CryoKRAQEN: Kernel-Regularized Annealing for Quantized Embedding Networks in Cryo-EM Heterogeneous Reconstruction

Supplementary Material

Overview

This supplementary document expands the methodological background and empirical evidence of the main paper. Sec. 1 introduces fundamental cryo-EM formulation and the core challenges of heterogeneous reconstruction, while Sec. 4 provides concise definitions used throughout the work. Further implementation details—including network parameterization, training procedures, and dataset-dependent hyperparameters—are presented in Sec. 2 to support clear reproducibility and ensure fair comparisons across baselines.

Extended experimental results are provided in Sec. 3. Sec. 3.1 reports controlled evaluations on CryoBench datasets, covering FSC-based quantification and latent-space analysis, whereas Sec. 3.2 presents complementary experiments on EMPIAR real data. To further aid interpretation of structural variability, we additionally provide videos of our 3D heterogeneous reconstructions for each dataset on the project page, allowing readers to visually examine the recovered conformational transitions.

1. Basics of Cryo-EM Heterogeneous Reconstruction

1.1. Problem Formulation

Background. Single-particle cryo-EM seeks to recover a 3D electron density ρ from a collection of noisy 2D particle images. Let X_i denote the Fourier transform of the i-th particle image. By the central slice theorem, X_i corresponds to a 2D slice of the Fourier-domain density $\widehat{\rho}$ orthogonal to the viewing direction. The image-formation model is expressed as

$$X_i(\boldsymbol{\xi}) = \mathrm{CTF}_i \, \mathcal{S}(\phi_i) \, \widehat{\rho}(\boldsymbol{\xi}) + \varepsilon_i, \tag{16}$$

where $\phi_i = (R_i, t_i)$ is the particle pose, $\mathcal{S}(\phi_i)$ extracts the slice specified by orientation R_i, CTF_i is the contrast transfer function, and ε_i denotes acquisition noise.

Classical heterogeneous reconstruction. To account for conformational variability, traditional approaches approximate the underlying structural ensemble with K discrete density maps $\{\rho_1, \ldots, \rho_K\}$. Each image is associated with a latent class index, leading to the objective

$$\arg\max_{\{\rho_j\}} \sum_{i=1}^{N} \log \left(\sum_{j=1}^{K} \pi_{ij} \, p(X_i \mid \rho_j) \right) + \sum_{j=1}^{K} \log p(\rho_j), \tag{17}$$

where π_{ij} are the class-assignment probabilities. Although effective for well-separated discrete states, this formulation struggles when the underlying heterogeneity is continuous, high-dimensional, or involves coupled domain motions.

Challenges. Because biomolecular dynamics seldom reduce to a small number of discrete conformations, modeling structural variability using multiple independently reconstructed volumes is inherently restrictive. Hierarchical refinement pipelines require repeated reinitialization, manual selection of class numbers, and substantial computational resources. These challenges become especially acute in the presence of flexible regions or mixed molecular compositions, where subtle conformational signals are easily confounded by noise, making robust heterogeneous modeling particularly difficult.

1.2. Comparison to Homogeneous Reconstruction

Homogeneous cryo-EM reconstruction assumes all particle images arise from a single density ρ, allowing images to be averaged once orientations are estimated. In contrast, heterogeneous reconstruction must disentangle multiple structural states—potentially forming a continuous manifold—while simultaneously estimating poses and class assignments. The key differences include:

- **Latent structure variability:** homogeneous reconstruction models a single ρ, whereas heterogeneous reconstruction must capture discrete or continuous conformational changes.
- **Ambiguity in image assignments:** particle images must be associated with distinct structural states, dramatically increasing the inference complexity.
- **Sensitivity to noise and sampling:** structural variability can be subtle and is often overshadowed by low SNR, making reliable classification and refinement difficult.
- **Propagation of pose errors:** misestimated orientations can confound structural separation, whereas homogeneous refinement is more tolerant.

These distinctions highlight why simple averaging or single-state refinement fails under substantial conformational diversity, motivating algorithmic frameworks capable of modeling high-dimensional structural heterogeneity.

2. Experimental Settings

2.1. Baselines

We compare CryoKRAQEN against several representative heterogeneous reconstruction methods, including RECOVAR [30], 3DFlex [23], and 3DVA [22], following the evaluation protocol of CryoBench [11]. It is important to note that, according to CryoBench, the reported results of these baselines in Table 1 are obtained with a *dataset-specific focus mask*, which is generated by dilating aggregated volumes with structural priors. This masking strategy effectively restricts evaluation to structurally relevant regions and can lead to non-trivial differences in quantitative metrics. Therefore, when making comparisons, these implementation details should be taken into careful consideration to ensure a fair and consistent interpretation of performance.

