond compositional sorting, CryoDECO excels at mapping continuous conformational landscapes. Using the IgG-1D and IgG-RL CryoBench datasets (Fig.5a,b)[27], we demonstrated that tuning the latent manifold topology correctly disentangles macromolecular dynamics (Fig. 5c,d). A tight bottleneck ($z = 4$) was optimal for resolving the simple one-dimensional motion of IgG-1D, producing a clear circular manifold (Fig. 5c,e and Extended Data Fig. 4). However, the highly flexible, randomized linker in IgG-RL required an expanded dimension ($z = 64$) to capture non-linear, high-dimensional motions (Fig. 5d,f and Extended Data Fig. 4).

This capability translates seamlessly to real-world data. For the tri-snRNP spliceosome complex, traversing the latent manifold visualized a smooth transition characterized by a counter-clockwise rotation of the head region (Fig. 6a,b and Supplementary Movie S1). Similarly, analysis of the $\alpha V\beta 8$ integrin dataset successfully captured the continuous flexibility of the protein's extended arm (Fig. 6c,d and Supplementary Movie S2).

Discussion

The primary challenge in current cryo-electron microscopy is no longer the pursuit of high resolution for rigid, purified proteins, but rather the interpretation of the inherent complexity within heterogeneous and native samples. Our work demonstrates that the fundamental bottleneck in resolving such samples—the *ab initio* circular dependency—is largely an artifact of the *tabula rasa* optimization strategy common in current pipelines. By introducing CryoDECO, we illustrate that the integration of foundation model priors

fundamentally reshapes the optimization landscape, allowing for the high-resolution determination of structures that were previously considered computationally inaccessible due to extreme compositional or conformational complexity.

This shift represents a fundamental move toward treating 3D reconstruction as a problem of semantic interpretation. While traditional *ab initio* learners treat 2D projection images as purely geometric entities to be aligned from a random start, CryoDECO leverages the massive biological knowledge extracted from the Cryo-IEF foundation model. Because the encoder has been pre-trained on an unprecedented and extensive corpus of particle images, it acts as a semantic interpreter that can identify specific macromolecular features amidst noise and orientation ambiguity before a 3D volume is even initialized. This effectively provides a "warm start" to the reconstruction loop, preventing the optimization from falling into the local minima that typically lead to job failure in high-cardinality mixtures like the Tomotwin-100 dataset.

The success of CryoDECO on native *C. thermophilum* extracts provides a practical proof-of-concept for the "Panoramic Structural Biology" paradigm. Historically, structural biology has been a reductive science requiring the purification of a single complex to near-homogeneity. CryoDECO shifts the burden of separation from the wet lab to the computational pipeline, enabling high-fidelity *in silico* purification. This approach not only significantly reduces biochemical labor but also preserves the native structural context. By resolving complexes directly from unpurified lysates, researchers can capture transient

interactions and conformational dynamics—the "dark matter" of the proteome—that are frequently lost under the harsh conditions of traditional stabilization or chromatography.

Beyond these practical applications, our findings reveal that the topology of the latent manifold is not merely a hyperparameter, but a mathematical reflection of a sample's physical degrees of freedom. We observe a "manifold matching" principle where discrete compositional heterogeneity requires high-dimensional orthogonality ($z = 128$) to prevent feature bleed-through between unrelated structures. Conversely, mapping continuous conformational motion requires tighter bottlenecks ($z = 4$ to $z = 64$) to enforce the manifold smoothness necessary to visualize dynamic transitions. That is, a wide bottleneck allows the model to memorize disjoint clusters into orthogonal sub-spaces, while a tight bottleneck acts as a strong regularizer that forces the network to map structural variations onto a continuous, interpolatable low-dimensional manifold. This flexible architectural interface allows CryoDECO to serve as a dual-purpose tool, capable of both sorting hyper-complex mixtures and mapping the subtle, non-linear flexibilities of proteins like the $\alpha V\beta 8$ integrin.

While CryoDECO significantly lowers the barrier for analyzing complex samples, current implementation requires user selection of the optimal latent dimensionality based on expected sample complexity. As we move toward a more automated systems structural biology, future developments will focus on adaptive manifold estimation to dynamically tune latent capacity based on the intrinsic data distribution. Furthermore, extending this prior-informed framework to 3D cryo-electron tomography represents a promising frontier

for resolving molecular structures within the even more complex environment of the intact cell. Ultimately, by breaking the circular dependency of reconstruction, CryoDECO establishes a new standard for high-throughput analysis, bringing the field closer to a truly "panoramic" understanding of the cellular machinery in its native state.

Methods

Prior-Guided Heterogeneous Reconstruction

The CryoDECO framework adopts an encoder-decoder architecture, but unlike traditional methods that learn features from scratch, it introduces a paradigm shift by injecting learned structural priors to guide the learning process. The encoder of CryoDECO is initialized with the pre-trained Cryo-IEF model[31], a foundation model retrained on a vast and diverse dataset of approximately 130 million cryo-EM particles. This initialization is not merely for robust feature extraction, it serves to regularize the highly non-convex optimization landscape of heterogeneous reconstruction. In conventional approaches, random initialization often leads to unstable convergence and "blind" pose searches. By contrast, our encoder leverages the universal structural knowledge encapsulated in Cryo-IEF to map 2D images directly to a semantically structured latent manifold from the onset. Although the decoder is initialized from scratch, this structured input acts as a strong regularizer, ensuring that the optimization begins within a valid structural manifold, thereby reducing the complexity of the subsequent optimization.

Retraining of the Cryo-IEF Foundation Model

To construct a more robust structural prior for CryoDECO, we retrained the Cryo-IEF foundation model on an expanded corpus of approximately 134 million particle images. While the original Cryo-IEF used a ViT-Base backbone[33], we employed a Vision Transformer Small (ViT-Small) architecture (12 layers, 384 embedding dimensions, 12 heads) by default to optimize computational efficiency (Extended Data Fig. 3a).

We trained the model from scratch using the Momentum Contrast (MoCo v3) self-supervised learning framework [34], following the hyperparameter configuration of the original Cryo-IEF study[31]. 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, horizontal flipping, and Gaussian blurring to enforce invariance to noise and orientation. 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 built upon a Vision Transformer (ViT-Small) architecture initialized with Cryo-IEF weights, serving as a "structural compass" for the downstream decoder. The backbone consists of a 12-layer transformer encoder with an embedding dimension of 384 and 12 attention heads, and it processes the input image using a patch size of 14 $\times$ 14. Evaluation of ViT variants confirmed that the Small architecture provides the optimal trade-off between computational efficiency and representation performance (Extended Data Fig. 3a).

Decoder as a neural field. The decoder is a 3-layer residual MLP with a hidden dimension of 256, functioning as a coordinate-based neural field in the Hartley domain. To synthesize a 2D projection based on the Central Slice Theorem, the network first generates a 2D grid of continuous frequency coordinates, denoted as x, situated on the origin plane ($z = 0$). These lattice coordinates are then multiplied by the particle's orientation matrix R, dynamically rotating the grid in 3D space to extract the specific 2D slice corresponding to the particle's viewing angle. We employ Gaussian random Fourier feature encoding, $\gamma(\cdot)$, to map these rotated 3D frequency coordinates into high-frequency feature space. Specifically, random frequency vectors are sampled from a Gaussian distribution and used to compute $k = (Rx) \cdot$ freqs, followed by concatenating [sin(k), cos(k)] to produce a 3 × 2 × pe_dim dimensional spatial encoding (e.g., 384D for pe_dim=64). The final input to the MLP is constructed by concatenating the positionally encoded frequency coordinates and conformation features z (from the encoder) (cat$[\gamma(Rx), z]$). This design ensures the MLP synthesizes fine structural details dictated by z along the precise slice defined by R, while in-plane translation t is subsequently handled by applying a phase shift to the generated Hartley transform. Incorporating Layer Normalization (LN) after each linear layer stabilized training and accelerated convergence compared to a standard architecture (Extended Data Fig. 3b).

