CRYOGAUSSIAN: A UNIFIED GAUSSIAN SPLATTING METHOD FOR HOMOGENEOUS AND HETEROGENEOUS CRYO-EM RECONSTRUCTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Single-particle cryo-electron microscopy (cryo-EM) resolves biomolecular structures at near-atomic resolution without crystallization. Reconstructing a biomolecule's 3D density from noisy 2D particle images presents two core challenges: (1) recovering as much reliable structural detail as the data support in both homogeneous and heterogeneous reconstruction; and (2) identifying rare and transient compositional states and tracing subtle, continuous conformational changes. To address both challenges, we propose CryoGaussian, a cryo-EM-native Gaussian splatting method that unifies homogeneous and heterogeneous reconstruction for high-fidelity structural recovery. Starting from randomly initialized Gaussians, CryoGaussian fits an explicit 3D density representation directly to particle images through orthographic line-integral projection modulated by the contrast transfer function. To model heterogeneity, CryoGaussian learns particle-image and per-Gaussian latents that jointly condition Gaussian attributes to capture compositional and conformational variation. The inferred states then guide optimization of the Gaussian representation for detailed per-state reconstruction. We validate CryoGaussian: (1) on all five synthetic CryoBench datasets, it achieves competitive performance in heterogeneity estimation and per-state reconstruction; (2) on homogeneous experimental data, it resolves structural detail in the RAG1–RAG2 complex, the SARS-CoV-2 spike on intact virions, and the 80S ribosome; and (3) on heterogeneous experimental data, it reveals previously validated structural features through unsupervised heterogeneity estimation, including a 50S assembly intermediate with rRNA helix 68 density but no central protuberance and continuous conformational variation in the SF3b and helicase regions of the pre-catalytic spliceosome. Sample maps and links are provided in the supplementary material; code available upon publication.

1 INTRODUCTION

Single-particle cryo-electron microscopy (cryo-EM) reconstructs three-dimensional biomolecular density at near-atomic resolution from noisy two-dimensional particle images without crystallization (Cheng et al., 2015; Yip et al., 2020). Resolving compositional and conformational heterogeneity reveals molecular organization and function, as illustrated by respiratory supercomplexes and proteasome states (Zheng et al., 2024; Dong et al., 2019). Such interpretation also depends on local map quality: resolving side chains and cofactors can distinguish chemical states. For example, a 1.9 Å yeast fatty acid synthase map revealed a distortion of flavin mononucleotide indicating its reduced state (Singh et al., 2023). Biological interpretation therefore requires both reliable structural detail and an account of the states present in a sample.

Recovering reliable structural detail from homogeneous and heterogeneous particle images poses challenges. Low signal-to-noise ratios obscure fine features, while overfitting can introduce misleading density (Scheres & Chen, 2012). For heterogeneous samples, discrete classification can miss rare states or average continuous variation (Rabuck-Gibbons et al., 2022; Punjani & Fleet, 2023). Density-preserving deformation models including 3DFlex capture motion but have limited applicability to compositional changes (Punjani & Fleet, 2023). Other continuous representations

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.

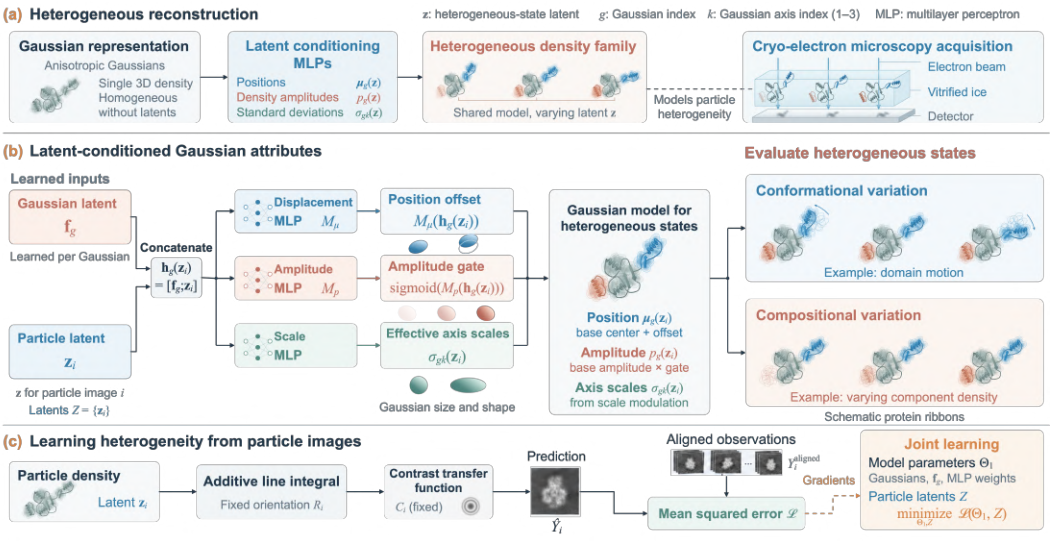

Figure 1: **CryoGaussian Overview.** Cryo-EM-native Gaussian splatting unifies homogeneous reconstruction, compositional/conformational heterogeneity estimation, and per-state reconstruction.

accommodate both heterogeneity types (Zhong et al., 2021), yet resolving heterogeneity alone does not ensure faithful per-state reconstruction. Even with identical inferred latents, reconstruction procedures recover different structural detail (Gilles & Singer, 2025). Recovering compositional and conformational variation while retaining reliable per-state detail remains challenging.

Gaussian mixtures provide physically motivated representations of molecular density and structural variation (Doyle & Turner, 1968; Peng et al., 1996; Chen & Ludtke, 2021). 3D Gaussian Splatting (3DGS) makes explicit Gaussians computationally attractive through differentiable rasterization for high-fidelity novel-view synthesis (Kerbl et al., 2023). However, cryo-EM records line-integral projections modulated by the contrast transfer function (CTF), and particles may differ compositionally and conformationally. Projected Gaussians add without occlusion, requiring rasterization adapted to cryo-EM image formation (Singer & Sigworth, 2020; Chen & Ling, 2026). Yet a physically appropriate forward model leaves Gaussian width and count undetermined for noisy, discretely sampled images. This motivates deriving representation constraints from particle-image sampling and spatial support, alongside modeling structural variation and optimizing the cryo-EM rasterizer.

We introduce CryoGaussian, a physically grounded framework unifying homogeneous and heterogeneous reconstruction through cryo-EM-native Gaussian splatting. With randomly initialized Gaussians, it uses data-derived width and count constraints to control the representation's detail and capacity. Homogeneous reconstruction fits one density; heterogeneous reconstruction introduces particle-image and per-Gaussian latents that jointly condition Gaussian attributes to capture compositional and conformational variation. The inferred states guide optimization of the Gaussian representation for detailed per-state reconstruction. A CUDA rasterizer optimized for cryo-EM image formation and distributed training supports both settings. These designs connect heterogeneity estimation with high-fidelity structural recovery within a common reconstruction framework.

Main contributions. (i) We introduce CryoGaussian, a physically grounded, cryo-EM-native Gaussian splatting method unifying homogeneous and heterogeneous reconstruction for high-fidelity structural recovery. (ii) CryoGaussian captures compositional and conformational heterogeneity and reconstructs detailed per-state densities, as evaluated on synthetic and experimental datasets. (iii) CryoGaussian shows potential for structural discovery through unsupervised heterogeneity estimation, recovering a previously validated 50S assembly intermediate with rRNA helix 68 density but no central protuberance and known conformational variation in the SF3b and helicase regions of the pre-catalytic spliceosome (Zhong et al., 2021). (iv) We optimize the CUDA rasterizer for cryo-EM image formation through per-Gaussian gradient accumulation and projection scheduling, and implement distributed training to accelerate reconstruction.

2 Related Work

Cryo-EM reconstruction. RELION, CryoSPARC, and EMAN2 reconstruct voxelized densities (Scheres, 2012b; Punjani et al., 2017; Tang et al., 2007), whereas neural fields represent density as a continuous function of Fourier or real-space coordinates (Zhong et al., 2021; Qu et al., 2025; Huang et al., 2024; Herreros et al., 2025). Empirical Bayesian methods regularize reconstruction using estimated signal and noise statistics (Scheres, 2012a), while Blush and CryoFM incorporate learned density priors through denoising and posterior sampling, respectively (Kimanius et al., 2024; Zhou et al., 2025). CryoGaussian uses an explicit Gaussian density representation for homogeneous and heterogeneous reconstruction without a pretrained density prior.

Heterogeneous reconstruction. Discrete classification reconstructs densities for particle classes (Scheres et al., 2007; Scheres, 2012b). Continuous descriptions use linear subspaces in 3DVA, covariance estimation in RECOVAR, or latent-conditioned density decoders in CryoDRGN and OPUS-DSD (Punjani & Fleet, 2021; Gilles & Singer, 2025; Zhong et al., 2021; Luo et al., 2023). Hydra and CryoHype model mixtures of distinct complexes with neural fields (Levy et al., 2024; Gu et al., 2026). Motion-based approaches use rigid-body transformations, deformation fields, or Gaussian displacements (Nakane et al., 2018; Punjani & Fleet, 2023; Schwab et al., 2024); CryoSPIRE additionally models compositional variation (Shekarforoush et al., 2025). CryoDRGN-AI directly optimizes particle-image latents, whereas cryoSTAR uses atomic-model constraints to infer latents for subsequent per-state reconstruction (Levy et al., 2025; Li et al., 2024). CryoGaussian infers compositional and conformational heterogeneity by fitting latent-conditioned Gaussians directly to particle images, then uses these latents for per-state reconstruction.

Gaussian mixtures in cryo-EM. Gaussian mixtures can fit density maps and infer heterogeneity from particle images (Kawabata, 2008; 2018; Chen & Ludtke, 2021; Chen et al., 2023). Initialization and count can incorporate structural information: e2gmm uses density-map fits at a target resolution or atom counts (Chen & Ludtke, 2021; Chen et al., 2024), DynaMight recommends two Gaussians per residue (Schwab et al., 2024), and CryoSPIRE initializes from reconstructed density (Shekarforoush et al., 2025). CryoGaussian initializes Gaussians randomly and derives the Gaussian count from a spatial capacity estimate based on particle images instead of prior structural knowledge.

Gaussian splatting. Classical volume splatting projects reconstruction kernels into image-plane footprints (Westover, 1990; Zwicker et al., 2001); 3DGS combines explicit Gaussians with differentiable rasterization and depth-ordered alpha compositing (Kerbl et al., 2023). Adaptations to projection imaging include X-ray image synthesis and tomographic reconstruction (Cai et al., 2025; Zha et al., 2024; Talegaonkar et al., 2025), and CryoSplat's homogeneous cryo-EM reconstruction from randomly initialized Gaussians using CTF-modulated orthographic projection (Chen & Ling, 2026). Mip-Splatting adapts Gaussian kernels to input-view sampling rates, while 3DGS-MCMC relocates low-opacity Gaussians (Yu et al., 2024; Kheradmand et al., 2024). CryoGaussian addresses homogeneous and heterogeneous reconstruction using 3DGS adapted to cryo-EM image formation.

3 Methods

CryoGaussian targets high-fidelity homogeneous and heterogeneous reconstruction from N particle images $\{Y_i\}_{i=1}^N$ with pixel size a and dimensions $H \times W$. We adopt the fixed-pose setting used by CryoDRGN and RECOVAR (Zhong et al., 2021; Gilles & Singer, 2025), treating orientations $R_i \in \mathrm{SO}(3)$, in-plane translations $\mathbf{t}_i$, and CTFs C_i as fixed inputs. Homogeneous reconstruction learns a single Gaussian density; heterogeneous reconstruction extends it with particle-image and per-Gaussian latents that condition Gaussian attributes to model compositional and conformational heterogeneity (Figure 1).

3.1 Gaussian density representation and cryo-EM image formation

We represent the biomolecular density conditioned on latent $\mathbf{z}$ as

$$\rho_\Theta(\mathbf{x};\mathbf{z}) = \sum_{g=1}^{G} p_g(\mathbf{z})\mathcal{N}_3(\mathbf{x};\boldsymbol{\mu}_g(\mathbf{z}),\boldsymbol{\Sigma}_g(\mathbf{z}))\,, \tag{1}$$

where G is the Gaussian count and $\mathcal{N}_3$ denotes a normalized three-dimensional Gaussian. Each Gaussian has a position $\boldsymbol{\mu}_g(\mathbf{z}) \in \mathbb{R}^3$, covariance $\boldsymbol{\Sigma}_g(\mathbf{z}) \succ 0$, and amplitude $p_g(\mathbf{z}) > 0$. The parameters Θ comprise the Gaussian parameters and, in heterogeneous reconstruction, per-Gaussian latents and Gaussian attribute decoder weights (Section 3.3). Homogeneous reconstruction omits $\mathbf{z}$.

Under the weak-phase-object and projection approximations, cryo-EM image formation comprises line integration along the electron beam followed by CTF modulation (Singer & Sigworth, 2020). Let $P = [I_2 \ \mathbf{0}]$ select the two image-plane coordinates. For physical image-plane coordinates $\mathbf{v}$, orthographic integration is analytic:

$$\begin{aligned}[\mathcal{P}_{R_i}\rho_\Theta](\mathbf{v};\mathbf{z}_i) &\equiv \int_{\mathbb{R}} \rho_\Theta\big(R_i^\mathsf{T}(v_1,v_2,w)^\mathsf{T};\mathbf{z}_i\big)\ dw \\ &= \sum_{g=1}^{G} p_g(\mathbf{z}_i)\mathcal{N}_2\big(\mathbf{v};PR_i\boldsymbol{\mu}_g(\mathbf{z}_i), PR_i\boldsymbol{\Sigma}_g(\mathbf{z}_i)R_i^\mathsf{T}P^\mathsf{T}\big)\,,\end{aligned} \tag{2}$$

where $\mathcal{N}_2$ is a normalized two-dimensional Gaussian. We evaluate these Gaussian footprints on the image grid with spacing a and apply CTF modulation:

$$\widehat{Y}_i^{\mathrm{pre}}(\mathbf{v}) = \kappa[\mathcal{P}_{R_i}\rho_\Theta](\mathbf{v};\mathbf{z}_i), \qquad \widehat{Y}_i = \mathcal{F}_2^{-1}\Big[C_i\,\mathcal{F}_2[\widehat{Y}_i^{\mathrm{pre}}]\Big]\,. \tag{3}$$

Here $\mathcal{F}_2$ is the discrete Fourier transform of the sampled image, and κ is the fixed conversion from projected density to image intensity under the chosen normalization. Gaussian contributions add independently of depth order, so our differentiable rasterizer uses additive accumulation in place of standard 3DGS's front-to-back alpha compositing (Kerbl et al., 2023).

Applying the supplied in-plane translations through the Fourier shift theorem gives aligned targets Y_i^{aligned}. We minimize the image-domain mean squared error, consistent with the squared-error reconstruction terms used in neural and Gaussian-based cryo-EM methods (Zhong et al., 2021; Shekarforoush et al., 2025; Chen & Ling, 2026):

$$\mathcal{L}(\Theta, Z) = \frac{1}{NHW}\sum_{i=1}^{N}\sum_{\mathbf{v}}\Big[\widehat{Y}_i(\mathbf{v};\Theta,\mathbf{z}_i) - Y_i^{\mathrm{aligned}}(\mathbf{v})\Big]^2, \qquad Z = \{\mathbf{z}_i\}_{i=1}^N, \tag{4}$$

where the inner sum runs over the image grid. For homogeneous reconstruction, $\mathcal{L}$ depends only on Θ, which we optimize subject to the two data-derived constraints in Section 3.2.

3.2 Data-derived constraints

The pixel size a defines the Nyquist frequency $1/(2a)$ (Shannon, 1949). We use the corresponding sampling scale $d = 2a$ for both the Gaussian FWHM constraint and Gaussian-count estimation.

Gaussian FWHM constraint. To prevent Gaussians from shrinking below the sampling scale $d = 2a$ when fitting noisy particle images, we adapt standard 3DGS (Kerbl et al., 2023) by constraining each Gaussian's full width at half maximum (FWHM). For every effective principal-axis scale σ (standard deviation),

$$\mathrm{FWHM} = 2\sqrt{2\ln 2}\,\sigma \geq d, \qquad d = 2a. \tag{5}$$

Figure 18 evaluates this constraint in heterogeneity estimation.

Gaussian count constraint. Gaussian-mixture cryo-EM methods commonly specify the Gaussian count before optimization (Chen et al., 2024; Schwab et al., 2024). We estimate this representation budget G from pixel size and spatial support in a CTF-weighted backprojection ρ_{BP} of the particle images (Appendix F.2). Let $\mathcal{X}$ denote its voxel centers on a grid with spacing a. For a relative density threshold $\tau \in (0,1)$, we define

$$\rho_+(\mathbf{x}) = \max(\rho_{\mathrm{BP}}(\mathbf{x}), 0), \qquad \mathcal{X}_\tau = \left\{\mathbf{x}\in\mathcal{X} : \rho_+(\mathbf{x}) \geq \tau \max_{\mathbf{x}'\in\mathcal{X}}\rho_+(\mathbf{x}')\right\}. \tag{6}$$

The estimated spatial support Ω_τ is the union of all above-threshold voxel cells, including disconnected regions. Its volume is $V_\tau = n_\tau a^3$, where $n_\tau = |\mathcal{X}_\tau|$.

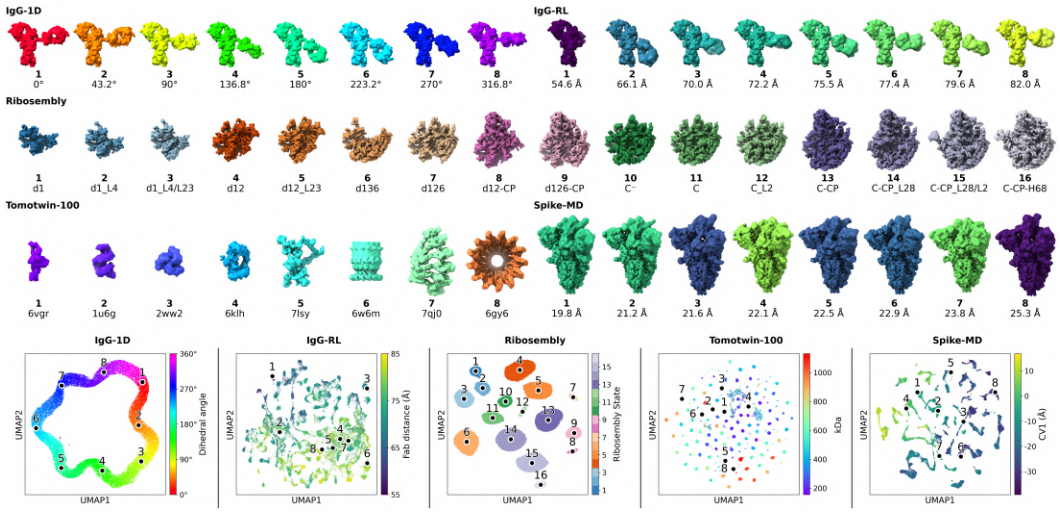

Figure 2: **Heterogeneous reconstruction on CryoBench.** Reconstructed densities (top) and UMAP projections of inferred latents (bottom) are linked by matching numbers. Colors indicate the reference IgG-1D dihedral angle, IgG-RL Fab distance, Ribosembly assembly state, Tomotwin-100 molecular mass, and Spike-MD CV1. Spike-MD density labels indicate receptor-binding domain opening. All 16 Ribosembly states are shown, with names following Qin et al. (2023).

The space–bandwidth product characterizes the approximate effective dimension of signals concentrated within a spatial region and a frequency band (Slepian, 1964; Simons & Wang, 2011). In cryo-EM, this concentration framework has been used to set the effective dimension of density representations (Katsevich et al., 2015). We use the rotationally symmetric sampling band $B_d = \{\boldsymbol{\xi} : \|\boldsymbol{\xi}\|_2 \le 1/d\}$, with frequencies in cycles per unit length, to define the spatial capacity estimate

$$\mathcal{S}_\tau = V_\tau |B_d| = \frac{4\pi}{3}\frac{V_\tau}{d^3}, \qquad G = \mathcal{R}(\mathcal{S}_\tau) = \mathcal{R}\Big(\frac{\pi}{6} n_\tau\Big), \tag{7}$$

where $\mathcal{R}$ rounds to an integer Gaussian count. This theory-motivated rule sets the budget of the nonlinear Gaussian density representation from estimated spatial support and the sampling scale used by the FWHM constraint. Appendix F derives the space–bandwidth relation and its application to Gaussian-count estimation. Section 4 gives the default τ; Section 4.4 and Figure 17 evaluate threshold sensitivity.

For homogeneous half-map reconstruction, we estimate G separately from each half-set's backprojection. Backprojection supplies only the count: Gaussian positions are initialized randomly in both settings and are not restricted to the estimated spatial support. The count is fixed throughout optimization. Gaussian relocation, when used, adjusts existing positions while preserving G, inspired by 3DGS-MCMC (Kheradmand et al., 2024) (Appendix E.3).

3.3 Heterogeneous reconstruction

Gaussian attribute decoding. Each particle image has a trainable latent $\mathbf{z}_i \in \mathbb{R}^{d_z}$, and each Gaussian has a trainable per-Gaussian latent $\mathbf{f}_g \in \mathbb{R}^{d_f}$. Separate multilayer perceptrons (MLPs) decode displacement, amplitude modulation, and scale modulation from their concatenation, $\mathbf{h}_g(\mathbf{z}) = [\mathbf{f}_g; \mathbf{z}]$. The displacement and amplitude decoders produce

$$\boldsymbol{\mu}_g(\mathbf{z}) = \boldsymbol{\mu}_g^{\text{base}} + M_\mu(\mathbf{h}_g(\mathbf{z})), \qquad p_g(\mathbf{z}) = p_g^{\text{base}}\ \text{sigmoid}(M_p(\mathbf{h}_g(\mathbf{z}))), \tag{8}$$

where $\boldsymbol{\mu}_g^{\text{base}}$ and p_g^{base} are the learned base position and amplitude. Particle-image latents parameterize variation across the dataset; per-Gaussian latents determine how each Gaussian's attributes respond to that variation. The scale decoder modulates the three principal-axis scales, giving the effective covariance

$$\begin{aligned} \boldsymbol{\Sigma}_g(\mathbf{z}) &= Q_g\, \text{diag}\big(\sigma_{g1}(\mathbf{z})^2, \sigma_{g2}(\mathbf{z})^2, \sigma_{g3}(\mathbf{z})^2\big)\, Q_g^{\mathsf{T}}, \\ 2\sqrt{2\ln 2}\,\sigma_{gk}(\mathbf{z}) &\ge d, \qquad k = 1, 2, 3, \end{aligned} \tag{9}$$

Table 1: **Per-image FSC AUC on CryoBench.** Mean (standard deviation) and median across evaluated particle images, with background masking; higher is better. Bold and underline denote best and second-best means and medians, including ties at the displayed precision. For starred methods, included values are taken directly from their publications; dashes mark comparisons not included. Evaluation and sources: Appendices D.2 and C.4.