2.2. Hyperparameters

CryoKRAQEN is implemented in PyTorch based on `cryoDRGN` version 3.0.0b. The Fourier-domain triplane codebook contains $K = 128$ embeddings for all datasets. Each codebook is processed by a 3-layer MLP with hidden width 1024. Reconstruction and vector-quantization losses are optimized jointly, with the latter weighted by a dataset-specific factor β (Table S1).

Dataset	$T_{\max}$	$T_{\min}$	β
IgG-1D	4.0	0.25	0.5
IgG-RL	4.0	0.25	0.5
Ribosembly	4.0	0.25	0.5
Tomotwin-100	8.0	0.25	0.1
Spike-MD	8.0	0.125	0.1

Table S1. Dataset-specific hyperparameters for soft assignment.

2.3. Model Efficiency and Computational Cost

To evaluate computational efficiency, we compare the parameter scale and training throughput of CryoKRAQEN with the widely adopted baseline CryoDRGN. Model sizes are reported by counting the total number of trainable parameters,

$$\sum_{p\in\Theta,\ p \text{ trainable}} \text{numel}(p),$$

following the standard convention used in prior work [31].

As summarized in Table S2, CryoKRAQEN adopts a significantly more compact architecture, containing only **4.21B** trainable parameters compared to CryoDRGN's **19.41B**, corresponding to a **4.6× reduction**. This reduction arises from our streamlined deformation module and lightweight latent-field decoder.

Training speed is benchmarked on a single NVIDIA A40 GPU using identical batch configurations. CryoKRAQEN requires only **0.3588 s** to process each 100-image batch, compared to **0.4929 s** for CryoDRGN, yielding a **27.2% improvement in throughput**. This demonstrates that our model achieves higher computational efficiency without sacrificing reconstruction quality.

Overall, CryoKRAQEN offers a substantially more efficient design—both in parameter count and per-batch compute time—making it more suitable for large-scale or high-throughput cryo-EM studies.

Method	**Parameters**	**Time / 100 images (s)**
CryoDRGN	19,414 M	0.4929
Ours	4,211 M	0.3588

Table S2. Comparison of model efficiency between CryoKRAQEN and CryoDRGN. The reported parameter and runtime statistics are computed using the IgG-1D dataset as a representative example, illustrating the substantially improved computational efficiency of CryoKRAQEN while maintaining reconstruction fidelity.

3. Additional Experimental Results

This section reports extended results on synthetic and real datasets, complementing the main paper. The supplementary package also includes a `videos/` directory, where each dataset-specific animation illustrates the reconstructed 3D heterogeneity for convenient visual inspection.

3.1. Synthetic Dataset Experiments

Additional synthetic dataset results are provided here, including detailed comparisons with baselines across all CryoBench benchmarks. We show per-image FSC distributions, volume slices, and latent space interpolations for each dataset.

Visualization of Representative Reconstructions. Figures S3–S4 illustrate CryoKRAQEN reconstructions alongside baseline methods for three representative datasets: IgG-RL, Ribosembly, and Spike-MD. For IgG-RL (Fig. S3), CryoKRAQEN successfully recovers both fixed and flexible Fab regions despite their random orientations, outperforming other methods in overall FSC and local structural fidelity. In Ribosembly (Fig. S1), CryoKRAQEN produces latent embeddings that more clearly separate ground-truth assembly states, enhancing neighborhood consistency and volumetric accuracy relative to classical 3D Class approaches. For Spike-MD (Fig. S4), CryoKRAQEN captures detailed domain orientations and subtle conformational features, maintaining global structural coherence and closely matching the molecular dynamics-derived ground truth.