Latent manifold regularization. The dimensionality of the latent particle feature (z) acts as a critical topological regularizer that affects the model's ability to capture underlying structural heterogeneity (Extended Data Fig. 4). On real-world datasets, we tailor z to the

intrinsic complexity of the heterogeneity. For discrete compositional heterogeneity, a high-dimensional space (e.g., $z = 128$) is necessary to provide sufficient orthogonality, allowing the model to encode disjoint structural states without forcing them into overlapping regions of the manifold. In the case of conformational dynamics, the optimal dimensionality depends on the system's degrees of freedom. We find that a tight information bottleneck (e.g., $z = 4$) is optimal for simple continuous motions, forcing the model to map dynamics onto a smooth, low-dimensional manifold rather than overfitting to noise. However, conformational landscapes characterized by high entropy and multiple degrees of freedom—such as the randomized flexible linker in the IgG-RL dataset (Extended Data Fig. 4)—demand expanded latent capacity (e.g., $z = 64$) to accurately capture the complex non-linear feature space.

Optimization Protocol

CryoDECO's optimization progressively adapts the universal Cryo-IEF priors to specific experimental data using a three-stage protocol[26].

Warm-up adaptation. The network first undergoes a brief adaptation phase using a subset of 10,000 images with fixed random poses. Our encoder leverages its pre-trained weights to project particle images onto a structured latent manifold, enabling the decoder to establish a valid coarse structural distribution before pose optimization begins. During this stage, network parameters are updated via gradient descent minimizing the reconstruction loss using the AdamW optimizer (learning rate 10^{-5} for encoder and 10^{-4} for decoder, batch size 64, learning rate and batch size are scaled by number of GPUs).

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. Leveraging the structural priors from the pre-trained Cryo-IEF smooths the optimization landscape, enabling this iterative search to converge more reliably toward the global optimum. In this stage batch size is reduced to 22 (scaled by number of GPUs) to accommodate memory constraints.

Refinement with SGD (Continuous optimization). Finally, we transition from discrete grid search to continuous optimization. In this stage, the pose parameters $\phi_i = (R_i, t_i)$ for each particle i are instantiated as learnable tensors stored in a lookup table. These variables are initialized with the best values from the hierarchical search. We then perform joint optimization of the network weights θ, latent codes z_i, and pose parameters ϕ_i using Stochastic Gradient Descent (SGD) with the Adam optimizer. 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 uses batch size 192 and runs for 100 epochs. The learning rate of lookup table is 10^{-4}. The learning rate and batch size are scaled by number of GPUs.

Objective Functions

Cryo-IEF pre-training. The Cryo-IEF encoder backbone is pre-trained using the InfoNCE contrastive loss function [35] with a temperature parameter $\tau = 0.5$. 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).

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:

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

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 $\mathcal{V}_\theta$ maps the latent embedding $\boldsymbol{z}_i$ to a 3D density volume in the Hartley domain. 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.

Latent Space Clustering

Gaussian Mixture Model (GMM)[36] is applied to extract discrete structural classes from the latent representations learned by CryoDECO. GMM is a probabilistic clustering approach that models the overall data distribution as a mixture of multivariate Gaussian distributions, each representing a distinct heterogeneous state. For datasets exhibiting compositional heterogeneity (e.g., the EM ladder and C. thermophilum native cell extract datasets), the number of mixture components was set to the expected number of distinct macromolecular targets. The GMM parameters were optimized via the Expectation-Maximization (EM) algorithm. Particles were subsequently assigned to discrete groups based on their maximum posterior probability, and these grouped particles were utilized for downstream homogeneous 3D refinement. To accelerate computation, particle features were downsampled to 16 dimensions using Uniform Manifold Approximation and Projection (UMAP)[37] if $z > 16$.

Data

Dataset for pre-training cryo-IEF. On the basis of the 117 datasets (~65 million particles) used in previous work[31], we further expanded the pre-training dataset for Cryo-IEF by incorporating an additional 308 publicly available cryo-EM datasets from EMPIAR[38], as summarized in Extended Data Table 1. The detailed processing steps for each dataset are as

435 described previously[31]. The expanded dataset comprises approximately 134 million particle
436 images, significantly enhancing the diversity and representativeness of the training data.
437
438 **CryoBench datasets[27].** The Ribosembly dataset contains 16 distinct structures of the
439 ribosome assembly intermediate. It presents a discrete heterogeneity problem where
440 structures share a common core that progressively acquires additional proteins and
441 ribosomal RNA. The particle distribution across states is non-uniform, totaling 335,240
442 particles, mimicking realistic experimental conditions. The data was simulated at a signal-
443 to-noise ratio (SNR) of 0.01, testing the method's ability to resolve and classify discrete
444 compositional states under typical noise levels. In addition to the full 16-state dataset, a
445 resampled sub-dataset comprising four distinct particle species was generated to evaluate
446 early-stage clustering dynamics (as shown in Fig. 1).
447
448 The Tomotwin-100 dataset represents an extreme case of compositional
449 heterogeneity[27]. This dataset contains a mixture of 100 different macromolecular complexes
450 from a curated cellular catalogue, with molecular weights ranging from approximately 200
451 kDa to 4 MDa. It comprises 100,000 particle images at an SNR of 0.01. It challenges the
452 method's capacity to perform ab initio reconstruction and classification in a highly complex
453 mixture without a common structural core, simulating a scenario akin to analyzing cellular
454 lysates.
455
456 The IgG-1D dataset features continuous conformational heterogeneity simulated from
457 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[32] and imported into cryoSPARC (v4.6.0)[24]. We performed CTF estimation (CTFFIND4[39]), particle picking (Blob Picker), particle extraction, and 2D classification. This process yielded 117,527 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. Following classification, 3D refinement of the four structures was performed separately using cryoSPARC's Homogeneous Refinement tool with appropriate symmetry settings: O for Apoferritin, D2 for β-galactosidase, and I for the PP7 virus-like particle. Before the final refinement, all particles are mapped back to their original 400-pixel box size in CryoSPARC. For TMV, helical refinement was used with a rise of 1.408 Å and a twist of 22.04°. The original study[32] reported high-resolution

reconstructions for Apoferritin (2.52 Å; 101,354 particles), β-galactosidase (3.09 Å; 5,696 particles), PP7 (3.40 Å; 2,351 particles), and TMV (2.85 Å; 4,972 particles).

Cryo-EM data from a *Chaetomium thermophilum* native cell extract. This dataset (EMPIAR-10892)[12] comprises 2,799 movies. The movies were processed using cryoSPARC[24], including motion correction (Patch Motion Correction) and CTF estimation (Patch CTF Estimation). From these, 1,846,404 particles were picked (Blob Picker) and extracted from all micrographs with a box size of 320 pixels. Several rounds of 2D classification and template-based picking yielded 78,100 particles for heterogeneous reconstruction. These particles were then downsampled to a 128-pixel box size for input into CryoDECO and mapped back to 320 pixels before final refinement in cryoSPARC.

Cryo-EM data of tri-snRNP. This dataset (EMPIAR-10073)[40] consists of 138,899 particles of the U4/U6.U5 tri-snRNP (tri-small nuclear ribonucleoprotein) complex from Saccharomyces cerevisiae. Particles were downsampled to a 128-pixel box size before analysis.

Cryo-EM data of $\alpha V\beta 8$ integrin. This dataset (EMPIAR-10345)[41] 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.

