CRYOHYPER: TOWARD RELIABLE CONFORMATIONAL SPACES IN CRYO-EM AND CRYO-ET

Anonymous authors
Paper under double-blind review

ABSTRACT

Resolving conformational heterogeneity is essential for understanding the functional mechanisms of macromolecular complexes. Existing continuous heterogeneity methods can generate smoothly varying structures but do not necessarily organize them according to structural relationships. As a result, structurally dissimilar conformations may become neighbors in the learned representation, while distinct motions can become entangled, complicating recovery of continuous conformational changes. This problem is compounded by noise and missing-wedge effects in cryo-electron tomography (cryo-ET), hindering atomic conformational analysis. We present CryoHyper, a framework for learning reliable conformational spaces from cryo-electron microscopy (cryo-EM) and cryo-ET. Guided by a reference atomic model, CryoHyper models domain motions in a hyperbolic subspace and local deformations in a Euclidean subspace. The learned conformational coordinates condition density generation, placing atomic models and density maps in the same conformational coordinate system. For cryo-ET, the framework combines learnable spectral weighting with reconstruction training using random rotations and artificial wedge masking to reduce the influence of noise and missing-wedge effects on conformational inference. Experiments on simulated and public datasets show that CryoHyper recovers physically plausible conformational changes and organizes them in closer agreement with the underlying structural relationships.

1 INTRODUCTION

Macromolecular complexes rarely function as single static structures. Instead, they undergo coordinated conformational changes during molecular recognition, catalysis, membrane transport, and macromolecular assembly (Nussinov et al., 2023; Toader et al., 2023). Conformational heterogeneity analysis should therefore recover not only a collection of possible structures, but also the relationships among them. In a reliable conformational space, nearby positions should represent structurally similar states, and continuous paths should follow ordered structural changes. Training data typically lack conformation labels, and image reconstruction alone does not enforce this correspondence. A model may generate smoothly varying structures while its latent distances remain affected by entangled motions or imaging variations. As a result, latent proximity may not faithfully represent structural differences (Jeon et al., 2024).

Cryo-electron microscopy (cryo-EM) and cryo-electron tomography (cryo-ET) are powerful tools for studying macromolecular conformational heterogeneity. Single particle cryo-EM uses large collections of particle images to reconstruct structures at high resolution, whereas cryo-ET obtains three-dimensional subtomograms from tilt series and enables macromolecules to be studied in environments closer to their native cellular context (Cheng, 2018; Bäuerlein & Baumeister, 2021). Recent learning methods infer a latent coordinate for each particle and decode it into a density map or structural change, substantially advancing continuous heterogeneity analysis (Zhong et al., 2021a; Punjani & Fleet, 2023; Toader et al., 2023). Methods based on reference atomic models further discourage implausible bond geometry, interatomic distances, and steric clashes (Li et al., 2024; Zhang et al., 2025). Nevertheless, constraining decoded structures does not by itself ensure that distances between latent coordinates reflect conformational differences. This challenge is greater in cryo-ET, where lower signal quality, fewer particles, and the missing-wedge (Radermacher, 1988) further complicate the extraction of conformational information (Turk & Baumeister, 2020). Cryo-ET still

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

lacks a method that integrates atomic structural constraints into conformational space learning while accounting for noise and missing-wedge effects.

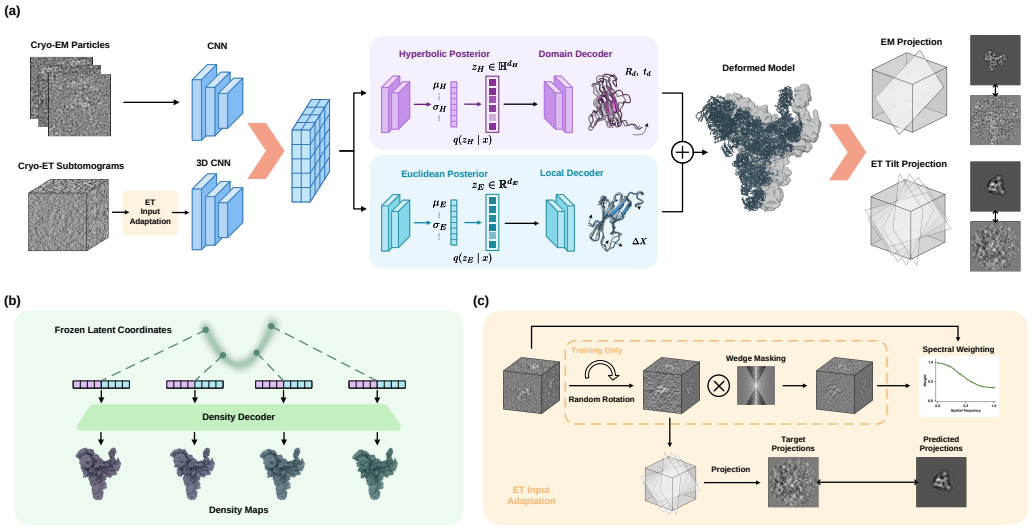


Figure 1: Overview of CryoHyper. (a) The atomic stage learns conformational coordinates from cryo-EM particle images or cryo-ET subtomograms, using hyperbolic and Euclidean subspaces for domain-level rigid motion and local atomic deformation, respectively, with projection supervision. (b) The density stage fixes the learned coordinates and generates atom-constrained density maps in the same conformational space. (c) ET input adaptation combines learnable spectral weighting with reconstruction training using random rotations and artificial wedge masking to reduce noise and missing-wedge effects.

CryoHyper learns a reliable conformational space whose latent distances reflect structural differences (Figure 1). Starting from a reference atomic model, CryoHyper jointly learns atomic deformations in hyperbolic and Euclidean subspaces. The hyperbolic and Euclidean subspaces represent domain motions and local atomic changes, respectively, together providing an inductive bias for organizing conformational variation at different scales. The resulting coordinates condition a density decoder, allowing atomic models and density maps to describe conformational variation in the same coordinate system. The framework accepts both cryo-EM particle images and cryo-ET subtomograms. For cryo-ET, ET Input Adaptation combines learnable spectral weighting with reconstruction training using random rotations and artificial wedge masking to reduce the influence of noise and missing-wedge effects on conformational inference. Experiments on simulated and public datasets show that CryoHyper recovers physically plausible conformational changes, learns latent distances that better agree with structural relationships, and supports particle selection for downstream refinement.

Our main contributions are summarized below.

- We introduce a joint hyperbolic and Euclidean representation guided by atomic structure. It captures large domain motions and local atomic changes while making latent distances more faithful to structural relationships between conformations.
- We apply this representation to both cryo-EM particle images and cryo-ET subtomograms, introducing ET Input Adaptation with learnable spectral weighting and artificial wedge masking during reconstruction training to reduce noise and missing-wedge effects on conformational inference.
- We construct a simulated cryo-ET benchmark with a known continuous conformational trajectory, enabling quantitative evaluation of conformational recovery and latent distances under limited particle counts, varying noise levels, and different tilt ranges.
- We establish a workflow linking atomic motion analysis, particle selection, and downstream refinement, providing more reliable structural evidence for conformational analysis of experimental cryo-EM and cryo-ET data.

2 Related Work

Continuous Heterogeneity Reconstruction. Classical cryo-EM analysis describes heterogeneity through discrete three-dimensional classification (Scheres, 2012) or dominant directions of structural variation (Punjani & Fleet, 2021). cryoDRGN (Zhong et al., 2021a) learns an implicit density representation, and cryoDRGN2 (Zhong et al., 2021b) incorporates pose estimation. OPUS-DSD (Luo et al., 2023) learns a continuous latent space through structural disentanglement. 3DFlex (Punjani & Fleet, 2023) describes motion through deformation fields, whereas RECOVAR (Gilles & Singer, 2025) estimates continuous heterogeneity using regularized covariance estimation and kernel regression. In cryo-ET, cryoDRGN-ET (Rangan et al., 2024) and OPUS-Tomo (Luo et al., 2024) learn heterogeneous density representations from tilt series and subtomograms, respectively. These approaches enable continuous structural recovery, but reconstruction quality alone does not ensure that latent distances reflect structural differences. CryoHyper focuses on this correspondence when learning conformational spaces.

Atomic Prior Conformational Modeling. Atomic structures constrain plausible molecular motion. Normal mode analysis (Tama et al., 2004) describes collective motions, while flexible fitting (Topf et al., 2008; Trabuco et al., 2008) fits atomic models to density maps. CryoSTAR (Li et al., 2024) learns structurally regularized atomic deformations and conditions density generation on the inferred conformational coordinates. cryoSPHERE (Ducrocq et al., 2025) jointly learns structural segmentation and approximately rigid segment motions. CryoDyna (Zhang et al., 2025) combines attention across image views with multiscale deformation to infer atomic motion. Atomic constraints alone do not determine how conformations are organized in latent space. CryoHyper jointly addresses atomic plausibility and conformational organization within a framework for cryo-EM and cryo-ET.

Cryo-ET Restoration. IsoNet (Liu et al., 2022) learns to restore density affected by the missing-wedge, while DeepDeWedge (Wiedemann & Heckel, 2024) jointly addresses denoising and missing-wedge reconstruction. Drawing on rotation and artificial wedge masking strategies for density restoration, CryoHyper incorporates these operations into atomic reconstruction training. ET Input Adaptation combines these strategies with learnable spectral weighting to limit the influence of noise and missing information on conformational inference.

Hyperbolic Representation Learning. Hyperbolic space supports representations of hierarchical relationships with low distortion, as demonstrated by Poincaré embeddings (Nickel & Kiela, 2017) and subsequent extensions to neural networks (Ganea et al., 2018) and generative models (Mathieu et al., 2019). CryoHyper couples hyperbolic domain coordinates with Euclidean local coordinates to provide an inductive bias for learning relationships among conformations with variation at different scales. The structural relationships captured by this representation are evaluated against known conformations and through downstream particle refinement.

3 Method

3.1 Problem Formulation

Given a reference structure $\mathbf{X}^0$ with N atoms, CryoHyper learns a conformational coordinate $\mathbf{z}_i$ for each cryo-EM particle image or cryo-ET subtomogram $\mathbf{y}_i$. Particle poses $\mathbf{P}_i$ and imaging parameters are provided as inputs. Our objective is to recover plausible atomic deformations and a conformational space whose distances reflect structural relationships. As shown in Figure 1, training proceeds in two stages. The atomic stage learns conformational coordinates and atomic deformations through reconstruction supervision and structural constraints. Subtomograms pass through ET Input Adaptation before entering the encoder. The density stage fixes the learned coordinates and trains a decoder to generate corresponding density maps in the same conformational coordinate system.

3.2 Conformational Space Learning

3.2.1 Latent Representation

An encoder for each input modality extracts features and estimates separate Euclidean and hyperbolic posteriors. The Euclidean component uses a diagonal Gaussian. The hyperbolic compo-

nent consists of two-dimensional Poincaré disks, whose distance scale increases toward the boundary (Nickel & Kiela, 2017). This provides an inductive bias for representing conformational differences at varying scales. For each disk, the encoder predicts a posterior location and dispersion, and we use reparameterized sampling during training. The concatenated coordinates $\mathbf{z}_i = [\mathbf{z}_i^E, \mathbf{z}_i^H]$ control local atomic changes and domain motions, respectively.

PCA and UMAP use the analysis coordinates

$$\bar{\mathbf{z}}_i = \left[\mathbf{z}_i^E, \log_{\mathbf{0}}(\mathbf{z}_i^H)\right], \tag{1}$$

where the logarithmic map acts on each Poincaré factor. Sampled trajectories are mapped back through the exponential map for decoding. Quantitative distance evaluation and particle selection use the full latent coordinates. The domain distance combines geodesic distances across Poincaré factors, and the local distance is Euclidean. Each component distance is normalized by its median nonzero value over sampled pairs of posterior means, with these scales held fixed. The combined distance is the square root of the sum of their squares, balancing the numerical scales of the two components.

3.2.2 Atomic Deformation

The two components control complementary deformations of the reference structure (Figure 1a). For atom j with reference coordinate $\mathbf{x}_j^0$, let $d = d(j)$ denote its domain and $\mathbf{c}_d$ the reference domain center. The hyperbolic branch predicts a rotation $\mathbf{R}_d \in SO(3)$ and translation $\mathbf{t}_d$, and the Euclidean branch predicts local displacements $\Delta\mathbf{X}$, with $\Delta\mathbf{x}_j$ for atom j. The decoded coordinate is

$$\mathbf{x}_j(\mathbf{z}) = \mathbf{R}_d(\bar{\mathbf{z}}^H)(\mathbf{x}_j^0 - \mathbf{c}_d) + \mathbf{c}_d + \mathbf{t}_d(\bar{\mathbf{z}}^H) + \Delta\mathbf{x}_j(\mathbf{z}^E). \tag{2}$$

The rigid transformation preserves distances within each domain, while local displacements represent finer flexibility. Structural regularization constrains the resulting atomic coordinates. Changes in the learned conformational coordinates can therefore be interpreted through the corresponding atomic motions.

3.3 ET Input Adaptation

Noise and incomplete angular sampling can obscure conformational differences in subtomograms. ET Input Adaptation combines learnable spectral weighting with reconstruction training under artificial wedge masking (Figure 1c).