Method	IgG-1D		IgG-RL		Ribosembly		Tomotwin-100		Spike-MD	
	Mean (std)	Median	Mean (std)	Median	Mean (std)	Median	Mean (std)	Median	Mean (std)	Median
CryoDRGN	0.351 (0.028)	0.356	0.331 (0.016)	0.333	0.412 (0.023)	0.415	0.316 (0.046)	0.321	0.340 (0.009)	0.340
CryoDRGN-AI-fixed	0.364 (0.002)	0.364	0.348 (0.012)	0.350	0.372 (0.032)	0.375	0.202 (0.044)	0.207	0.301 (0.012)	0.303
OPUS-DSD	0.335 (0.026)	0.339	0.343 (0.016)	0.346	0.362 (0.083)	0.382	0.237 (0.049)	0.251	0.229 (0.027)	0.242
3DFlex	0.335 (0.003)	0.335	0.337 (0.007)	0.337	–	–	–	–	0.304 (0.011)	0.306
3DVA	0.349 (0.004)	0.350	0.333 (0.014)	0.335	0.375 (0.038)	0.375	0.088 (0.040)	0.077	0.324 (0.010)	0.323
RECOVAR	0.386 (0.005)	0.388	0.363 (0.011)	0.363	0.429 (0.018)	**0.432**	0.258 (0.109)	0.254	0.362 (0.011)	0.365
3D Class	0.297 (0.019)	0.291	0.309 (0.010)	0.307	0.289 (0.081)	0.288	0.046 (0.026)	0.037	0.307 (0.023)	0.308
CryoSPIRE*	0.396 (0.013)	**0.400**	**0.375 (0.020)**	**0.391**	0.422 (0.015)	0.421	–	–	–	–
CryoHype*	–	–	–	–	–	–	**0.346 (0.033)**	**0.353**	–	–
SOLVAR	0.388 (0.003)	0.389	0.369 (0.009)	0.368	0.431 (0.017)	**0.432**	0.246 (0.104)	0.238	0.354 (0.012)	0.355
CryoKRAQEN*	0.375 (0.004)	0.376	0.352 (0.010)	0.352	0.418 (0.019)	0.422	0.335 (0.021)	0.337	0.328 (0.010)	0.329
CryoGaussian (Ours)	**0.397 (0.004)**	0.398	0.366 (0.020)	0.370	**0.432 (0.014)**	**0.432**	0.335 (0.035)	0.341	**0.364 (0.012)**	**0.366**

where $Q_g \in \mathrm{SO}(3)$ is the learned Gaussian rotation. Appendix E defines the Gaussian parameterization and decoders.

Heterogeneity estimation and per-state reconstruction. To estimate heterogeneity, we jointly optimize particle-image latents Z and representation parameters Θ_1 from randomly initialized Gaussians. For per-state reconstruction, we hold the inferred latents $Z^\star$ fixed and fit an independently initialized representation Θ_2 to the same images. Both optimizations minimize the projection error in Equation equation 4 under the data-derived constraints:

$$\underset{\Theta_1, Z}{\text{minimize}}\ \mathcal{L}(\Theta_1, Z), \qquad \text{followed by} \qquad \underset{\Theta_2}{\text{minimize}}\ \mathcal{L}(\Theta_2, Z^\star). \tag{10}$$

Per-state reconstruction. For a selected latent $\mathbf{z}$, the final decoders specify the positions, amplitudes, and scales of the Gaussians shared across states, defining $\rho_{\Theta_2}(\mathbf{x}; \mathbf{z})$. We evaluate this density on a three-dimensional grid without additional optimization.

3.4 Efficient cryo-EM Gaussian splatting

Additive cryo-EM projection accumulates overlapping Gaussian contributions without standard 3DGS's opacity-based early termination. Our differentiable CUDA rasterizer uses per-Gaussian gradient accumulation to reduce atomic contention from concurrent pixel-wise updates. Projection scheduling balances the workload, while distributed training parallelizes optimization. These optimizations preserve the image-formation model. Appendix G describes the partitioning of Gaussian lists for image tiles and corresponding gradient reductions.

4 Experiments

We evaluate CryoGaussian on CryoBench's five synthetic datasets: IgG-1D, IgG-RL, Ribosembly, Tomotwin-100, and Spike-MD (Jeon et al., 2024). Homogeneous reconstruction is evaluated on the RAG1–RAG2 complex (EMPIAR-10049) (Ru et al., 2015), SARS-CoV-2 spike on intact virions (EMPIAR-10492) (Ke et al., 2020), and the 80S ribosome (EMPIAR-10028) (Wong et al., 2014); heterogeneous reconstruction, on the assembling 50S ribosomal subunit (EMPIAR-10076) (Davis et al., 2016) and the pre-catalytic spliceosome (EMPIAR-10180) (Plaschka et al., 2017).

We set the base learning rates to 2×10^{-4} for Gaussian positions; 10^{-3} for Gaussian scales, rotations, base-amplitude parameters, and trainable particle-image latents; and 10^{-4} for per-Gaussian latents and the displacement, amplitude, and scale MLPs. These rates are defined for a global batch size of one and scaled by the square root of the global batch size, consistent with the learning-rate scaling in Grendel and gsplat (Zhao et al., 2025; Ye et al., 2025). The Gaussian position learning

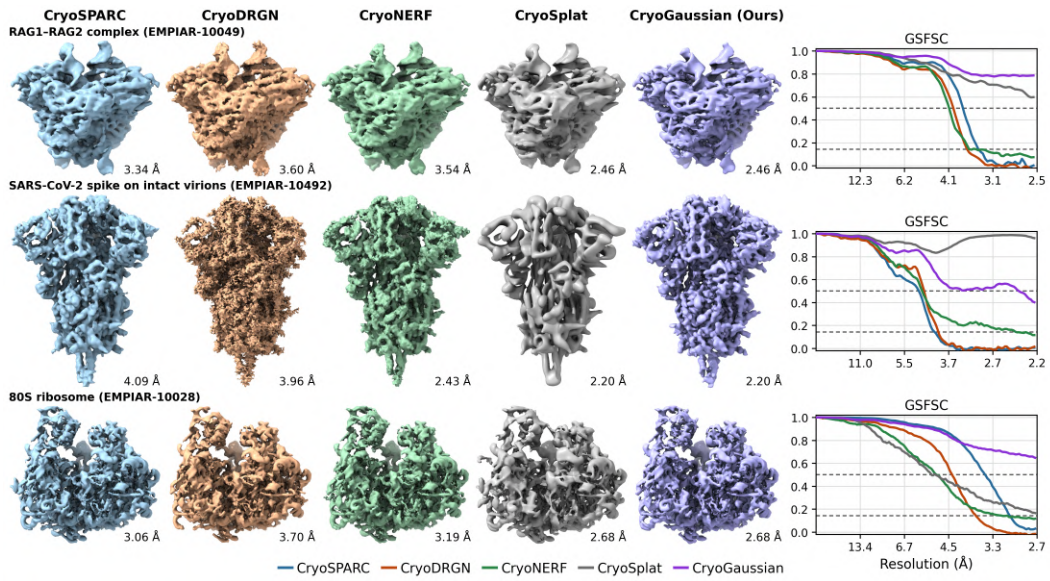

Figure 3: **Homogeneous reconstruction on experimental data.** Reconstructed densities and gold-standard FSC for RAG1–RAG2 (EMPIAR-10049, top), SARS-CoV-2 spike on intact virions (EMPIAR-10492, middle), and the 80S ribosome (EMPIAR-10028, bottom). Dashed lines mark FSC values of 0.5 and 0.143. Resolutions use the 0.143 crossing, or the Nyquist limit if FSC remains above 0.143 throughout the sampled frequency range.

rate decays to 1% of its initial value, following 3DGS and X-Gaussian (Kerbl et al., 2023; Cai et al., 2025). We use Adam (Kingma & Ba, 2015) with momentum coefficients $(0.9, 0.999)$, $\epsilon = 10^{-15}$, and zero global weight decay. Homogeneous reconstruction uses 20,000 iterations; heterogeneity estimation and per-state reconstruction each use 30,000 iterations. The latent dimensions are $d_z = 4$ and $d_f = 24$; MLP configurations are given in Appendix E. Gaussian-count estimation uses $\tau = 0.02$, with the rationale discussed in Appendix F.3. Inspired by 3DGS-MCMC (Kheradmand et al., 2024), we relocate Gaussians every 100 iterations starting at iteration 500, stopping at iteration 16,700 for homogeneous reconstruction and 25,000 for heterogeneous reconstruction. Dataset-specific settings are provided in Appendix C.2. Data preparation, evaluation protocols, and baseline configurations are detailed in Appendices C.1, D, and C.4, respectively. Computational costs are reported in Appendix C.3.

4.1 CryoBench evaluation

IgG-1D, IgG-RL, and Spike-MD assess conformational heterogeneity through one-dimensional Fab motion, randomized linker conformations, and a molecular-dynamics ensemble, respectively. Ribosembly and Tomotwin-100 assess compositional heterogeneity through 16 unequally populated assembly states and a mixture of 100 distinct complexes, respectively. Across both forms of heterogeneity, CryoGaussian achieves competitive per-image and per-conformation FSC AUC (Tables 1 and 2), following the CryoBench evaluation protocols described in Appendix D.2. Figure 2 summarizes reconstructed densities and inferred latents. Detailed reconstruction–reference comparisons and FSC curves are provided in Figures 8–12, while Figure 13 shows representative densities selected by latent clustering. Additional IgG-1D reconstructions at SNR 0.005 and 0.001 are shown in Figures 24 and 25. Table 3 reports comprehensive masked and unmasked statistics. Mean FSC AUC differs little with masking across the evaluated conditions, except for IgG-1D at SNR 0.001, where masking increases the mean by approximately 0.010 under both evaluation protocols.

4.2 Homogeneous reconstruction on experimental data

We evaluate the RAG1–RAG2 complex (EMPIAR-10049), SARS-CoV-2 spike on intact virions (EMPIAR-10492), and the emetine-bound *Plasmodium falciparum* 80S ribosome (EMPIAR-

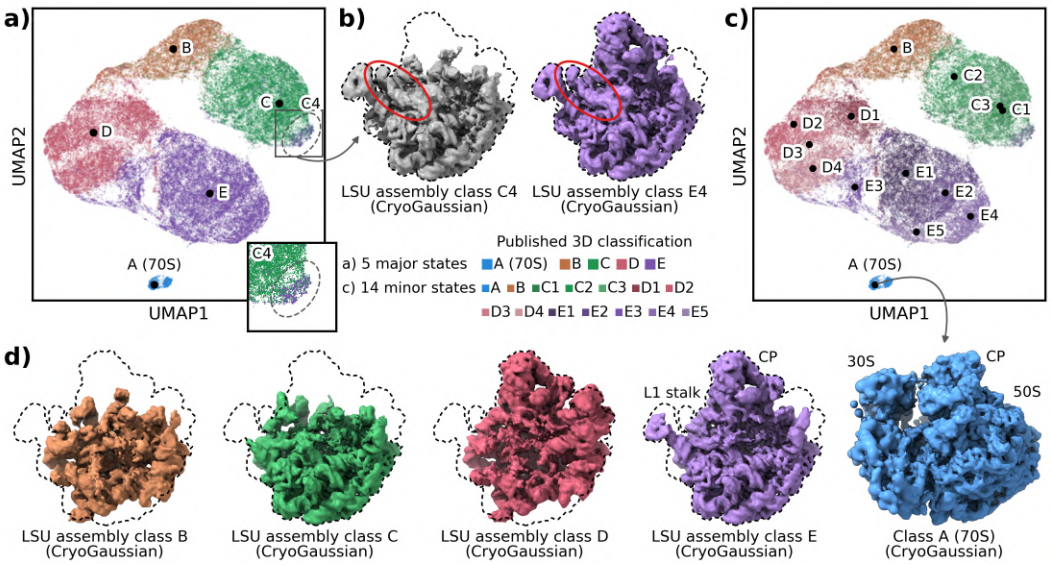

Figure 4: **Compositional heterogeneity in the assembling 50S ribosomal subunit (EMPIAR-10076). (a, c)** UMAP projections colored by published major and minor class assignments (Davis et al., 2016); black points mark class centers. The inset locates the previously reported C4 selection (Zhong et al., 2021). **(b)** C4 and E4 densities. **(d)** Major assembly classes B–E and the 70S class A. Dashed outlines indicate the mature 50S envelope; red ellipses mark rRNA helix 68 (H68).

10028). Figure 3 shows reconstructed densities and gold-standard FSC (GSFSC), using CryoSPARC's auto-tight corrected curves computed from independently reconstructed half-maps in the Validation (FSC) job with mask auto-tightening and noise-substitution correction (Appendix D.3). CryoSplat and CryoGaussian yield the highest GSFSC values among the compared methods. Since half-map agreement alone does not establish reconstruction accuracy (Chen & Ling, 2026), we additionally reconstruct the three complexes from particle images simulated from PDB-derived densities, with CTF modulation and white noise at SNR 0.05. Map-to-map FSC against these reference densities reaches 0.5 at 2.98, 2.54, and 3.29 Å, respectively (Figure 7). These comparisons provide complementary evidence of agreement with the reference densities under the simulated imaging conditions. Simulation settings are provided in Appendix C.1. Additional viewing directions are shown in Figure 6. Figure 26 additionally reports CryoSPARC GSFSC curves for the same CryoGaussian half-map pairs without a mask and with spherical, loose, and tight masks.

4.3 Heterogeneous reconstruction on experimental data

Compositional heterogeneity (EMPIAR-10076). We analyze the assembling 50S ribosomal subunit from an L17-depleted sample using the CryoDRGN-filtered particle subset. The inferred latents distinguish five major classes and capture finer variation among 14 minor classes, including 70S class A (Figure 4). CryoGaussian recapitulates an assembly intermediate with resolved rRNA helix 68 (H68) density but no resolved central protuberance, first identified with CryoDRGN (Zhong et al., 2021). This is consistent with the reported docking of H68 without an assembled central protuberance. Map-to-map FSC assesses agreement with deposited assembly-state reconstructions (Figure 14). Homogeneous CryoSPARC refinement of particles identified by the released CryoDRGN indices for this intermediate recovers the same features (Figure 15(a)), supporting the presence of this intermediate in the selected population. Additional assembly-state comparisons are presented in Appendix A.3. Together, these results support CryoGaussian's potential to characterize compositional heterogeneity in experimental data.

Conformational heterogeneity (EMPIAR-10180). We analyze the *Saccharomyces cerevisiae* precatalytic spliceosome using the CryoDRGN-filtered particle subset. Reconstructions along the first two principal components of the inferred latents show progressive rearrangements of the SF3b and helicase regions relative to the body and foot (Figure 5). These changes are consistent with

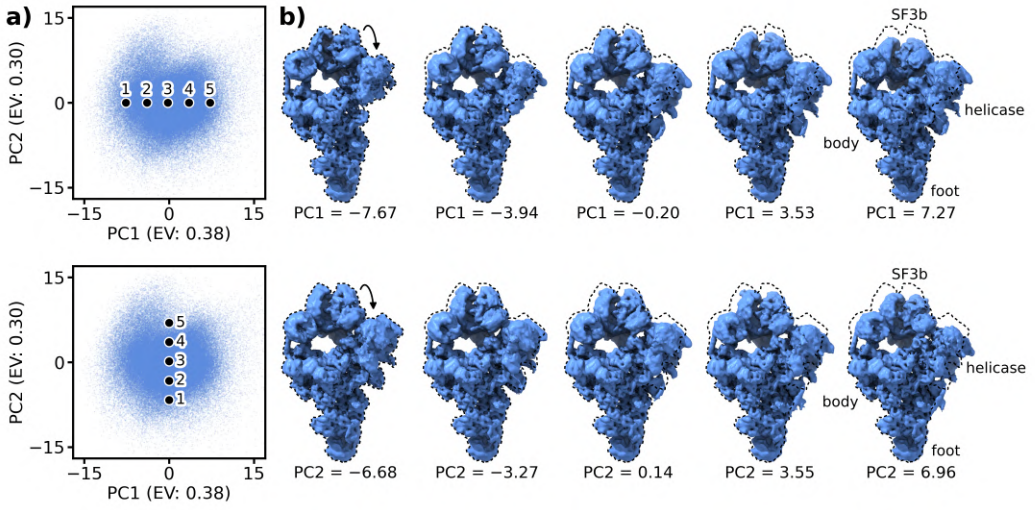

Figure 5: **Conformational variation in the pre-catalytic spliceosome (EMPIAR-10180). (a)** Principal-component projections of inferred latents, with five numbered positions along PC1 (top) and PC2 (bottom). EV denotes the fraction of latent variance explained by each component. **(b)** Densities reconstructed at positions 1–5, ordered from left to right in each row. Dashed outlines reproduce each row's first density; arrows indicate the displayed direction of conformational change.

the reported mobility of the SF3b-containing U2 region and the helicase within the pre-catalytic complex (Plaschka et al., 2017; Zhong et al., 2021). A separate full-stack analysis distinguishes particle populations lacking resolved SF3b density or showing additional density near the U2 core (Figure 16), recapitulating structural features reported by CryoDRGN (Zhong et al., 2021). Representative-latent selection and density comparison are detailed in Appendix D.4. Together, these results support CryoGaussian's potential to identify structural subpopulations without supervision and characterize conformational heterogeneity in experimental data.

4.4 Ablations and robustness

We examine sensitivity to Gaussian count and the relative density threshold used for Gaussian-count estimation, ablate the Gaussian FWHM constraint, and compare random Gaussian initialization with consensus initialization for heterogeneity estimation (Appendices B.1, B.2, and B.3). We further test robustness to particle selection by recovering the previously reported 50S assembly intermediate from particles filtered with CryoGaussian; homogeneous CryoSPARC refinement of particles retained from the published state selection corroborates its H68 density and lack of a resolved central protuberance (Figure 15(b); Appendix B.4). We also assess robustness to image noise and sensitivity to evaluation masks, observing competitive reconstruction performance across the tested noise levels and generally limited masking sensitivity (Appendices B.5 and B.6).

5 Conclusion

We presented CryoGaussian, a cryo-EM-native Gaussian splatting method that unifies homogeneous and heterogeneous reconstruction. By fitting explicit Gaussians directly to particle images, CryoGaussian estimates compositional and conformational heterogeneity without supervision and reconstructs the corresponding densities. Across all five CryoBench datasets, it achieves competitive performance in heterogeneity estimation and per-state reconstruction. On experimental data, it resolves structural detail and recovers previously validated features of 50S ribosome assembly and continuous conformational variation in the pre-catalytic spliceosome. These results highlight CryoGaussian's potential to support structural discovery by connecting inferred heterogeneity with biologically interpretable three-dimensional reconstructions.

AI USE STATEMENT

In accordance with the conference guidelines, we acknowledge the use of Large Language Models (LLMs) during the preparation of this manuscript. We utilized LLMs for paper writing assistance, specifically for language polishing and improving the clarity and readability of our work. The LLMs assisted in refining linguistic expression, ensuring proper grammar and academic writing style, and enhancing the overall flow of technical content.

All core research contributions, including the novel methodology, experimental design, theoretical analysis, and scientific insights presented in this work, were developed independently by the authors. The LLMs were used solely as writing assistance tools (i.e., for polishing writing) and did not contribute to the conceptual development, experimental validation, or interpretation of results.

ETHICS STATEMENT

Our work aims to improve homogeneous and heterogeneous cryo-EM reconstruction without introducing any additional ethical concerns or resolving existing ones.

REPRODUCIBILITY STATEMENT

We have made comprehensive efforts to ensure the reproducibility of our work. Our complete methodology is detailed in Section 3, with the Gaussian-count estimation, Gaussian parameterization and decoder architecture, and renderer gradient derivations provided in Appendices F, E and G, respectively. Experiments are described in Section 4, with training settings in Appendix C.2. All datasets, including EMPIAR and CryoBench (Jeon et al., 2024), are publicly available, and their preprocessing is described in Appendix C.1. Evaluation protocols and baseline configurations are provided in Appendices D and C.4. Sample EMPIAR-10076 density maps are available at `https://zenodo.org/records/22976183`, including the previously reported 50S assembly intermediate with rRNA helix 68 density but no resolved central protuberance, recovered using particles independently filtered with CryoGaussian. All source code will be made publicly available upon publication to facilitate reproducibility and future research.

REFERENCES

Daniil A. Boiko, Robert MacKnight, Ben Kline, and Gabe Gomes. Autonomous chemical research with large language models. *Nature*, 624(7992):570–578, 2023. doi: 10.1038/s41586-023-06792-0. URL https://doi.org/10.1038/s41586-023-06792-0.

Yuanhao Cai, Yixun Liang, Jiahao Wang, Angtian Wang, Yulun Zhang, Xiaokang Yang, Zongwei Zhou, and Alan Yuille. Radiative gaussian splatting for efficient X-Ray novel view synthesis. In *Computer Vision – ECCV 2024*, volume 15059 of *Lecture Notes in Computer Science*, pp. 283–299. Springer, 2025. doi: 10.1007/978-3-031-73232-4_16.

Muyuan Chen and Steven J. Ludtke. Deep learning-based mixed-dimensional Gaussian mixture model for characterizing variability in cryo-EM. *Nature Methods*, 18(8):930–936, 2021. doi: 10.1038/s41592-021-01220-5.

Muyuan Chen, Bogdan Toader, and Roy Lederman. Integrating molecular models into CryoEM heterogeneity analysis using scalable high-resolution deep Gaussian mixture models. *Journal of Molecular Biology*, 435(9):168014, 2023. doi: 10.1016/j.jmb.2023.168014.

Muyuan Chen, Michael F. Schmid, and Wah Chiu. Improving resolution and resolvability of single-particle cryoEM structures using Gaussian mixture models. *Nature Methods*, 21(1):37–40, 2024. doi: 10.1038/s41592-023-02082-9.

Suyi Chen and Haibin Ling. CryoSplat: Gaussian splatting for Cryo-EM homogeneous reconstruction. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/5f173562e7662b14fb5c5695f225ea46-Abstract-Conference.html.

Yifan Cheng, Nikolaus Grigorieff, Pawel A. Penczek, and Thomas Walz. A primer to single-particle cryo-electron microscopy. *Cell*, 161(3):438–449, 2015. doi: 10.1016/j.cell.2015.03.050.

Joseph H. Davis, Yong Zi Tan, Bridget Carragher, Clinton S. Potter, Dmitry Lyumkis, and James R. Williamson. Modular assembly of the bacterial large ribosomal subunit. *Cell*, 167(6):1610–1622.e15, 2016. doi: 10.1016/j.cell.2016.11.020.

Yuanchen Dong, Shuwen Zhang, Zhaolong Wu, Xuemei Li, Wei Li Wang, Yanan Zhu, Svetla Stoilova-McPhie, Ying Lu, Daniel Finley, and Youdong Mao. Cryo-EM structures and dynamics of substrate-engaged human 26S proteasome. *Nature*, 565(7737):49–55, 2019. doi: 10.1038/s41586-018-0736-4.

P. A. Doyle and P. S. Turner. Relativistic Hartree–Fock X-ray and electron scattering factors. *Acta Crystallographica Section A*, 24(3):390–397, 1968. doi: 10.1107/S0567739468000756.

Wenyuan Gao, Yutan Wu, and Xuming He. CryoKRAQEN: Kernel-regularized annealing for quantized embedding networks in Cryo-EM heterogeneous reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 28298–28307, 2026.