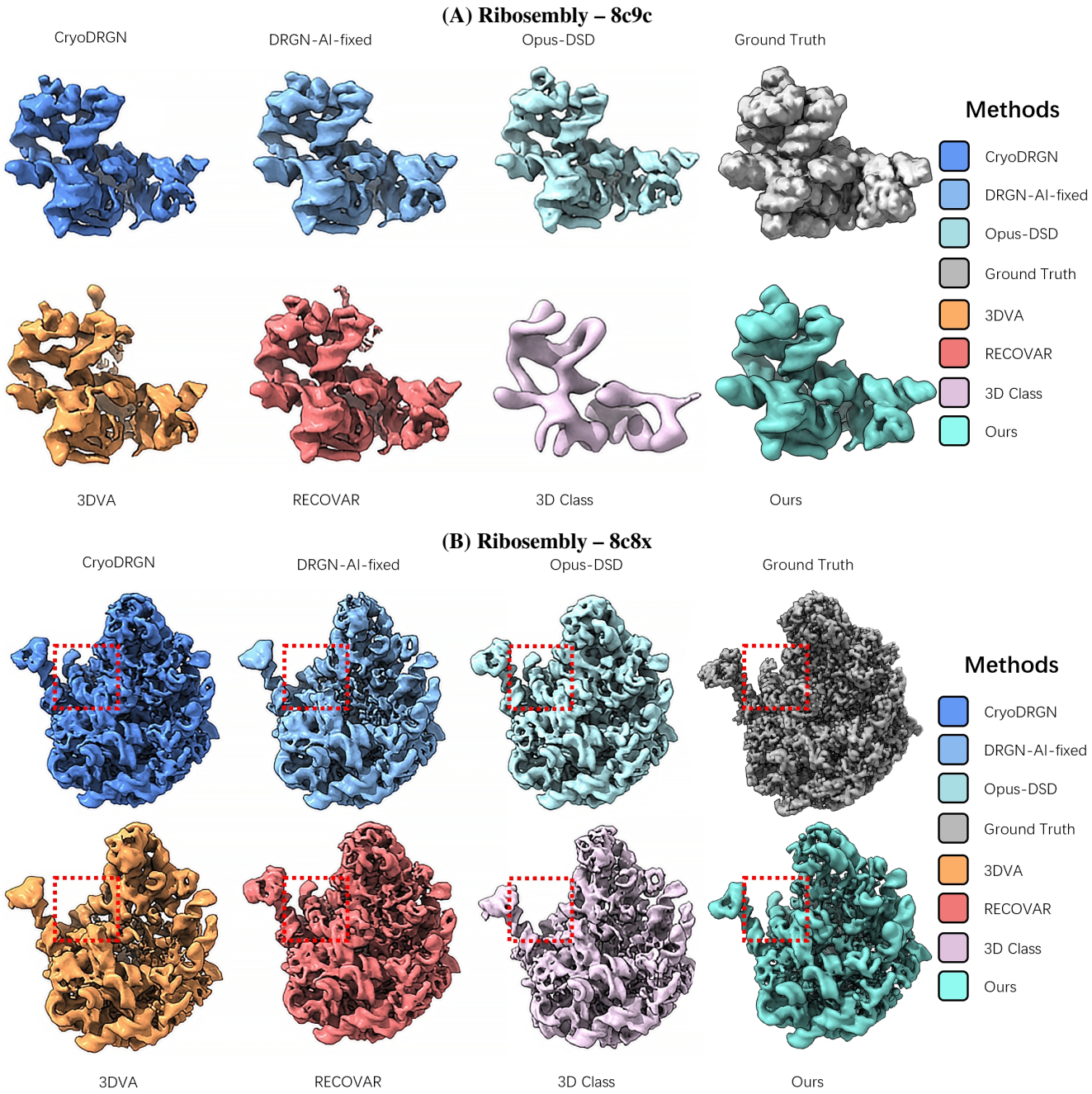


Figure S1. **Comparison on Ribosembly (compositional heterogeneity).** Visualization of CryoKRAQEN and baseline methods across two representative Ribosembly conformations. Most approaches successfully reconstruct the underlying assemblies; however, latent embeddings from several baselines do not fully separate the ground-truth assembly states. CryoKRAQEN demonstrates improved clustering structure and enhanced neighborhood consistency, effectively capturing compositional structural variations. Notably, the α-helical segment highlighted by the red dashed box is either missing or severely degraded in baseline reconstructions, whereas CryoKRAQEN preserves this flexible structural element with substantially higher fidelity.

Analysis of FSC Comparisons. Across the four CryoBench datasets, CryoKRAQEN achieves higher FSC values than cryoDRGN in four cases, demonstrating improved reconstruction fidelity and better capture of conformational and compositional heterogeneity, as shown in Figure S6. The only exception is Spike-MD, where CryoKRAQEN shows a slight decrease (-2.16%) relative to cryoDRGN. We note that the CryoDRGN results shown here were obtained from our own runs using the official CryoDRGN 3.0.0b implementation and exhibit minor differences compared to the