3.3.1 Learnable Spectral Weighting

Before the 3D encoder, we apply radial weights to the input spectrum. For a volume $\mathbf{V}$,

$$\mathcal{S}(\mathbf{V}) = \mathcal{F}^{-1}[w\,\mathcal{F}(\mathbf{V})], \qquad w(r) = 1 - \alpha + \alpha\,\sigma\big(a(r)\big), \tag{3}$$

where $\mathcal{F}$ is the Fourier transform. Products in the Fourier domain are applied elementwise. The weight $w(r)$ depends on radial spatial frequency r, $a(r)$ interpolates learnable radial parameters, and σ is the sigmoid function. The fixed bound $0 \le \alpha < 1$ limits attenuation, with $w(0) = 1$ set to preserve the input mean. Joint reconstruction training learns the frequency weights for conformational inference. Weighting is applied during training and inference, with reconstruction targets unchanged.

3.3.2 Training with Artificial Wedges

We augment a subset of training subtomograms through random rotation and artificial wedge masking. Let $\mathbf{M}_i$ describe the original Fourier support, $\mathcal{R}_i$ be a random rotation operator, and $\mathbf{A}_i$ be an artificial mask with the original fixed tilt support. The encoder input and its effective frequency mask are

$$\widetilde{\mathbf{V}}_i = \mathcal{F}^{-1}[\mathbf{A}_i\,\mathcal{F}(\mathcal{R}_i\mathbf{V}_i)], \qquad \widetilde{\mathbf{M}}_i = \mathbf{A}_i\,(\mathcal{R}_i\mathbf{M}_i). \tag{4}$$

The volume passes through $\mathcal{S}$, and $\widetilde{\mathbf{M}}_i$ conditions posterior estimation through encoder feature modulation. Rotation changes the original missing-wedge orientation relative to the artificial mask. The target $\mathcal{R}_i\mathbf{V}_i$ retains the information available before artificial masking. Supervision covers its original frequency support $\mathcal{R}_i\mathbf{M}_i$, including measured information removed from the input by $\mathbf{A}_i$.

3.4 Reconstruction and Optimization

The atomic coordinates decoded from $\mathbf{z}_i$ are rendered as a Gaussian mixture density ρ_i. For a cryo-EM image $\mathbf{I}_i$, projection at the particle pose and contrast transfer function (CTF) modulation (Rohou & Grigorieff, 2015) give

$$\widehat{\mathbf{I}}_i = \mathrm{CTF}_i \, \Pi_{\mathbf{P}_i}(\rho_i). \tag{5}$$

For cryo-ET, $\mathcal{T}_i$ transforms the density into the subtomogram frame using pose $\mathbf{P}_i$, and $\mathcal{W}_i$ applies the original Fourier mask $\mathbf{M}_i$. At a sampled tilt angle θ, the prediction and target are

$$\widehat{\mathbf{u}}_{i,\theta} = \mathrm{CTF}_{i,\theta} \, \Pi_\theta[\mathcal{W}_i \mathcal{T}_i(\rho_i)], \qquad \mathbf{u}_{i,\theta} = \Pi_\theta(\mathbf{V}_i). \tag{6}$$

Here, Π denotes projection and $\mathrm{CTF}_{i,\theta}$ is the two-dimensional CTF for sample i at tilt angle θ. Tilt directions are sampled over the original available range. Reconstruction uses a masked correlation loss with frequency-restricted supervision.

The atomic training objective is

$$\mathcal{L}_{\mathrm{atom}} = \mathcal{L}_{\mathrm{rec}} + \mathcal{L}_{\mathrm{str}} + \beta \mathcal{L}_{\mathrm{KL}}. \tag{7}$$

$\mathcal{L}_{\mathrm{rec}}$ denotes image or projection reconstruction loss. $\mathcal{L}_{\mathrm{str}}$ collects weighted connectivity, secondary structure distance, reference pairwise distance, and steric clash constraints. $\mathcal{L}_{\mathrm{KL}}$ regularizes the Euclidean and hyperbolic posteriors against their respective priors.

3.5 Density Generation

After atomic training, we map the posterior means to analysis coordinates $\bar{\boldsymbol{\mu}}_i$ and keep them fixed. The density decoder generates a map conditioned on these coordinates (Figure 1b),

$$\widehat{\mathbf{D}}_i = G_{\theta_D}(\bar{\boldsymbol{\mu}}_i). \tag{8}$$

The decoder uses an implicit Fourier representation and is optimized separately through density reconstruction. Supervision uses Fourier slices determined by particle poses and, for cryo-ET, tilt angles. Fixed coordinates preserve the conformational organization learned during atomic training and allow the density maps to describe additional signal beyond the reference model.

4 Experiments

4.1 Experimental Setup

We evaluate conformational space recovery on simulated 3NVQ cryo-ET data and structural interpretation and refinement on EMPIAR-10073 (Nguyen et al., 2016) and EMPIAR-10453 (Turoňová et al., 2020). Simulations separately reduce the tilt range from $\pm 60^\circ$ to $\pm 45^\circ$ or the signal-to-noise ratio (SNR) from 0.1 to 0.01, with a 3° tilt step throughout. Additional experiments (Appendix E) use EMPIAR-10180 (Plaschka et al., 2017) to assess atomic motion learning when flexibility prevents local density reconstruction, EMPIAR-10059 (Gao et al., 2016) to assess robustness to reference model mismatch, and CryoBench Spike-MD (Jeon et al., 2024) to validate conformational spaces. Cryo-EM baselines include CryoSTAR (Li et al., 2024), CryoDyna (Zhang et al., 2025), cryoSPHERE (Ducrocq et al., 2025), cryoDRGN (Zhong et al., 2021a), RECOVAR (Gilles & Singer, 2025), OPUS-DSD (Luo et al., 2023), and cryoSPARC 3D classification (Punjani et al., 2017). Cryo-ET baselines are cryoDRGN-ET (Rangan et al., 2024) and OPUS-Tomo (Luo et al., 2024).

We assess conformational neighborhoods using Balanced Neighborhood Similarity (BNS) and bidirectional Information Imbalance (Glielmo et al., 2022). Atom-density consistency (AD-FSC) evaluates atomic and density agreement using Fourier shell correlation (FSC) at 0.5, while latent guided refinement resolution (R-FSC) uses gold-standard FSC (Scheres & Chen, 2012) to assess whether latent neighborhoods support refinement. Map correlation coefficient (Map CC), structural similarity (SSIM), and backprojection agreement (BP-FSC) provide complementary evaluations. CryoHyper uses four hyperbolic and four Euclidean latent dimensions, with AdamW at a learning rate of 10^{-4} in both stages. Atomic training uses 25 epochs for cryo-EM and 50 for cryo-ET, followed by 20 density epochs. Details of dataset preparation, implementation, and evaluation protocols are provided in Appendices B, C, and D, respectively.

4.2 Conformational Space Recovery

The simulated benchmark uses 100 conformations sampled along PC1 of DynaRepo trajectory A005K (Mokhtari et al., 2026), based on the 3NVQ Sema7A–PlexinC1 complex (Liu et al., 2010). Labels are used only for evaluation. CryoHyper preserves a continuous progression of ground truth conformations in the UMAP visualization (Figure 2b). The cryoDRGN-ET embedding separates into disconnected groups, while OPUS-Tomo shows greater overlap between source conformations. CryoHyper's decoded atomic models and densities reproduce the approximately symmetric opening and closing of the PlexinC1 subunits around the Sema7A dimer in the source trajectory (Figure 2a,c). BNS corrects neighborhood overlap for chance and weights source conformations equally. Under the baseline condition, CryoHyper achieves the highest BNS-AUC$_{20}$ and lowest Information Imbalance in both directions (Figure 3), showing that latent neighbors are close in conformation and conformational neighbors remain close in latent space. The visualization and metrics show that the learned representation preserves conformational order and neighborhood relationships.

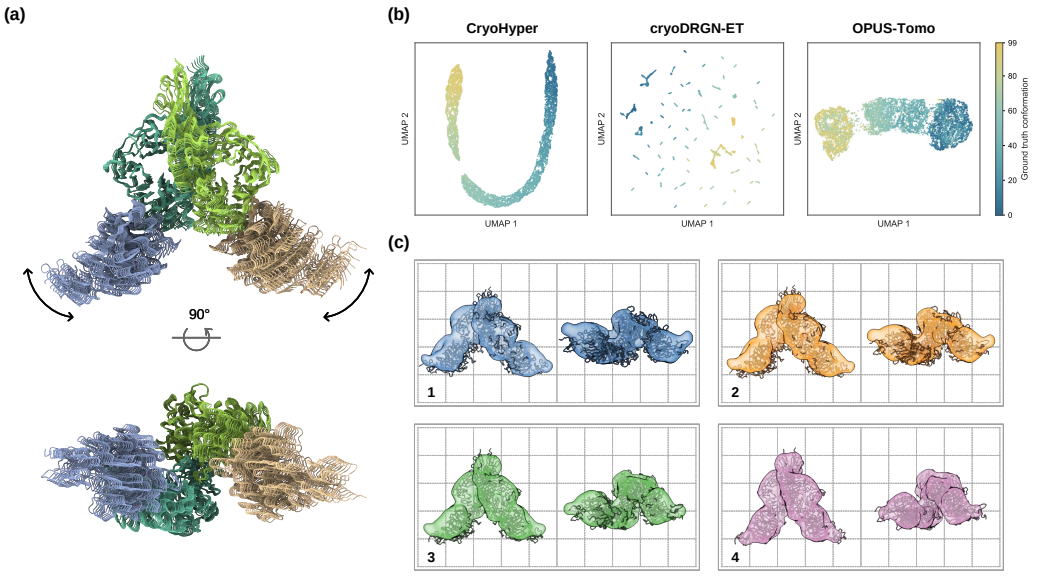


Figure 2: Conformational space recovery on the simulated 3NVQ cryo-ET dataset. (a) Atomic motion decoded along PC1, showing the opening and closing of the PlexinC1 subunits from two views. (b) Latent UMAP embeddings of different methods colored by ground truth conformation. (c) Atomic models overlaid with corresponding density maps at selected PC1 sampling points.

The conditions share source conformations, particle positions, and poses, providing a controlled comparison of sensitivity to tilt range and noise. At $\pm 45°$, CryoHyper retains the lowest Information Imbalance in both directions, preserving neighborhood relationships despite reduced tilt coverage. At SNR 0.01, its BNS-AUC$_{20}$ reaches 41.88, compared with 24.35 for cryoDRGN-ET and 22.09 for OPUS-Tomo, while Information Imbalance remains close to the baseline values. CryoHyper retains 95.5% of its baseline BNS-AUC$_{20}$ despite a tenfold reduction in SNR. These results demonstrate robust conformational neighborhood recovery under reduced tilt range and increased noise. On CryoBench Spike-MD, CryoHyper also achieves higher Neighborhood Similarity than CryoSTAR and CryoDyna, extending conformational neighborhood recovery to simulated cryo-EM data (Appendix E).

4.3 Structural Analysis and Refinement

On EMPIAR-10073, CryoHyper recovers head and foot motion of the yeast U4/U6.U5 tri-snRNP complex consistent with reported flexibility (Häcker et al., 2008; Li et al., 2024; Punjani & Fleet, 2023) (Figure 4a,b). The decoded atomic models and density maps show consistent motion. Across

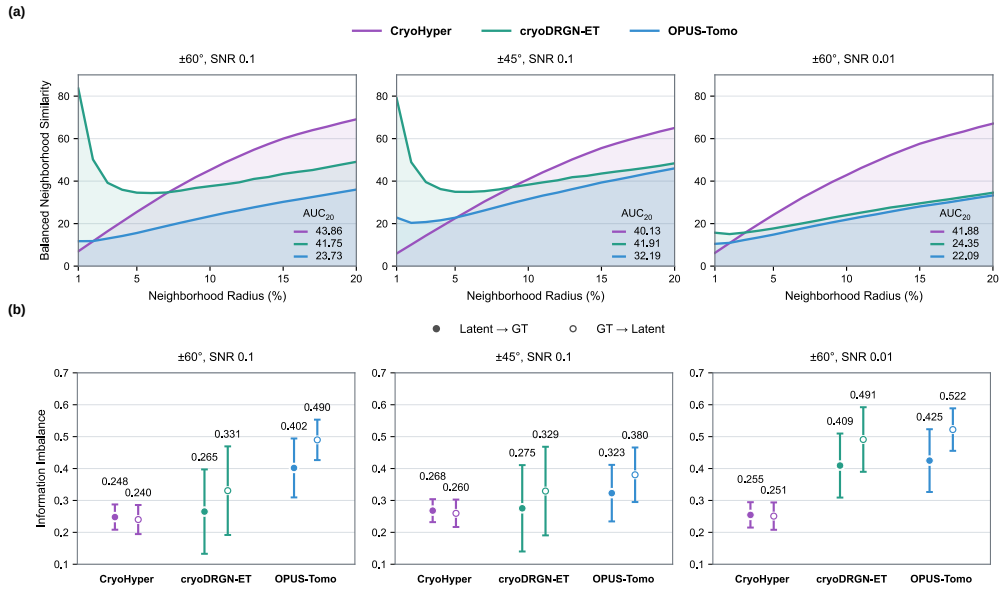


Figure 3: Conformational space recovery under different tilt ranges and noise levels on the simulated 3NVQ cryo-ET dataset. (a) Balanced Neighborhood Similarity across neighborhood scales, with AUC_{20} reported for each method. (b) Bidirectional Information Imbalance, with points and error bars indicating means and standard deviations.