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.

Jeffrey Gu, Minkyu Jeon, Ambri Ma, Serena Yeung-Levy, and Ellen D. Zhong. CryoHype: Reconstructing a thousand cryo-EM structures with transformer-based hypernetworks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 35280–35290, 2026. URL https://openaccess.thecvf.com/content/CVPR2026/html/Gu_CryoHype_Reconstructing_a_thousand_cryo-EM_structures_with_transformer-based_hypernetworks_CVPR_2026_paper.html.

David Herreros, Carlos Perez Mata, Chari Noddings, Deli Irene, James Krieger, David A. Agard, Ming-Daw Tsai, Carlos Oscar Sanchez Sorzano, and Jose Maria Carazo. Real-space heterogeneous reconstruction, refinement, and disentanglement of CryoEM conformational states with HetSIREN. *Nature Communications*, 16(1):3751, 2025. doi: 10.1038/s41467-025-59135-0.

Yue Huang, Chengguang Zhu, Xiaokang Yang, and Manhua Liu. High-resolution real-space reconstruction of cryo-EM structures using a neural field network. *Nature Machine Intelligence*, 6(8): 892–903, 2024. doi: 10.1038/s42256-024-00870-2.

Rob J. Hyndman. Computing and graphing highest density regions. *The American Statistician*, 50 (2):120–126, 1996. doi: 10.1080/00031305.1996.10474359.

Kiarash Jamali, Lukas Käll, Rui Zhang, Alan Brown, Dari Kimanius, and Sjors H. W. Scheres. Automated model building and protein identification in cryo-EM maps. *Nature*, 628(8007): 450–457, 2024. doi: 10.1038/s41586-024-07215-4. URL https://doi.org/10.1038/s41586-024-07215-4.

Minkyu Jeon, Rishwanth Raghu, Miro Astore, Geoffrey Woollard, Ryan Feathers, Alkin Kaz, Sonya M. Hanson, Pilar Cossio, and Ellen D. Zhong. CryoBench: Diverse and challenging datasets for the heterogeneity problem in cryo-EM. In *Advances in Neural Information Processing Systems*, volume 37, 2024. doi: 10.52202/079017-2840.

Eugene Katsevich, Alexander Katsevich, and Amit Singer. Covariance matrix estimation for the Cryo-EM heterogeneity problem. *SIAM Journal on Imaging Sciences*, 8(1):126–185, 2015. doi: 10.1137/130935434.

Takeshi Kawabata. Multiple subunit fitting into a low-resolution density map of a macromolecular complex using a Gaussian mixture model. *Biophysical Journal*, 95(10):4643–4658, 2008. doi: 10.1529/biophysj.108.137125.

Takeshi Kawabata. Gaussian-input Gaussian mixture model for representing density maps and atomic models. *Journal of structural biology*, 203(1):1–16, 2018.

Zunlong Ke, Joaquin Oton, Kun Qu, Mirko Cortese, Vojtech Zila, Lesley McKeane, Takanori Nakane, Jasenko Zivanov, Christopher J. Neufeldt, Berati Cerikan, John M. Lu, Julia Peukes, Xiaoli Xiong, Hans-Georg Kräusslich, Sjors H. W. Scheres, Ralf Bartenschlager, and John A. G. Briggs. Structures and distributions of SARS-CoV-2 spike proteins on intact virions. *Nature*, 588 (7838):498–502, 2020. doi: 10.1038/s41586-020-2665-2.

Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3D gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics*, 42(4), 2023. doi: 10.1145/3592433.

Shakiba Kheradmand, Daniel Rebain, Gopal Sharma, Weiwei Sun, Yang-Che Tseng, Hossam Isack, Abhishek Kar, Andrea Tagliasacchi, and Kwang Moo Yi. 3D Gaussian Splatting as Markov Chain Monte Carlo. In *Advances in Neural Information Processing Systems*, volume 37, 2024. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/93be245fce00a9bb2333c17ceae4b732-Abstract-Conference.html.

Dari Kimanius, Kiarash Jamali, Max E. Wilkinson, Sofia Lövestam, Vaithish Velazhahan, Takanori Nakane, and Sjors H. W. Scheres. Data-driven regularization lowers the size barrier of cryo-EM structure determination. *Nature Methods*, 21(7):1216–1221, 2024. doi: 10.1038/s41592-024-02304-8.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations*, 2015.

Axel Levy, Rishwanth Raghu, David Shustin, Adele Peng, Huan Li, Oliver Clarke, Gordon Wetzstein, and Ellen Zhong. Mixture of neural fields for heterogeneous reconstruction in cryo-em. *Advances in Neural Information Processing Systems*, 37:56988–57017, 2024.

Axel Levy, Rishwanth Raghu, J. Ryan Feathers, Michal Grzadkowski, Frédéric Poitevin, Jake D. Johnston, Francesca Vallese, Oliver Biggs Clarke, Gordon Wetzstein, and Ellen D. Zhong. CryoDRGN-AI: neural ab initio reconstruction of challenging cryo-EM and cryo-ET datasets. *Nature Methods*, 22(7):1486–1494, 2025. doi: 10.1038/s41592-025-02720-4.

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.

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.

Takanori Nakane, Dari Kimanius, Erik Lindahl, and Sjors H. W. Scheres. Characterisation of molecular motions in cryo-EM single-particle data by multi-body refinement in RELION. *eLife*, 7: e36861, 2018. doi: 10.7554/eLife.36861.

NIST/SEMATECH. Critical values of the chi-square distribution. e-Handbook of Statistical Methods, Section 1.3.6.7.4, n.d. URL https://www.itl.nist.gov/div898/handbook/eda/section3/eda3674.htm. Accessed September 2026.

L. M. Peng, G. Ren, S. L. Dudarev, and M. J. Whelan. Robust parameterization of elastic and absorptive electron atomic scattering factors. *Acta Crystallographica Section A: Foundations of Crystallography*, 52(2):257–276, 1996. doi: 10.1107/S0108767395014371.

Xingang Peng, Ruihan Guo, Fenglin Guo, Ziyi Wang, Jiayu Sun, Jiaqi Guan, Yinjun Jia, Yan Xu, Yanwen Huang, Muhan Zhang, Jian Peng, Xinquan Wang, Chuanhui Han, Zihua Wang, and Jianzhu Ma. Unified modeling of 3D molecular generation via atomic interactions with PocketXMol. *Cell*, 189(7):1904–1922.e28, 2026. doi: 10.1016/j.cell.2026.01.003. URL https://doi.org/10.1016/j.cell.2026.01.003.

Clemens Plaschka, Pei-Chun Lin, and Kiyoshi Nagai. Structure of a pre-catalytic spliceosome. *Nature*, 546(7660):617–621, 2017. doi: 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. doi: 10.1016/j.jsb.2021.107702.

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.

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.

Ali Punjani, Haowei Zhang, and David J. Fleet. Non-uniform refinement: adaptive regularization improves single-particle cryo-EM reconstruction. *Nature Methods*, 17:1214–1221, 2020. doi: 10.1038/s41592-020-00990-8.

Bo Qin, Simon M. Lauer, Annika Balke, Carlos H. Vieira-Vieira, Jörg Bürger, Thorsten Mielke, Matthias Selbach, Patrick Scheerer, Christian M. T. Spahn, and Rainer Nikolay. Cryo-EM captures early ribosome assembly in action. *Nature Communications*, 14(1):898, 2023. doi: 10.1038/s41467-023-36607-9.

Huaizhi Qu, Xiao Wang, Yuanyuan Zhang, Sheng Wang, William Stafford Noble, and Tianlong Chen. CryoNeRF: reconstruction of homogeneous and heterogeneous cryo-EM structures using neural radiance field. bioRxiv preprint, 2025.

Jessica N. Rabuck-Gibbons, Dmitry Lyumkis, and James R. Williamson. Quantitative mining of compositional heterogeneity in cryo-EM datasets of ribosome assembly intermediates. *Structure*, 30(4):498–509.e4, 2022. doi: 10.1016/j.str.2021.12.005.

Heng Ru, Melissa G. Chambers, Tian-Min Fu, Alexander B. Tong, Maofu Liao, and Hao Wu. Molecular mechanism of V(D)J recombination from synaptic RAG1-RAG2 complex structures. *Cell*, 163(5):1138–1152, 2015. doi: 10.1016/j.cell.2015.10.055.

Sjors H. W. Scheres. A bayesian view on Cryo-EM structure determination. *Journal of Molecular Biology*, 415(2):406–418, 2012a. doi: 10.1016/j.jmb.2011.11.010.

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.

Sjors H. W. Scheres, Haixiao Gao, Mikel Valle, Gabor T. Herman, Paul P. B. Eggermont, Joachim Frank, and Jose-Maria Carazo. Disentangling conformational states of macromolecules in 3D-EM through likelihood optimization. *Nature Methods*, 4(1):27–29, 2007. doi: 10.1038/nmeth992.

Sjors HW Scheres. RELION: implementation of a bayesian approach to cryo-em structure determination. *Journal of structural biology*, 180(3):519–530, 2012b.

Johannes Schwab, Dari Kimanius, Alister Burt, Tom Dendooven, and Sjors H. W. Scheres. DynaMight: estimating molecular motions with improved reconstruction from cryo-EM images. *Nature Methods*, 21:1855–1862, 2024. doi: 10.1038/s41592-024-02377-5.

Claude E. Shannon. Communication in the presence of noise. *Proceedings of the IRE*, 37(1):10–21, 1949. doi: 10.1109/JRPROC.1949.232969.

Shayan Shekarforoush, David B. Lindell, Marcus A. Brubaker, and David J. Fleet. Reconstructing heterogeneous biomolecules via hierarchical gaussian mixtures and part discovery. In *Advances in Neural Information Processing Systems*, volume 38, 2025. URL `https://openreview.net/forum?id=CCbXvvcdF9`.

Frederik J. Simons and Dong V. Wang. Spatiospectral concentration in the Cartesian plane. *GEM - International Journal on Geomathematics*, 2(1):1–36, 2011. doi: 10.1007/s13137-011-0016-z.

Amit Singer and Fred J. Sigworth. Computational methods for single-particle electron cryomicroscopy. *Annual Review of Biomedical Data Science*, 3:163–190, 2020. doi: 10.1146/annurev-biodatasci-021020-093826.

Kashish Singh, Georg Bunzel, Benjamin Graf, Ka Man Yip, Meina Neumann-Schaal, Holger Stark, and Ashwin Chari. Reconstruction of a fatty acid synthesis cycle from acyl carrier protein and cofactor structural snapshots. *Cell*, 186(23):5054–5067.e16, 2023. doi: 10.1016/j.cell.2023.10.009.

David Slepian. Prolate spheroidal wave functions, fourier analysis and uncertainty—IV: Extensions to many dimensions; generalized prolate spheroidal functions. *Bell System Technical Journal*, 43 (6):3009–3057, 1964. doi: 10.1002/j.1538-7305.1964.tb01037.x.

Chinmay Talegaonkar, Yash Belhe, Ravi Ramamoorthi, and Nicholas Antipa. Volumetrically consistent 3D Gaussian rasterization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10953–10963, 2025. doi: 10.1109/CVPR52734.2025.01023.

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.

Ziwen Wang, Jiajun Fan, Ruihan Guo, Thao Nguyen, Heng Ji, and Ge Liu. ProteinZero: Self-improving protein generation via online reinforcement learning. *Transactions on Machine Learning Research*, 2026. ISSN 2835-8856. URL https://openreview.net/forum?id=pVvm0CQAr6.

Lee Westover. Footprint evaluation for volume rendering. In *Proceedings of the 17th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '90, pp. 367–376. Association for Computing Machinery, 1990. doi: 10.1145/97879.97919.

Miłosz Wieczór, Phu K. Tang, Modesto Orozco, and Pilar Cossio. Omicron mutations increase interdomain interactions and reduce epitope exposure in the SARS-CoV-2 spike. *iScience*, 26(2): 105981, 2023. doi: 10.1016/j.isci.2023.105981.

Wilson Wong, Xiao-chen Bai, Alan Brown, Israel S. Fernandez, Eric Hanssen, Melanie Condron, Yan Hong Tan, Jake Baum, and Sjors H. W. Scheres. Cryo-EM structure of the Plasmodium falciparum 80S ribosome bound to the anti-protozoan drug emetine. *eLife*, 3:e03080, 2014. doi: 10.7554/eLife.03080.

Roey Yadgar, Roy R. Lederman, and Yoel Shkolnisky. SOLVAR: Fast covariance-based heterogeneity analysis with pose refinement for cryo-EM. *arXiv preprint arXiv:2602.17603*, 2026.

Vickie Ye, Ruilong Li, Justin Kerr, Matias Turkulainen, Brent Yi, Zhuoyang Pan, Otto Seiskari, Jianbo Ye, Jeffrey Hu, Matthew Tancik, and Angjoo Kanazawa. gsplat: An open-source library for Gaussian Splatting. *Journal of Machine Learning Research*, 26(34):1–17, 2025. URL https://jmlr.org/papers/v26/24-1476.html.

Ka Man Yip, Niels Fischer, Elham Paknia, Ashwin Chari, and Holger Stark. Atomic-resolution protein structure determination by cryo-EM. *Nature*, 587(7832):157–161, 2020. doi: 10.1038/s41586-020-2833-4. URL https://doi.org/10.1038/s41586-020-2833-4.

Zehao Yu, Anpei Chen, Binbin Huang, Torsten Sattler, and Andreas Geiger. Mip-Splatting: Alias-free 3D Gaussian Splatting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 19447–19456, June 2024. doi: 10.1109/CVPR52733.2024.01839.

Ruyi Zha, Tao Jun Lin, Yuanhao Cai, Jiwen Cao, Yanhao Zhang, and Hongdong Li. R^2-Gaussian: Rectifying radiative gaussian splatting for tomographic reconstruction. In *Advances in Neural Information Processing Systems*, volume 37, 2024. doi: 10.52202/079017-1427.

Hexu Zhao, Haoyang Weng, Daohan Lu, Ang Li, Jinyang Li, Aurojit Panda, and Saining Xie. On scaling up 3D Gaussian Splatting training. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=pQqeQpMkE7.

Wan Zheng, Pengxin Chai, Jiapeng Zhu, and Kai Zhang. High-resolution in situ structures of mammalian respiratory supercomplexes. *Nature*, 631(8019):232–239, 2024. doi: 10.1038/s41586-024-07488-9.

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, 2021. doi: 10.1038/s41592-020-01049-4.

Yi Zhou, Yilai Li, Jing Yuan, and Quanquan Gu. CryoFM: A flow-based foundation model for Cryo-EM densities. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/file/6450ea28ebbc8437bc38775157818172-Paper-Conference.pdf.

Matthias Zwicker, Hanspeter Pfister, Jeroen van Baar, and Markus Gross. EWA volume splatting. In *Proceedings of IEEE Visualization 2001*, pp. 29–36. IEEE, 2001. doi: 10.1109/VISUAL.2001.964490.

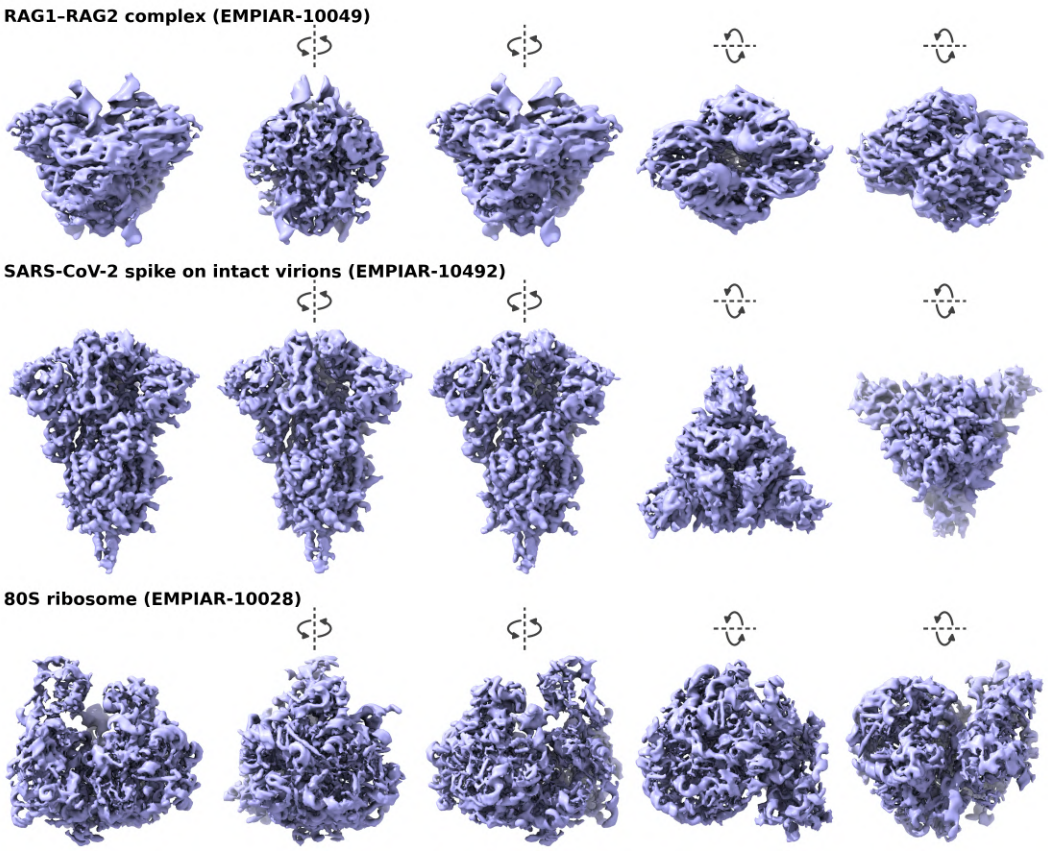

Figure 6: **CryoGaussian homogeneous reconstructions from different viewing directions.** Reconstruction results are detailed in Section 4.2.

A Additional reconstruction results

A.1 Homogeneous reconstruction

The experimental comparison in Figure 3 evaluates agreement between separately reconstructed half-maps. Additional viewing directions are shown in Figure 6. To complement this evaluation with known reference densities, we reconstruct simulated particle images for the same three biomolecular systems (Figure 7). The reference densities are generated from atomic models of the RAG1–RAG2 complex (PDB 3JBX), SARS-CoV-2 spike (PDB 6ZWV), and the 80S ribosome (PDB 3J79 and 3J7A).

Map-to-map FSC against the simulation input densities reaches 0.5 at 2.98, 2.54, and 3.29 Å, respectively. These comparisons assess reconstruction against the density used to simulate the particle images, complementing the half-map consistency measured on experimental data. Figure 26 provides the experimental half-map FSC curves under different masks.

A.2 CryoBench reconstruction

Table 2 complements the per-image comparison by evaluating one representative latent for each reference state. For CryoGaussian, mean per-conformation FSC AUC is 0.402 on IgG-1D, 0.387 on IgG-RL, 0.432 on Ribosembly, and 0.336 on Tomotwin-100. For Spike-MD, the table repeats the per-image result for reference, as specified in Appendix D.2. The standard deviations describe variation across the evaluated reference comparisons within a run.

Figures 8–12 pair per-image reconstructed densities with their corresponding reference densities and show masked and unmasked FSC curves. The paired displays relate per-state reconstruction to the

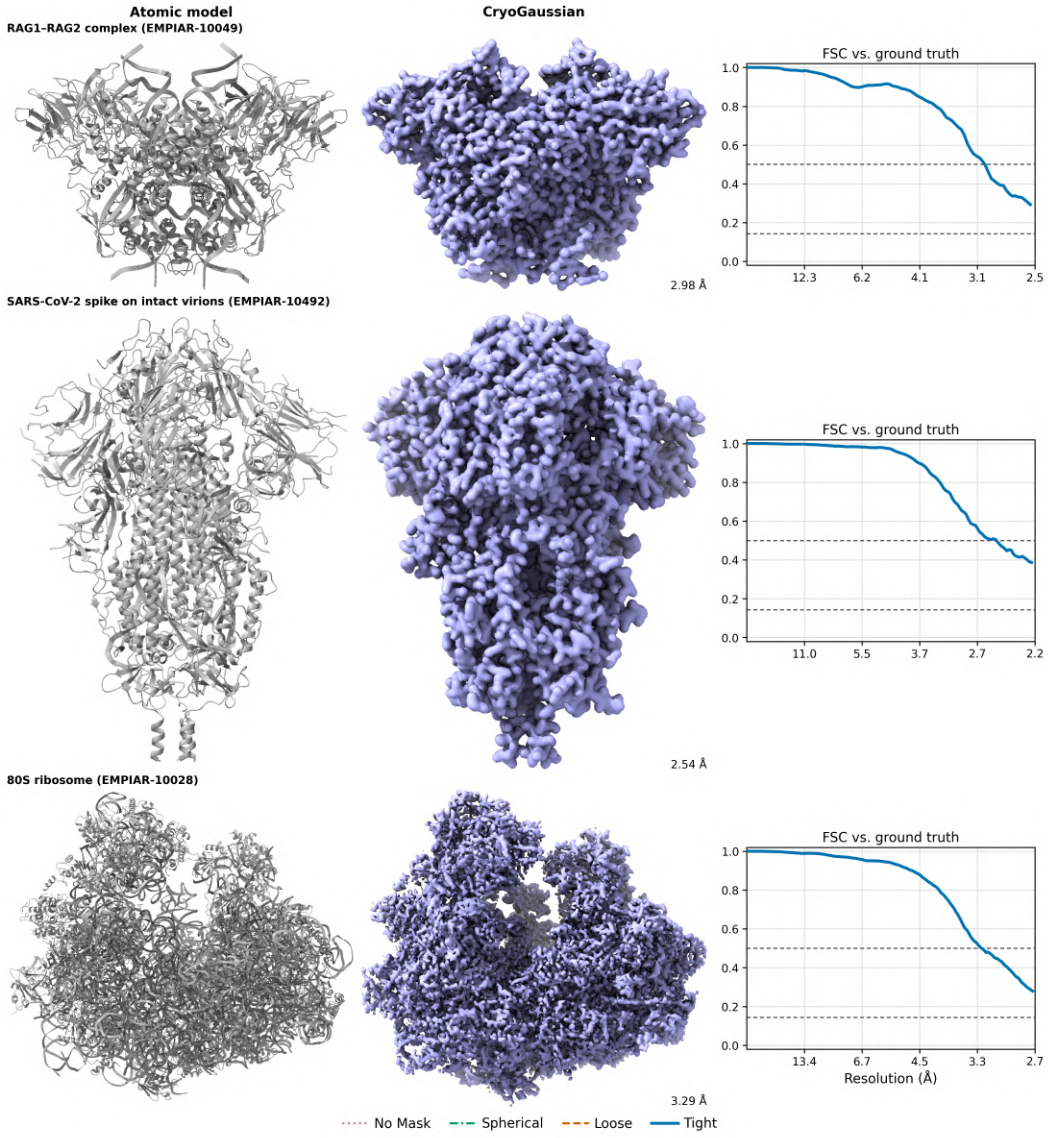

Figure 7: **CryoGaussian homogeneous reconstructions from simulated particle images.** Rows show the RAG1–RAG2 complex (PDB 3JBX), SARS-CoV-2 spike (PDB 6ZWV), and the 80S ribosome (PDB 3J79 and 3J7A), from top to bottom. Columns show the atomic models used to generate the reference densities (left), reconstructed densities (center), and map-to-map FSC against the reference densities (right). Particle images were simulated in CryoSPARC with CTF modulation and white noise at SNR 0.05. FSC curves computed without a mask and with spherical, loose, and tight masks overlap. Horizontal dashed lines indicate FSC values of 0.5 and 0.143. Resolutions at FSC = 0.5 are 2.98, 2.54, and 3.29 Å, respectively; all curves remain above 0.143 up to the Nyquist frequency.