Metrics

K-NN scores. We utilize the K-NN score to quantitatively evaluate how well the extracted features separate particles from different heterogeneous states. We employ a Leave-One-Out (LOO) evaluation strategy. For each data point (feature vector) $\boldsymbol{f}_i$ in the entire dataset, we identify its k nearest neighbors by computing the dot product similarity S_{ij} with all other data points $\boldsymbol{f}_j$ (where $j \neq i$). Instead of a simple majority vote, we perform a weighted classification. The contribution of each of the k neighbors is weighted based on its similarity score S_{ij} and a temperature parameter T. The score for a specific class c for sample i is computed by summing the weights of its neighbors belonging to that class:

$$\text{Score}_i(c) = \sum_{r=1}^{k} exp(S_{ir}/T) \cdot 1(l_{ir} = c) \tag{3}$$

where l_{ir} is the label of the r-th nearest neighbor for sample i (found from the set where $j \neq i$), and $\mathbf{1}(\cdot)$ is the indicator function. In our experiments, we set $k = 10$ and $T = 1$. We report the Top-1 and Top-10 accuracy. Let y_i be the true class for sample i, and let $(\hat{y}_i^{(1)}, \hat{y}_i^{(2)}, \ldots, \hat{y}_i^{(k')})$ be the list of predicted classes ranked by $\text{Score}_i(c)$, where $k' = min(k, C)$ and C is the total number of classes. The Top-1 score is the standard LOO accuracy, checking if the highest-scoring predicted class matches the true class:

$$\text{Top-1 Score} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}\left(\hat{y}_i^{(1)} = y_i\right) \tag{4}$$

where N is the total number of samples in the dataset. The Top-10 score measures if the true class y_i is present within the top-10 ranked predictions:

$$\text{Top-10 Score} = \frac{1}{N}\sum_{i=1}^{N}\left[\sum_{j=1}^{min(10,k,C)} \mathbf{1}\left(\hat{y}_i^{(j)} = y_i\right)\right] \tag{5}$$

Adjusted Rand Index (ARI). The Adjusted Rand Index (ARI) is used to measure the similarity between the true class assignments (ground truth U) and the predicted clustering results (V). Given a set of N samples, let $U = \{u_1, u_2, \ldots, u_R\}$ be the true class assignments (with R classes), and $V = \{v_1, v_2, \ldots, v_C\}$ be the predicted cluster assignments (with C clusters). We can construct a contingency table where n_{ij} represents the number of samples that are in both true class u_i and predicted cluster v_j. Let $a_i = \sum_j n_{ij}$ be the total number of samples in true class u_i, and $b_j = \sum_i n_{ij}$ be the total number of samples in predicted cluster v_j. The ARI is calculated as:

$$\mathrm{ARI} = \frac{\sum_{ij}\binom{n_{ij}}{2} - \left[\sum_i \binom{a_i}{2}\sum_j \binom{b_j}{2}\right] / \binom{N}{2}}{\frac{1}{2}\left[\sum_i \binom{a_i}{2} + \sum_j \binom{b_j}{2}\right] - \left[\sum_i \binom{a_i}{2}\sum_j \binom{b_j}{2}\right] / \binom{N}{2}} \tag{6}$$

Adjusted Mutual Information (AMI). The Adjusted Mutual Information (AMI) is another metric used to assess the agreement between two clusterings (true labels U and predicted labels V). The AMI is defined as:

$$\mathrm{AMI}(U,V) = \frac{\mathrm{MI}(U,V) - E[\mathrm{MI}(U,V)]}{max(H(U), H(V)) - E[\mathrm{MI}(U,V)]} \tag{7}$$

where $\mathrm{MI}(U,V)$ is the mutual information between the two clusterings, $H(U)$ and $H(V)$ are the entropies of the respective clusterings, and $E[\mathrm{MI}(U,V)]$ is the expected mutual information between two random clusterings with the same number of clusters and data distribution.

CryoDRGN-AI Analysis.

CryoDRGN-AI[26] 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.

Computational Resources.

All particles were downsampled to a 128-pixel box size prior to training. Experiments were conducted using NVIDIA A40 (40 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 (111,733 particles; EMPIAR-11693): ~14 hours on 2x NVIDIA A40 GPUs. C. thermophilum (78,100 particles; EMPIAR-10892): ~9 hours on 2x NVIDIA A40 GPUs.

Inclusion and ethics statement

All collaborators in this study meet the authorship criteria required by Nature Portfolio journals and have been duly included as authors. Roles and responsibilities were agreed upon by all collaborators prior to the commencement of the research. This study does not result in stigmatization, incrimination, discrimination, or any other personal risk to participants. Additionally, the research does not pose health, safety, security, or other risks

565 to the researchers involved. We have discussed benefit-sharing measures and ensured that
566 local and regional research relevant to this study is appropriately cited.
567
568 **Reporting summary**
569 Further information on research design is available in the Nature Portfolio Reporting
570 Summary linked to this article.
571
572 **Data availability.**
573 The raw micrographs are available in the Electron Microscopy Public Image Archive
574 https://www.ebi.ac.uk/empiar/ under accession code EMPIAR-11693, EMPIAR–10892,
575 EMPIAR-10073 and EMPIAR-10345. The IgG-D and IgG-RL datasets from CryoBench
576 are available at https://zenodo.org/records/11629428, and the Ribosembly and Tomotwin-
577 100 datasets from CryoBench are available at https://zenodo.org/records/12528292.
578
579 **Code availability.**
580 The codes and pretrained models for Cryo-IEF are available at https://github.com/westlake-
581 repl/Cryo-IEF. The codes for CryoDECO are available at
582 https://github.com/yanyang1998/CryoDECO.
583

References

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* https://doi.org/10.1126/science.aat4346 (2018) doi:10.1126/science.aat4346.
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. 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. 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).
15. 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).
16. 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).
17. 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).
18. 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).
19. Punjani, A. & Fleet, D. J. 3DFlex: determining structure and motion of flexible proteins from cryo-EM. *Nat. Methods* **20**, 860–870 (2023).
20. 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).

21. Gilles, M. A. & Singer, A. Cryo-EM heterogeneity analysis using regularized covariance estimation and kernel regression. *Proc. Natl. Acad. Sci.* **122**, e2419140122 (2025).
22. Sigworth, F. J. A Maximum-Likelihood Approach to Single-Particle Image Refinement. *J. Struct. Biol.* **122**, 328–339 (1998).
23. Scheres, S. H. W. RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
24. 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).
25. 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 4066–4075 (2021).
26. Levy, A. *et al.* CryoDRGN-AI: neural ab initio reconstruction of challenging cryo-EM and cryo-ET datasets. *Nat. Methods* **22**, 1486–1494 (2025).
27. Jeon, M. *et al.* CryoBench: Diverse and challenging datasets for the heterogeneity problem in cryo-EM. Preprint at http://arxiv.org/abs/2408.05526 (2024).
28. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
29. Rives, A. *et al.* Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. *Proc. Natl. Acad. Sci.* **118**, e2016239118 (2021).
30. Lin, Z. *et al.* Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science* **379**, 1123–1130 (2023).
31. Yan, Y., Fan, S., Yuan, F. & Shen, H. A comprehensive foundation model for cryo-EM image processing. *Nat. Methods* 1–8 (2025) doi:10.1038/s41592-025-02916-8.
32. 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).
33. Dosovitskiy, A. *et al.* An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. in (2020).
34. 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.
35. 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).
36. Reynolds, D. Gaussian Mixture Models. in *Encyclopedia of Biometrics* 659–663 (Springer, Boston, MA, 2009). doi:10.1007/978-0-387-73003-5_196.
37. McInnes, L., Healy, J., Saul, N. & Großberger, L. UMAP: Uniform Manifold Approximation and Projection. *J. Open Source Softw.* **3**, 861 (2018).
38. Iudin, A., Korir, P. K., Salavert-Torres, J., Kleywegt, G. J. & Patwardhan, A. EMPIAR: a public archive for raw electron microscopy image data. *Nat. Methods* **13**, 387–388 (2016).
39. Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).
40. 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).