Table 1: Quantitative comparison on EMPIAR-10073 using atom-density consistency (AD-FSC), latent guided refinement resolution (R-FSC), backprojection agreement (BP-FSC), map correlation coefficient (Map CC), and structural similarity (SSIM). Values are means and standard deviations across eight sampled states. cryoSPHERE has no separate density reconstruction stage. Best results are shown in bold.

Method	AD-FSC (Å) ↓	R-FSC (Å) ↓	BP-FSC (Å) ↓	Map CC ↑	SSIM ↑
cryoDRGN	N/A	5.04 ± 0.41	9.25 ± 0.32	0.83 ± 0.01	0.47 ± 0.03
RECOVAR	N/A	5.32 ± 0.17	$\mathbf{9.10 \pm 0.08}$	0.76 ± 0.01	0.38 ± 0.01
OPUS-DSD	N/A	5.44 ± 0.34	41.14 ± 2.18	0.44 ± 0.02	0.21 ± 0.01
cryoSPHERE	N/A	5.47 ± 0.22	N/A	N/A	N/A
CryoSTAR	12.56 ± 4.48	6.22 ± 1.50	13.07 ± 4.08	0.72 ± 0.06	0.36 ± 0.04
CryoDyna	10.13 ± 2.82	6.04 ± 1.01	11.77 ± 3.60	0.92 ± 0.03	0.59 ± 0.06
CryoHyper	$\mathbf{9.53 \pm 0.61}$	$\mathbf{5.02 \pm 0.46}$	9.56 ± 0.49	$\mathbf{0.93 \pm 0.01}$	$\mathbf{0.63 \pm 0.03}$

eight sampled states, CryoHyper achieves the best mean atom-density consistency and lowest variation among the compared atomic methods (Figure 4d, Table 1). For each method, eight subsets of 10,000 nearest particles in the full latent space undergo a common refinement protocol. With particle count and refinement procedure fixed, resolution differences primarily reflect how effectively each representation groups particles with compatible structures. CryoHyper achieves slightly better mean resolution than cryoDRGN and outperforms the other methods. Its resolutions span 4.57–5.73 Å, a narrower range than CryoSTAR and CryoDyna (Figure 4f). All eight subsets support refinement, whereas four of eight 3D classes are too sparsely populated (Figure 4e). CryoHyper distinguishes conformations grouped into the same 3D class. Latent guided refinement reveals differences in head and foot positions that are not resolved in the class averages (Appendix E). Refined maps reproduce the decoded motion (Figure 4c), validating neighborhood coherence independently of the density decoder. When evaluated against maps directly backprojected from the selected particle images us-

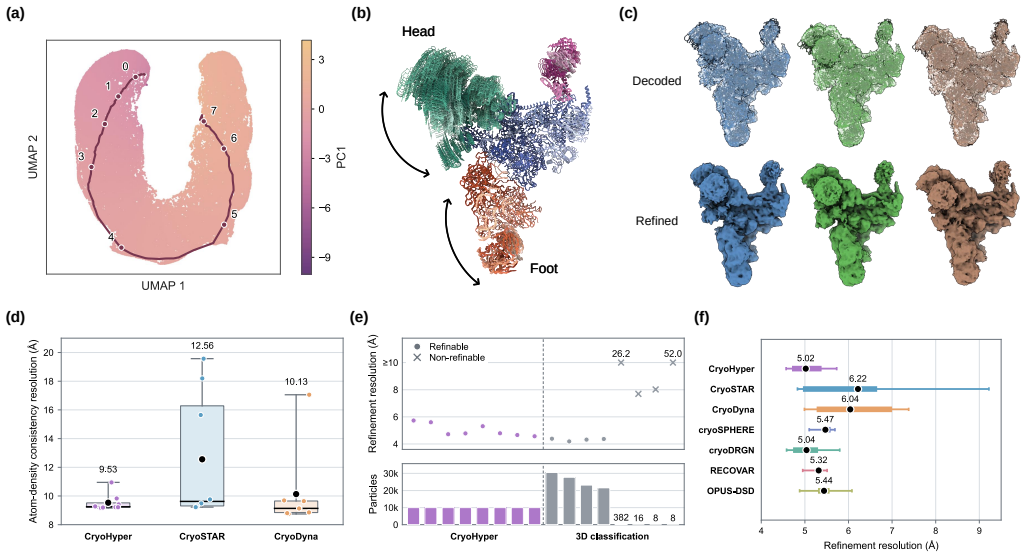


Figure 4: Structural analysis and refinement on EMPIAR-10073. (a) Latent UMAP colored by PC1, with the sampling trajectory. (b) Atomic motion decoded along PC1. (c) Decoded atomic models overlaid with density maps and corresponding refined maps at three sampled states. (d) Atom-density consistency evaluated by FSC at the 0.5 threshold. (e) Particle counts and refinement resolutions compared with 3D classification. (f) Latent guided refinement resolution across methods.

ing their poses, CryoHyper achieves better mean BP-FSC than the other atomic methods and the highest Map CC and SSIM (Table 1), supporting agreement between decoded densities and the particle neighborhoods used for refinement. On EMPIAR-10180, CryoHyper recovers atomic motion in flexible regions that remain poorly resolved in the reconstructed density maps. Using a reference from a different ligand state, CryoHyper recovers consistent decoded and refined motion on EMPIAR-10059 (Appendix E).

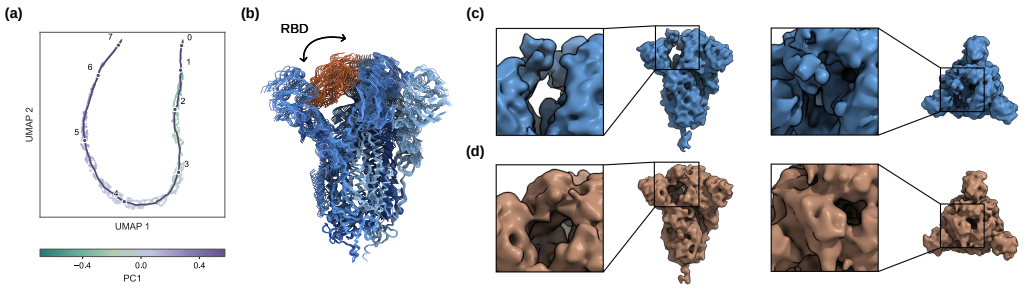


Figure 5: Structural analysis and refinement on EMPIAR-10453. (a) Latent UMAP colored by PC1, with the sampling trajectory. (b) Atomic motion decoded along PC1, showing the opening and closing of a receptor-binding domain (RBD). (c, d) Maps independently refined from 1,185 particles at each trajectory endpoint, reaching resolutions of 10.98 and 10.63 Å, respectively.

On real SARS-CoV-2 spike subtomograms from EMPIAR-10453, CryoHyper recovers opening and closing of one receptor-binding domain (RBD), consistent with reported conformational variability (Turoňová et al., 2020), despite low signal quality, limited particle counts, and the missing-wedge (Figure 5a,b). Independent RELION refinement of 1,185 particles per PC1 endpoint resolves open and closed states consistent with the decoded motion (Figure 5c,d). This conformational organization supports refinement of distinct structural states from limited particle subsets in real cryo-ET.

4.4 Ablation Study

H and E denote hyperbolic and Euclidean dimensions, respectively. On EMPIAR-10073, H4E4 improves both mean atom-density consistency and refinement resolution over E4E4 while retaining the encoder, both deformation branches, total latent dimension, and training settings (Figure 6). With the architecture fixed, the improvements in both atom-density consistency and refinement resolution support the hyperbolic domain representation as an effective inductive bias for structural decoding and particle selection. H4E4 also outperforms H2E6 and H6E2 on both mean resolutions and reduces variation in atom-density consistency, favoring equal allocation to domain motion and local deformation.

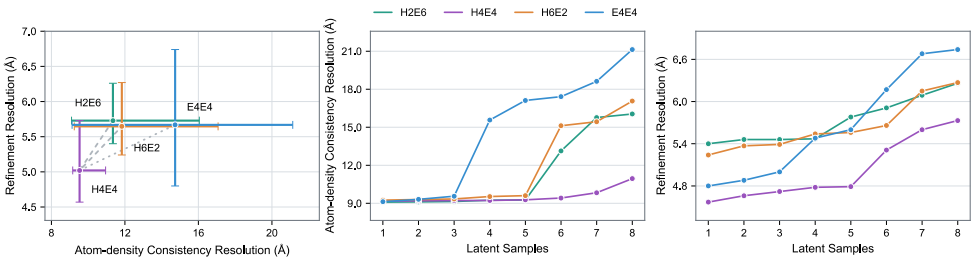


Figure 6: Latent space allocation ablation on EMPIAR-10073. From left to right, joint resolution comparison, atom-density consistency, and latent guided refinement. H and E denote hyperbolic and Euclidean dimensions. E4E4 uses Euclidean coordinates in both decoder branches.

Table 2: Ablation of ET input adaptation on the simulated 3NVQ cryo-ET dataset.

Tilt / SNR	Model	BNS-AUC$_{20}$ ↑	$\Delta_{\mathrm{L}\to\mathrm{GT}}$ ↓	$\Delta_{\mathrm{GT}\to\mathrm{L}}$ ↓
±60° / 0.1	w/o adaptation	43.35 ± 0.12	0.252 ± 0.039	0.242 ± 0.048
	Full model	$\mathbf{43.86 \pm 0.11}$	$\mathbf{0.248 \pm 0.040}$	$\mathbf{0.240 \pm 0.046}$
±45° / 0.1	w/o adaptation	39.42 ± 0.09	0.271 ± 0.034	0.263 ± 0.040
	Full model	$\mathbf{40.13 \pm 0.14}$	$\mathbf{0.268 \pm 0.036}$	$\mathbf{0.260 \pm 0.043}$
±60° / 0.01	w/o adaptation	37.52 ± 0.09	0.281 ± 0.034	0.276 ± 0.039
	Full model	$\mathbf{41.88 \pm 0.09}$	$\mathbf{0.255 \pm 0.040}$	$\mathbf{0.251 \pm 0.043}$

Disabling ET input adaptation removes artificial wedge training and learnable spectral weighting. Both models are trained on the same subtomograms, with the remaining architecture and training schedule held fixed. Across all three imaging conditions, the full model achieves higher BNS-AUC$_{20}$ and lower Information Imbalance in both directions (Table 2), supporting the contribution of ET input adaptation to conformational neighborhood recovery.

5 Conclusion

CryoHyper learns conformational spaces that better preserve structural relationships in cryo-EM and cryo-ET. Guided by a reference atomic model, its joint hyperbolic and Euclidean representation organizes domain motions and local atomic changes. Shared conformational coordinates retain this organization during density generation. ET Input Adaptation combines learnable spectral weighting and artificial wedge training to reduce the influence of noise and missing-wedge effects on conformational inference. Experiments on simulated and public datasets demonstrate more accurate conformational neighborhoods, physically plausible atomic motion, and robustness to reduced angular coverage and increased noise. Independent refinement of particles selected from these neighborhoods recovers distinct structural states, demonstrating that the learned conformational organization supports both atomic interpretation and downstream structural reconstruction.

REPRODUCIBILITY STATEMENT

Once the OpenReview discussion forums open, we will share an anonymous code repository through a comment visible only to reviewers and area chairs. Upon acceptance, we will release the code as open source and make the simulated cryo-ET benchmark publicly available. Section 3 and Appendix A describe the model architecture, latent distances, structural constraints, and ET Input Adaptation. Appendix B documents dataset sources, preprocessing, and construction of the simulated cryo-ET benchmark. Appendix C provides training schedules, hyperparameters, baseline and ablation configurations, parameter counts, training times, and the hardware and software environment. Appendix D defines the evaluation metrics and specifies neighborhood sampling, particle selection, and refinement protocols used to assess conformational organization and structural agreement.

AI USE STATEMENT

In this work, we used generative AI tools to assist with translation. We did not use generative AI tools for other applicable tasks requiring disclosure. Additionally, we used generative AI tools to improve the readability of the manuscript, including grammar, clarity, and concision. We reviewed all language revisions for accuracy and consistency with the intended meaning. We take responsibility for the final content of this work.

REFERENCES

Felix J.B. Bäuerlein and Wolfgang Baumeister. Towards visual proteomics at high resolution. *Journal of Molecular Biology*, 433(20):167187, 2021. ISSN 0022-2836. doi: https://doi.org/10.1016/j.jmb.2021.167187. URL https://www.sciencedirect.com/science/article/pii/S0022283621004204. From Protein Sequence to Structure at Warp Speed: How Alphafold Impacts Biology.