Table 2: **Per-conformation FSC AUC on CryoBench.** Mean (standard deviation) and median across evaluated reference states, with background masking; higher is better. Bold and underline denote best and second-best means and medians, including ties at the displayed precision. Baseline coverage follows publicly documented per-conformation evaluations. For Spike-MD, † repeats the per-image result for reference and is excluded from per-conformation rankings. For starred methods, included values are taken directly from their publications. Dashes indicate results not included for the corresponding dataset and protocol. Evaluation and sources: Appendices D.2 and C.4.

Method	IgG-1D		IgG-RL		Ribosembly		Tomotwin-100		Spike-MD	
	Mean (std)	Median	Mean (std)	Median	Mean (std)	Median	Mean (std)	Median	Mean (std)	Median
CryoDRGN	0.366 (0.003)	0.366	0.349 (0.008)	0.348	0.415 (0.019)	0.415	0.321 (0.034)	0.322	0.340 (0.009)	0.340
CryoDRGN-AI-fixed	0.366 (0.001)	0.366	0.355 (0.007)	0.354	0.372 (0.032)	0.374	0.206 (0.041)	0.212	0.301 (0.012)	0.304
OPUS-DSD	0.340 (0.006)	0.341	0.346 (0.029)	0.349	0.372 (0.046)	0.382	0.256 (0.038)	0.266	0.228 (0.030)	0.242
3DFlex	0.336 (0.002)	0.336	0.339 (0.007)	0.339	–	–	–	–	0.304 (0.010)	0.306
3DVA	0.351 (0.003)	0.351	0.341 (0.006)	0.341	0.375 (0.038)	0.372	0.095 (0.040)	0.083	0.324 (0.010)	0.325
RECOVAR	0.391 (0.001)	0.391	0.372 (0.008)	0.371	0.430 (0.016)	**0.432**	0.306 (0.093)	0.330	**0.363 (0.011)**	**0.366**
3D Class	0.297 (0.019)	0.291	0.309 (0.010)	0.307	0.289 (0.081)	0.288	0.046 (0.026)	0.037	0.307 (0.023)	0.308
CryoSPIRE*	**0.402 (0.002)**	**0.402**	0.386 (0.014)	**0.389**	0.427 (0.014)	0.424	–	–	–	–
CryoHype*	–	–	–	–	–	–	–	–	–	–
SOLVAR	–	–	–	–	–	–	–	–	–	–
CryoKRAQEN*	–	–	–	–	–	–	–	–	–	–
CryoGaussian (Ours)	**0.402 (0.001)**	**0.402**	**0.387 (0.010)**	0.387	**0.432 (0.016)**	**0.432**	**0.336 (0.034)**	**0.343**	0.364 (0.012)†	0.366†

reference heterogeneity represented by the inferred latents. Figure 13 provides a complementary selection based on latent clustering, with each representative density decoded at an observed latent nearest to a cluster center. Lower-SNR IgG-1D results are discussed separately in Appendix B.5.

A.3 Heterogeneous reconstruction on experimental data

Assembling 50S ribosomal subunit. Figure 14 shows the reconstructed minor assembly states and their map-to-map FSC against deposited densities using the CryoDRGN-filtered particle subset. Published class assignments are used to select representative latents after training. Figure 15 compares the intermediate with H68 density but no resolved central protuberance with homogeneous CryoSPARC reconstructions using the released CryoDRGN state selection (a) and its intersection with the CryoGaussian-filtered particle set (b). The separate particle-selection analysis is described in Appendix B.4.

Pre-catalytic spliceosome. The principal-component traversals in Figure 5 describe conformational variation within the CryoDRGN-filtered subset. Figure 16 separately examines the full particle stack using the preprocessing in Appendix C.1. Its cluster-selected densities include differences in SF3b density and additional density near the U2 core. This analysis characterizes the broader particle stack and does not define the particle selection used for the main reconstruction.

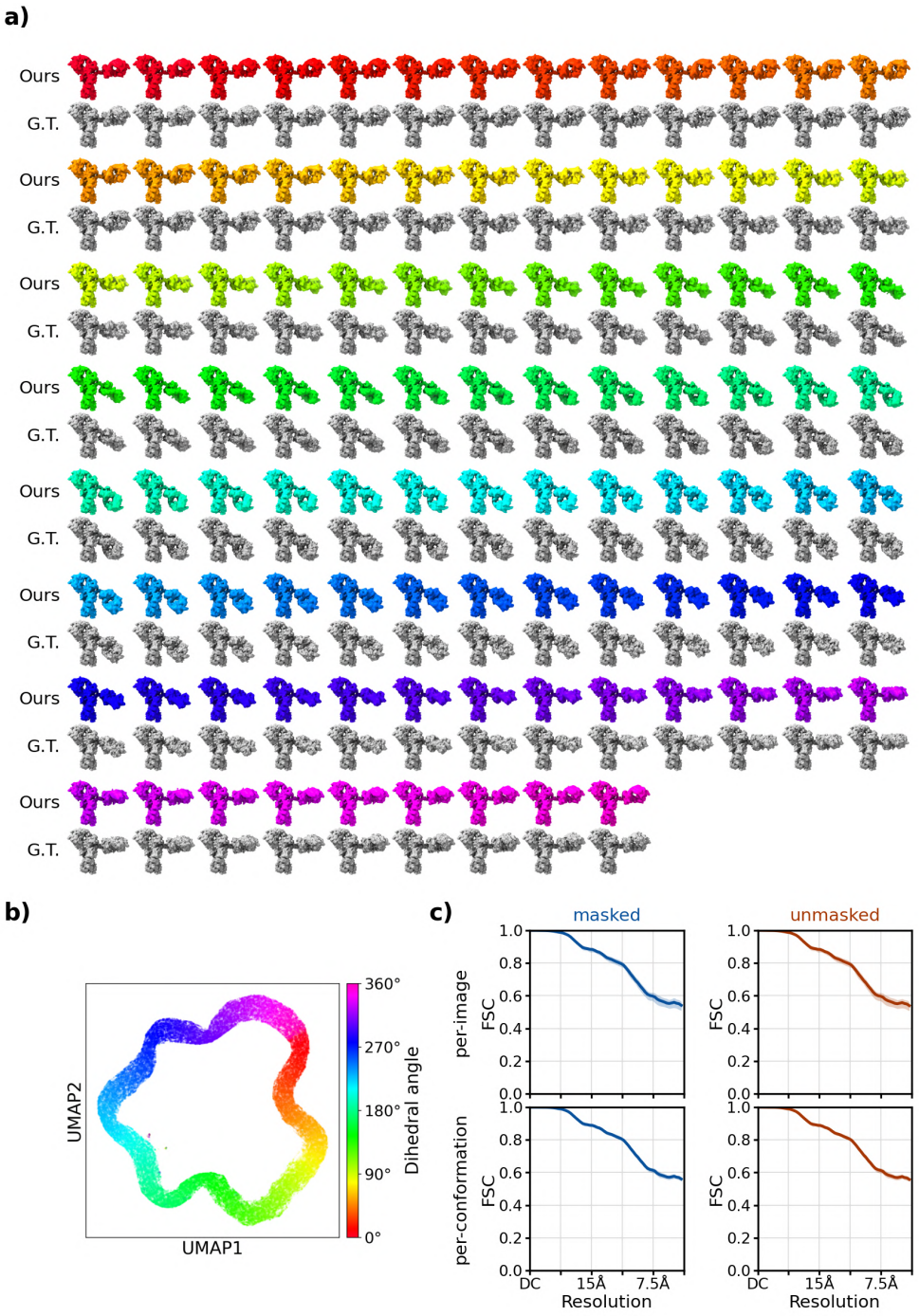

Figure 8: **IgG-1D reconstruction at SNR 0.01.** **(a)** CryoGaussian densities reconstructed at the inferred latents for 100 particle images (Ours), each paired with its reference density (G.T., gray). Reconstructed densities are colored by the reference dihedral angle. **(b)** UMAP projection of inferred latents with the same color scale. **(c)** Per-image FSC (top) and per-conformation FSC (bottom), evaluated with background masking (left) and without masking (right). Solid curves show the mean, and shaded regions span the 5th–95th percentiles across 100 evaluated particle images or 100 reference states, respectively.

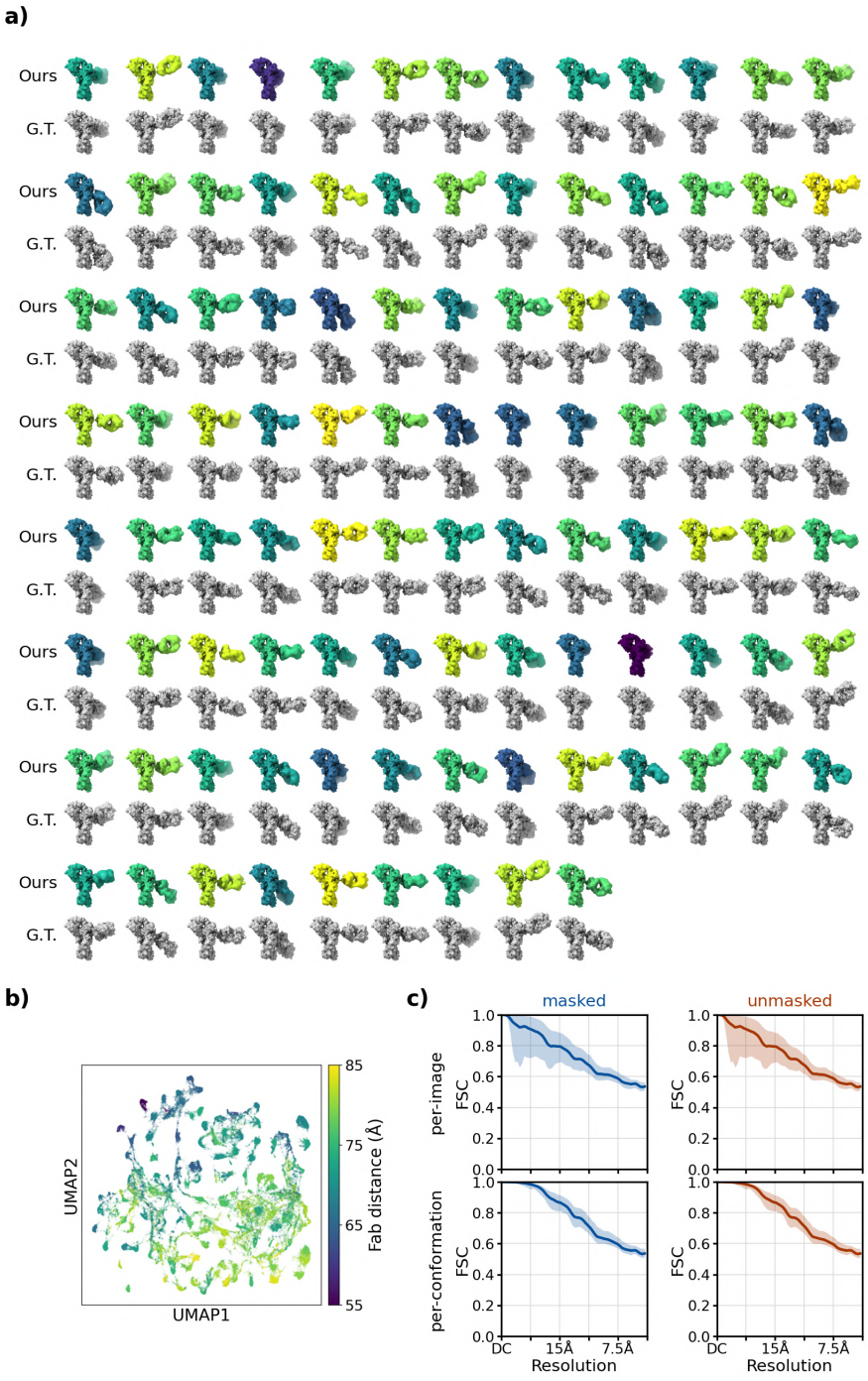

Figure 9: **IgG-RL reconstruction.** **(a)** CryoGaussian densities reconstructed at the inferred latents for 100 particle images (Ours), paired with their reference densities (G.T., gray). Colors indicate the distance from the Fab domain to the center of mass of the remaining structure. **(b)** UMAP projection of inferred latents with the same color scale. **(c)** Per-image FSC (top) and per-conformation FSC (bottom), with background masking (left) and without masking (right). Solid curves show the mean, and shaded regions span the 5th–95th percentiles across 100 evaluated particle images or 100 reference states, respectively.

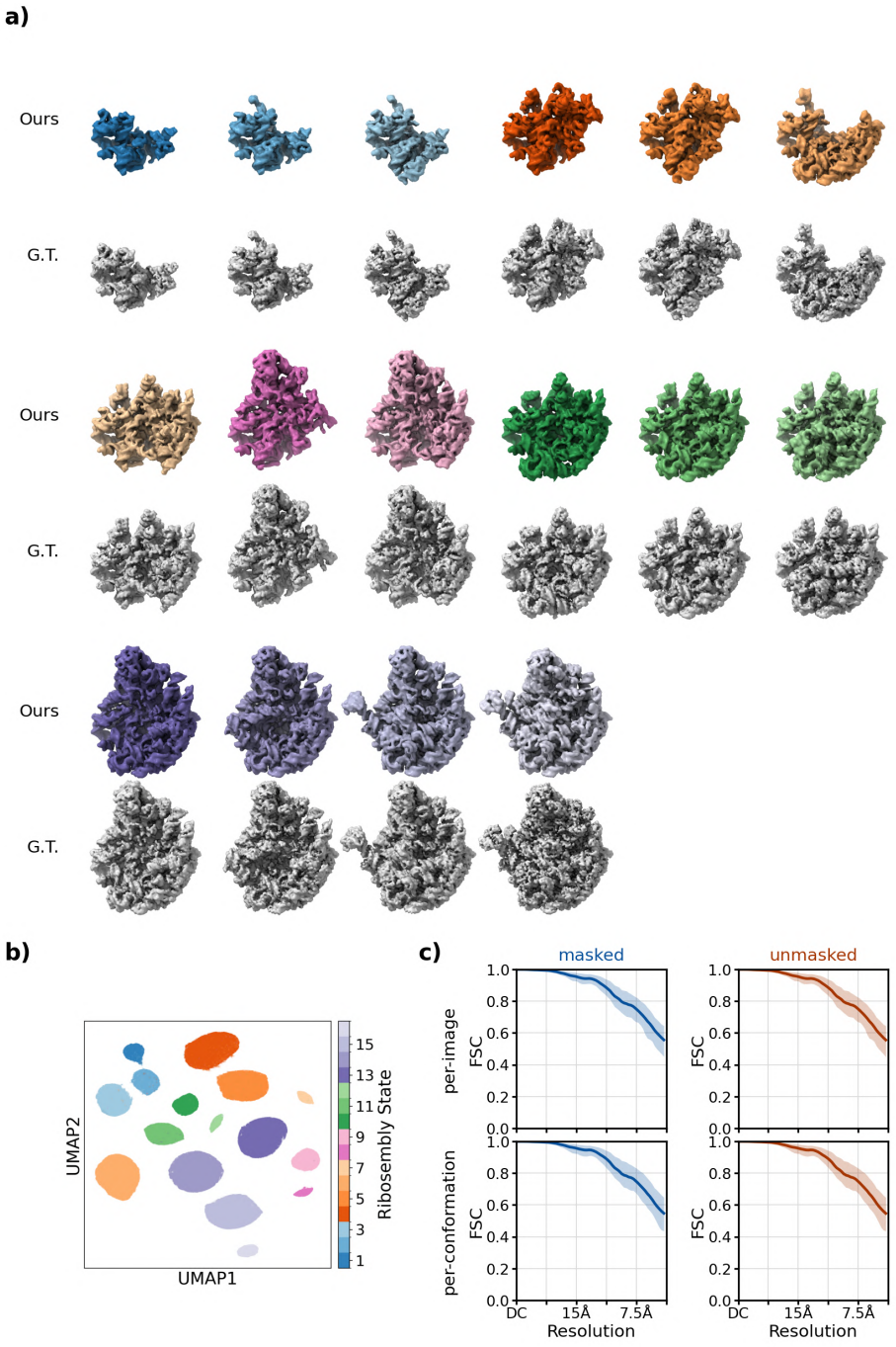

Figure 10: **Ribosembly reconstruction.** **(a)** CryoGaussian densities reconstructed at inferred latents for particle images spanning the 16 reference assembly states (Ours), paired with their reference densities (G.T., gray). **(b)** UMAP projection of inferred latents. Colors in (a) and (b) identify reference assembly states. **(c)** Per-image FSC (top) and per-conformation FSC (bottom), with background masking (left) and without masking (right). Solid curves show the mean, and shaded regions span the 5th–95th percentiles across 101 evaluated particle images in the top row and 16 reference states in the bottom row.

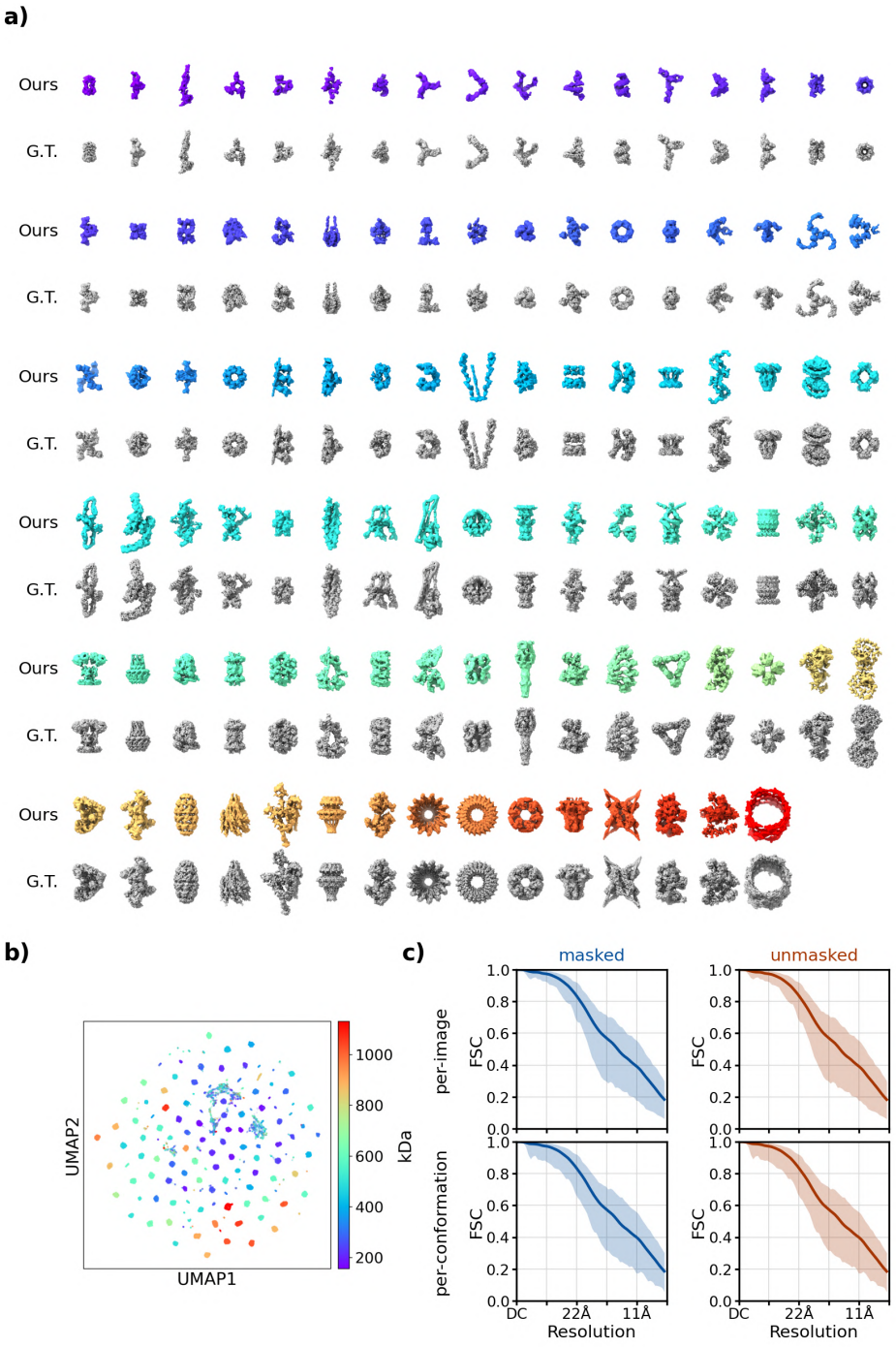

Figure 11: **Tomotwin-100 reconstruction.** **(a)** Per-image CryoGaussian densities (Ours) and paired reference densities (G.T., gray) for 100 distinct complexes. **(b)** UMAP projection of inferred latents; colors in (a) and (b) indicate molecular mass in kDa. **(c)** Per-image FSC (top) and per-conformation FSC (bottom), with background masking (left) and without masking (right). Solid curves show the mean, and shaded regions span the 5th–95th percentiles across 100 evaluated particle images or 100 reference complexes, respectively.