675 41. Campbell, M. G. *et al.* Cryo-EM Reveals Integrin-Mediated TGF-β Activation
676 without Release from Latent TGF-β. *Cell* **180**, 490-501.e16 (2020).
677 42. Pettersen, E. F. *et al.* UCSF ChimeraX: Structure visualization for researchers,
678 educators, and developers. *Protein Sci.* **30**, 70–82 (2021).
679
680

Acknowledgments

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

Authors and affiliations

Zhejiang University, Hangzhou, Zhejiang, China

Yang Yan & Shiqi Fan

School of Engineering, Westlake University, Hangzhou, Zhejiang, China

Yang Yan, Yanwanyu Xi, Shiqi Fan & Fajie Yuan

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

Yifei Wang, Ziyun Tang & Huaizong Shen

Westlake Laboratory of Life Sciences and Biomedicine, Hangzhou, Zhejiang, China

Yifei Wang, Ziyun Tang & Huaizong Shen

Institute of Biology, Westlake Institute for Advanced Study, Hangzhou, Zhejiang, China

Yifei Wang, Ziyun Tang & Huaizong Shen

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 Cryo-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.

Corresponding authors

Correspondence to Fajie Yuan or Huaizong Shen.

Ethics declarations

Competing interests

The authors declare no competing interests.

Figures

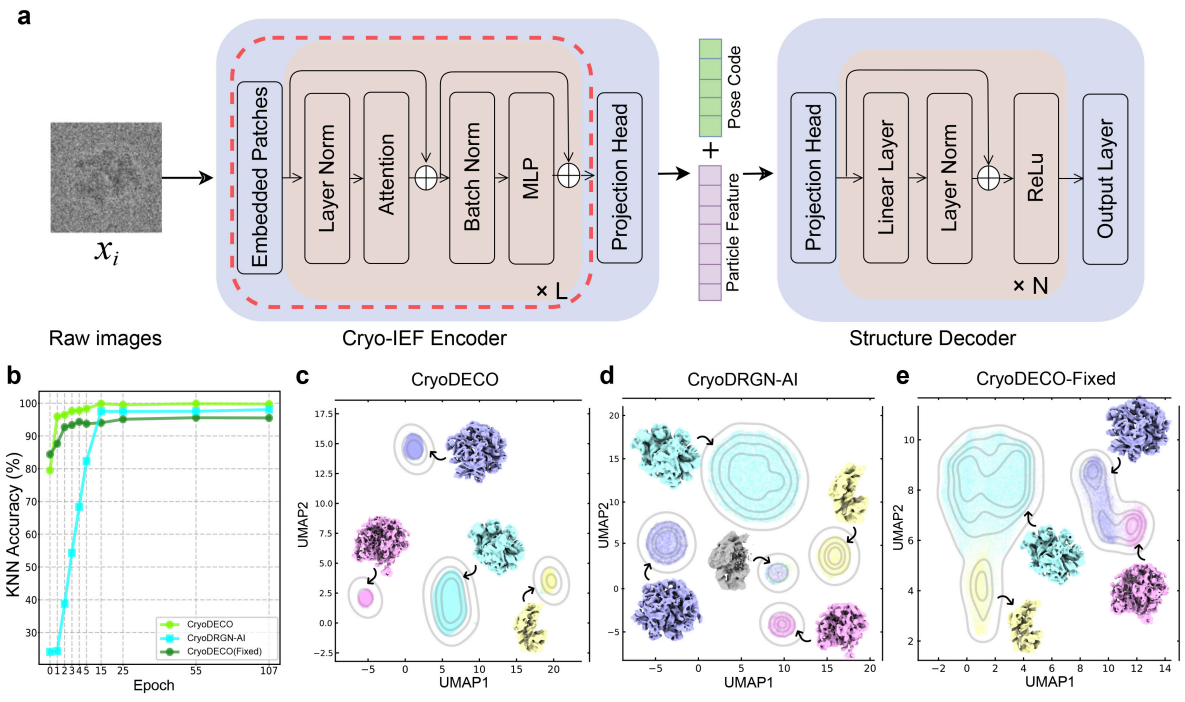


Fig. 1 | Overview and validation of the CryoDECO framework. (**a**) CryoDECO utilizes a prior-informed encoder-decoder architecture to break the *ab initio* circular dependency. The encoder, a self-supervised pre-trained cryo-IEF model, acts as a semantic interpreter and extracts features onto a structured manifold, while the decoder reconstructs 3D structures and searches for orientations. (**b**) Comparison of training dynamics (k-NN score vs. epoch) between CryoDECO and CryoDRGN-AI on the test dataset of the resampled CryoBench Ribosembly[27]. (**c-e**) UMAP visualizations of the feature space at the final training epoch for CryoDECO (**c**), CryoDRGN-AI(**d**), and CryoDECO-Fixed (**e**). Data points are color-coded by their ground-truth structures shown in Extended Data Fig 1. Insets show structures generated by the decoder from the center of each feature cluster.

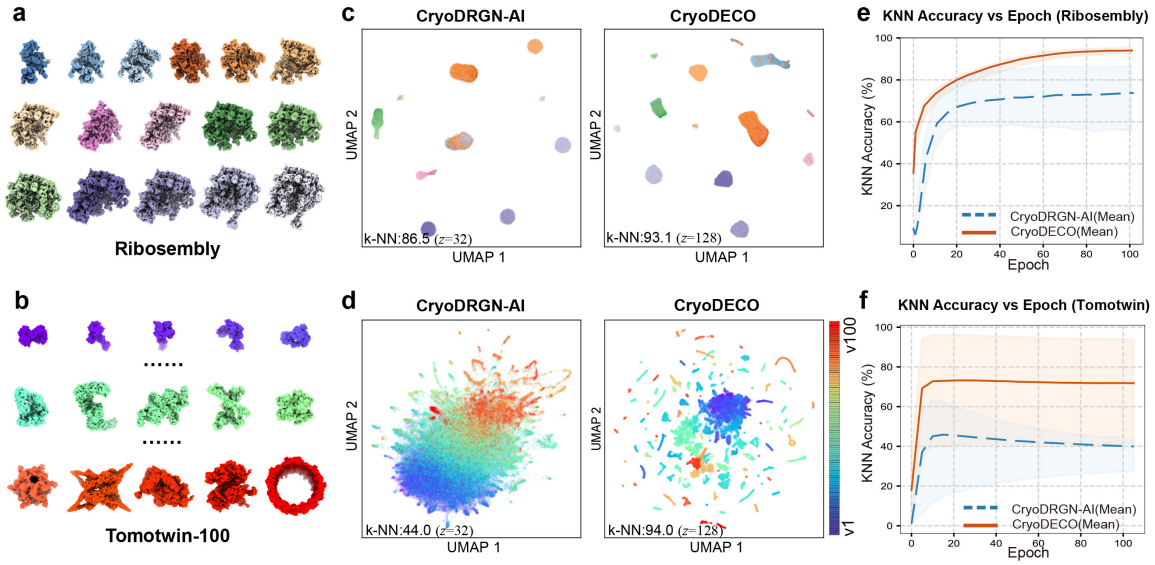


Fig. 2 | CryoDECO's classification performance on CryoBench datasets of extreme compositional heterogeneity. (**a**) The 16 ground-truth structures in the CryoBench Ribosembly dataset[27]. (**b**) Examples of the 100 ground-truth structures in the Tomotwin-100 dataset[27]. (**c,d**) UMAP visualizations of particle feature spaces from the Ribosembly (**c**) and Tomotwin-100 (**d**) datasets. Features from CryoDRGN-AI (left panels) are shown for comparison. Data points are color-coded by their ground-truth structures. Optimal latent dimensions (z) are shown (Ribosembly: z = 32 for CryoDRGN-AI, z = 128 for CryoDECO; Tomotwin-100: z = 32 for CryoDRGN-AI, z = 128 for CryoDECO). (**e,f**) k-NN scores during training for the Ribosembly (**e**) and Tomotwin-100 (**f**) datasets. Scores are averaged over three independent runs (CryoDECO: orange line; CryoDRGN-AI: blue dashed line) with varying particle feature dimensions (z=4, 16, 32, 64, 128). Shaded areas represent the score range across all z dimensions.