Hua Cheng, R. Dustin Schaeffer, Yuxing Liao, Lisa N. Kinch, Jimin Pei, Shuoyong Shi, Bong-Hyun Kim, and Nick V. Grishin. Ecod: An evolutionary classification of protein domains. *PLOS Computational Biology*, 10(12):e1003926, 2014. doi: 10.1371/journal.pcbi.1003926. URL https://app.dimensions.ai/details/publication/pub.1023790270. https://doi.org/10.1371/journal.pcbi.1003926.

Yifan Cheng. Single-particle cryo-em—how did it get here and where will it go. *Science*, 361(6405): 876–880, 2018.

Gabriel Claude Jean Ducrocq, Lukas Grunewald, Sebastian Westenhoff, and Fredrik Lindsten. cryosphere: Single-particle heterogeneous reconstruction from cryo em. In Y. Yue, A. Garg, N. Peng, F. Sha, and R. Yu (eds.), *International Conference on Learning Representations*, volume 2025, pp. 16774–16787, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/file/2ac2eac5098dba08208807b65c5851cc-Paper-Conference.pdf.

Octavian-Eugen Ganea, Gary Bécigneul, and Thomas Hofmann. Hyperbolic neural networks, 2018. URL https://arxiv.org/abs/1805.09112.

Yuan Gao, Erhu Cao, David Julius, and Yifan Cheng. Trpv1 structures in nanodiscs reveal mechanisms of ligand and lipid action. *Nature*, 534(7607):347–351, 2016. doi: 10.1038/nature17964. URL https://doi.org/10.1038/nature17964.

Marc Aurèle Gilles and Amit Singer. Cryo-em heterogeneity analysis using regularized covariance estimation and kernel regression. *Proceedings of the National Academy of Sciences*, 122(9): e2419140122, 2025. doi: 10.1073/pnas.2419140122. URL https://www.pnas.org/doi/abs/10.1073/pnas.2419140122.

Aldo Glielmo, Claudio Zeni, Bingqing Cheng, Gábor Csányi, and Alessandro Laio. Ranking the information content of distance measures. *PNAS Nexus*, 1(2), April 2022. ISSN 2752-6542. doi: 10.1093/pnasnexus/pgac039. URL http://dx.doi.org/10.1093/pnasnexus/pgac039.

Irina Häcker, Bjoern Sander, Monika M Golas, Elmar Wolf, Elif Karagöz, Berthold Kastner, Holger Stark, Patrizia Fabrizio, and Reinhard Lührmann. Localization of prp8, brr2, snu114 and u4/u6 proteins in the yeast tri-snrnp by electron microscopy. *Nature Structural & Molecular Biology*, 15(11):1206–1212, 2008. doi: 10.1038/nsmb.1506. URL https://doi.org/10.1038/nsmb.1506.

Minkyu Jeon, Rishwanth Raghu, Miro Astore, Geoffrey Woollard, Ryan Feathers, Alkin Kaz, Sonya M. Hanson, Pilar Cossio, and Ellen Zhong. Cryobench: Diverse and challenging datasets for the heterogeneity problem in cryo-em. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang (eds.), *Advances in Neural Information Processing Systems*, volume 37, pp. 89468–89512. Curran Associates, Inc., 2024. doi: 10.52202/079017-2840. URL https://proceedings.neurips.cc/paper_files/paper/2024/file/a2ef5ba272df8f168dc38037cc946be0-Paper-Datasets_and_Benchmarks_Track.pdf.

Wolfgang Kabsch and Christian Sander. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, 22(12):2577–2637, 1983. doi: https://doi.org/10.1002/bip.360221211. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/bip.360221211.

Yilai Li, Yi Zhou, Jing Yuan, Fei Ye, and Quanquan Gu. Cryostar: leveraging structural priors and constraints for cryo-em heterogeneous reconstruction. *Nature Methods*, 21(12):2318–2326, 2024. doi: 10.1038/s41592-024-02486-1. URL https://doi.org/10.1038/s41592-024-02486-1.

Heli Liu, Z. Sean Juo, Ann Hye-Ryong Shim, Pamela J. Focia, Xiaoyan Chen, K. Christopher Garcia, and Xiaolin He. Structural basis of semaphorin-plexin recognition and viral mimicry from sema7a and a39r complexes with plexinc1. *Cell*, 142(5):749–761, 2010. doi: 10.1016/j.cell.2010.07.040. URL https://doi.org/10.1016/j.cell.2010.07.040. doi: 10.1016/j.cell.2010.07.040.

Yun-Tao Liu, Heng Zhang, Hui Wang, Chang-Lu Tao, Guo-Qiang Bi, and Z. Hong Zhou. Isotropic reconstruction for electron tomography with deep learning. *Nature Communications*, 13(1):6482, 2022. doi: 10.1038/s41467-022-33957-8. URL https://doi.org/10.1038/s41467-022-33957-8.

Zhenwei Luo, Fengyun Ni, Qinghua Wang, and Jianpeng Ma. Opus-dsd: deep structural disentanglement for cryo-em single-particle analysis. *Nature Methods*, 20(11):1729–1738, 2023. doi: 10.1038/s41592-023-02031-6. URL https://doi.org/10.1038/s41592-023-02031-6.

Zhenwei Luo, Xiangru Chen, Qinghua Wang, and Jianpeng Ma. Opus-tomo: Towards resolving dynamics and compositional heterogeneities of biomolecules with cryo-electron tomography, 2024. URL https://doi.org/10.1101/2024.06.30.601442.

Emile Mathieu, Charline Le Lan, Chris J. Maddison, Ryota Tomioka, and Yee Whye Teh. Continuous hierarchical representations with poincaré variational auto-encoders. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL https://proceedings.neurips.cc/paper_files/paper/2019/file/0ec04cb3912c4f08874dd03716f80df1-Paper.pdf.

Omid Mokhtari, Emmanuelle Bignon, Hamed Khakzad, and Yasaman Karami. Dynarepo: the repository of macromolecular conformational dynamics. *Nucleic Acids Research*, 54(D1):D393–D401, 01 2026. ISSN 1362-4962. doi: 10.1093/nar/gkaf1130. URL https://doi.org/10.1093/nar/gkaf1130.

Takanori Nakane, Dari Kimanius, Erik Lindahl, and Sjors HW Scheres. Characterisation of molecular motions in cryo-em single-particle data by multi-body refinement in relion. *eLife*, 7:e36861, jun 2018. ISSN 2050-084X. doi: 10.7554/eLife.36861. URL https://doi.org/10.7554/eLife.36861.

Thi Hoang Duong Nguyen, Wojciech P. Galej, Xiao-chen Bai, Chris Oubridge, Andrew J. Newman, Sjors H. W. Scheres, and Kiyoshi Nagai. Cryo-em structure of the yeast u4/u6.u5 tri-snrnp at 3.7 Å resolution. *Nature*, 530(7590):298–302, 2016. doi: 10.1038/nature16940. URL https://doi.org/10.1038/nature16940.

Maximillian Nickel and Douwe Kiela. Poincaré embeddings for learning hierarchical representations. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper_files/paper/2017/file/59dfa2df42d9e3d41f5b02bfc32229dd-Paper.pdf.

Ruth Nussinov, Yonglan Liu, Wengang Zhang, and Hyunbum Jang. Protein conformational ensembles in function: roles and mechanisms. *RSC Chemical Biology*, 4(11):850–864, 11 2023. ISSN 2633-0679. doi: 10.1039/d3cb00114h. URL https://doi.org/10.1039/d3cb00114h.

Clemens Plaschka, Pei-Chun Lin, and Kiyoshi Nagai. Structure of a pre-catalytic spliceosome. *Nature*, 546(7660):617–621, 2017. doi: 10.1038/nature22799. URL https://doi.org/10.1038/nature22799.

Ali Punjani and David J. Fleet. 3d variability analysis: Resolving continuous flexibility and discrete heterogeneity from single particle cryo-em. *Journal of Structural Biology*, 213(2):107702, 2021. ISSN 1047-8477. doi: https://doi.org/10.1016/j.jsb.2021.107702. URL https://www.sciencedirect.com/science/article/pii/S1047847721000071.

Ali Punjani and David J. Fleet. 3dflex: determining structure and motion of flexible proteins from cryo-em. *Nature Methods*, 20(6):860–870, 2023. doi: 10.1038/s41592-023-01853-8. URL https://doi.org/10.1038/s41592-023-01853-8.

Ali Punjani, John L Rubinstein, David J Fleet, and Marcus A Brubaker. cryosparc: algorithms for rapid unsupervised cryo-em structure determination. *Nature Methods*, 14(3):290–296, 2017. doi: 10.1038/nmeth.4169. URL https://doi.org/10.1038/nmeth.4169.

Michael Radermacher. Three-dimensional reconstruction of single particles from random and nonrandom tilt series. *Journal of electron microscopy technique*, 9(4):359–394, 1988.

Ramya Rangan, Ryan Feathers, Sagar Khavnekar, Adam Lerer, Jake D. Johnston, Ron Kelley, Martin Obr, Abhay Kotecha, and Ellen D. Zhong. Cryodrgn-et: deep reconstructing generative networks for visualizing dynamic biomolecules inside cells. *Nature Methods*, 21(8):1537–1545, 2024. doi: 10.1038/s41592-024-02340-4. URL https://doi.org/10.1038/s41592-024-02340-4.

Alexis Rohou and Nikolaus Grigorieff. Ctffind4: Fast and accurate defocus estimation from electron micrographs. *Journal of Structural Biology*, 192(2):216–221, 2015. ISSN 1047-8477. doi: https://doi.org/10.1016/j.jsb.2015.08.008. URL https://www.sciencedirect.com/science/article/pii/S1047847715300460. Recent Advances in Detector Technologies and Applications for Molecular TEM.

Sjors H W Scheres and Shaoxia Chen. Prevention of overfitting in cryo-em structure determination. *Nature Methods*, 9(9):853–854, 2012. doi: 10.1038/nmeth.2115. URL https://doi.org/10.1038/nmeth.2115.

Sjors H.W. Scheres. Relion: Implementation of a bayesian approach to cryo-em structure determination. *Journal of Structural Biology*, 180(3):519–530, 2012. ISSN 1047-8477. doi: https://doi.org/10.1016/j.jsb.2012.09.006. URL https://www.sciencedirect.com/science/article/pii/S1047847712002481.

Florence Tama, Osamu Miyashita, and Charles L. Brooks III. Normal mode based flexible fitting of high-resolution structure into low-resolution experimental data from cryo-em. *Journal of Structural Biology*, 147(3):315–326, 2004. ISSN 1047-8477. doi: https://doi.org/10.1016/j.jsb.2004.03.002. URL https://www.sciencedirect.com/science/article/pii/S1047847704000590. Time-Resolved Imaging of Macromolecular Processes and Interactions.

Guang Tang, Liwei Peng, Philip R. Baldwin, Deepinder S. Mann, Wen Jiang, Ian Rees, and Steven J. Ludtke. Eman2: An extensible image processing suite for electron microscopy. *Journal of Structural Biology*, 157(1):38–46, 2007. ISSN 1047-8477. doi: https://doi.org/10.1016/j.jsb.2006.05.009. URL https://www.sciencedirect.com/science/article/pii/S1047847706001894. Software tools for macromolecular microscopy.

Dimitry Tegunov and Patrick Cramer. Real-time cryo-electron microscopy data preprocessing with warp. *Nature Methods*, 16(11):1146–1152, 2019. doi: 10.1038/s41592-019-0580-y. URL https://doi.org/10.1038/s41592-019-0580-y.

Bogdan Toader, Fred J. Sigworth, and Roy R. Lederman. Methods for cryo-em single particle reconstruction of macromolecules having continuous heterogeneity. *Journal of Molecular Biology*, 435(9):168020, 2023. ISSN 0022-2836. doi: https://doi.org/10.1016/j.jmb.2023.168020. URL https://www.sciencedirect.com/science/article/pii/S0022283623000761. New Frontier of Cryo-Electron Microscopy Technology.

Maya Topf, Keren Lasker, Ben Webb, Haim Wolfson, Wah Chiu, and Andrej Sali. Protein structure fitting and refinement guided by cryo-em density. *Structure*, 16(2):295–307, 2008. doi: 10.1016/j.str.2007.11.016. URL https://doi.org/10.1016/j.str.2007.11.016. doi: 10.1016/j.str.2007.11.016.

Leonardo G. Trabuco, Elizabeth Villa, Kakoli Mitra, Joachim Frank, and Klaus Schulten. Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. *Structure*, 16(5):673–683, 2008. doi: 10.1016/j.str.2008.03.005. URL https://doi.org/10.1016/j.str.2008.03.005. doi: 10.1016/j.str.2008.03.005.

Martin Turk and Wolfgang Baumeister. The promise and the challenges of cryo-electron tomography. *FEBS letters*, 594(20):3243–3261, 2020.

Beata Turoňová, Mateusz Sikora, Christoph Schürmann, Wim J. H. Hagen, Sonja Welsch, Florian E. C. Blanc, Sören von Bülow, Michael Gecht, Katrin Bagola, Cindy Hörner, Ger van Zandbergen, Jonathan Landry, Nayara Trevisan Doimo de Azevedo, Shyamal Mosalaganti, Andre Schwarz, Roberto Covino, Michael D. Mühlebach, Gerhard Hummer, Jacomine Krijnse Locker, and Martin Beck. In situ structural analysis of sars-cov-2 spike reveals flexibility mediated by three hinges. *Science*, 370(6513):203–208, 2020. doi: 10.1126/science.abd5223. URL https://www.science.org/doi/abs/10.1126/science.abd5223.