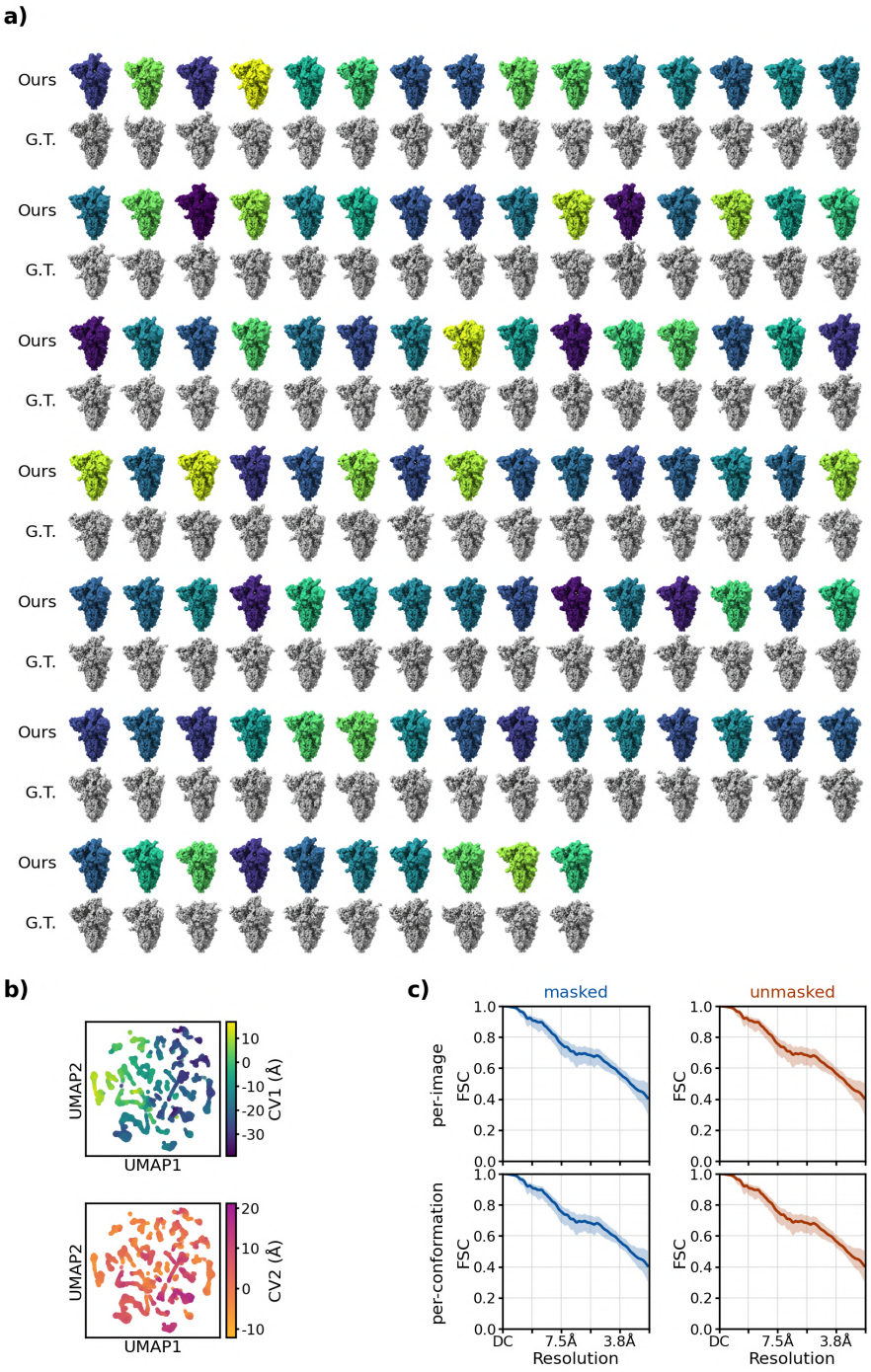

Figure 12: **Spike-MD reconstruction.** **(a)** CryoGaussian densities reconstructed at the inferred latents for 100 selected particle images (Ours), paired with their reference densities (G.T., gray) and colored by collective variable CV1. **(b)** UMAP projections of inferred latents, colored by CV1 (top) and CV2 (bottom), with collective-variable definitions from Wieczór et al. (2023). **(c)** FSC with background masking (left) and without masking (right). Both rows show the same per-image comparisons (Appendix D.2). Solid curves show the mean, and shading spans the 5th–95th percentiles across the reference comparisons.

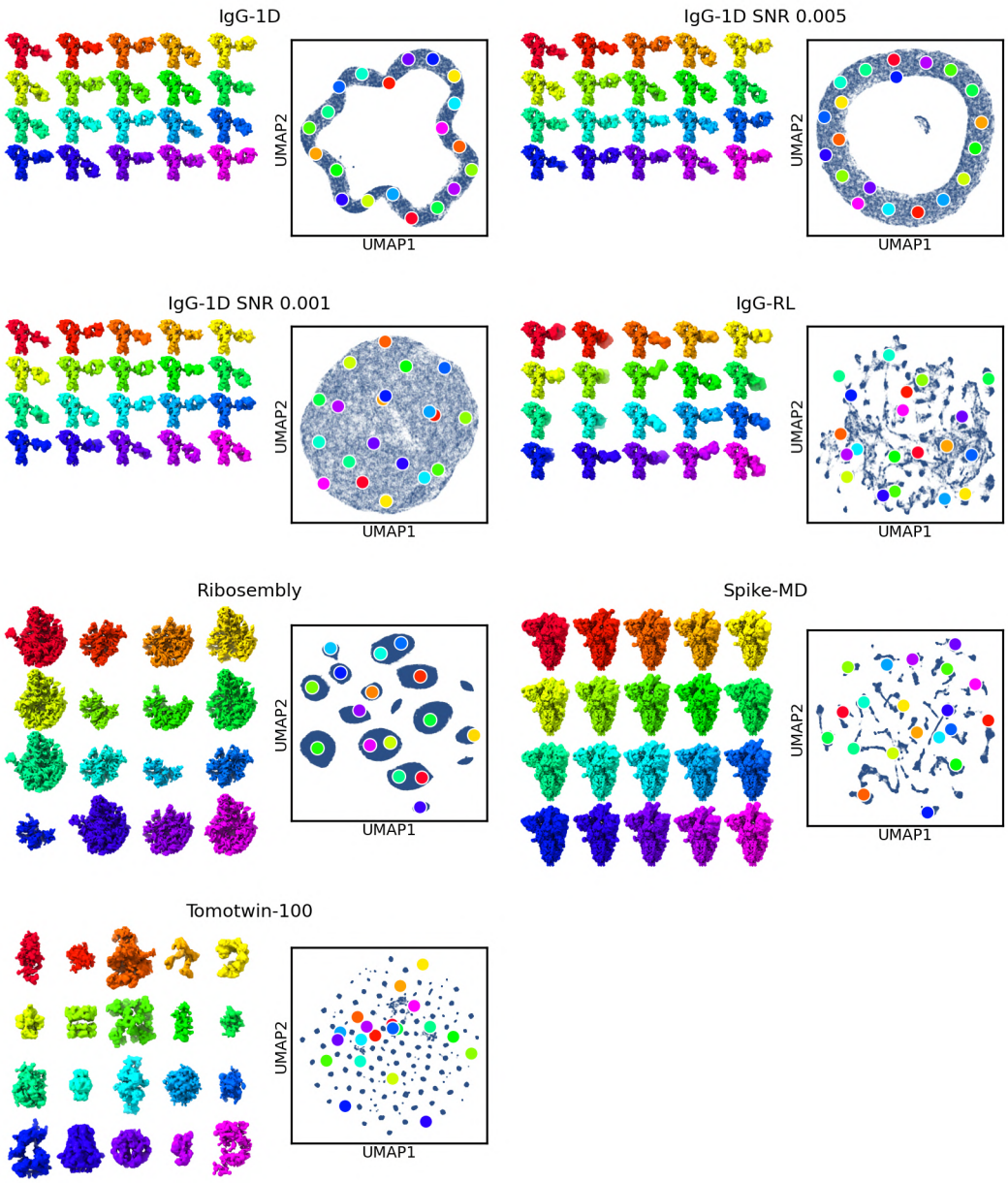

Figure 13: **Representative densities selected by latent clustering.** For each dataset and noise condition, inferred latents are partitioned by k-means into 20 clusters, or 16 for Ribosembly. The observed latent nearest to each cluster center is decoded to obtain a representative density. Each dataset block shows these densities alongside the UMAP projection, with colored markers locating the selected latents. Matching colors link each density to its marker and indicate cluster identity within that dataset.

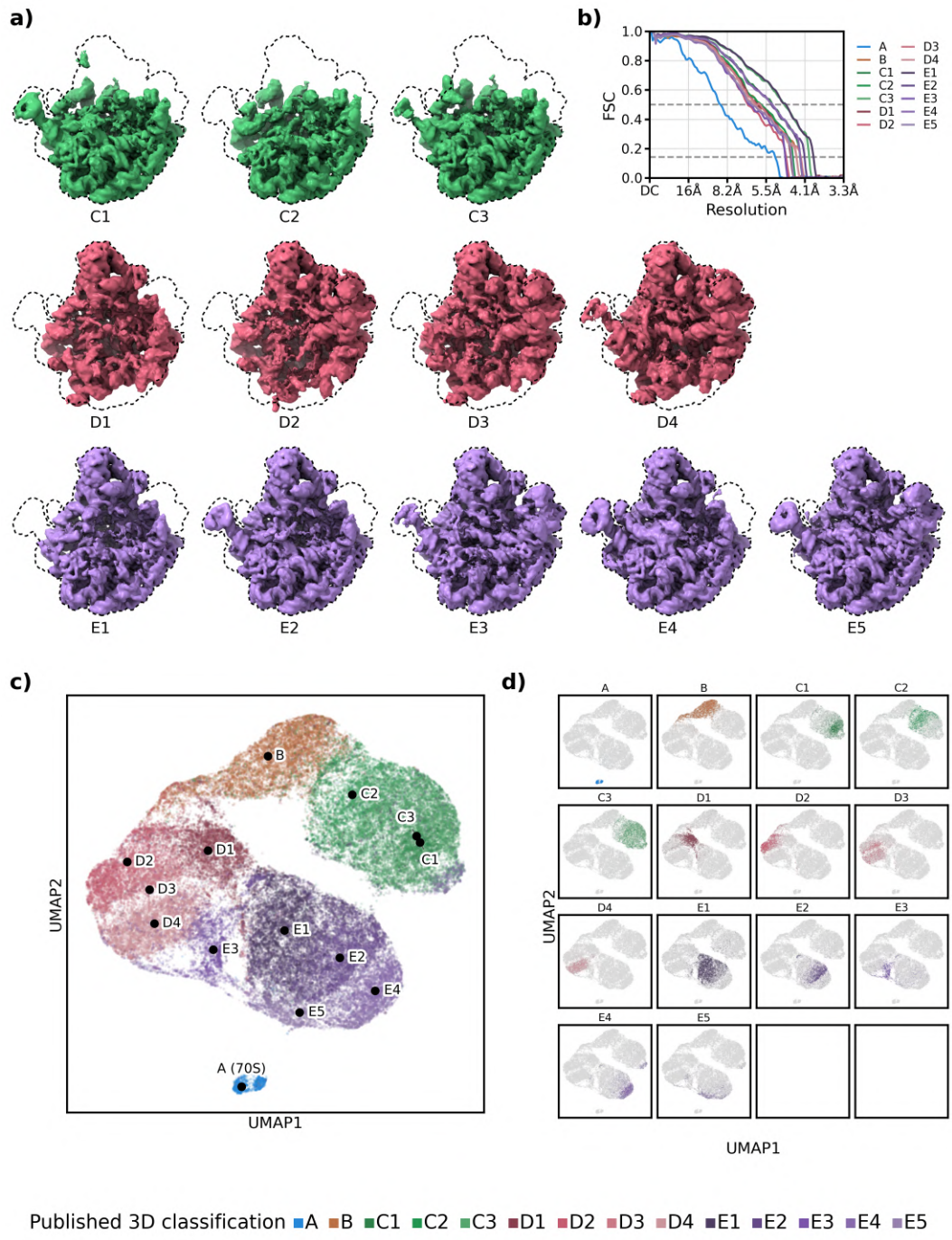

Figure 14: **Assembly-state reconstruction using the CryoDRGN particle selection for EMPIAR-10076.** The model is trained on 87,327 particle images. **(a)** Reconstructed densities for the twelve minor classes C1–C3, D1–D4, and E1–E5, with dashed outlines indicating the mature 50S envelope. **(b)** Map-to-map FSC between CryoGaussian densities and the corresponding deposited reconstructions (Davis et al., 2016), for class A and the thirteen assembly classes B–E5. Horizontal dashed lines indicate FSC values of 0.5 and 0.143. **(c)** UMAP projection of inferred latents colored by published class assignments. **(d)** The same projection with each class highlighted separately. Colors are shared across panels (b)–(d). Representative-latent selection and map comparison are described in Appendix D.4.

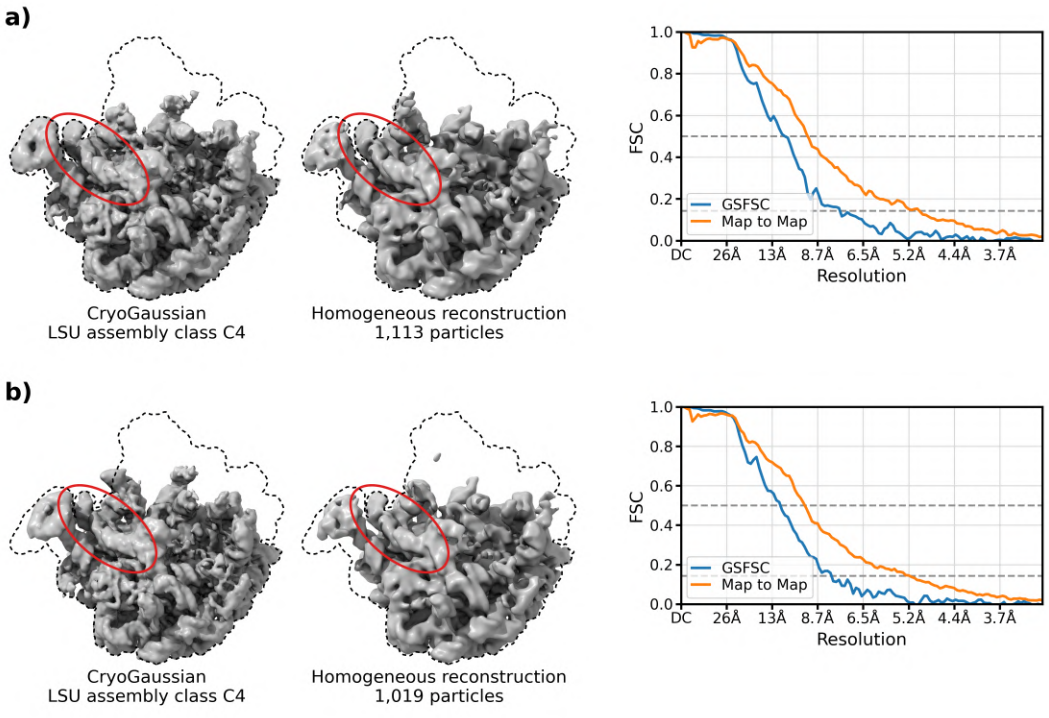

Figure 15: **Homogeneous refinement of the previously reported 50S assembly intermediate across particle selections. (a, b)** CryoSPARC refinement uses the 1,113 particles identified by the released CryoDRGN indices for this intermediate (a) and their intersection with the CryoGaussian-filtered particle set (1,019 particles; b) (Zhong et al., 2021). Each row shows the CryoGaussian density, the CryoSPARC density obtained by homogeneous refinement of the corresponding selected particle images, and FSC curves. Blue curves show gold-standard FSC between independent CryoSPARC half-maps; orange curves show map-to-map FSC between the displayed CryoGaussian and CryoSPARC densities. Horizontal dashed lines indicate FSC values of 0.5 and 0.143. Dashed outlines indicate the mature 50S envelope, and red ellipses mark rRNA helix 68 (H68).

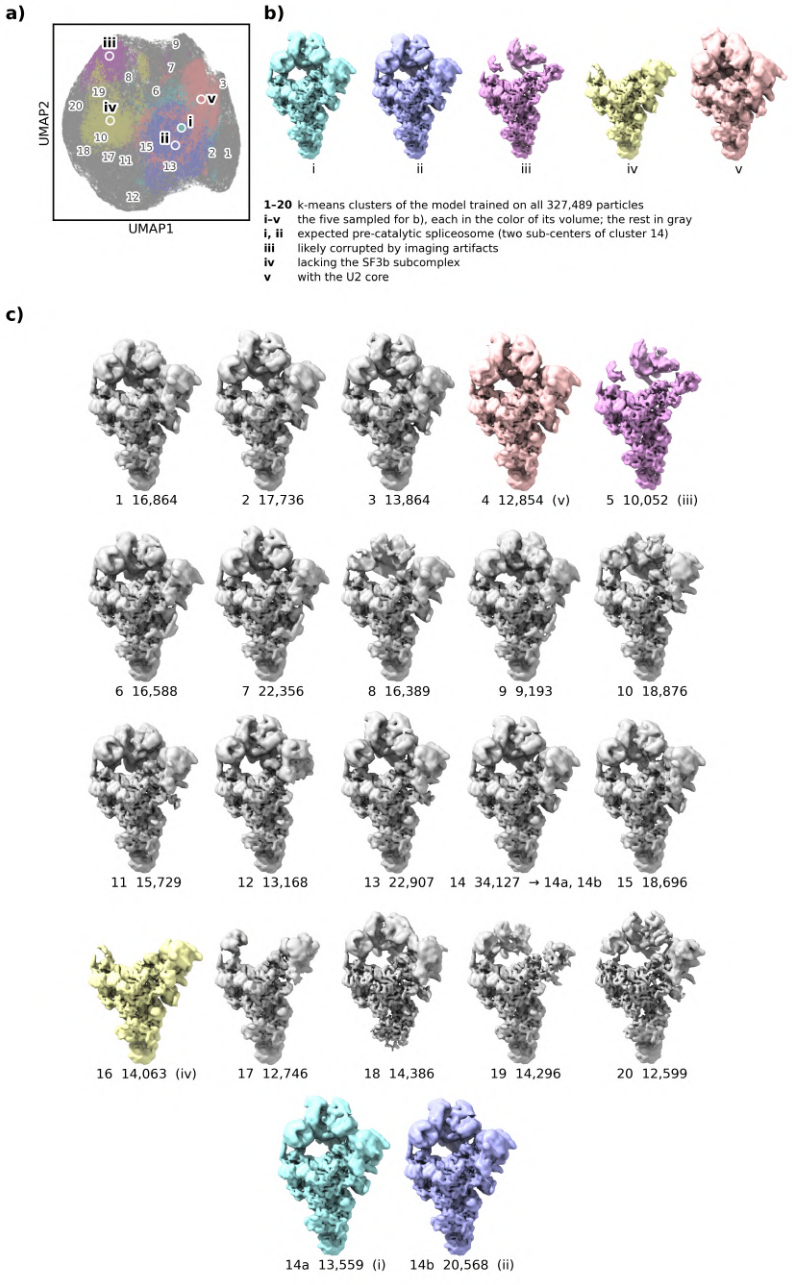

Figure 16: **Latent-space analysis of the full EMPIAR-10180 particle stack.** A four-dimensional model is trained on 327,489 particle images at 128×128 pixels. **(a)** UMAP projection of inferred latents, with 20 k-means clusters numbered 1–20. Five selected groups are colored to match the reconstructed densities in (b); the remaining groups are gray. **(b)** Representative densities from two subdivisions of the largest cluster (i, ii), a cluster with poorly resolved density (iii), a cluster lacking density in the SF3b region (iv), and a cluster with additional density near the U2 core (v). **(c)** Representative densities for all 20 clusters and the two subdivisions 14a and 14b, with particle counts below each density. Representative-latent selection is described in Appendix D.4. This full-stack analysis is separate from the CryoDRGN-filtered particle subset used in Figure 5.

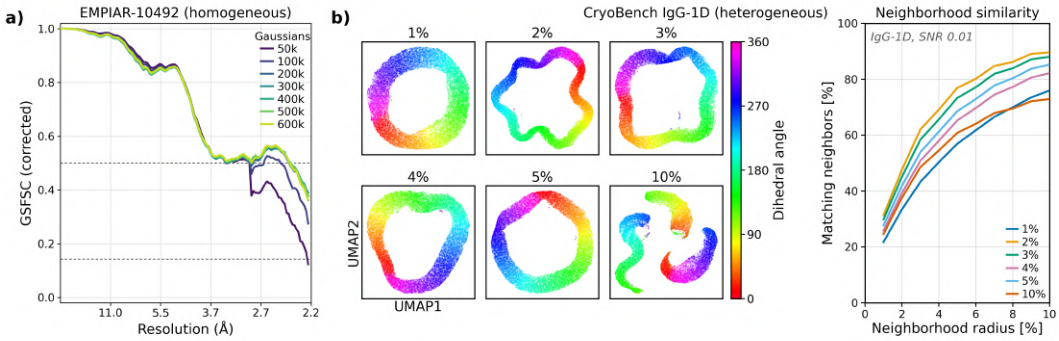

Figure 17: **Sensitivity to Gaussian count and relative density threshold.** **(a)** Corrected gold-standard FSC for homogeneous reconstruction of EMPIAR-10492. The Gaussian count varies from 50,000 to 600,000 as indicated in the legend. FSC is evaluated in CryoSPARC using an automatically generated tight mask with noise substitution. **(b)** Sensitivity to the relative density threshold used for Gaussian-count estimation on IgG-1D at SNR 0.01, with $\tau \in \{0.01, 0.02, 0.03, 0.04, 0.05, 0.10\}$. UMAP projections of inferred latents, colored by the reference dihedral angle, are shown on the left; corresponding neighborhood-similarity curves are shown on the right, following Appendix D.1. The default threshold $\tau = 0.02$ yields the highest neighborhood similarity among the tested values. The UMAP projections retain an ordered loop for thresholds from 0.01 to 0.05, whereas $\tau = 0.10$ produces a fragmented projection and lower neighborhood similarity.

B Ablations and Robustness

B.1 Gaussian Count and Relative Density Threshold

Figure 17 examines the Gaussian count directly in homogeneous reconstruction and its estimation through τ in heterogeneous reconstruction. In panel (a), the EMPIAR-10492 particle stack is reconstructed with 50,000, 100,000, 200,000, 300,000, 400,000, 500,000, and 600,000 Gaussians. Gains in high-frequency half-map agreement diminish as the count increases, with little change from 500,000 to 600,000 Gaussians. The estimate at $\tau = 0.02$, approximately 5.3×10^5 Gaussians (Table 6), lies between these configurations, supporting it as a practical representation budget for this dataset.

In panel (b), we vary τ over $\{0.01, 0.02, 0.03, 0.04, 0.05, 0.10\}$ on IgG-1D at SNR 0.01. Although $\tau = 0.02$ gives the highest neighborhood similarity, thresholds of 0.01–0.05 retain conformational ordering in the UMAP projections, indicating robustness around the default rather than dependence on a narrowly tuned value. Increasing τ to 0.10 reduces the estimated support and Gaussian count, yielding a fragmented projection and lower neighborhood similarity. Conversely, lowering τ to 0.01 increases the count without improving neighborhood similarity over the default. These comparisons show that increasing the Gaussian count alone does not ensure better heterogeneity estimation, complementing the space–bandwidth motivation in Appendix F.

B.2 Gaussian FWHM Constraint

We compare heterogeneity estimation on IgG-1D with and without the Gaussian FWHM constraint at SNR 0.01 and 0.001 (Figure 18). The constraint in Equation equation 5 sets a lower bound of $2a$ on Gaussian FWHM, preventing Gaussian widths from decreasing below this sampling scale. At SNR 0.01, the inferred latents retain an ordered loop in the UMAP projection with the constraint, whereas removing it produces folded, overlapping branches. At SNR 0.001, both projections become diffuse. Neighborhood similarity, evaluated in the inferred latent space against the reference heterogeneity embedding, is higher with the constraint across the displayed neighborhood sizes at both noise levels. These comparisons support constraining Gaussian FWHM when learning conformational heterogeneity from noisy particle images.