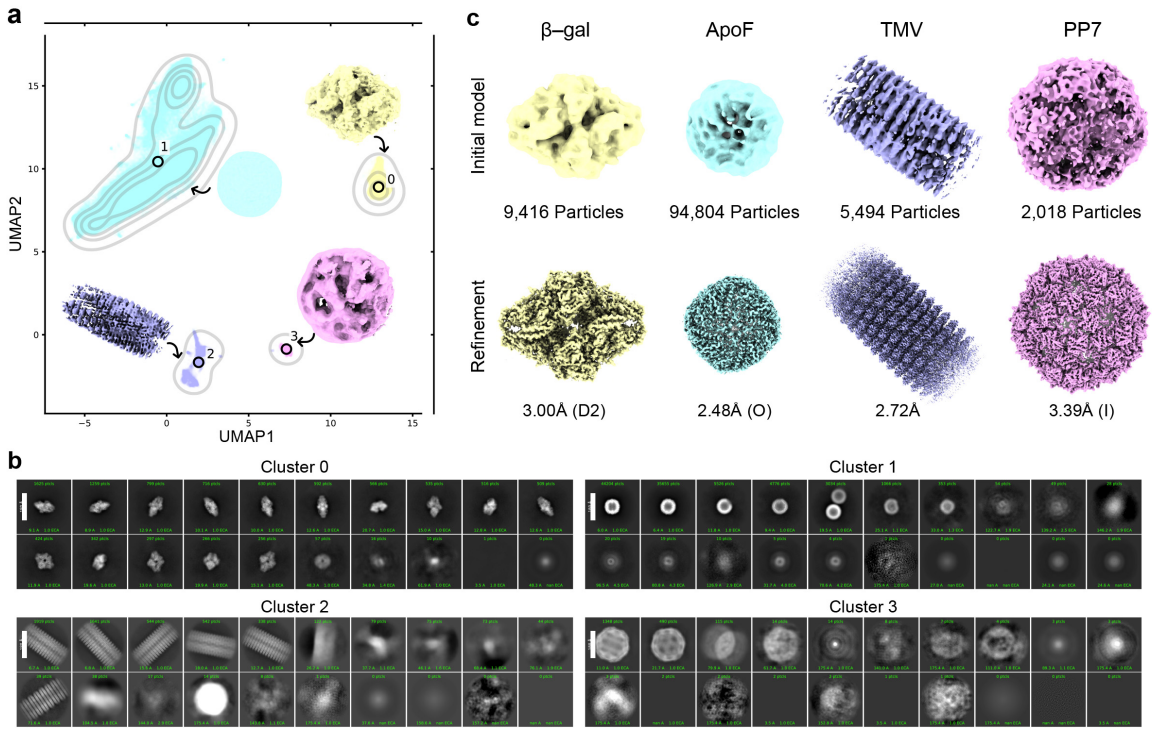


Fig. 3 | CryoDECO classifies four distinct structures from the EM ladder dataset. (**a**) UMAP visualization of the particle feature space extracted by CryoDECO. It demonstrates that the framework successfully disentangles four distinct complexes from a mixture. Particles are clustered into four groups (color-coded) using Gaussian Mixture Model (GMM). Insets show the initial 3D structures generated by the decoder from the center of each feature cluster. (**b**) 2D class averages for each cluster of particles, confirming the homogeneity. (**c**) Final 3D reconstructions of the four clusters of particles after ab-initio reconstruction and non-uniform refinement in CryoSPARC. Structures are color-coded as in (**a**). All 3D visualizations were created using ChimeraX[42].

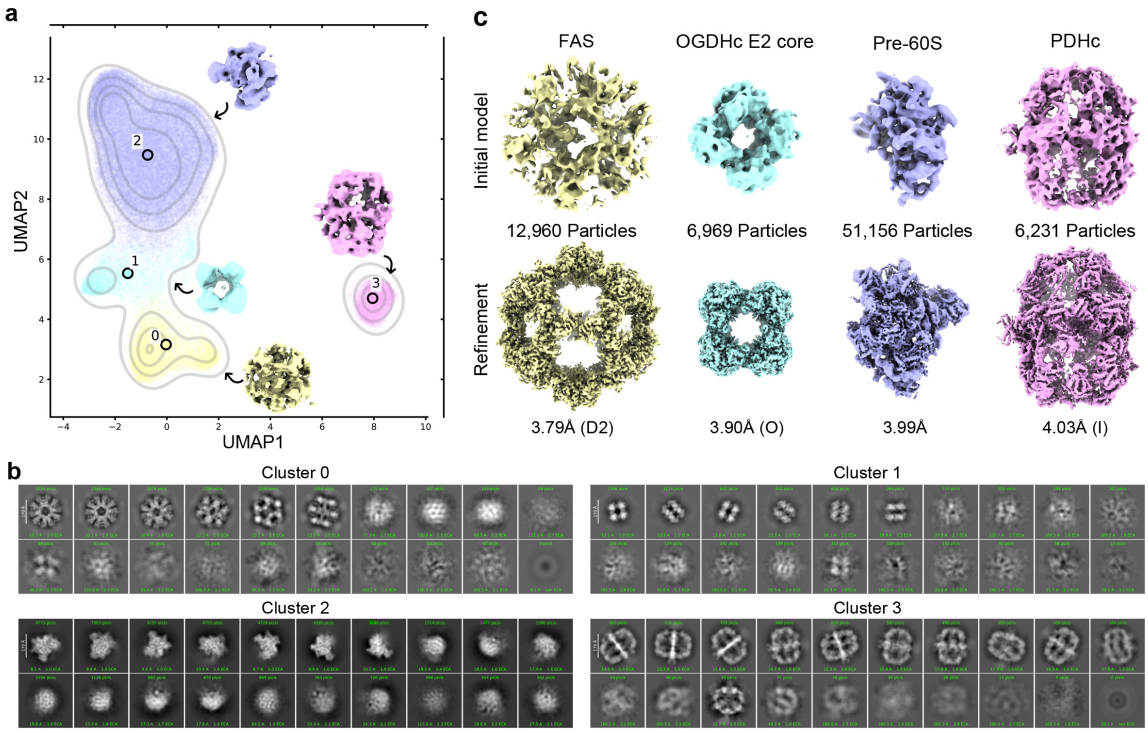


Fig. 4 | CryoDECO classifies endogenous protein communities from *C. thermophilum* native cell extracts. (**a**) UMAP visualization of the particle feature space extracted by CryoDECO. It demonstrates that the framework successfully deconvolves unpurified *C. thermophilum* lysates into four discrete clusters. Particles are clustered into four groups (color-coded) using Gaussian Mixture Model (GMM). Insets show the initial 3D structures generated by the decoder from the center of each feature cluster. (**b**) 2D class averages for each cluster of particles, confirming the homogeneity. (**c**) Final 3D reconstructions of the four clusters of particles after ab-initio reconstruction and non-uniform refinement in CryoSPARC. Structures are color-coded as in (**a**).