Marin van Heel and Michael Schatz. Fourier shell correlation threshold criteria. *Journal of Structural Biology*, 151(3):250–262, 2005. ISSN 1047-8477. doi: https://doi.org/10.1016/j.jsb.2005.05.009. URL https://www.sciencedirect.com/science/article/pii/S1047847705001292.

Nguyen Xuan Vinh, Julien Epps, and James Bailey. Information theoretic measures for clusterings comparison: is a correction for chance necessary? In *Proceedings of the 26th Annual International Conference on Machine Learning*, ICML '09, pp. 1073–1080, New York, NY, USA, 2009. Association for Computing Machinery. ISBN 9781605585161. doi: 10.1145/1553374.1553511. URL https://doi.org/10.1145/1553374.1553511.

Zhou Wang, Eero P Simoncelli, and Alan C Bovik. Multiscale structural similarity for image quality assessment. In *The thrity-seventh asilomar conference on signals, systems & computers, 2003*, volume 2, pp. 1398–1402. Ieee, 2003.

Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.

Simon Wiedemann and Reinhard Heckel. A deep learning method for simultaneous denoising and missing wedge reconstruction in cryogenic electron tomography. *Nature Communications*, 15(1):8255, 2024. doi: 10.1038/s41467-024-51438-y. URL https://doi.org/10.1038/s41467-024-51438-y.

Chengwei Zhang, Shimian Li, Yihao Niu, Zhen Zhu, Sihao Yuan, Sirui Liu, and Yi Qin Gao. Cryodyna: Multiscale end-to-end modeling of cryo-em macromolecule dynamics with physics-aware neural network, 2025. URL https://arxiv.org/abs/2510.16510.

Ellen D. Zhong, Tristan Bepler, Bonnie Berger, and Joseph H. Davis. Cryodrgn: reconstruction of heterogeneous cryo-em structures using neural networks. *Nature Methods*, 18(2):176–185, 2021a. doi: 10.1038/s41592-020-01049-4. URL https://doi.org/10.1038/s41592-020-01049-4.

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

A Method Details

A.1 Latent Representation and Distances

CryoHyper separates domain motion and local atomic deformation through hyperbolic and Euclidean latent components (Figure 7c). For an even domain dimension d_H, the hyperbolic component consists of $d_H/2$ two-dimensional Poincaré disks (Nickel & Kiela, 2017) with fixed curvature -1. The joint space is $\mathcal{Z} = \mathbb{R}^{d_E} \times \prod_{m=1}^{d_H/2} \mathbb{D}^2_{-1}$. For two coordinates $\mathbf{u}, \mathbf{v}$ in a unit Poincaré disk, the geodesic distance is

$$d_{\mathbb{D}}(\mathbf{u}, \mathbf{v}) = \operatorname{arcosh}\left(1 + \frac{2\|\mathbf{u} - \mathbf{v}\|_2^2}{(1 - \|\mathbf{u}\|_2^2)(1 - \|\mathbf{v}\|_2^2)}\right). \tag{9}$$

For particles i and j, the domain and local distances are

$$\begin{aligned} d_H(i,j) &= \left[\sum_{m=1}^{d_H/2} d_{\mathbb{D}}\big(\mathbf{z}_{i,m}^H, \mathbf{z}_{j,m}^H\big)^2\right]^{1/2}, \\ d_E(i,j) &= \|\mathbf{z}_i^E - \mathbf{z}_j^E\|_2. \end{aligned} \tag{10}$$

Let s_H and s_E be the median nonzero distances of the respective components over sampled pairs of posterior means. Both scales are estimated once for a trained representation and held fixed. The combined distance is

$$d_{\text{latent}}(i,j) = \sqrt{\left(\frac{d_H(i,j)}{s_H}\right)^2 + \left(\frac{d_E(i,j)}{s_E}\right)^2}. \tag{11}$$

Normalization balances the numerical scales of the two components. The same scales are used for particle comparisons and distances from sampled trajectory coordinates. BNS, Information Imbalance, and particle selection use this distance in the full latent space. In the E4E4 ablation, the domain distance is Euclidean, with the same component normalization and combination rule.

PCA and UMAP use $\bar{\mathbf{z}}_i = [\mathbf{z}_i^E, \log_{\mathbf{0}}(\mathbf{z}_i^H)]$, where the logarithmic map acts separately on each Poincaré factor. Trajectory samples are mapped back through the corresponding exponential maps for atomic decoding and distance evaluation. UMAP displays the resulting organization and particle selections. Its two-dimensional coordinates are used exclusively for visualization. Density generation retains the fixed analysis coordinates learned in the atomic stage, preserving the association between particles, atomic models, and density maps.

A.2 Structural Regularization

Structural regularization acts on the final atomic coordinates after combining domain transformations and local displacements (Figure 7d). Let $r_{jk}(\mathbf{z}) = \|\mathbf{x}_j(\mathbf{z}) - \mathbf{x}_k(\mathbf{z})\|_2$ and $r_{jk}^0 = \|\mathbf{x}_j^0 - \mathbf{x}_k^0\|_2$ denote decoded and reference distances. Following CryoSTAR (Li et al., 2024), connectivity, secondary structure, and reference distance constraints share the pairwise penalty

$$\ell_{\text{pair}}(\mathbf{z}; \mathcal{E}) = \frac{1}{|\mathcal{E}|} \sum_{(j,k)\in\mathcal{E}} \big[r_{jk}(\mathbf{z}) - r_{jk}^0\big]^2. \tag{12}$$

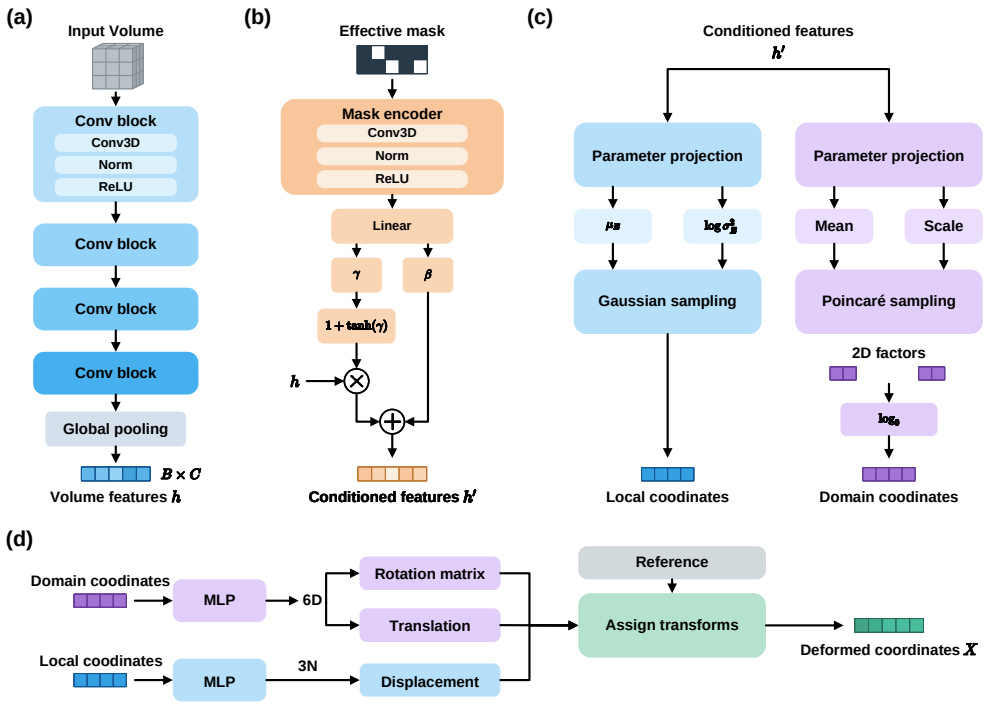


Figure 7: Network modules for conformational inference and atomic deformation in CryoHyper. (a) Volume encoding with convolutional blocks and global pooling. (b) Feature modulation conditioned on the effective Fourier mask. (c) Euclidean and Poincaré posterior sampling for local and domain coordinates, with the latter mapped to the tangent space for decoding. (d) Domain rotations and translations combined with local displacements to deform the reference atomic model.

The pair set $\mathcal{E}_{\text{con}}$ links consecutive modeled residues along each chain. We construct $\mathcal{E}_{\text{sse}}$ from DSSP (Kabsch & Sander, 1983) secondary structure assignments of the reference model. Within each annotated α-helix, Cα pairs span three or four modeled residues. Between annotated β-strands, nearest Cα partners are selected in both directions for each strand pair, retaining reference distances of at most 8 Å. These pairs and their reference distances remain fixed during training. The set $\mathcal{E}_{\text{dist}}$ contains additional pairs selected by reference distances. These sets respectively constrain chain continuity, local structural organization, and the surrounding atomic arrangement. For reference distance regularization, a small fraction of intra-chain pairs with the greatest variation in squared distance error across the training batch is excluded, allowing flexibility while retaining the remaining constraints.

Clash regularization penalizes candidate pairs that become closer than a minimum separation $r_{\min}$. With $\mathcal{C}(\mathbf{z}) = \{(j,k) \in \mathcal{E}_{\text{clash}} : r_{jk}(\mathbf{z}) < r_{\min}\}$, the penalty is

$$\ell_{\text{clash}}(\mathbf{z}) = \frac{\sum_{(j,k)\in\mathcal{C}(\mathbf{z})} \left[r_{\min} - r_{jk}(\mathbf{z})\right]^2}{\max(1, |\mathcal{C}(\mathbf{z})|)}. \tag{13}$$

It is zero when no candidate pair violates the separation threshold. The structural objective is the batch average of the weighted penalties

$$\mathcal{L}_{\text{str}} = \mathbb{E}_{\mathbf{z}}\left[\sum_{t\in\{\text{con,sse,dist}\}} \lambda_t \ell_{\text{pair}}(\mathbf{z}; \mathcal{E}_t) + \lambda_{\text{clash}} \ell_{\text{clash}}(\mathbf{z})\right]. \tag{14}$$

The same constraints apply to structures inferred from original and augmented inputs. They complement reconstruction supervision by constraining the structural changes represented at each latent coordinate.

A.3 ET Input Adaptation

Learnable spectral weighting adjusts the frequency content supplied to the encoder. We parameterize $a(r)$ by linear interpolation between trainable values at uniformly spaced radial frequencies, normalized to the axial Nyquist frequency of the encoder input. The fixed attenuation bound $0 \leq \alpha < 1$ constrains the weights to $1 - \alpha \leq w(r) \leq 1$, with $w(0) = 1$ preserving the input mean. Radial weights are shared across directions and particles and learned jointly with the conformational model. This bounds attenuation while allowing reconstruction supervision to determine which frequencies contribute to inference. The weighting remains active at inference.

Previous studies have demonstrated the effectiveness of rotation and artificial wedge masking for cryo-ET density restoration (Liu et al., 2022; Wiedemann & Heckel, 2024). In CryoHyper, artificial wedge training exposes the encoder to different patterns of missing information. For a selected subtomogram $\mathbf{V}_i$, we sample a proper cube rotation $\mathcal{R}_i$ from the 24 rotations implemented by axis permutations and flips. These rotations preserve both the voxel grid and handedness. Let $\mathbf{M}_i$ denote the original Fourier support and $\mathbf{A}_i$ the artificial mask with the dataset's original tilt support. The augmented input $\widetilde{\mathbf{V}}_i$ is obtained by applying the artificial mask to the rotated volume in Fourier space. The artificial mask is fixed in the augmented volume frame, while the original support rotates with the particle volume. Their intersection defines the effective mask $\widetilde{\mathbf{M}}_i$, identifying frequencies retained in the encoder input. The fraction of augmented inputs increases during warmup, and each augmented input replaces its original counterpart within the training batch.

The volume encoder extracts features through convolutional blocks and global pooling (Figure 7a). A mask encoder conditions these features on the effective Fourier support before posterior estimation (Figure 7b). For input feature h_j, the conditioned feature is

$$h'_j = h_j \left[1 + \tanh g_j(\widetilde{\mathbf{M}}_i)\right] + b_j(\widetilde{\mathbf{M}}_i), \tag{15}$$

where g_j and b_j are learned scale and shift functions. For unaugmented inputs, conditioning uses $\mathbf{M}_i$. This provides the encoder with the frequency support of each input when estimating conformational coordinates.

The reconstruction target is the rotated volume $\mathcal{R}_i\mathbf{V}_i$ before artificial masking, with supervision over $\mathcal{R}_i\mathbf{M}_i$. Consequently, frequencies removed by the artificial mask but present in the original data remain part of the target. Atomic predictions stay in the reference frame, and projection directions account for the augmentation rotation. Equivalently, reconstruction can be evaluated in the original subtomogram frame by composing the projection operator with the inverse rotation. Original and augmented inputs share the reconstruction, structural, and posterior regularization objectives. At inference, the encoder receives the original subtomogram and its Fourier support, with random rotation and artificial masking disabled.