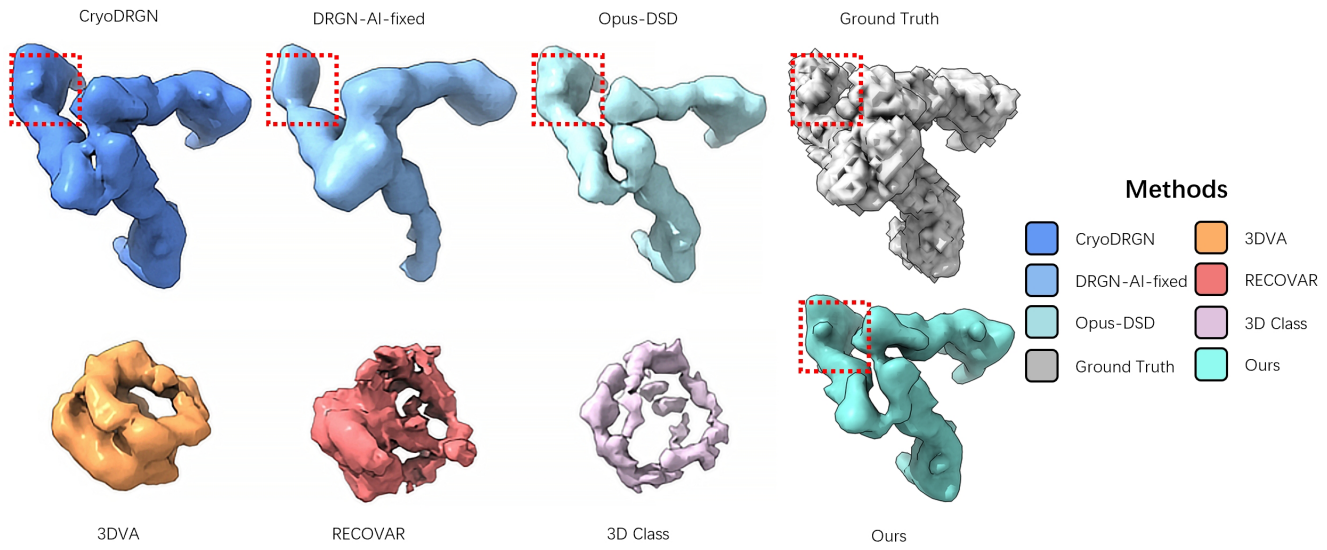


Figure S2. **Comparison on Tomotwin-100 (compositional heterogeneity).** Qualitative evaluation of CryoKRAQEN alongside baseline methods on the Tomotwin-100 synthetic dataset, which contains 100 distinct macromolecular complexes. Classical approaches (3DVA, RECOVAR, and 3D Class) often produce collapsed or severely distorted reconstructions that deviate substantially from the ground-truth structures. Neural reconstruction baselines such as CryoDRGN and DRGN-AI also exhibit limited effective resolution, frequently generating oversmoothed density maps in flexible or low-occupancy regions, as highlighted in the red dashed boxes. CryoKRAQEN preserves fine structural details while maintaining clear separation among heterogeneous components, demonstrating strong robustness and accuracy under complex multi-component mixtures.

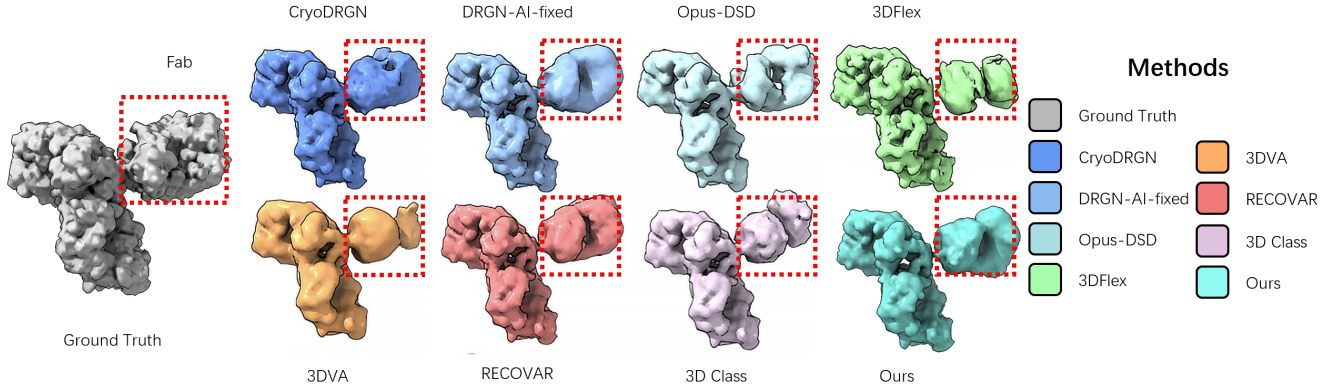


Figure S3. **Comparison on IgG-RL (conformational heterogeneity).** Visualization of CryoKRAQEN and baseline methods on a representative IgG-RL conformation. The flexible Fab arms (highlighted by red dashed boxes) exhibit large orientation variability, making high-resolution recovery particularly difficult. CryoKRAQEN accurately reconstructs both the rigid core and these flexible regions, preserving domain orientations and local structural details.

FSC values reported in CryoBench.

Visualization of Latent Space To further evaluate the quality of the learned structural manifold, we visualize the latent embeddings using UMAP across five Cryo-EM datasets (Fig. S5). The upper row shows embeddings from CryoDRGN, while the lower row displays those generated by CryoKRAQEN.

Across IgG-1D, IgG-RL, Ribosembly, Tomotwin-100, and Spike-MD, CryoKRAQEN produces noticeably more structured and physically meaningful latent spaces. For datasets characterized by continuous flexibility (e.g., IgG variants, Spike-MD), our model captures smooth trajectories corresponding to conformational continua. These trajectories reflect the annealing evolution of the neural deformation field—importantly, they emerge *without* imposing any Gaussian prior on the latent variables. As a consequence, CryoKRAQEN does not artificially constrain the manifold geometry, in contrast to CryoDRGN, whose latent embeddings (e.g., Ribosembly) often form nearly circular clusters due to the influence of isotropic Gaussian priors.