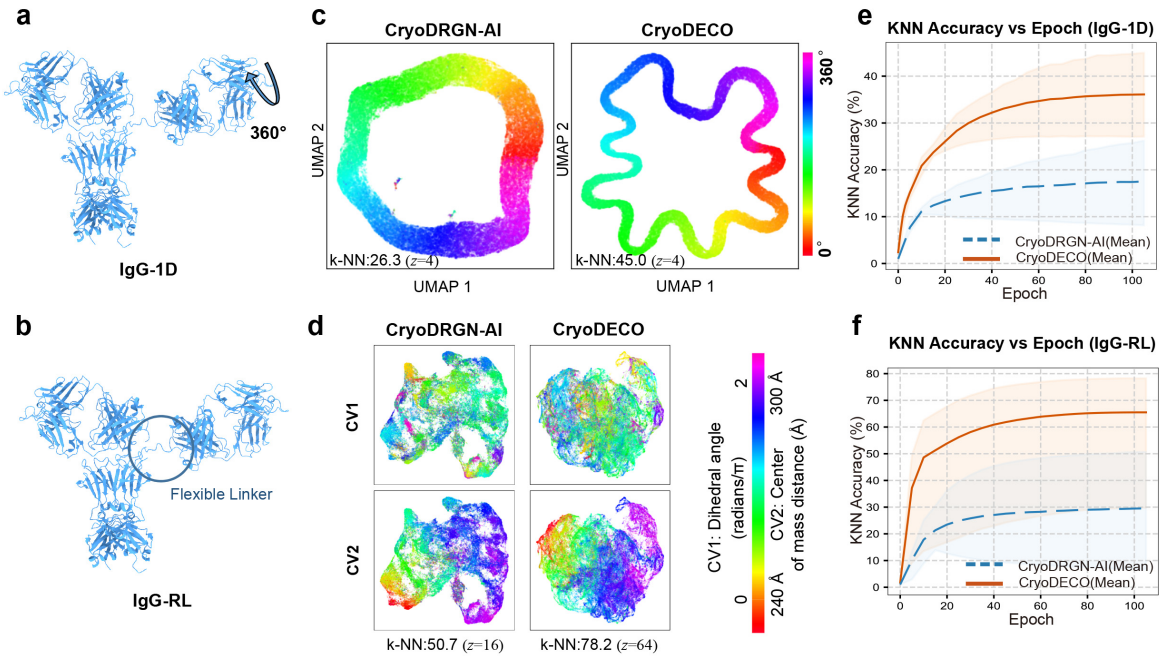


Fig. 5 | CryoDECO's classification performance on CryoBench datasets of conformational heterogeneity. (**a**) The IgG-1D dataset's heterogeneity is generated by rotating a dihedral angle in one Fab domain[27]. (**b**) The IgG-RL dataset's heterogeneity is generated by randomizing a linker region. Both datasets contain 100 conformations with 1,000 particles each (SNR = 0.01). (**c,d**) UMAP visualizations of particle feature spaces from IgG-1D (**c**) and IgG-RL (**d**), color-coded by conformation. Features from CryoDRGN-AI (left panels) are shown for comparison. Optimal latent particle feature dimensions (z) are indicated (IgG-1D: z = 4 for both models; IgG-RL: z = 16 for CryoDRGN-AI, z = 64 for CryoDECO). (**e,f**) k-NN scores during training for IgG-1D (**e**) and IgG-RL (**f**). Scores are averaged over runs (CryoDECO: orange line; CryoDRGN-AI: blue dashed line) with varying particle feature dimensions (z=4, 16, 32, 64, 128). Shaded areas represent the score range across all z dimensions.

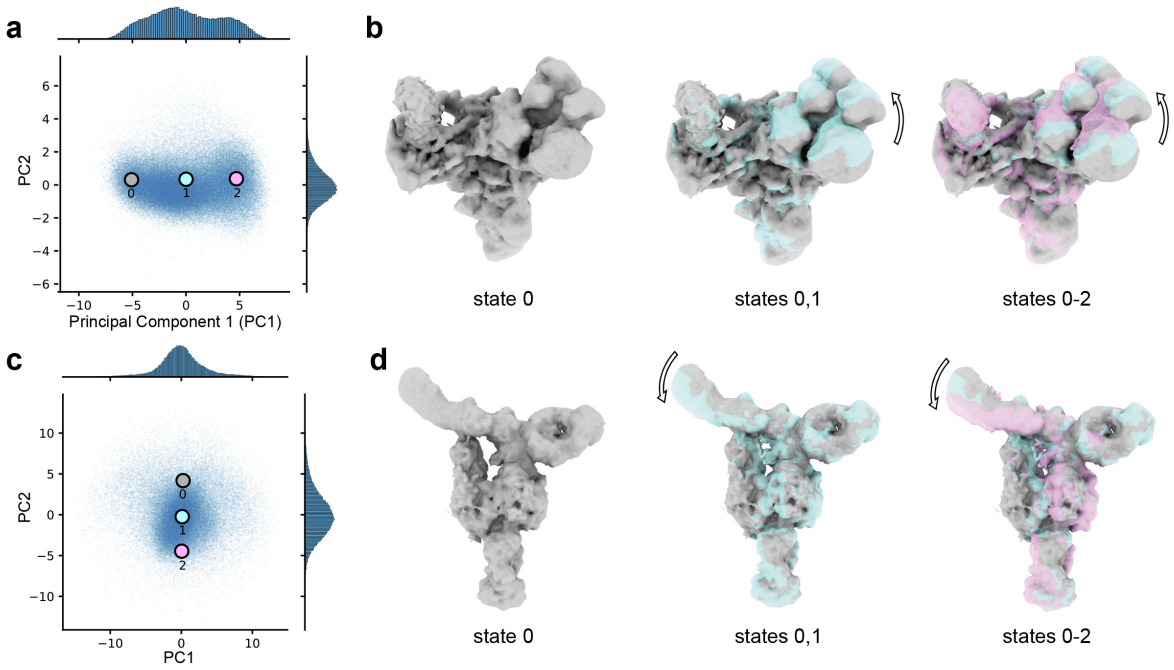


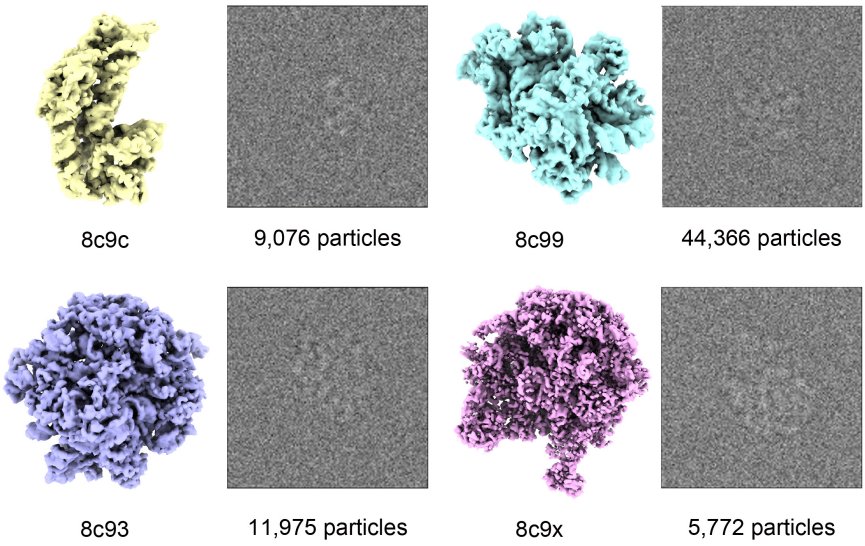
Fig. 6 | Analysis of the continuous conformational heterogeneity of tri-snRNP and *αVβ8* integrin datasets. (**a,c**) PCA visualization of the feature space for tri-snRNP (**a**) and *αVβ*8 integrin (**c**). Three points were selected along PC1 for tri-snRNP and PC2 for *αVβ*8 to represent distinct conformations (gray, cyan, violet dots). (**b,d**) Reconstructions from the selected data points for tri-snRNP (**b**) and *αVβ*8 (**d**), respectively. Data points are color-coded to indicate their different states. Superimposition of different states of structures are displayed to illustrate the sequential conformational changes. All 3D visualizations were created using ChimeraX.