B Datasets

B.1 Public Datasets

We obtained the public experimental datasets from EMPIAR. EMPIAR-10073 contains single particle cryo-EM images of the yeast U4/U6.U5 tri-snRNP complex (Nguyen et al., 2016) and is used for structural analysis and refinement. Additional experiments use EMPIAR-10180, a flexible precatalytic spliceosome complex (Plaschka et al., 2017), and EMPIAR-10059, which contains TRPV1 bound to DkTx and RTX in a lipid nanodisc (Gao et al., 2016), with surrounding density absent from the reference atomic model. The reference model for EMPIAR-10059 is apo TRPV1, determined in amphipol.

EMPIAR-10453 contains cryo-ET data of the SARS-CoV-2 spike protein (Turoňová et al., 2020), collected over a tilt range of $\pm 60^\circ$ with a 3° step. Subtomograms are obtained using Warp 2.0 (Tegunov & Cramer, 2019) and RELION 5. Table 3 summarizes particle counts, pixel sizes, box sizes, and reference models for the experimental datasets. Atomic training uses downsampled inputs, while density training retains the original box size. Domain assignments use ECOD annotations (Cheng et al., 2014) obtained through the PDB.

Table 3: Particle counts, preprocessing settings, and reference models for the experimental datasets. Box sizes are given in pixels.

Dataset	Particles	Pixel size (Å)	Input box	Atomic box	Density box	PDB
EMPIAR-10073	103,388	1.43	380	256	380	5GAN
EMPIAR-10180	54,002	1.70	320	256	320	5NRL
EMPIAR-10059	206,219	1.22	256	192	256	3J5P
EMPIAR-10453	3,619	1.33	256	128	256	6VSB

CryoBench Spike-MD (Jeon et al., 2024) contains 100,000 simulated single particle cryo-EM images of the SARS-CoV-2 spike protein, with source conformations sampled from molecular dynamics simulations. The original images have a pixel size of 1.5 Å and a box size of 256×256 pixels. The dataset provides source atomic structures, particle poses, and two collective variables describing conformational variation.

B.2 Simulated Cryo-ET Benchmark

The benchmark combines a known continuous conformational trajectory with particle labels and controlled tilt ranges and noise levels, allowing latent neighborhoods to be evaluated against the source conformational order. The molecular dynamics trajectory is obtained from DynaRepo with accession A005K (Mokhtari et al., 2026) and is based on the 3NVQ Sema7A–PlexinC1 complex (Liu et al., 2010). We project the trajectory onto its first principal component and export 100 atomic models at evenly spaced positions along PC1. Each source conformation is converted into a centered three dimensional density map using EMAN2 (Tang et al., 2007), with a target resolution of 4 Å and a voxel size of 2.0 Å.

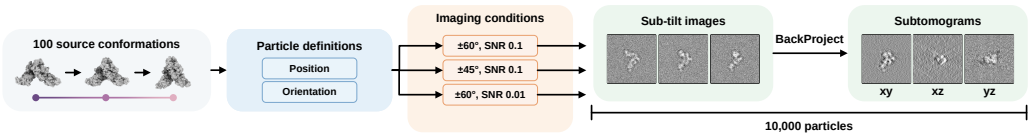


Figure 8: Generation of the simulated cryo-ET benchmark from 100 source conformations. Particle conformations, positions, and orientations are shared across three imaging conditions. Sub-tilt images are combined by Fourier backprojection to produce 10,000 subtomograms per condition.

Each particle is assigned a source conformation, a position in a virtual tomogram, and an orientation sampled uniformly over three dimensional rotations. These assignments are stored in a truth STAR file and shared across all imaging conditions. Each condition contains 10,000 particles with the same source conformations, positions, and orientations, enabling controlled comparisons of the effects of tilt range and noise on the recovered representation.

Subtomograms are generated directly at the particle level (Figure 8). For each particle, the assigned source density is projected at the prescribed tilt angles using its known orientation. Particle positions determine the depth contribution to defocus at each tilt. CTF modulation and additive Gaussian noise are applied to the sub-tilt images, which are then combined by Fourier backprojection to produce a three-dimensional subtomogram. Each output subtomogram has a voxel size of 2.0 Å and a box size of $160 \times 160 \times 160$ voxels. The finite tilt range leaves part of Fourier space unsampled and produces the missing-wedge. The simulation operates on local particle projections and does not require complete tomogram volumes or full tilt images. The baseline condition uses a tilt range of $\pm 60^\circ$ and SNR of 0.1. Two additional conditions separately reduce the tilt range to $\pm 45^\circ$ at SNR 0.1 and reduce SNR to 0.01 at $\pm 60^\circ$. Matched particle slices and Fourier amplitude sections illustrate the effects of these conditions (Figure 9). All conditions use a 3° tilt step. SNR is defined at the projection level as the ratio of signal variance to noise variance. For each particle, signal variance is

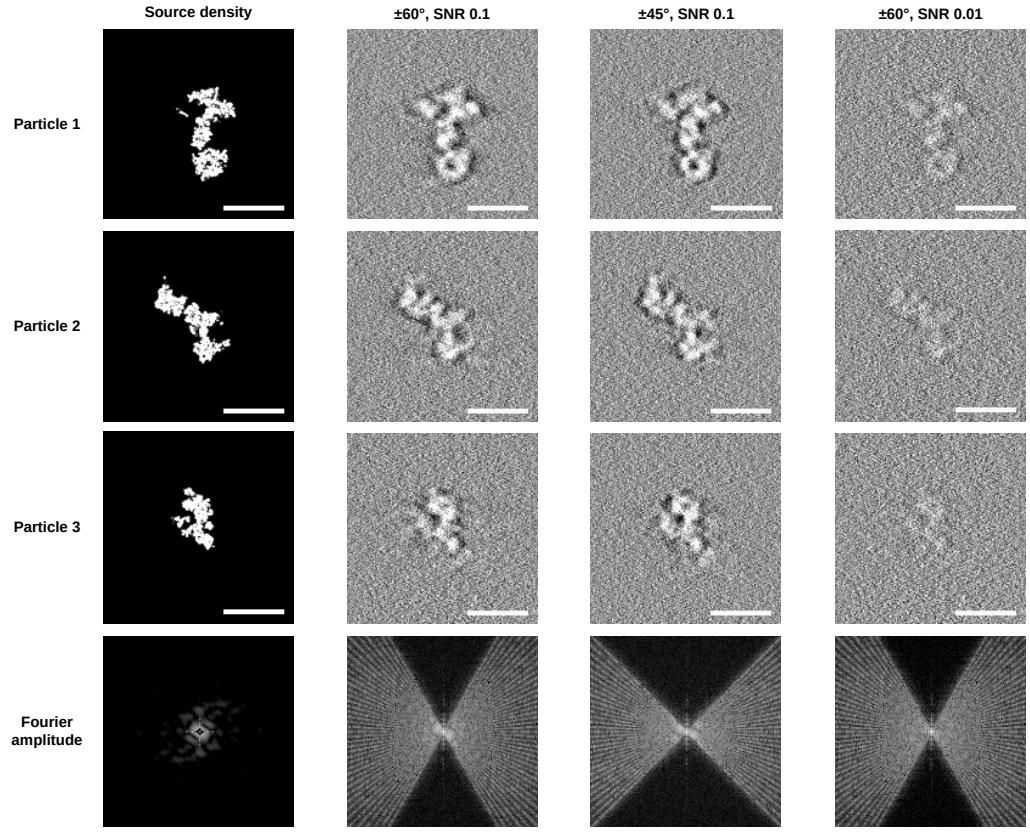


Figure 9: Source density and corresponding subtomogram slices for three particles under different tilt ranges and noise levels. Each particle retains the same source conformation and orientation across conditions. Representative Fourier amplitude sections are shown in the bottom row. Scale bars, 10 nm.

measured in the central particle region of the clean projection closest to zero tilt, and the resulting noise variance is held constant across its tilt images.

The benchmark comprises 100 source conformation PDB models and their density maps, 10,000 subtomograms per condition, corresponding particle tilt images, STAR files, CTF information, and particle conformation labels. CryoHyper and OPUS-Tomo use the subtomograms, while cryoDRGN-ET uses the corresponding particle tilt images. STAR files record particle identities, coordinates, Euler angles, and imaging parameters, with CTF values for individual tilts stored in associated files. Source labels are used exclusively for evaluation. The same particle correspondence is maintained across methods and imaging conditions.

C Implementation Details

C.1 Training Configuration

CryoHyper uses eight latent dimensions, with four allocated to domain motion and four to local atomic deformation. Cryo-EM images and cryo-ET subtomograms are encoded by a CNN and a 3D CNN, respectively. The domain decoder uses rotation and translation scales of 0.1 and 1.0. Both training stages use AdamW with a learning rate of 10^{-4} and 16-bit mixed precision. Training schedules are summarized in Table 4.

Table 4: CryoHyper training schedules.

Modality	Stage	Epochs	Batch size
Cryo-EM	Atomic	25	16
Cryo-EM	Density	20	3
Cryo-ET	Atomic	50	8
Cryo-ET	Density	20	3

The atomic reconstruction loss applies a 20 Å low-pass cutoff to the projection images being compared. Encoder inputs are downsampled particle images or subtomograms. Structural regularization and posterior settings are listed in Table 5. The KL weight increases during the first 10,000 steps to an upper value of 0.5.

Table 5: Atomic reconstruction and regularization settings.

Parameter	Value
Reconstruction weight	1.0
Connectivity weight	1.0
Secondary structure weight	1.0
Reference distance weight	1.0
Clash weight	1.0
Distance pair retention ratio	0.99
Within-chain distance cutoff	12 Å
Minimum clash cutoff	4 Å
KL upper weight	0.5
Free bits per dimension	1.5
KL warmup steps	10,000

ET Input Adaptation is enabled for cryo-ET. Its default spectral weighting uses six learnable radial parameters with an attenuation bound of $\alpha = 0.3$. The sigmoid outputs are initialized to 0.9, giving an initial frequency weight of 0.97, with the zero-frequency weight fixed at one. The fraction of particles receiving rotation and artificial wedge masking increases over the first 40,000 steps to 0.125. Augmentation replaces selected inputs within each batch and uses the same atomic training objective. Each particle undergoes one encoder and decoder pass. The original tilt range determines the artificial mask and the projection directions used for supervision.

Table 6: Computing infrastructure and software environment.

Item	Specification
GPU	NVIDIA A100 PCIe 40 GB
CPU	AMD EPYC 7742
Operating system	Ubuntu 22.04.4 LTS
PyTorch	2.8.0+cu128
CUDA	12.8
RELION	5.0-beta
Warp	2.0.0dev39
cryoSPARC	4.4.1

The density stage fixes the analysis coordinates obtained from the atomic posterior means. Its decoder has a hidden width of 1024 and uses Gaussian positional encoding and an implicit Fourier representation. Training minimizes the mean squared difference between predicted and target Fourier coefficients. Particle poses and tilt parameters determine the sampled Fourier slices. The learned conformational coordinates remain fixed throughout density optimization.

UMAP uses 15 neighbors, a minimum distance of 0.1, and random state 0. CryoHyper and comparison experiments use the same host, whose hardware and software environment are listed in Table 6.

C.2 Baselines and Ablations

Cryo-EM comparisons use the official implementations of CryoSTAR (Li et al., 2024), CryoDyna (Zhang et al., 2025), cryoDRGN (Zhong et al., 2021a), RECOVAR (Gilles & Singer, 2025), OPUS-DSD, and cryoSPARC 3D classification (Punjani et al., 2017). CryoSTAR and CryoDyna use 25 atomic epochs with batch size 16, followed by 20 density epochs with batch size 3. cryoDRGN uses 20 epochs with batch size 3. OPUS-DSD uses 25 epochs with batch size 3, an eight-dimensional latent representation, and a learning rate of 10^{-4}. RECOVAR uses its native pipeline with the particle STAR file and mask, retaining the original image size. All methods use the same EMPIAR-10073 particles for the comparison. Particle selection and refinement follow a common downstream protocol.

Cryo-ET comparisons use cryoDRGN-ET (Rangan et al., 2024) and OPUS-Tomo (Luo et al., 2024). Both methods are evaluated on the same simulated particles as CryoHyper under each tilt range and noise condition. cryoDRGN-ET uses particle tilt images, and OPUS-Tomo uses subtomograms. Ground truth labels are matched to particle identities across these input formats and used only for evaluation. Baseline input preparation and density generation follow the respective implementations.

The latent allocation ablation fixes the total dimension at eight and compares H2E6, H4E4, and H6E2 on EMPIAR-10073. H and E denote hyperbolic and Euclidean dimensions. The E4E4 control replaces the four-dimensional hyperbolic domain representation with a four-dimensional Euclidean representation, retaining the remaining network architecture, training data, training schedule, and random seed. Euclidean posterior sampling, KL regularization, and distances replace their hyperbolic counterparts. Both deformation branches and their parameter counts are preserved, isolating the contribution of the hyperbolic representation within the same architecture.