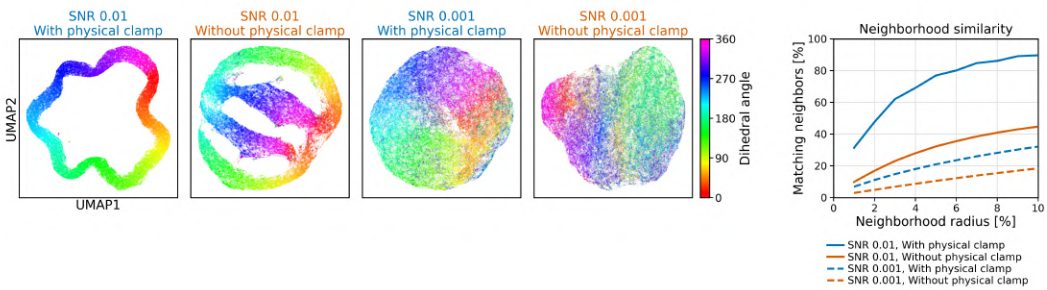

Figure 18: **Ablation of the Gaussian FWHM constraint on IgG-1D.** Heterogeneity estimation is evaluated at SNR 0.01 and 0.001 with and without the constraint in Equation equation 5, denoted "physical clamp" in the figure. **Left:** UMAP projections of inferred latents, colored by the reference dihedral angle. **Right:** Neighborhood similarity, measured as the percentage of matching neighbors (pMN) between the inferred latents and the reference heterogeneity embedding, following Appendix D.1. The constraint yields higher pMN across the displayed neighborhood sizes at both noise levels. Colors denote the constraint setting, and line styles denote SNR.

B.3 Gaussian initialization

IgG-1D provides a known one-dimensional conformational motion for evaluating whether inferred latents preserve reference conformational neighborhoods. We compare random and consensus initialization for heterogeneity estimation at SNR 0.01 and 0.001, giving four runs in total (Figure 19).

Random initialization follows Appendix E. For consensus initialization, we first obtain a consensus density by homogeneous reconstruction of all particle images. We initialize Gaussian positions by density-weighted sampling without replacement over thresholded voxels. Only the positions are initialized from this density; the remaining Gaussian attributes use the same initialization as in the random-initialization runs. The positions remain learnable during optimization.

Each run jointly learns latents for the particle images and the Gaussian density representation. We compute neighborhood similarity in the inferred latent space following Appendix D.1; UMAP projections visualize the latents colored by the reference dihedral angle.

Averaged over 20 neighborhood sizes (1–20% of the evaluation subset), pMN is 82.73% for random initialization and 55.84% for consensus initialization at SNR 0.01. At SNR 0.001, the corresponding values are 30.89% and 27.10%. Random initialization yields higher pMN across all evaluated neighborhood sizes at both noise levels. Thus, the tested consensus initialization does not improve agreement with reference conformational neighborhoods in these comparisons.

B.4 Particle selection

We examine EMPIAR-10076 using both the CryoDRGN-filtered and CryoGaussian-filtered particle subsets. Figure 20 documents CryoGaussian filtering by intersecting selections from one- and four-dimensional models. Of the 97,651 retained particles, a random 90% (87,885 particles) are used for subsequent heterogeneous reconstruction. The four-dimensional model uses the same latent dimension as this reconstruction. Figure 21 shows the reconstructed minor assembly states and their map-to-map FSC against deposited densities for the CryoGaussian-filtered subset. Published class assignments are used to select representative latents after training. Figure 22 compares reconstructed densities for major assembly classes B–E with CryoDRGN and across the two particle selections used for CryoGaussian reconstruction. Both particle selections recover the intermediate with H68 density but no resolved central protuberance. Clustering the inferred latents without assembly-state labels also yields a cluster enriched for particles from this previously reported intermediate. Homogeneous CryoSPARC refinement of the intersection between the released state selection and the CryoGaussian-filtered particle set supports agreement on these structural features (Figure 15(b)).

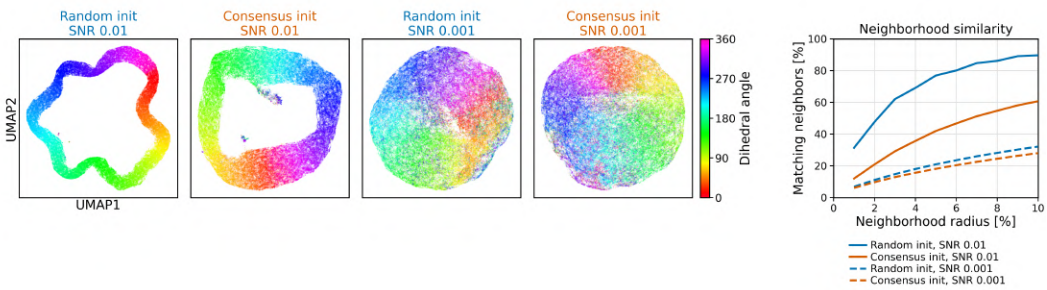

Figure 19: **Ablation of Gaussian initialization on IgG-1D.** Heterogeneity estimation is evaluated at SNR 0.01 and 0.001 using random initialization or Gaussian positions sampled from a consensus density obtained by homogeneous reconstruction. Consensus initialization uses density-weighted sampling without replacement over thresholded voxels, as detailed in Appendix B.3. **Left:** UMAP projections of inferred latents, colored by the reference dihedral angle. **Right:** Neighborhood similarity, measured as the percentage of matching neighbors (pMN) between the inferred latents and the reference heterogeneity embedding, plotted against neighborhood size as a percentage of the 20,000-image evaluation subset (Appendix D.1). Higher pMN indicates better preservation of reference conformational neighborhoods. Colors denote Gaussian initialization, and line styles denote SNR. Random initialization yields higher pMN across the evaluated neighborhood sizes at both noise levels.

B.5 Sensitivity to image noise

We use the three released IgG-1D noise conditions at SNR 0.01, 0.005, and 0.001 (Jeon et al., 2024). Figure 23 combines latent-space evaluation and per-image FSC for these conditions. Mean masked per-image FSC AUC is 0.397, 0.391, and 0.302, respectively. CryoGaussian achieves the highest mean per-image FSC AUC among the compared methods at SNR 0.01 and 0.005 and remains competitive at SNR 0.001. The decrease is more pronounced at SNR 0.001; neighborhood similarity also decreases as noise increases. The corresponding mean per-conformation FSC AUC values are 0.402, 0.398, and 0.328 (Table 3).

Gaussian-count estimates for these noise conditions are reported in Table 6. Figures 24 and 25 show the lower-SNR reconstructed densities and FSC curves in detail. Baseline sources and latent-space evaluation procedures for the composite comparison are specified in Appendices C.4 and D.1.

B.6 Sensitivity to masking

Homogeneous reconstruction. Figure 26 compares gold-standard FSC under four masking conditions for each homogeneous reconstruction. The same pair of half-maps is used throughout each comparison. The curves differ by at most 0.03 over the evaluated spatial frequencies, indicating limited sensitivity of half-map FSC to the tested masks. All curves remain above 0.143 up to the respective Nyquist frequencies.

Heterogeneous reconstruction. We assess the sensitivity of per-image and per-conformation FSC AUC to background masking by evaluating the same reconstructed densities with and without the evaluation mask (Table 3), following the masked and unmasked comparison in CryoBench (Jeon et al., 2024). Removing the mask changes the mean FSC AUC by at most 0.001 at the displayed precision, except for IgG-1D at SNR 0.001, where the mean decreases by approximately 0.010 for both per-image and per-conformation evaluation. These results indicate limited sensitivity of the reported FSC AUC to background masking, with a larger effect at the lowest SNR.

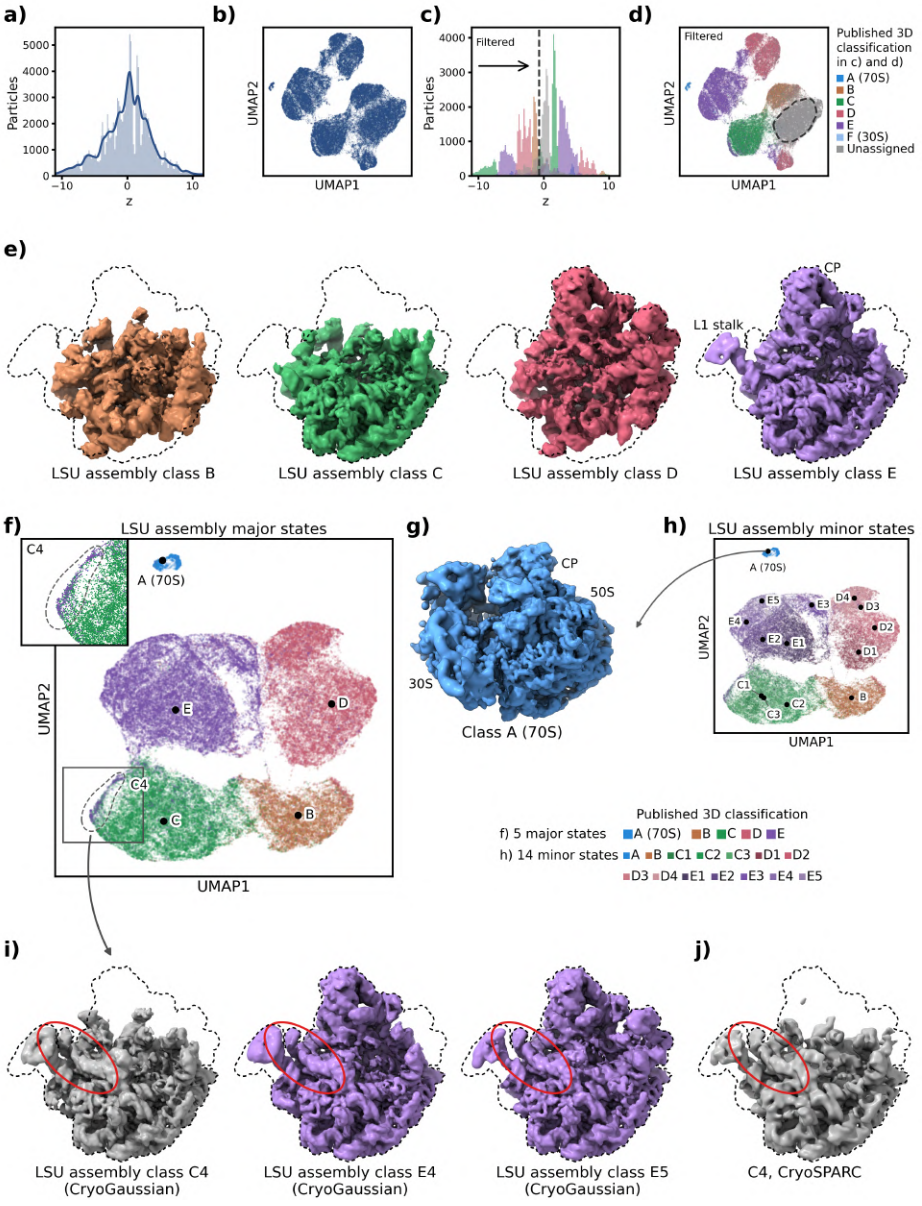

Figure 20: **Particle filtering and reconstruction of EMPIAR-10076 with CryoGaussian. (a, b)** Inferred latents from one- and four-dimensional models trained on all 131,899 particle images: a latent histogram with a kernel-density estimate (a) and a UMAP projection (b). **(c, d)** The same representations colored by published class assignments (Davis et al., 2016). The vertical dashed line in (c) marks the lower boundary of the excluded latent interval; the dashed contour in (d) identifies the excluded Gaussian-mixture component. The intersection of the two retained particle sets contains 97,651 images, of which a random 90% (87,885 images) are used for subsequent reconstruction. The four-dimensional model matches the latent dimension used in our heterogeneous reconstruction. **(e, g)** Reconstructed densities for assembly classes B–E and the 70S class A. **(f, h)** Inferred latents colored by major and minor class assignments; the inset identifies the previously reported C4 selection (Zhong et al., 2021). **(i)** C4, E4, and E5 densities. **(j)** Homogeneous reconstruction of the selected C4 particle images with CryoSPARC. Dashed outlines indicate the mature 50S envelope, and red ellipses mark H68. Filtering and reconstruction settings are given in Appendices C.1 and D.4.

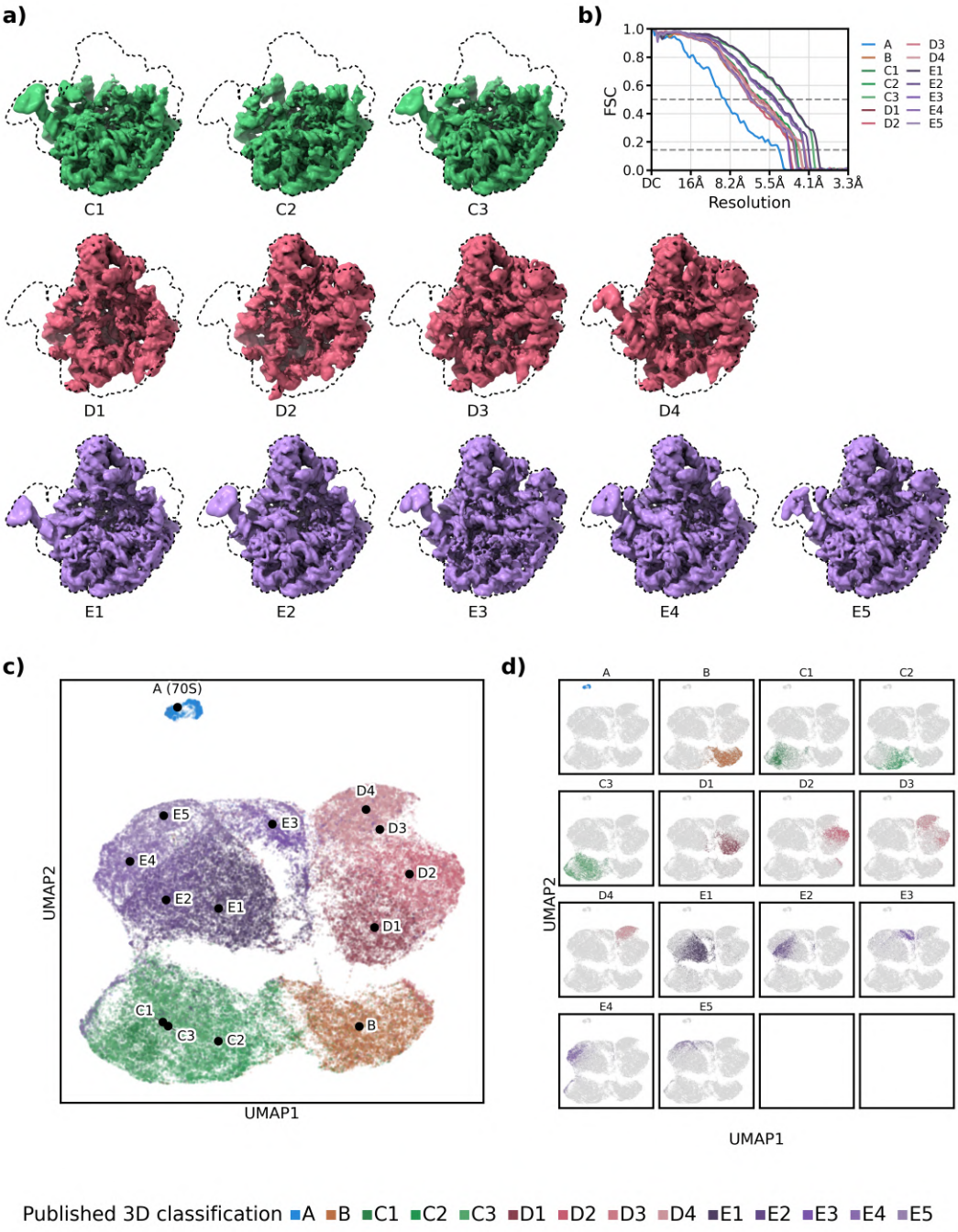

Figure 21: **Assembly-state reconstruction using the CryoGaussian particle selection for EMPIAR-10076.** The model is trained on 87,885 particle images selected as described in Figure 20. **(a)** Reconstructed densities for C1–C3, D1–D4, and E1–E5; dashed outlines indicate the mature 50S envelope. **(b)** Map-to-map FSC against the corresponding deposited reconstructions (Davis et al., 2016), including classes A and B. Horizontal dashed lines indicate FSC values of 0.5 and 0.143. **(c, d)** UMAP projections colored by published class assignments, shown jointly (c) and with one class highlighted at a time (d). Colors are shared across panels (b)–(d). Evaluation follows Figure 14.

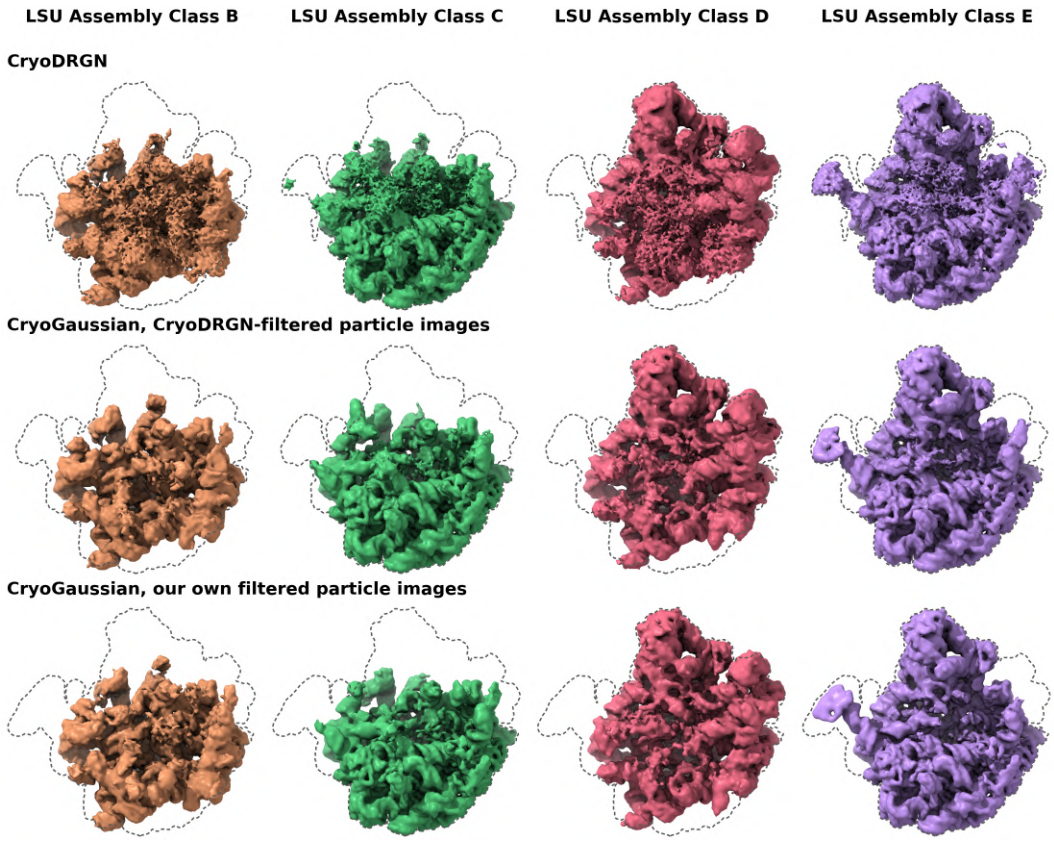

Figure 22: **Assembly-state comparison on EMPIAR-10076.** Columns show reconstructed densities for major assembly classes B–E. Rows show CryoDRGN (top), CryoGaussian using the CryoDRGN-filtered subset (middle), and CryoGaussian using the CryoGaussian-filtered subset (bottom). Dashed outlines indicate the mature 50S envelope. Particle selection is described in Appendix C.1.

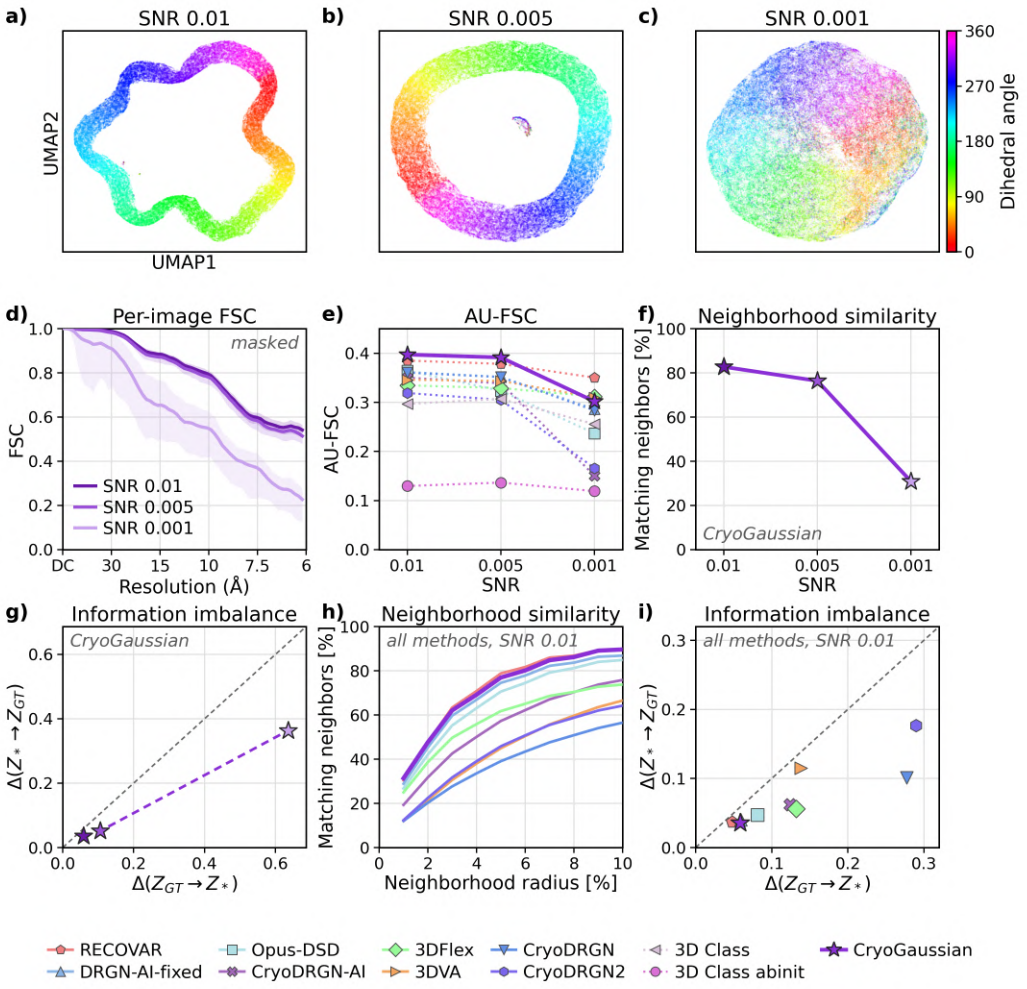

Figure 23: **Noise sensitivity of CryoGaussian on IgG-1D. (a–c)** UMAP projections of inferred latents at SNR 0.01, 0.005, and 0.001, respectively, colored by the reference dihedral angle. **(d)** Masked per-image FSC between reconstructed and reference densities. Solid curves show the mean, and shaded regions span the 5th–95th percentiles across 100 evaluated particle images. **(e)** Mean per-image FSC AUC as a function of SNR; baseline values are digitized from Figure 3a of CryoBench (Jeon et al., 2024). **(f, g)** Neighborhood similarity and information imbalance of the inferred latents at the three noise levels. **(h, i)** Corresponding comparisons at SNR 0.01, evaluated using the CryoBench protocol and released baseline latents; neighborhood similarity in (h) is plotted against neighborhood size as a percentage of the evaluation subset (Appendix D.1). Information imbalance uses reference heterogeneity embeddings regenerated following the CryoBench protocol. The legend is shared by (e), (h), and (i), with stars denoting CryoGaussian.