Table 1 Clustering Accuracy

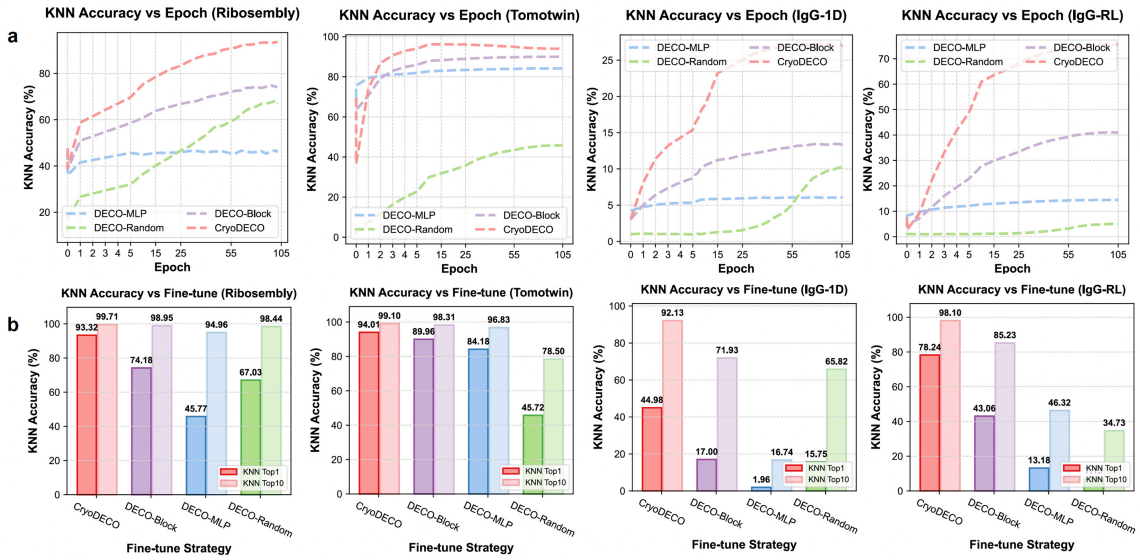
Method		Ribosembly		Tomotwin-100	
		ARI	AMI	ARI	AMI
With prior poses†	CryoDRGN[17]*	0.873	0.935	0.956	0.983
	CryoDRGN-AI-fixed[26]*	0.624	0.771	0.791	0.906
	Opus-DSD[20]*	0.891	0.934	0.500	0.781
	3DVA[16]*	0.666	0.823	0.058	0.335
	RECOVAR[21]*	0.968	0.976	0.315	0.649
	CryoDECO	**0.980**	**0.977**	**0.999**	**0.999**
ab initio‡	CryoDRGN2[25]*	0.529	0.618	0.116	0.374
	CryoDRGN-AI[26]*	0.644	0.729	0.086	0.275
	CryoDECO	**0.860**	**0.908**	**0.622**	**0.853**

† Reconstruction with fixed (ground truth) poses.
‡ *ab initio* reconstruction where no input poses are provided.
* Results from CryoBench paper[27].

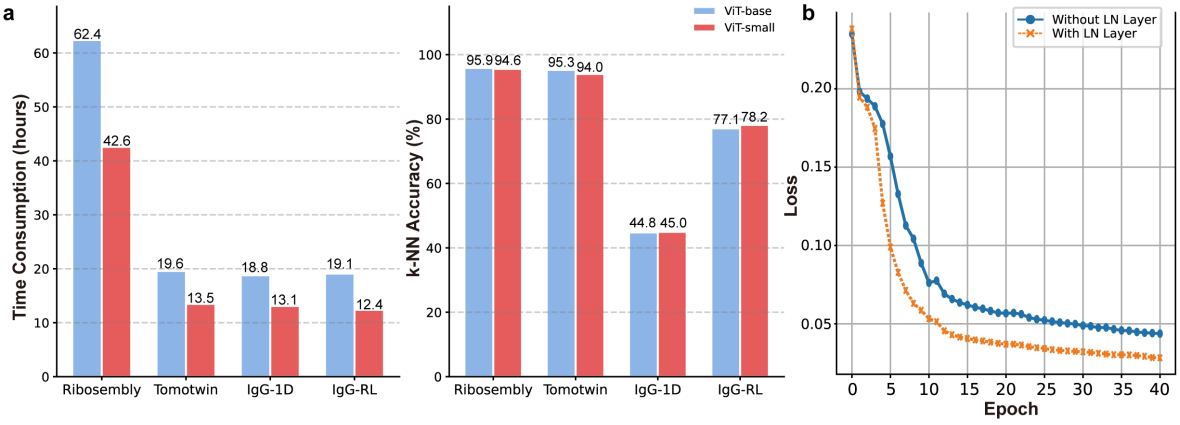
Extended Data Figures and Tables



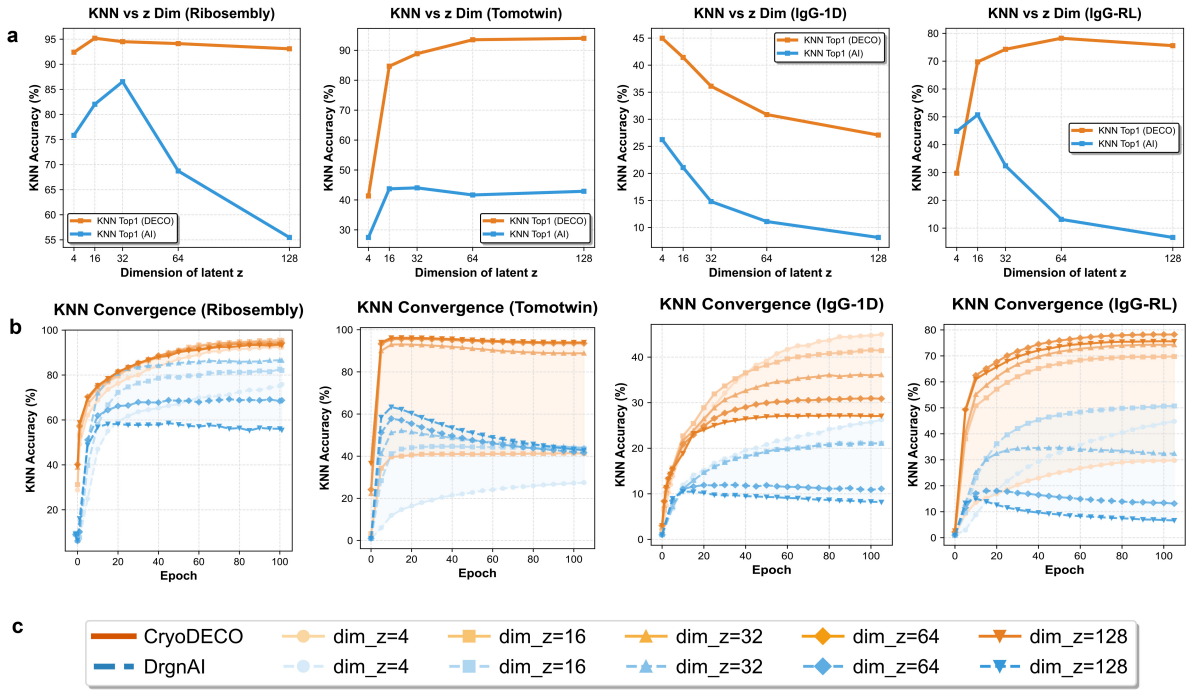
Extended Data Fig. 1 | The resampled CryoBench Ribosembly dataset. This dataset contains simulated particles from four distinct structures with SNR=0.01.



Extended Data Fig. 2 | Ablation study of different fine-tuning strategies on the encoder architecture. An ablation study and convergence analysis were conducted to evaluate the impact of different fine-tuning strategies on the encoder architecture of CryoDECO. Experiments were performed on four synthetic datasets: Ribosembly, Tomotwin, IgG-1D, and IgG-RL[27]. (**a**) k-NN classification accuracy convergence curves over epochs for CryoDECO under four different fine-tuning strategies on the encoder architecture: CryoDECO (fine-tuning the full network), DECO-Block (fine-tuning the last 3 transformer blocks), DECO-MLP (fine-tuning the projection head), and DECO-Random (randomly initializing the encoder). This panel illustrates the training dynamics of each strategy. (**b**) Final top-1 and top-10 k-NN classification accuracy and convergence dynamics of the CryoDECO model. The solid, more saturated bars represent top-1 accuracy, while the lighter, semi-transparent bars represent top-10 accuracy. This panel provides a direct comparison of the effectiveness of each strategy after training completion. The dimensions of the latent particle feature (z) are set to the same value as in Figs. 2 and 6.