The ET Input Adaptation ablation compares the full model with the adaptation disabled under the baseline simulation condition of $\pm 60^\circ$, a 3° step, and SNR 0.1. The ablated model receives subtomograms without learnable spectral weighting or artificial wedge augmentation. The remaining network architecture is unchanged, and both models use the same particles and training schedule.

C.3 Computational Cost

Tables 7 and 8 report parameter counts and computation times on EMPIAR-10073 and the simulated 3NVQ cryo-ET dataset, respectively, using a single NVIDIA A100 GPU. Parameter counts sum the atomic and density networks for methods with two stages. On EMPIAR-10073, CryoHyper completes both stages in 67.77 hours, comparable to CryoSTAR and shorter than CryoDyna, while achieving better atom-density consistency and refinement resolution (Table 1).

Table 7: Computational cost on EMPIAR-10073. RECOVAR time covers its native pipeline.

Method	Parameters (M)	Atomic (h)	Density (h)	Total (h)
cryoDRGN	123.58	N/A	37.73	37.73
RECOVAR	N/A	N/A	11.53	11.53
OPUS-DSD	61.42	N/A	280.82	280.82
cryoSPHERE	34.04	47.58	N/A	47.58
CryoSTAR	53.76	10.73	56.61	67.34
CryoDyna	57.87	24.60	73.79	98.39
CryoHyper	54.08	27.72	40.06	67.77

On simulated cryo-ET data, CryoHyper uses 19.05 million parameters across both stages, fewer than either baseline, and completes training in 26.82 hours (Table 8). Its total training time is below that of OPUS-Tomo, with stronger conformational neighborhood recovery (Figure 3).

Table 8: Computational cost on the simulated 3NVQ cryo-ET dataset.

Method	Parameters (M)	Atomic (h)	Density (h)	Total (h)
cryoDRGN-ET	31.25	N/A	7.32	7.32
OPUS-Tomo	61.40	N/A	28.45	28.45
CryoHyper	19.05	12.45	14.37	26.82

D Evaluation Protocols

D.1 Conformational Neighborhoods

Conformational neighborhoods are evaluated at the particle level. For the simulated benchmark, particle i has source conformation label c_i, and its ground truth distance to particle j is $d_{\mathrm{GT}}(i,j) = |c_i - c_j|$. Labels follow the order of the 100 source conformations along PC1. CryoHyper uses the full latent distance defined in Equation (11). Methods with a Euclidean latent representation use Euclidean distance in their original latent coordinates. UMAP coordinates are used only for visualization.

Balanced Neighborhood Similarity. Balanced Neighborhood Similarity (BNS) measures neighborhood overlap after correcting for chance and balancing contributions from source conformations. Let N be the number of evaluated particles, C the number of source conformations, and I_c the particles belonging to conformation c. Denote the k nearest neighbors of particle i in latent and ground truth space by $\mathcal{N}_i^{\mathrm{L}}(k)$ and $\mathcal{N}_i^{\mathrm{GT}}(k)$, excluding the query particle. The balanced overlap is

$$O(k) = \frac{1}{C}\sum_{c=1}^{C}\frac{1}{|I_c|}\sum_{i\in I_c}\frac{|\mathcal{N}_i^{\mathrm{L}}(k)\cap\mathcal{N}_i^{\mathrm{GT}}(k)|}{k}. \tag{16}$$

The expected overlap of random neighborhoods is $q(k) = k/(N-1)$. The corrected score is

$$\mathrm{BNS}(k) = 100\,\frac{O(k)-q(k)}{1-q(k)}. \tag{17}$$

A score of zero corresponds to chance overlap, and 100 indicates identical neighborhoods. We evaluate neighborhood radii from 1% to 20%, setting $k_n = \mathrm{round}(Nn/C)$ for $n = 1,\ldots,20$ and $C = 100$. BNS-AUC$_{20}$ is the trapezoidal area under the curve divided by the interval length from 1% to 20%. Higher values indicate stronger neighborhood preservation. Evaluation uses 20 repeats, each retaining 80% of the particles without replacement within every source conformation. Neighborhood sizes are recomputed for each retained subset. Curves and AUC values are reported as means and standard deviations across repeats.

Bidirectional Information Imbalance. Information Imbalance (Glielmo et al., 2022) measures how informative neighborhoods in one space are about neighborhoods in another. Let r_{ij}^{A} be the rank of particle j among the neighbors of particle i in space A, starting at one and excluding i. For N evaluated particles, the directional estimator is

$$\Delta(A\to B;k) = \frac{2}{N^2 k}\sum_{i=1}^{N}\sum_{j:\, r_{ij}^{A}\le k} r_{ij}^{B}. \tag{18}$$

We report both $\Delta(\mathrm{L}\to\mathrm{GT})$ and $\Delta(\mathrm{GT}\to\mathrm{L})$. Lower values in the first direction indicate that latent neighbors are close along the source trajectory. Lower values in the reverse direction indicate that source trajectory neighbors remain close in the learned representation. Each evaluation samples 2,000 particles without replacement from the 10,000 particles, with five repeats and $k \in \{25, 125, 250, 500\}$. The reported mean and standard deviation aggregate the 20 values from the four neighborhood sizes and five repeats.

D.2 Atomic and Density Agreement

Atom-density consistency. Atom-density consistency (AD-FSC) compares the density rendered from a decoded atomic model with the density generated at the same latent coordinate. Atomic models are rendered as Gaussian mixture maps with element-dependent amplitudes and a Gaussian width of 2.0 Å. The paired maps share the same grid and pixel size. For Fourier coefficients F_1 and F_2, Fourier shell correlation (FSC) (van Heel & Schatz, 2005) at spatial frequency s is

$$\mathrm{FSC}(s) = \frac{\mathrm{Re}\sum_{\mathbf{k}\in\mathcal{S}(s)} F_1(\mathbf{k})F_2^*(\mathbf{k})}{\sqrt{\sum_{\mathbf{k}\in\mathcal{S}(s)} |F_1(\mathbf{k})|^2 \sum_{\mathbf{k}\in\mathcal{S}(s)} |F_2(\mathbf{k})|^2}}, \tag{19}$$

where $\mathcal{S}(s)$ is the Fourier shell and * denotes complex conjugation. AD-FSC reports the map-to-map FSC resolution at the 0.5 threshold. A lower resolution value indicates agreement extending to finer spatial detail. This metric is evaluated for methods that produce both atomic models and density maps.

Backprojection agreement. Backprojection agreement (BP-FSC) compares a decoded map with a map obtained by directly backprojecting particle images. For each of eight sampled PC1 coordinates on EMPIAR-10073, we generate a density map with the corresponding method and select its 10,000 nearest particles in the full latent space. Their images are backprojected in cryoSPARC using the existing particle poses. BP-FSC is the map-to-map FSC resolution at the 0.5 threshold for this pair. The same decoded and backprojected maps are used for the following real-space metrics. Let u_q and v_q denote the backprojected and decoded densities at voxel q, and m_q the corresponding mask weight.

Map correlation coefficient. Map correlation coefficient (Map CC) measures voxel intensity correlation within the mask,

$$\mathrm{CC}(u,v) = \frac{\sum_q m_q(u_q-\bar{u})(v_q-\bar{v})}{\sqrt{\sum_q m_q(u_q-\bar{u})^2 \sum_q m_q(v_q-\bar{v})^2}}, \tag{20}$$

where $\bar{u} = \sum_q m_q u_q / \sum_q m_q$ and $\bar{v}$ is defined analogously. Higher Map CC indicates stronger agreement in the spatial distribution of density.

Structural similarity. Structural similarity (SSIM) (Wang et al., 2004) combines local intensity, contrast, and structural agreement. For local means μ_u, μ_v, variances σ_u^2, σ_v^2, and covariance σ_{uv},

$$\mathrm{SSIM}_{\mathrm{local}} = \frac{(2\mu_u\mu_v + C_1)(2\sigma_{uv} + C_2)}{(\mu_u^2 + \mu_v^2 + C_1)(\sigma_u^2 + \sigma_v^2 + C_2)}. \tag{21}$$

Local statistics use three dimensional Gaussian neighborhoods with a standard deviation of 1.5 voxels. The decoded map is intensity normalized to the backprojected map, and both maps are scaled to a common unit range for SSIM, with $C_1 = 0.01^2$ and $C_2 = 0.03^2$. The reported score is the mask-weighted average of local values. Multiscale structural similarity (MS-SSIM) (Wang et al., 2003) combines contrast and structure across five successively downsampled scales with the intensity term at the coarsest scale. Higher SSIM and MS-SSIM indicate stronger agreement.

Additional map metrics. Normalized mean squared error (NMSE) uses the intensity normalized decoded map $\tilde{v}$,

$$\mathrm{NMSE}(u,\tilde{v}) = \frac{\sum_q m_q(u_q-\tilde{v}_q)^2}{\sum_q m_q(u_q-\bar{u})^2}. \tag{22}$$

Lower NMSE indicates smaller density error relative to the intensity variation of the backprojected map. Normalized mutual information (NMI) (Vinh et al., 2009) measures dependence between paired voxel intensities,

$$\mathrm{NMI}(u,v) = \frac{2I(u;v)}{H(u)+H(v)}. \tag{23}$$

Here, I denotes mutual information and H denotes entropy, estimated from a joint intensity histogram with 64 bins per map and voxel contributions weighted by the mask. Higher NMI indicates greater shared information. For the EMPIAR-10073 comparison, map metrics are summarized by the mean and standard deviation across eight sampled states.

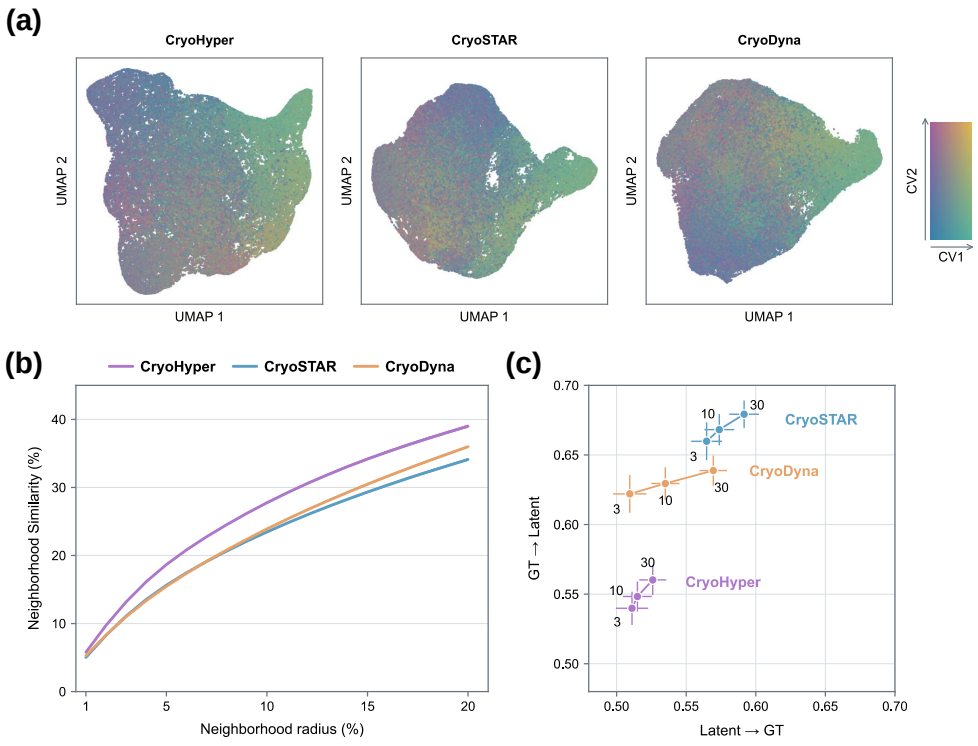


Figure 10: Conformational space recovery on CryoBench Spike-MD. (a) Latent UMAPs jointly colored by the two ground truth collective variables, CV1 and CV2. (b) Neighborhood Similarity across neighborhood radii. (c) Bidirectional Information Imbalance at $k = 3, 10, 30$, with point labels indicating k. Shading and error bars denote standard deviations across evaluation subsets.

D.3 Particle Selection and Refinement

Cryo-EM refinement. Latent guided refinement resolution (R-FSC) evaluates whether local particle neighborhoods support refinement. On EMPIAR-10073, the 10,000 nearest particles to each of eight sampled PC1 coordinates are selected independently using distances in the full latent space. Subsets can therefore overlap. Particle identities and their original STAR entries are retained, and all subsets are processed with a common cryoSPARC refinement protocol (Punjani et al., 2017). Resolution is determined by gold-standard half-map FSC at the 0.143 threshold (Scheres & Chen, 2012). Means and standard deviations are calculated across the eight subsets.

3D classification. CryoSPARC 3D classification partitions the same particle population into eight classes. Each class with sufficient particles undergoes homogeneous refinement in cryoSPARC. We compare class populations and attainable refinement resolutions with those of the eight latent neighborhoods. Particle counts are retained in the comparison because the latent neighborhoods have a fixed size, while class populations vary. Classes that do not support refinement are indicated separately in the figure.

Cryo-ET refinement. On EMPIAR-10453, 1,185 particles are selected from each endpoint of the PC1 trajectory. The two subsets are processed independently with RELION 5, and their resolutions are evaluated by gold-standard half-map FSC at 0.143. The resulting maps are compared with each other and with the decoded motion to assess whether the particle subsets support the recovered conformational differences.