For compositional heterogeneity datasets (e.g., Ri-

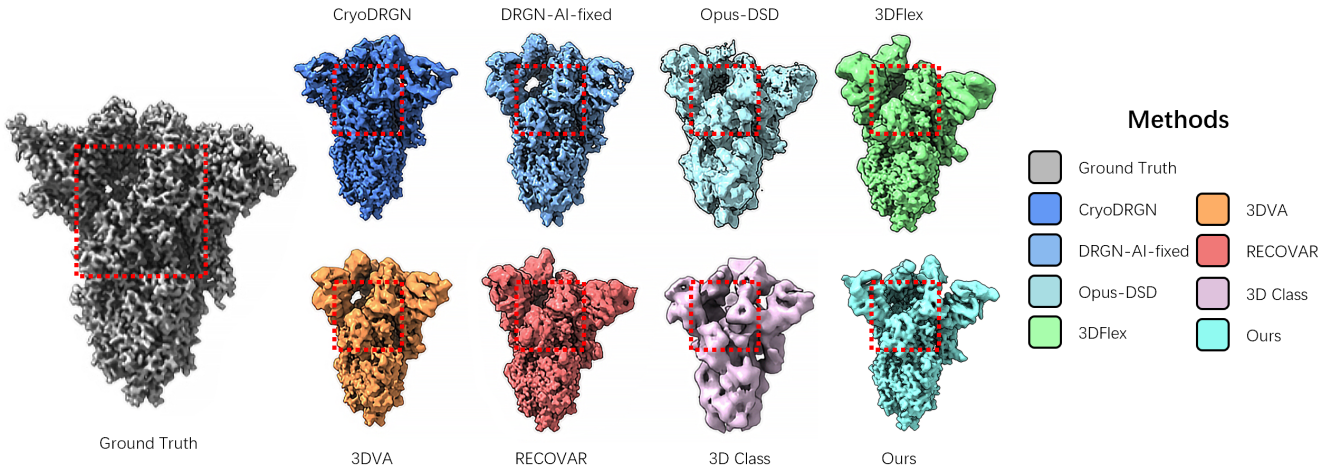


Figure S4. **Comparison on Spike-MD.** Visualization of CryoKRAQEN and baseline methods on a representative Spike-MD conformation, derived from molecular dynamics simulations. Classical methods and existing neural approaches (e.g., CryoDRGN, DRGN-AI) exhibit noticeable artifacts in the regions highlighted by the red dashed boxes, including blurred domain boundaries and over-smoothed local features. CryoKRAQEN, by contrast, captures fine-grained domain motions and subtle conformational variations, producing a reconstruction that closely aligns with the ground-truth density while maintaining both local detail and global structural coherence.

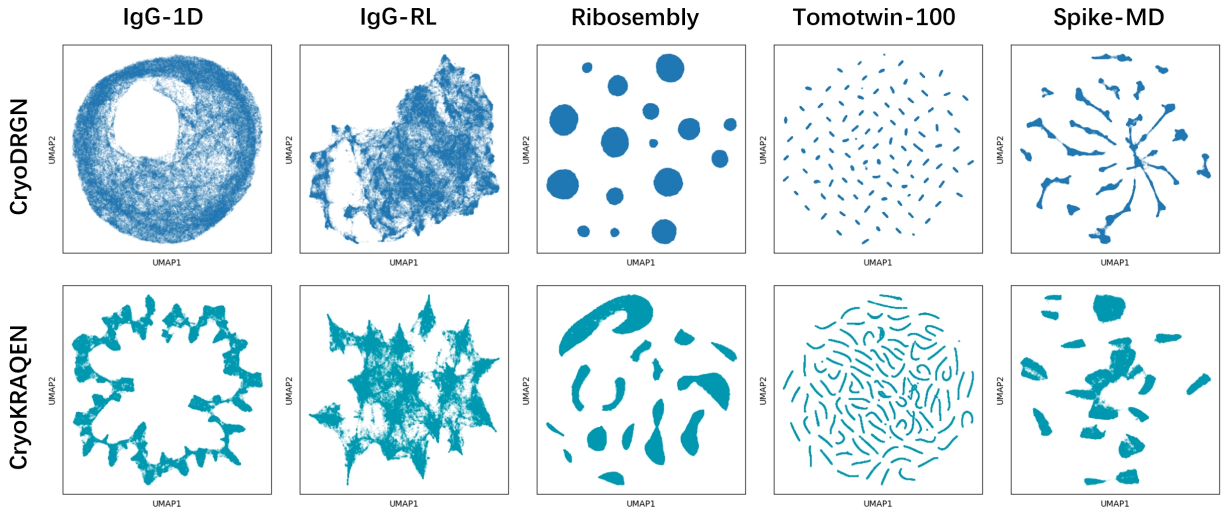


Figure S5. **Latent embedding comparison across five Cryo-EM datasets.** The top row shows embeddings generated by CryoDRGN, while the bottom row shows embeddings from CryoKRAQEN. Each column corresponds to a different dataset: IgG-1D, IgG-RL, Ribosembly, Tomotwin-100, and Spike-MD. CryoKRAQEN produces more structured latent spaces with better separation of conformational states, demonstrating improved representation of heterogeneous structural variability relative to CryoDRGN.