Extended Data Fig. 3 | Ablation study on encoder scaling and decoder normalization in CryoDECO. (**a**) Performance comparison of encoder scaling across four datasets. The bar charts display the time consumption (left) and latent space quality (Top-1 k-NN accuracy, right) for ViT-base (blue bars) and ViT-small (red bars) encoders. Note that increasing the encoder size to ViT-base significantly raises the computational cost with marginal or negligible gains in classification accuracy compared to ViT-small. (**b**) Reconstruction loss profiles comparing CryoDECO with (orange dashed line) and without (blue solid line) layer normalization in the decoder. Training was performed on the cryoDRGN synthetic toy dataset[12] with 10 epochs for pose searching.



838

839 **Extended Data Fig. 4 | Effects of latent particle feature dimensionality (*z*) on the**

840 **performance of CryoDECO and CryoDRGN-AI.** A parameter analysis of the

841 CryoDECO (DECO) and CryoDRGN-AI (AI) models was conducted on the Ribosembly,

842 Tomotwin, IgG-1D, and IgG-RL synthetic datasets[27], specifically examining the effect of

843 different latent particle feature dimensions (*z*) on k-NN classification accuracy. (**a**) Top-1 k-

844 NN classification accuracy as a function of the latent particle feature dimension (*z*). For

845 each dataset, both models were trained and the final accuracies were evaluated. (**b**) Top-1

846 k-NN classification accuracy convergence curves over epochs for different latent particle

847 feature dimensions. The orange lines represent CryoDECO (DECO) and the blue lines

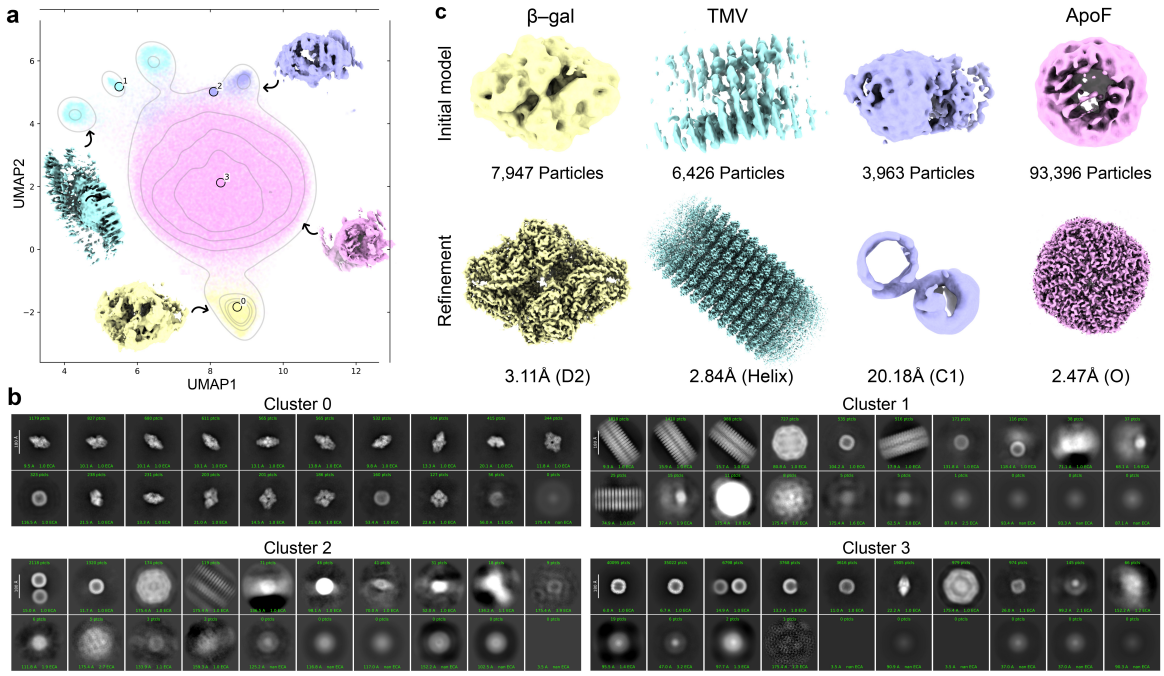
848 represent CryoDRGN-AI (AI). Lighter shades correspond to lower dimensions, while

849 darker shades correspond to higher dimensions, as detailed in the legend. (**c**) Legend

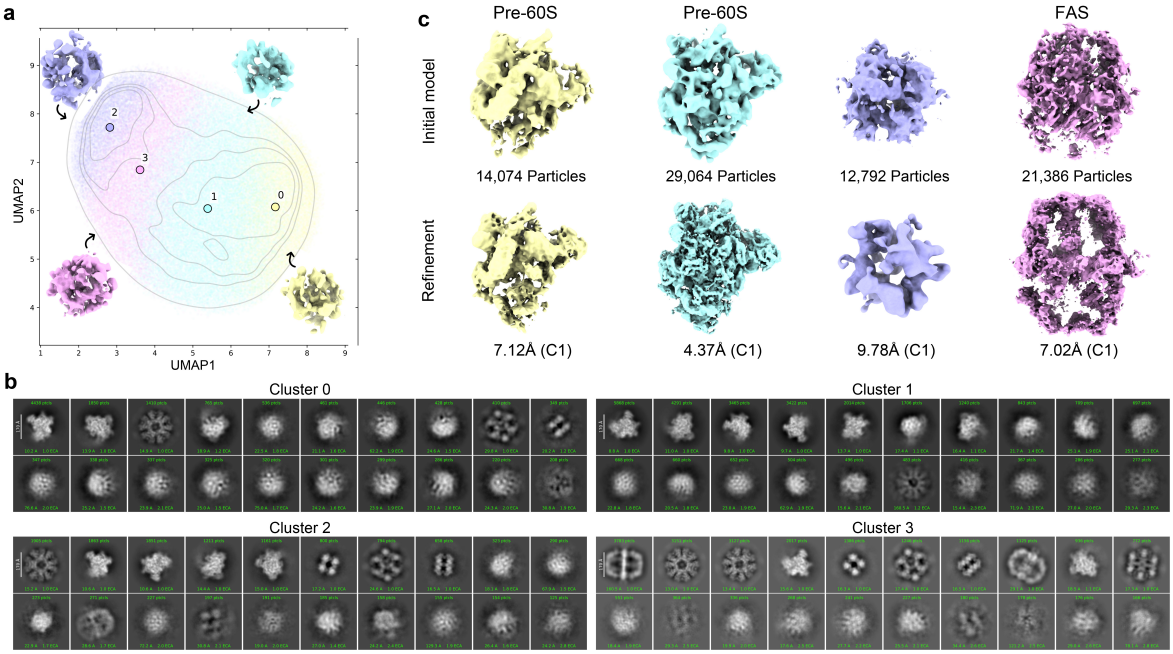
850 detailing the marker and color scheme used in panel (**b**) for different latent particle feature

851 dimensions.

852



Extended Data Fig. 5 | CryoDRGN-AI's classification performance on the EM ladder dataset. (**a**) UMAP visualization of the particle feature space extracted by CryoDRGN-AI. Particles were clustered into four groups (color-coded) using Gaussian Mixture Model (GMM). Insets show 3D structures generated by the decoder from cluster centers. Note the poor separation of clusters compared to Fig. 3a. (**b**) 2D class averages for each cluster, showing the mixing of particle types. (**c**) Final 3D reconstructions of the four clusters after ab-initio reconstruction and non-uniform refinement in CryoSPARC. Note that the process failed to reconstruct correct structure of PP7. All 3D visualizations were created using ChimeraX.



Extended Data Fig. 6 | CryoDRGN-AI's classification performance on the *C. thermophilum* native cell extracts. (**a**) UMAP visualization of the particle feature space. Particles were clustered into four groups (color-coded) using Gaussian Mixture Model (GMM). Insets show 3D structures generated by the decoder from cluster centers. Note the significant overlap between clusters compared to Fig. 4a. (**b**) 2D class averages for each cluster, confirming significant structural mixing. (**c**) Final 3D reconstructions of the four clusters. Note that the process failed to reconstruct correct structures of OGDHc E2 core and PDHc. All 3D visualizations were created using ChimeraX.

Extended Data Table 1 | Summary of Expanded Data Downloaded from EMPIAR for Cryo-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		