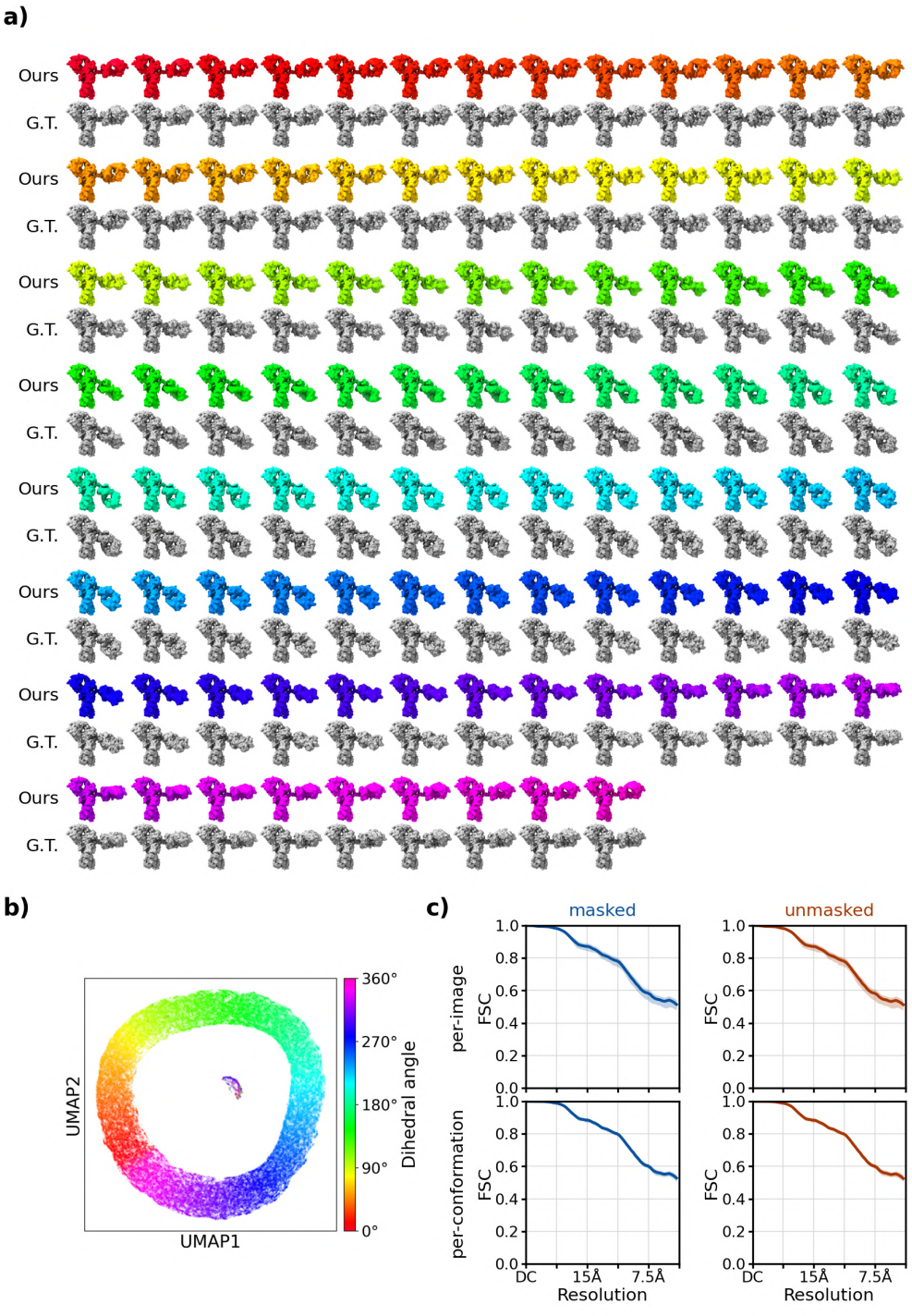

Figure 24: **IgG-1D reconstruction at SNR 0.005.** **(a)** Per-image reconstructed densities and paired reference densities. **(b)** UMAP projection of inferred latents, colored by the reference dihedral angle as in Figure 8. **(c)** Per-image FSC (top) and per-conformation FSC (bottom), with background masking (left) and without masking (right). Curves and shading show the mean and 5th–95th percentiles across 100 evaluated particle images or 100 reference states, respectively.

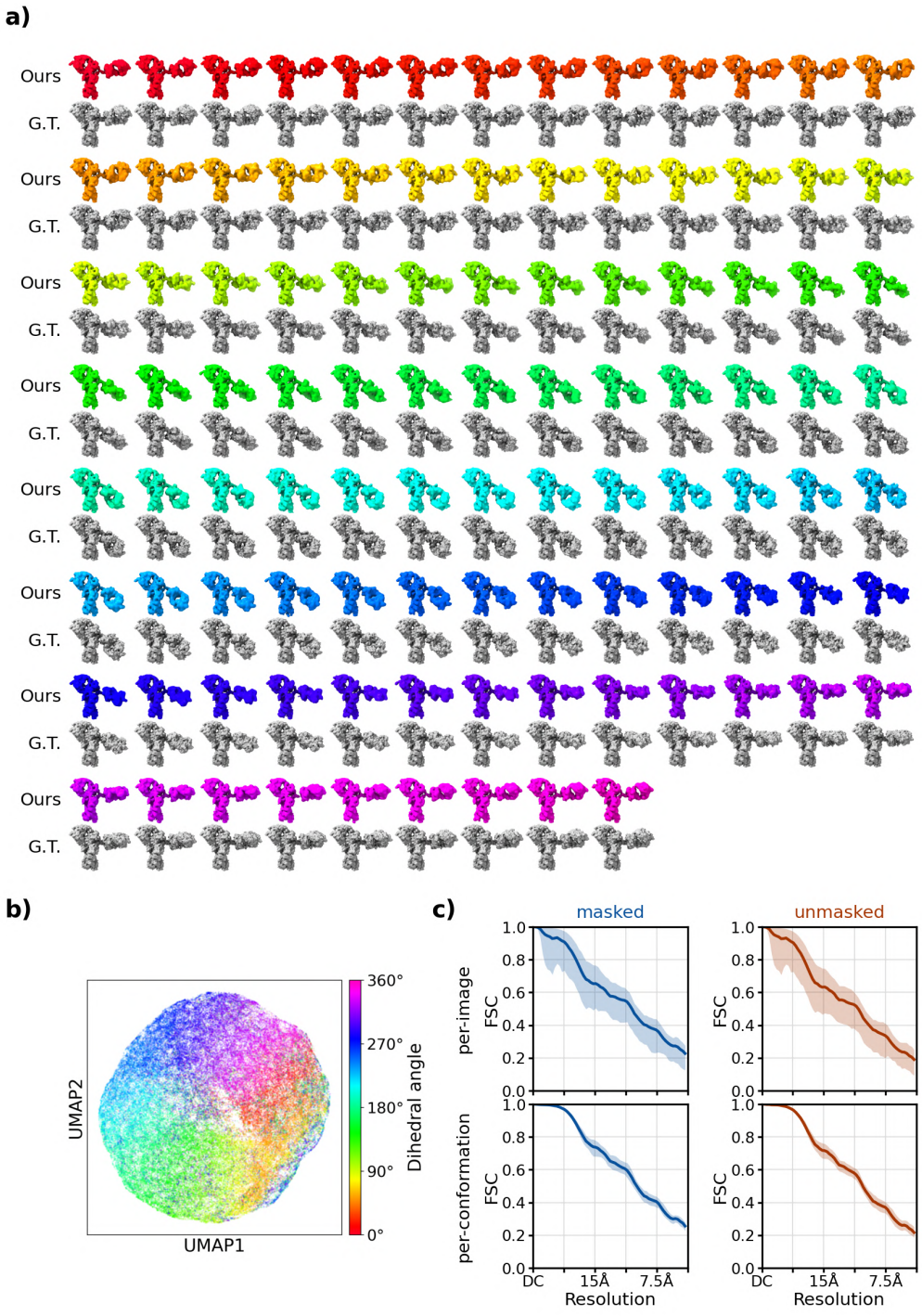

Figure 25: **IgG-1D reconstruction at SNR 0.001.** **(a)** Per-image reconstructed densities and paired reference densities. **(b)** UMAP projection of inferred latents, colored by the reference dihedral angle as in Figure 8. **(c)** Per-image FSC (top) and per-conformation FSC (bottom), with background masking (left) and without masking (right). Curves and shading show the mean and 5th–95th percentiles across 100 evaluated particle images or 100 reference states, respectively.

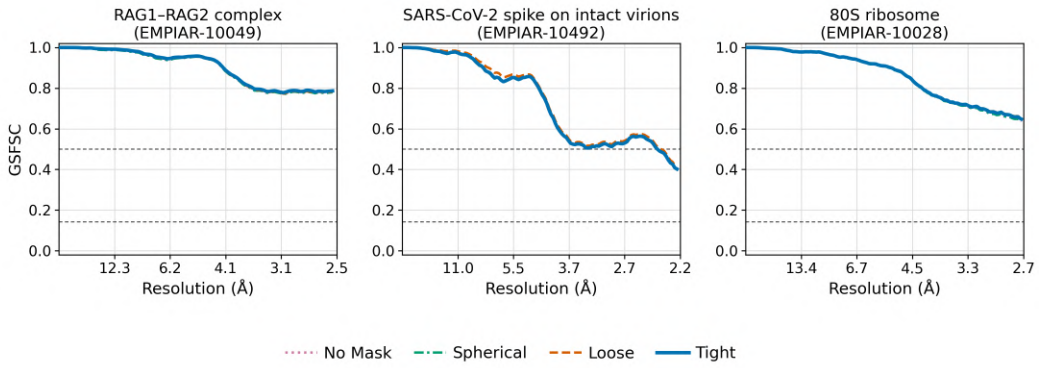

Figure 26: **Gold-standard FSC of CryoGaussian homogeneous reconstructions.** Panels show the RAG1–RAG2 complex (EMPIAR-10049), SARS-CoV-2 spike on intact virions (EMPIAR-10492), and the 80S ribosome (EMPIAR-10028), from left to right. Within each panel, the same pair of independently reconstructed half-maps is evaluated in CryoSPARC without a mask and with spherical, loose, and tight masks. Horizontal dashed lines indicate FSC values of 0.5 and 0.143. All curves remain above 0.143 up to the Nyquist frequency, corresponding to Nyquist limits of 2.46, 2.20, and 2.68 Å, respectively.

Table 3: **Masked and unmasked FSC AUC for CryoGaussian on CryoBench.** We report the mean (standard deviation), median, minimum, and maximum across evaluated particle images for per-image FSC and reference states for per-conformation FSC.

Dataset	Evaluation	Masked			Unmasked		
		Mean (std)	Median	Min–max	Mean (std)	Median	Min–max
IgG-1D (SNR 0.01)	Per-image	0.397 (0.004)	0.398	0.382–0.404	0.397 (0.004)	0.398	0.382–0.403
	Per-conformation	0.402 (0.001)	0.402	0.398–0.404	0.401 (0.001)	0.401	0.397–0.404
IgG-1D (SNR 0.005)	Per-image	0.391 (0.012)	0.394	0.296–0.400	0.391 (0.012)	0.394	0.296–0.399
	Per-conformation	0.398 (0.001)	0.398	0.394–0.402	0.397 (0.002)	0.397	0.393–0.401
IgG-1D (SNR 0.001)	Per-image	0.302 (0.030)	0.316	0.205–0.336	0.292 (0.030)	0.303	0.200–0.331
	Per-conformation	0.328 (0.005)	0.328	0.315–0.339	0.318 (0.005)	0.318	0.302–0.329
IgG-RL	Per-image	0.366 (0.020)	0.370	0.318–0.400	0.366 (0.020)	0.369	0.318–0.400
	Per-conformation	0.387 (0.010)	0.387	0.352–0.410	0.387 (0.010)	0.387	0.352–0.410
Ribosembly	Per-image	0.432 (0.014)	0.432	0.386–0.452	0.432 (0.014)	0.432	0.386–0.452
	Per-conformation	0.432 (0.016)	0.432	0.391–0.454	0.432 (0.016)	0.432	0.391–0.454
Tomotwin-100	Per-image	0.335 (0.035)	0.341	0.217–0.387	0.335 (0.035)	0.341	0.217–0.387
	Per-conformation	0.336 (0.034)	0.343	0.234–0.390	0.336 (0.034)	0.342	0.234–0.389
Spike-MD	Per-image	0.364 (0.012)	0.366	0.326–0.386	0.364 (0.012)	0.366	0.326–0.386

Table 4: Experimental particle-image datasets. Particle selection and preprocessing are described below. The first three datasets are used for homogeneous reconstruction and the last two for heterogeneous reconstruction. Pixel sizes refer to the deposited particle images.

Dataset	Image dimensions (pixels)	Pixel size (Å/pixel)
EMPIAR-10049 (Ru et al., 2015)	192×192	1.23
EMPIAR-10028 (Wong et al., 2014)	360×360	1.34
EMPIAR-10492 (Ke et al., 2020)	256×256	1.098
EMPIAR-10076 (Davis et al., 2016)	320×320	1.31
EMPIAR-10180 (Plaschka et al., 2017)	320×320	1.699

C Experimental settings

C.1 Datasets and preprocessing

Table 4 summarizes the source particle-image datasets; particle selection and preprocessing are described below. Table 5 summarizes the CryoBench particle images used in our experiments.

Experimental data. For EMPIAR-10049, EMPIAR-10028, and the main EMPIAR-10076 and EMPIAR-10180 reconstructions, we follow the particle selection and preprocessing procedures provided with CryoDRGN (Zhong et al., 2021), using the corresponding pose and CTF metadata. EMPIAR-10049 and EMPIAR-10028 retain their deposited image dimensions. For EMPIAR-10492, we use the particle selection and corresponding pose and CTF metadata from Ke et al. (2020), without additional selection or downsampling. For heterogeneous reconstruction, we Fourier-downsample the 320×320 images from EMPIAR-10076 and EMPIAR-10180 to 256×256, following CryoDRGN (Zhong et al., 2021). This preserves the physical field of view and yields pixel sizes of 1.6375 and 2.12375 Å/pixel, respectively. EMPIAR-10076 contains 131,899 source images; its CryoDRGN-provided selection contains 87,327 images. For EMPIAR-10180, the main reconstruction uses 139,720 particle images from the CryoDRGN-filtered subset. We exclude corrupted particle 308902 from the 327,490-image source stack, leaving 327,489 images for the separate full-stack analysis in Figure 16, Fourier-downsampled to 128×128 at 4.2475 Å/pixel following Zhong et al. (2021). Particle images, pose and CTF metadata, and selection indices remain in corresponding order throughout preprocessing.

Simulated homogeneous data. We generate reference densities from atomic models of the RAG1–RAG2 complex (PDB 3JBX), SARS-CoV-2 spike (PDB 6ZWV), and the 80S ribosome (PDB 3J79 and 3J7A). Image dimensions and pixel sizes match the corresponding experimental data. Defocus ranges, determined from the mean defocus values in the experimental metadata, are 11,327–23,913, 3,969–37,180, and 8,131–38,652 Å, respectively. The corresponding spherical aberration coefficients are 2.0, 2.7, and 2.0 mm; all simulations use an accelerating voltage of 300 kV and amplitude contrast of 0.1. CryoSPARC generates particle images by projecting the reference densities, applying CTF modulation, and adding Gaussian white noise at SNR 0.05. Reconstructed densities are evaluated by map-to-map FSC against the reference densities used to generate the particle images (Appendix D.3).

Half-set assignments for homogeneous reconstruction. We retain the half-set assignments recorded in the upstream particle metadata. The supplied particle selection is intersected with each half-set, preserving particle identities and their corresponding imaging metadata. Particles are not reassigned between half-sets. Each half-map is initialized and optimized independently, without sharing learned parameters or intermediate maps (Appendix D.3).

CryoGaussian particle selection for EMPIAR-10076. For the separate filtering analysis in Figure 20, we follow the filtering strategy of Zhong et al. (2021) with one- and four-dimensional models trained on all 131,899 particle images. The four-dimensional model uses the same latent dimension $d_z = 4$ as our heterogeneous reconstruction. For the one-dimensional latents, we form a Gaussian

Table 5: CryoBench datasets used in the main comparison. Dimensions and pixel sizes describe the processed images; the signal-to-noise ratio identifies the noise condition of the released dataset. For Spike-MD, reference states are the 46,789 distinct molecular-dynamics configurations reported by CryoBench.

Dataset	Number of particle images	Reference states	Image dimensions (pixels)	Pixel size (Å/pixel)	Signal-to-noise ratio
IgG-1D	100,000	100	128×128	3.0	0.01
IgG-RL	100,000	100	128×128	3.0	0.01
Ribosembly	335,240	16	128×128	3.0	0.01
Tomotwin-100	100,000	100	128×128	4.5	0.01
Spike-MD	100,000	46,789	256×256	1.5	0.1

kernel density estimate and exclude the interval between the two local minima adjacent to its highest peak, retaining 100,528 images. For the four-dimensional latents, we fit a five-component Gaussian mixture and exclude the component with the largest mean latent norm, retaining 105,211 images. The intersection of the two retained sets contains 97,651 images. We train the subsequent reconstruction on a random 90% of this intersection, comprising 87,885 images. Published class labels are used only for post-hoc evaluation of the selection.

CryoBench data. We use the released CryoBench datasets (Jeon et al., 2024) and their supplied pose and CTF metadata, without additional particle selection. IgG-1D, IgG-RL, and Tomotwin-100 each contain 1,000 particle images per state; Ribosembly has unequal state populations. IgG-1D, IgG-RL, and Ribosembly are Fourier-downsampled from 256×256 to 128×128, and Tomotwin-100 from 384×384 to 128×128. In each case, the physical field of view is preserved. Spike-MD is used at its released image dimensions. Its reference-state count follows the sampled molecular-dynamics configurations rather than a partition into discrete structural classes.

Shared settings. Heterogeneity estimation and per-state reconstruction use the same particle images, pose and CTF metadata, and particle selections.

C.2 Training settings

Shared training settings are specified in Section 4. For homogeneous reconstruction, we use global batch sizes of 16 for EMPIAR-10049 and EMPIAR-10028, and 4 for the smaller EMPIAR-10492 particle stack. For IgG-1D, IgG-RL, Tomotwin-100, Spike-MD, EMPIAR-10076, and EMPIAR-10180, global batch sizes are 256 for heterogeneity estimation and 128 for per-state reconstruction. For the larger Ribosembly particle stack, we use 512 and 256, respectively.

C.3 Computational cost

We run CryoGaussian on eight NVIDIA RTX PRO 6000 Blackwell Server Edition GPUs. Table 6 reports global batch sizes, step times, training times, forward throughput, and peak allocated memory per GPU for homogeneous or heterogeneous reconstruction.

C.4 Baseline configurations

The FSC AUC comparisons include CryoDRGN, CryoDRGN-AI-fixed, OPUS-DSD, 3DFlex, 3DVA, RECOVAR, 3D Classification, CryoSPIRE, CryoHype, SOLVAR (Yadgar et al., 2026), and CryoKRAQEN. Dataset coverage follows Tables 1 and 2 under the CryoBench evaluation framework (Jeon et al., 2024). We obtain the SOLVAR scores by running its released source code. For CryoSPIRE (Shekarforoush et al., 2025), CryoHype (Gu et al., 2026), and CryoKRAQEN (Gao et al., 2026), the included scores are taken directly from their respective publications. These methods carry the same asterisk in both tables; dashes denote results not included for a given dataset and protocol. For fair comparison under a consistent evaluation protocol, the per-conformation comparison includes baseline scores only where a per-conformation evaluation is explicitly documented in

Table 6: **Computational cost of CryoGaussian on eight NVIDIA RTX PRO 6000 Blackwell Server Edition GPUs.** We report global batch sizes. Gaussian-count estimates at $\tau = 0.02$ are rounded to the nearest integer.

Dataset	Box	Estimated Gaussians	Global batch	Step time (ms)	Training time (h)	Forward (particles/s)	Peak allocated (GiB/GPU)
EMPIAR-10492	256	530,767	4	11.945	0.066	–	–
EMPIAR-10049	192	126,886	16	12.275	0.068	2,159.9	0.107
EMPIAR-10028	360	787,490	16	39.195	0.218	616.3	0.558
EMPIAR-10180	256	448,148	256	376.644	6.825	956.3	12.142
EMPIAR-10076 (CryoDRGN-filtered)	256	513,705	256	520.190	8.500	841.1	14.141
EMPIAR-10076 (CryoGaussian-filtered)	256	463,802	256	464.008	7.632	950.5	12.776
IgG-1D (SNR 0.01)	128	10,316	256	17.313	0.299	27,250.4	0.332
IgG-1D (SNR 0.005)	128	10,736	256	17.943	0.310	25,501.7	0.344
IgG-1D (SNR 0.001)	128	31,338	256	36.510	0.644	9,738.0	0.897
IgG-RL	128	11,836	256	34.330	0.460	24,235.4	0.372
Ribosembly	128	33,033	512	73.886	1.213	12,287.0	1.881
Tomotwin-100	128	57,818	256	72.897	1.217	4,942.5	1.654
Spike-MD	256	115,877	256	216.492	2.788	3,680.6	3.220

the benchmark or method publication. Baseline per-image scores are reported separately and are not substituted for per-conformation scores.

We rerun all methods in the homogeneous comparison, including CryoSPARC (Punjani et al., 2017), CryoDRGN (Zhong et al., 2021), CryoNERF (Qu et al., 2025), and CryoSplat (Chen & Ling, 2026). CryoSPARC uses non-uniform refinement for EMPIAR-10049 and EMPIAR-10028, and homogeneous refinement for EMPIAR-10492. The EMPIAR-10049 CryoSPARC reconstruction uses C2 symmetry; the other two CryoSPARC reconstructions use C1. CryoDRGN homogeneous reconstructions use `train_nn` in version 3.4.4 for 25 epochs with batch size 8. On EMPIAR-10492, all five methods use the same particle selection.

For Figure 23, baseline FSC AUC values in panel (e) are digitized from Figure 3a of CryoBench. The latent-space comparisons in panels (h) and (i) are recomputed from the released baseline latents using the CryoBench evaluation code.

D EVALUATION PROTOCOLS

The synthetic-data evaluation distinguishes heterogeneity estimation from per-state reconstruction, following the evaluation framework of Jeon et al. (2024). Per-image and per-conformation FSC on CryoBench and half-map FSC on experimental data have different evaluation targets and are described separately below.

D.1 HETEROGENEITY ESTIMATION

We compare inferred latents with reference heterogeneity embeddings using neighborhood similarity (percentage of matching neighbors, pMN) and information imbalance, as defined in CryoBench (Jeon et al., 2024).