bosembly, Tomotwin-100), CryoKRAQEN further achieves cleaner separation between discrete structural states. The clusters adopt irregular, data-dependent shapes that reflect genuine structural relationships rather than prior-driven regularization. This indicates that CryoKRAQEN more faithfully preserves the topology of the underlying data distribution, allowing both continuous motions and discrete states to be embedded in a unified manifold without distortion.

Overall, these results highlight that our model learns a more disentangled and physically interpretable latent space: continuous conformations form coherent trajectories, discrete states remain well separated, and no artificial circularity or isotropy is introduced. This leads to improved structural representation and downstream reconstruction fidelity relative to CryoDRGN.

Ablation Study on Compositionally Heterogeneous Dataset. As shown in Table S3, the full method consistently outperforms all ablated variants. To properly evaluate the contributions of the kernel component and the triplet loss in discrete settings, we conduct controlled ablations by selectively removing each component. The results show that both the kernel design and the triplet loss remain effec-

tive in the discrete regime, rather than being limited to continuous settings, and contribute substantially to the overall performance. This highlights their robustness and complementary roles in improving representation quality and reconstruction fidelity.

Model Variant	Mean↑	Median↑
Full CryoKRAQEN	**0.335**	**0.337**
w/ Gaussian kernel	0.320	0.321
w/o kernel	0.272	0.271
w/o triplet loss	0.246	0.253

Table S3. **Ablation study on the Tomotwin-100 dataset.**

3.2. Real Dataset Experiments

To evaluate robustness under real experimental conditions, we test CryoKRAQEN on two challenging in-the-wild cryo-EM datasets: EMPIAR-10180 [28] and EMPIAR-10028 [21].

For EMPIAR-10180 (Fig. S7), CryoKRAQEN reconstructs a coherent density map with well-preserved global structure and identifiable local features. The method produces smooth conformational transitions and stable domain shapes, demonstrating strong generalization from synthetic and semi-synthetic datasets.

For EMPIAR-10028 (Fig. S8), CryoKRAQEN reconstructs multiple density maps corresponding to both outlier particles (map 0) and three distinct conformational states (maps 1–3) of the pre-catalytic spliceosome. The reconstructions capture continuous bending motions and preserve both global structural integrity and local domain details, effectively separating heterogeneous conformations in the latent space. Black outlines indicate the extended spliceosome for reference.

These results demonstrate that CryoKRAQEN generalizes well to real cryo-EM datasets, handling particle heterogeneity and noise while producing interpretable and biologically meaningful density maps across diverse experimental conditions.

3.3. Clustering Evaluation and Codebook Size

As shown in Table S4, our method achieves consistently high ARI and AMI on both datasets. With a moderate codebook size ($K = 128$), annealing automatically prunes redundant codes in a data-dependent manner: Ribosembly (16 structures) exhibits a 57.03% dead-code rate with global entropy 4.98 (max 7.00), while Tomotwin-100 (100 structures) shows a lower dead-code rate of 20.31% with higher entropy 6.66. These statistics indicate that assignments neither collapse onto a few codes nor spread uniformly, but instead adapt to dataset complexity, resulting in stable and non-degenerate usage of active codes. Consistent with this behavior, Table S5 shows that the codebook size affects performance: overly small K limits representation capacity, while excessively large K hampers convergence. Setting K slightly above the expected number of structures or key states provides modest redundancy, while the annealing process adaptively prunes the codebook, reducing sensitivity to the exact choice of K.

Method	Ribosembly		Tomotwin-100	
	ARI↑	AMI↑	ARI↑	AMI↑
Ours	**0.886**	**0.958**	**0.981**	**0.994**
CryoDRGN	0.873	0.935	0.956	0.983

Table S4. **Clustering evaluation.**

K	IgG-1D	Tomotwin-100
64	0.362 (0.007)	0.210 (0.070)
128	**0.375** (0.004)	**0.335** (0.021)
256	0.367 (0.003)	0.328 (0.024)