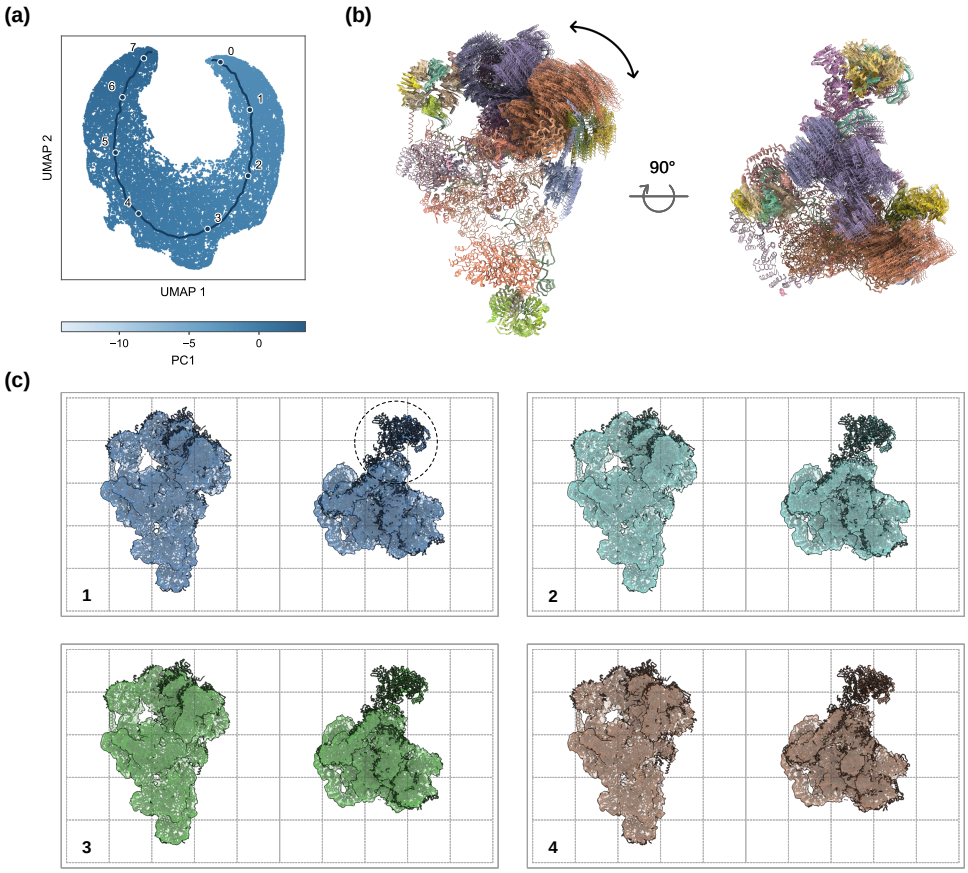


Figure 11: CryoHyper results on EMPIAR-10180. (a) Latent UMAP colored by PC1 with the sampled trajectory. (b) Atomic motion decoded along PC1 in two views. (c) Sampled atomic models overlaid with corresponding density maps. The dashed circle highlights density loss due to flexibility.

E Additional Experiments

E.1 Additional Cryo-EM Results

CryoBench Spike-MD (Jeon et al., 2024) evaluates conformational space recovery from simulated cryo-EM images against two ground truth collective variables (CVs). We compare CryoHyper, CryoSTAR, and CryoDyna using the official Neighborhood Similarity metric and bidirectional Information Imbalance (Figure 10). CryoHyper achieves the highest neighborhood similarity throughout the evaluated radius range. It also attains the lowest GT-to-latent imbalance at all three neighborhood sizes, with lower imbalance than both baselines in both directions at $k = 10$ and 30. The stronger neighborhood preservation shows that proximity in the source conformational space is better retained in the learned representation, extending the evidence for conformational space recovery to simulated cryo-EM data.

EMPIAR-10180 contains a large, flexible pre-catalytic spliceosome complex (Plaschka et al., 2017). The dataset allows evaluation of atomic motion learning when pronounced flexibility blurs local structural signal during particle averaging and reconstruction. CryoHyper recovers conformational motion consistent with previous analysis of its structural variability (Nakane et al., 2018) (Figure 11). Across eight sampled states, atom-density consistency ranges from 19.19 to 19.96 Å, with a mean of 19.57 Å. U2 snRNP flexibility blurs the local density, affecting its agreement with the atomic model

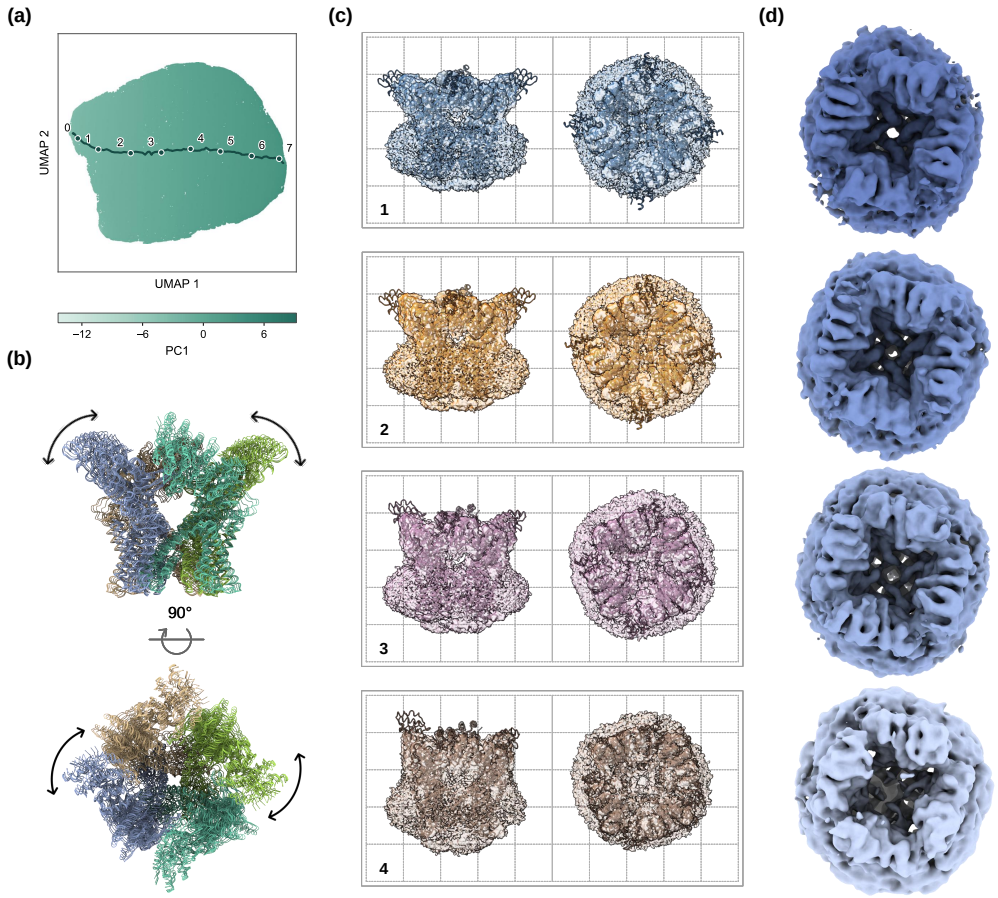


Figure 12: CryoHyper results on EMPIAR-10059. (a) Latent UMAP colored by PC1 with the sampled trajectory. (b) Atomic motion decoded along PC1 in two views. (c) Representative atomic models overlaid with their decoded density maps. (d) Representative results of latent guided refinement using 10,000 particles per sampled PC1 coordinate.

(dashed circle in Figure 11c). Latent guided refinement reaches a mean resolution of 9.55 Å and a median of 9.54 Å. The learned particle neighborhoods retain sufficient conformational coherence to support refinement of this flexible complex.

EMPIAR-10059 contains TRPV1 bound to DkTx and RTX in a lipid nanodisc (Gao et al., 2016). The apo reference model 3J5P differs from the target state and does not cover all surrounding density, providing a test of robustness to reference model mismatch. CryoHyper recovers an ordered trajectory of atomic motion and corresponding density variation (Figure 12). Refinement resolutions across eight sampled states range from 4.31 to 6.59 Å, with a mean of 5.02 Å and a median of 4.38 Å. Representative refined maps show a motion trend consistent with the decoded structures (Figure 12d). These results show that a reference from a different ligand state can support conformational learning and particle selection for refinement, with density recovery extending beyond the modeled structure.

On EMPIAR-10073, latent guided refinement resolves conformational differences within particle populations grouped together by 3D classification (Figure 13). The four classes supporting homogeneous refinement occupy broad, overlapping regions of the CryoHyper latent space, and each contributes particles to multiple sampled states (Figure 13b,c). Maps refined along PC1 reveal progressive head and foot displacements, providing finer conformational differentiation than the four class averages (Figure 13a). These changes are supported by refinement of the selected particle

images and agree with the decoded motion in Figure 4. The particle distributions and refined maps show how the recovered conformational organization supports particle selection that resolves structural variation within conventional 3D classes.

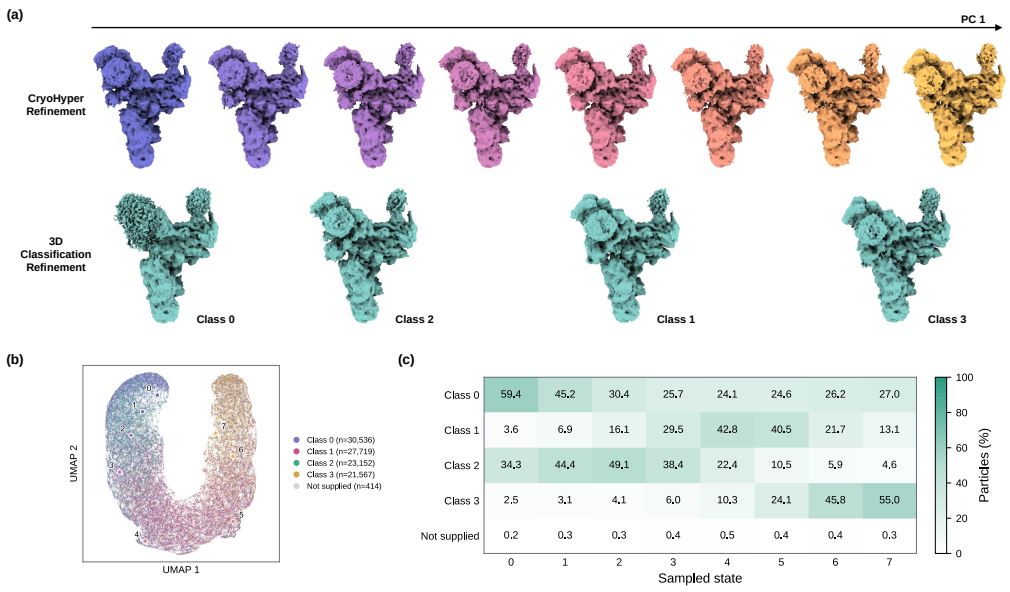


Figure 13: Latent guided refinement and 3D classification on EMPIAR-10073. (a) Maps refined from 10,000 particles at each of eight sampled PC1 coordinates, compared with maps obtained by homogeneous refinement of four 3D classes. (b) CryoHyper UMAP colored by 3D class membership, with the eight sampling points marked. (c) Class composition of each selected particle subset, expressed as percentages of its 10,000 particles. Individual classes contribute to multiple sampled states.

E.2 Extended Metric Comparison

Table 9: Additional map agreement metrics on EMPIAR-10073 using normalized mutual information (NMI), normalized mean squared error (NMSE), and multiscale structural similarity (MS-SSIM). Values are means and standard deviations across eight sampled states. Best results are shown in bold.

Method	NMI ↑	NMSE ↓	MS-SSIM ↑
cryoDRGN	0.153 ± 0.007	0.305 ± 0.019	0.875 ± 0.022
RECOVAR	0.106 ± 0.004	0.421 ± 0.013	0.813 ± 0.015
OPUS-DSD	0.040 ± 0.005	0.805 ± 0.014	0.498 ± 0.013
CryoSTAR	0.100 ± 0.019	0.480 ± 0.079	0.740 ± 0.077
CryoDyna	0.227 ± 0.031	0.158 ± 0.051	0.928 ± 0.022
CryoHyper	$\mathbf{0.252 \pm 0.017}$	$\mathbf{0.128 \pm 0.021}$	$\mathbf{0.942 \pm 0.010}$

The EMPIAR-10073 comparison additionally evaluates normalized mutual information (NMI), normalized mean squared error (NMSE), and multiscale structural similarity (MS-SSIM) between decoded maps and maps obtained by backprojecting the corresponding 10,000 nearest particle images using their existing poses (Table 9). These metrics use the same eight sampled states and map pairs as the main comparison. CryoHyper achieves the highest NMI of 0.252 ± 0.017 and MS-SSIM of 0.942 ± 0.010, together with the lowest NMSE of 0.128 ± 0.021.