Table S5. **Effect of codebook size K on AUC-FSC.**

4. Cryo-EM Terminology and Definitions

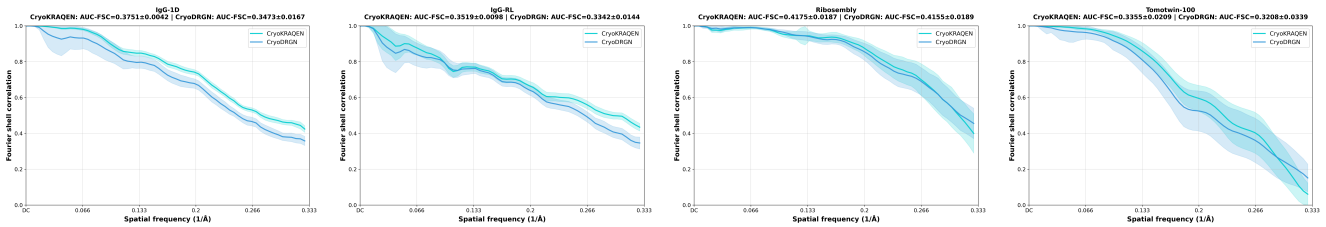
4.1. Sample

Cryo-EM samples consist of vitrified biological macromolecules whose intrinsic properties and organizational forms influence particle variability and reconstruction difficulty.

Biomolecules. Biomolecules are the primary subjects imaged in cryo-EM, forming the structural units from which particles arise. They mainly include:

- **Proteins:** Amino-acid polymers that fold into functional 3D structures. Their intrinsic flexibility—such as loop motions or domain rearrangements—introduces continuous structural variation across particles.
- **Nucleic acids:** RNA and DNA polymers with diverse folds. Their relatively low density results in lower image contrast, making particle identification and alignment more challenging.

Complex. Macromolecular complexes consist of multiple biomolecular components. Dynamic assembly, partial occupancy, or variable stoichiometry introduce compositional differences, while large-scale structural changes lead to conformational diversity. Complexes therefore serve as central subjects in heterogeneity analysis and reconstruction.



(a) **IgG-1D.** CryoKRAQEN (0.37508) outperforms cryoDRGN (0.34734) with a +7.99% improvement.

(b) **IgG-RL.** CryoKRAQEN (0.35187) improves over cryoDRGN (0.33425) by +5.27%.

(c) **Ribosembly.** CryoKRAQEN (0.41753) slightly outperforms cryoDRGN (0.41552) with +0.49%.

(d) **Tomotwin-100.** CryoKRAQEN (0.33545) exceeds cryoDRGN (0.32076) by +4.58%.

Figure S6. **FSC comparison results across datasets.** CryoKRAQEN consistently matches or outperforms cryoDRGN across different types of structural heterogeneity, demonstrating improved reconstruction fidelity and robustness.

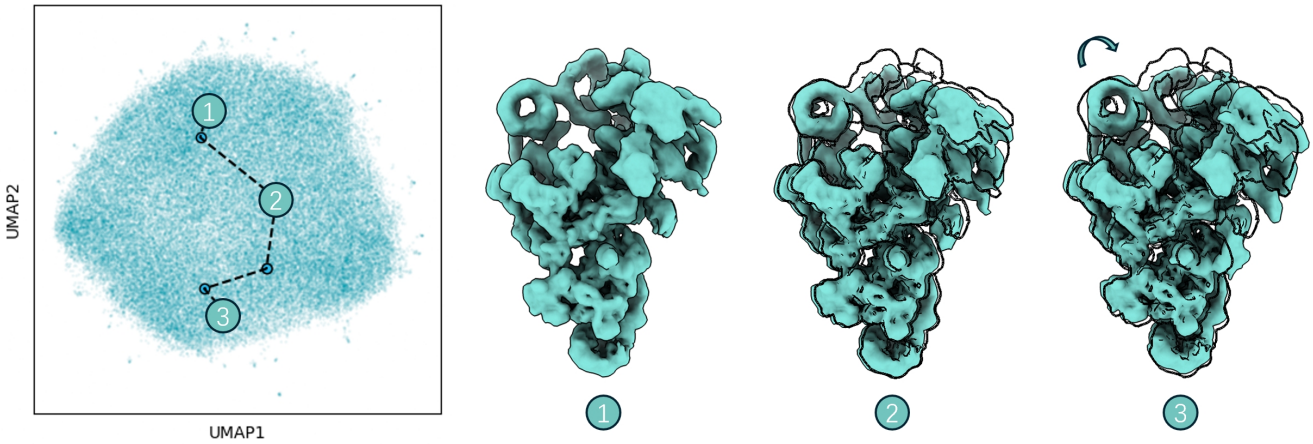


Figure S7. **Heterogeneity reconstruction on the EMPIAR-10180 Dataset (pre-catalytic spliceosome).** Qualitative reconstruction results of CryoKRAQEN on the real cryo-EM dataset EMPIAR-10180, which contains the pre-catalytic spliceosome exhibiting a known continuous bending motion. Prior work (e.g., DRGN-AI) reports that this dataset displays a smooth conformational trajectory—visible as a bending transition across reconstructed states—while particle images corresponding to broken or denatured molecules tend to form a distinct cluster in latent space. CryoKRAQEN similarly recovers the global bending motion while providing sharper and more stable density in flexible regions. The three representative density maps illustrate a coherent progression of the bending conformation. Outlines of the extended spliceosome (black contours) are overlaid to highlight the deformation path and to demonstrate that CryoKRAQEN maintains global architectural integrity alongside improved local structural fidelity.

4.2. Heterogeneity

Heterogeneity refers to the structural variability observed across particles in cryo-EM datasets.

Conformation. A conformation is a specific 3D arrangement adopted by a biomolecule or complex. Biomolecules naturally fluctuate among multiple conformations, forming ensembles of related structural states observable across particles.

Heterogeneity. Heterogeneity refers to structural variability observed across particle images, arising because different particles may not represent the same underlying 3D structure. This variability complicates both alignment and reconstruction. Two major forms are commonly analyzed:

- **Conformational heterogeneity:** Variability caused by changes in molecular shape, including continuous or discrete motions such as hinge rotations, domain bending, or global rearrangements within a complex.
- **Compositional heterogeneity:** Variability arising from differences in molecular composition, such as missing subunits, partial occupancy, additional RNA fragments, or alternative binding partners present only in subsets of particles.

4.3. Representation

Biomolecular structures are represented as 3D models, enabling reconstruction from 2D images.

Volume. A discrete 3D voxel grid representing the structural density of the molecule. Each voxel encodes local contributions to the overall shape.

Density Map. A continuous scalar field over 3D space, typically reflecting Coulombic potential. Density maps

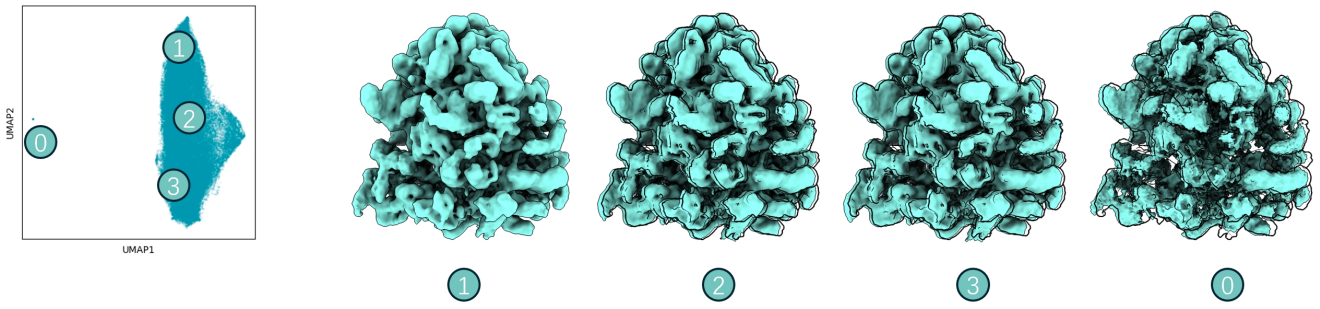


Figure S8. **Heterogeneity reconstruction on the EMPIAR-10028 Dataset (Pf80S ribosome).** CryoKRAQEN recovers multiple conformational states and outlier particles. Density map 0 corresponds to outlier particles (broken or denatured), while maps 1–3 capture distinct conformations of the Pf80S ribosome. All density maps are contoured consistently to enable direct visual comparison. The results demonstrate CryoKRAQEN's ability to separate heterogeneous conformations, preserve local and global structural features, and recover continuous motions without imposing an explicit Gaussian prior in the latent space, consistent with heterogeneity analyses reported for previous work.

capture both high-resolution details and broader conformational states.

Triplane. A factorized representation storing features on three orthogonal planes. It reduces memory use and computational cost while retaining spatial expressiveness for 3D reconstruction.

4.4. Data and Analysis

Key cryo-EM data terms and metrics for evaluating reconstruction quality, structural variability, and latent-space representations.

Particle. A single-molecule image extracted from micrographs, typically centered on one biomolecule or complex. Particles form the fundamental inputs for 3D reconstruction.

Reconstruction. A 3D volumetric map estimated from many particle images, representing either a homogeneous state or a heterogeneous ensemble.

Fourier Shell Correlation (FSC). A frequency-dependent measure of similarity between reconstructed and reference volumes, providing a standard estimate of structural resolution.

Latent-space visualization. Dimensionality reduction methods such as PCA or UMAP reveal continuous conformational trajectories and clusters corresponding to discrete compositional states.

Neighborhood similarity and information imbalance. Metrics that quantify how well local particle relationships are preserved in latent space, indicating the quality of learned heterogeneity structure.

5. Reproducibility and Data Availability

5.1. Code Release

The source code for this work will be released after publication to facilitate reproduction and benchmarking.

5.2. Data Availability

Synthetic datasets are publicly available through the CryoBench repository [11], while real datasets are accessible via the EMPIAR [10] database. All preprocessing scripts and metadata used in this work will be made available with the code release.