D.2 PER-IMAGE AND PER-CONFORMATION FSC EVALUATION

For density maps u and v, Fourier shell correlation is defined as (Jeon et al., 2024)

$$\mathrm{FSC}_{u,v}(\nu_j) = \frac{\mathrm{Re}\left[\sum_{\mathbf{k}\in S_j} \widehat{u}(\mathbf{k})\widehat{v}(\mathbf{k})^*\right]}{\sqrt{\left(\sum_{\mathbf{k}\in S_j} |\widehat{u}(\mathbf{k})|^2\right)\left(\sum_{\mathbf{k}\in S_j} |\widehat{v}(\mathbf{k})|^2\right)}}. \tag{11}$$

Here, S_j is the Fourier shell at radial spatial frequency ν_j; hats denote Fourier transforms, and * denotes complex conjugation. For masked comparisons, the evaluation mask is applied to both maps before the Fourier transform. Following CryoBench's per-image evaluation implementation (Jeon et al., 2024), we compute FSC AUC by trapezoidal integration:

$$\mathrm{FSC}_{\mathrm{AUC}} = \sum_{j=0}^{J-1} \frac{c_j + c_{j+1}}{2}(\nu_{j+1} - \nu_j), \tag{12}$$

where $\nu_0, \ldots, \nu_J$ are the sampled frequencies in cycles/voxel and $c_j = |\,\mathrm{FSC}_{u,v}(\nu_j)|$. We retain this frequency scale without rescaling AUC to a unit interval; its theoretical maximum over the full frequency range $[0, 1/2]$ is 0.5 (Jeon et al., 2024).

Per-image FSC compares a density reconstructed at an evaluated particle's inferred latent with that particle's reference density. It therefore retains the correspondence between an image and its underlying state. Per-conformation FSC instead selects one representative latent for each evaluated reference state: the latent nearest to the mean of the observed latents assigned to that state is decoded and compared with the corresponding reference density. Reference-state assignments are used for this evaluation selection. These two evaluations differ in their sampling and reference pairing (Jeon et al., 2024).

For CryoGaussian, densities are evaluated from the final Gaussian density representation ρ_{Θ_2}. Per-image and per-conformation FSC and their AUC summaries follow the CryoBench evaluation protocol and implementation (Jeon et al., 2024), with background masking unless otherwise stated. The density panels in Figures 8–12 use inferred latents for particle images; the FSC panels report the two evaluations separately, except for the Spike-MD reuse described below. Solid curves show the arithmetic mean, and shaded regions span the 5th–95th percentiles across the evaluated images or reference states. These bands describe variation across reference comparisons within the displayed reconstruction, rather than variation across independent training runs. Figure 13 instead selects an observed latent nearest to each k-means cluster center, using 20 clusters except for Ribosembly, which uses 16. Colors identify cluster membership within each dataset. For Spike-MD, we generate reference densities from the corresponding molecular-dynamics structures using ChimeraX 1.6.1 `molmap`, following the dataset-generation protocol of Jeon et al. (2024). For this dataset, we use the same image–reference pairs in both displayed evaluations and repeat the per-image result in the per-conformation table. This is our reporting convention, not a separate per-conformation evaluation; the repeated entry is excluded from per-conformation rankings.

D.3 HALF-MAP EVALUATION ON EXPERIMENTAL DATA

For homogeneous reconstruction, we preserve the inherited half-set assignments through particle selection (Appendix C.1). Each half-set provides the particle images for its own CTF-weighted backprojection, Gaussian-count estimation, and reconstruction. The two Gaussian density representations are randomly initialized and optimized separately, without sharing Gaussian parameters, optimizer states, or intermediate density maps (Scheres & Chen, 2012).

Gold-standard FSC uses Equation equation 11 to compare the independently reconstructed half-maps $u^{(1)}$ and $u^{(2)}$:

$$\mathrm{GSFSC}(\nu_j) = \mathrm{FSC}_{u^{(1)},u^{(2)}}(\nu_j). \tag{13}$$

The gold-standard designation refers to the independence of the two reconstructions (Scheres & Chen, 2012).

We compute FSC between the resulting half-maps using the Validation (FSC) job in CryoSPARC. The reported half-map resolution uses the FSC threshold of 0.143. When the FSC remains above this threshold up to the Nyquist frequency, we report the Nyquist limit of the particle-image sampling. Figure 3 and the Gaussian-count sweep in Figure 17(a) use the corrected FSC with an automatically generated tight mask and noise substitution. Appendix B.6 evaluates the same half-map pairs under different masking conditions.

For the simulated homogeneous reconstructions, particle images are generated in CryoSPARC from reference densities with CTF modulation and white noise at SNR 0.05. The FSC threshold of 0.5 is used for this map-to-map FSC comparison (Figure 7).

D.4 HETEROGENEOUS RECONSTRUCTION ON EXPERIMENTAL DATA

Assembly-state selection and comparison. We train CryoGaussian on the full EMPIAR-10076 particle stack for particle selection (Figure 20; Appendix B.4). Assembly-state comparisons use both the resulting CryoGaussian-filtered subset and the CryoDRGN-filtered subset. Published class assignments (Davis et al., 2016) identify the particle images associated with each reference state. We decode the observed latent nearest to the class-mean latent to obtain a representative density, following the selection procedure of Zhong et al. (2021). Map-to-map FSC compares these densities with the corresponding deposited reconstructions. The published assignments are used to select and evaluate representative densities after training. For the previously reported assembly intermediate with H68 density but no resolved central protuberance (Zhong et al., 2021), the CryoDRGN-filtered reconstruction uses a single representative latent: the observed particle latent nearest to the mean inferred latent of the 1,113 particles identified by the released CryoDRGN state indices. We decode this latent using the same representative-latent convention and density rendering pipeline as the other assembly classes. We additionally perform homogeneous refinement in CryoSPARC. Panel (a) of Figure 15 uses all 1,113 particles identified by the released CryoDRGN state indices; panel (b) uses their intersection with the 97,651-particle set retained by CryoGaussian filtering, yielding 1,019 particles. This intersection uses the full filtered set before the random 90% training subsampling. Gold-standard FSC assesses consistency between the independent CryoSPARC half-maps, while map-to-map FSC assesses agreement between the CryoGaussian and CryoSPARC densities.

Spliceosome latent selection. For the main EMPIAR-10180 reconstruction using the CryoDRGN-filtered subset, we apply principal component analysis to the inferred latents and decode the five positions marked along each of the first two principal components in Figure 5. Separately, we train CryoGaussian on the full particle stack (Figure 16), fit 20 k-means clusters to the inferred latents, and subdivide the largest cluster into two groups. Representative densities are decoded at the observed latents nearest to the cluster centers. This full-stack analysis does not determine the particle selection used for the main reconstruction.

E GAUSSIAN PARAMETERIZATION

E.1 BASE GAUSSIAN ATTRIBUTES

The base amplitude is parameterized as $p_g^{\text{base}} = \text{sigmoid}(\theta_g)$. Homogeneous reconstruction uses $p_g = p_g^{\text{base}}$ and $\boldsymbol{\mu}_g = \boldsymbol{\mu}_g^{\text{base}}$. Its covariance is

$$\boldsymbol{\Sigma}_g = Q_g \,\text{diag}(\sigma_{g1}^2, \sigma_{g2}^2, \sigma_{g3}^2) Q_g^{\mathsf{T}}, \qquad Q_g \in \text{SO}(3), \tag{14}$$

where each effective principal-axis scale satisfies the Gaussian FWHM constraint in Equation equation 5. We express scales in physical units throughout, using $s_{gk}^{\text{base}} > 0$ for a learned base scale before applying the constraint.

Gaussian positions are sampled from an isotropic Gaussian distribution centered in the reconstruction box. We initialize all three base scales to $2a$ and each rotation matrix Q_g to the identity matrix I_3.

E.2 GAUSSIAN ATTRIBUTE DECODERS

The displacement and amplitude decoders follow Equation equation 8. The scale decoder produces log-scale modulation $\Delta\boldsymbol{\ell}_g(\mathbf{z}) \in \mathbb{R}^3$ from $[\mathbf{f}_g; \mathbf{z}]$ and modulates each base scale separately:

$$\widetilde{\sigma}_{gk}(\mathbf{z}) = s_{gk}^{\text{base}} \exp[\Delta\ell_{gk}(\mathbf{z})], \qquad k = 1, 2, 3. \tag{15}$$

Applying the Gaussian FWHM constraint to each principal-axis scale gives the effective scales $\sigma_{gk}(\mathbf{z})$ used in Equation equation 9.

Decoder configuration. Each Gaussian attribute decoder takes $[\mathbf{f}_g; \mathbf{z}]$ as input and uses one hidden layer of width 32 with a ReLU activation. Latent dimensions are specified in Section 4. Separate decoders produce displacement, amplitude modulation, and scale modulation; all three are trained during heterogeneity estimation and per-state reconstruction. The displacement decoder predicts a three-dimensional displacement that is added directly to the base Gaussian position. Amplitude modulation follows the sigmoid parameterization in Equation equation 8.

E.3 GAUSSIAN RELOCATION AT A FIXED GAUSSIAN COUNT

Inspired by the relocation strategy of 3DGS-MCMC (Kheradmand et al., 2024), we redistribute existing Gaussians while retaining the predetermined Gaussian count G. Gaussians are selected for relocation when they have low base amplitudes. For each selected Gaussian, we sample a Gaussian from those not selected for relocation, with replacement and with probabilities proportional to their base amplitudes. The relocated Gaussian copies the base position, scales, and rotation of the sampled Gaussian. If Gaussian h is sampled m_h times, its base amplitude before relocation is divided equally among Gaussian h and the m_h relocated Gaussians, so that each receives $p_h^{\text{base}}/(m_h + 1)$.

F GAUSSIAN COUNT ESTIMATION

This appendix gives the theoretical basis of the spatial capacity estimate, the backprojection used to estimate spatial support, and the interpretation of the relative density threshold.

F.1 SPACE–BANDWIDTH BASIS

We derive the spatial capacity estimate in Equation equation 7 using the multidimensional space–frequency concentration problem (Slepian, 1964; Katsevich et al., 2015). Let $\Omega \subset \mathbb{R}^3$ be a measurable spatial region of finite volume and $B \subset \mathbb{R}^3$ a frequency region of finite volume. We use the Fourier convention

$$\widehat{f}(\boldsymbol{\xi}) = \int_{\mathbb{R}^3} f(\mathbf{x}) e^{-2\pi \mathrm{i} \boldsymbol{\xi}^{\mathsf{T}} \mathbf{x}} \, d\mathbf{x}. \tag{16}$$

The concentration operator on $L^2(\Omega)$ is

$$[T_{\Omega,B} f](\mathbf{x}) = \int_{\Omega} K_B(\mathbf{x}, \mathbf{y}) f(\mathbf{y}) \, d\mathbf{y}, \qquad K_B(\mathbf{x}, \mathbf{y}) = \int_B e^{2\pi \mathrm{i} \boldsymbol{\xi}^{\mathsf{T}} (\mathbf{x}-\mathbf{y})} \, d\boldsymbol{\xi}. \tag{17}$$

To justify the trace, define $\mathcal{A} : L^2(\Omega) \to L^2(B)$ by

$$[\mathcal{A} f](\boldsymbol{\xi}) = \int_{\Omega} f(\mathbf{x}) e^{-2\pi \mathrm{i} \boldsymbol{\xi}^{\mathsf{T}} \mathbf{x}} \, d\mathbf{x}, \qquad \boldsymbol{\xi} \in B. \tag{18}$$

Then $T_{\Omega,B} = \mathcal{A}^* \mathcal{A}$. Zero extension and Plancherel's theorem give $\|\mathcal{A}\| \leq 1$, while $\|\mathcal{A}\|_{\mathrm{HS}}^2 = \int_B \int_\Omega 1 \, d\mathbf{x} \, d\boldsymbol{\xi} = |\Omega||B| < \infty$. Thus $T_{\Omega,B}$ is positive and trace class, has eigenvalues $\lambda_j \in [0, 1]$, and its trace is

$$\mathcal{S}(\Omega, B) = \sum_j \lambda_j = \int_\Omega K_B(\mathbf{x}, \mathbf{x}) \, d\mathbf{x} \tag{19}$$

$$= \int_\Omega \int_B 1 \, d\boldsymbol{\xi} \, d\mathbf{x} = |\Omega| \, |B|. \tag{20}$$

The trace is the space–bandwidth product. It provides a nominal effective dimension for spatially and spectrally concentrated signals; its interpretation as a count of well-concentrated modes is approximate, whereas the trace identity is exact (Simons & Wang, 2011).

For the ball $B_d = \{\boldsymbol{\xi} : \|\boldsymbol{\xi}\|_2 \leq 1/d\}$,

$$|B_d| = \int_0^{1/d} 4\pi r^2 \, dr = \frac{4\pi}{3d^3}. \tag{21}$$

Treating the above-threshold voxels as a union of physical voxel cells gives $|\Omega_\tau| = V_\tau = n_\tau a^3$. At $d = 2a$,

$$\mathcal{S}_\tau = \mathcal{S}(\Omega_\tau, B_d) = \frac{4\pi}{3} \frac{V_\tau}{d^3} = \frac{4\pi}{3} \frac{n_\tau a^3}{(2a)^3} \tag{22}$$

$$= \frac{4\pi}{3} \frac{n_\tau a^3}{8a^3} = \frac{\pi}{6} n_\tau. \tag{23}$$

The factor $4\pi/3$ is the volume factor of the frequency-domain ball; the spatial support need not be spherical or connected. We obtain the Gaussian count $G = \mathcal{R}(\mathcal{S}_\tau)$ by rounding the spatial capacity estimate to an integer.

The concentration identity concerns a linear space of signals. Using its value as the Gaussian count in a nonlinear Gaussian representation gives a Gaussian-count estimate motivated by this theory, rather than a claim that individual Gaussians correspond one-to-one to concentration eigenfunctions. The relative density threshold τ specifies the estimated spatial support. The sampling bandwidth specifies the spatial scale at which capacity is measured. We use this estimate to set the Gaussian count; reconstruction accuracy is assessed experimentally. The Gaussian FWHM constraint uses the same spatial scale without assuming that Gaussian functions are strictly bandlimited.

F.2 BACKPROJECTION FOR SPATIAL SUPPORT ESTIMATION

We compute a CTF-weighted backprojection using regularized gridding in CryoDRGN (version 3.4.4) (Zhong et al., 2021). For heterogeneous reconstruction, the backprojection uses all particle images in the corresponding dataset; for homogeneous half-map reconstruction, it uses only the particle images in the respective half-set. Particle images are corrected for the supplied in-plane translations. Fourier downsampling, when used, preserves the physical field of view. After any downsampling, the pixel size a is also the voxel spacing used for spatial support estimation. The supplied orientations are used to backproject the particle images. The resulting density ρ_{BP} defines the estimated spatial support in Equation equation 6, assuming $\max_{\mathbf{x}\in\mathcal{X}} \rho_+(\mathbf{x}) > 0$. Only the Gaussian count derived from this support is used in subsequent reconstruction; Gaussian attributes are not initialized from the backprojected density.

F.3 INTERPRETATION OF THE RELATIVE DENSITY THRESHOLD

The relative density threshold τ defines the estimated spatial support Ω_τ used in the spatial capacity estimate; its numerical value is reported in Section 4. Thresholding relative to the positive maximum makes this estimate invariant to a positive multiplicative rescaling of the backprojected density. A lower threshold includes more weak-density regions in the estimated spatial support.

Support-volume interpretation. Let $\rho_{\max} = \max_{\mathbf{x}\in\mathcal{X}} \rho_+(\mathbf{x}) > 0$. The voxel set $\mathcal{X}_\tau$ admits an equivalent characterization as a maximizer of a discrete excess-mass criterion:

$$\mathcal{X}_\tau \in \arg\max_{\mathcal{U}\subseteq\mathcal{X}} \left[\sum_{\mathbf{x}\in\mathcal{U}} \frac{\rho_+(\mathbf{x})}{\rho_{\max}} - \tau|\mathcal{U}| \right]. \tag{24}$$

Each voxel contributes $\rho_+(\mathbf{x})/\rho_{\max} - \tau$; selecting all nonnegative contributions gives $\mathcal{X}_\tau$. This criterion balances retained peak-normalized density against support size. Since $V_\tau = a^3 n_\tau$ and $\mathcal{S}_\tau = (\pi/6)n_\tau$, the support-size penalty also controls the spatial capacity estimate used to set G.

Gaussian reference. A single three-dimensional Gaussian provides an analytical reference for $\tau = 0.02$. Its density superlevel sets are highest-density regions: each encloses a given probability mass within the minimum volume (Hyndman, 1996). Let $\varphi(\mathbf{r}) = \mathcal{N}_3(\mathbf{r}; \mathbf{0}, I_3)$ and let $F_{\chi^2_3}$ denote the cumulative distribution function of a chi-square random variable with three degrees of freedom. Then

$$\frac{\varphi(\mathbf{r})}{\varphi(\mathbf{0})} \geq \tau \quad \Longleftrightarrow \quad \|\mathbf{r}\|_2^2 \leq -2\ln\tau, \tag{25}$$

$$\int_{\varphi(\mathbf{r})\geq\tau\varphi(\mathbf{0})} \varphi(\mathbf{r})\, d\mathbf{r} = F_{\chi^2_3}(-2\ln\tau). \tag{26}$$

The 95-percent mass ellipsoid has relative boundary density $\exp[-F^{-1}_{\chi^2_3}(0.95)/2] \approx 0.0201$; the relevant chi-square quantile is approximately 7.815 (NIST/SEMATECH, n.d.). Thus, $\tau = 0.02$ approximates this relative boundary density and encloses approximately 0.9502 of the Gaussian mass. An invertible affine coordinate transformation gives the same relation for any nondegenerate three-dimensional Gaussian. This calculation supplies an analytical reference for the fixed threshold; the fraction of the integrated positive backprojected density enclosed by Ω_τ remains data dependent.

Dependence on the threshold. For a fixed backprojection and voxel grid, two thresholds $0 < \tau_1 < \tau_2 < 1$ satisfy

$$n_{\tau_1} - n_{\tau_2} = \left| \left\{ \mathbf{x} \in \mathcal{X} : \tau_1 \leq \frac{\rho_+(\mathbf{x})}{\rho_{\max}} < \tau_2 \right\} \right|. \tag{27}$$

The change in estimated spatial support is therefore determined exactly by the voxels whose normalized densities lie between the two thresholds. Its effect on the Gaussian count follows through $G = \mathcal{R}((\pi/6)n_\tau)$. The space–bandwidth identity gives $\mathcal{S}_\tau$ conditional on Ω_τ; it does not determine τ. In CryoGaussian, this support is used only to set the Gaussian count; the learned positions remain unconstrained by Ω_τ. Figure 17 evaluates the empirical effect of varying τ on heterogeneity estimation; the corresponding results are discussed in Appendix B.1.

G EFFICIENT CRYO-EM GAUSSIAN SPLATTING

This appendix describes the gradient reductions and the partitioning of Gaussian lists for image tiles underlying Section 3.4. We use the same Gaussian density, CTF modulation, and reconstruction objective as in Section 3.1. Let $\psi_{ig}(\mathbf{v})$ denote the contribution of Gaussian g to the predicted image $\widehat{Y}_i^{\mathrm{pre}}$ before CTF modulation: it is the gth summand in Equation equation 2, multiplied by κ in Equation equation 3. Rasterization groups the projected Gaussian footprints into image tiles. The following reductions refer to the Gaussian–pixel interactions selected by this assignment.

G.1 GRADIENTS THROUGH CTF MODULATION

The CTF-modulated prediction is related to the projection by a linear operator. Applying its adjoint to the image residual gives the gradient supplied to the rasterizer:

$$\delta_i \equiv \frac{\partial \mathcal{L}}{\partial \widehat{Y}_i^{\mathrm{pre}}} = \frac{2}{NHW} \mathcal{F}_2^{-1}\left[\overline{C_i}\, \mathcal{F}_2\left[\widehat{Y}_i - Y_i^{\mathrm{aligned}}\right]\right], \tag{28}$$

where the overline denotes complex conjugation; the CTFs used here are real-valued. The factor $2/(NHW)$ follows from Equation equation 4. CTF modulation couples spatial locations in the image domain, while the rasterizer receives δ_i after this operation has been differentiated. It can therefore accumulate gradient contributions over each projected Gaussian footprint using δ_i.

G.2 PER-GAUSSIAN GRADIENT ACCUMULATION

For differentiation, regard the effective Gaussian attributes for each particle image as separate intermediate variables, $\boldsymbol{\vartheta}_{ig} = (p_g(\mathbf{z}_i), \boldsymbol{\mu}_g(\mathbf{z}_i), \boldsymbol{\Sigma}_g(\mathbf{z}_i))$. These attributes are functions of Θ and, in heterogeneous reconstruction, $\mathbf{z}_i$, as defined in Section 3.1. Let $\mathcal{V}_{ig}$ be the image-grid locations assigned to Gaussian g by the rasterizer. With these interactions held fixed, the gradient reduction is

$$\nabla_{\boldsymbol{\vartheta}_{ig}} \mathcal{L} = \sum_{\mathbf{v} \in \mathcal{V}_{ig}} \delta_i(\mathbf{v})\, \nabla_{\boldsymbol{\vartheta}_{ig}} \psi_{ig}(\mathbf{v}). \tag{29}$$

A pixel-wise scatter computes contributions at each pixel and uses atomic updates to accumulate them into Gaussian gradients. Our gather implementation assigns one thread block to each Gaussian for each particle image. Threads in that block accumulate contributions to the projected Gaussian attributes and reduce them before writing the result, avoiding concurrent writes from different blocks to that Gaussian's gradient in this reduction. The projection derivatives then propagate these gradients to the three-dimensional Gaussian attributes in Equation equation 29.

For attributes shared across particle images, the resulting gradients are summed across those images. Gradients for attributes decoded separately for each particle image remain separate until they are propagated through the Gaussian attribute decoders to the trainable representation parameters. The same per-Gaussian gradient accumulation supports homogeneous and heterogeneous reconstruction.

G.3 PARTITIONED ADDITIVE PROJECTION

Let $\mathcal{G}_i(\mathcal{T})$ be the Gaussian list assigned to image tile $\mathcal{T}$ for particle image i. A long list can be divided into disjoint consecutive sublists $\mathcal{G}_i^{(q)}(\mathcal{T})$ whose union is $\mathcal{G}_i(\mathcal{T})$. For each pixel location $\mathbf{v}$ in that tile, linearity gives

$$\sum_{g \in \mathcal{G}_i(\mathcal{T})} \psi_{ig}(\mathbf{v}) = \sum_q \sum_{g \in \mathcal{G}_i^{(q)}(\mathcal{T})} \psi_{ig}(\mathbf{v}). \tag{30}$$

Each inner sum forms a partial projection that can be evaluated independently; the partial projections are then accumulated. This exposes additional parallelism when a small number of tiles contain long Gaussian lists. Batching particle images also supplies parallel work and amortizes preprocessing and scheduling overhead. We retain batched projection when the batch supplies sufficient parallelism and partition long tile lists when parallelism across particle images is limited.

The partitioning of Gaussian lists for image tiles and the gradient reductions above retain the same Gaussian–pixel interactions.

H EXTENDED RELATED WORK AND FUTURE DIRECTIONS

H.1 RECONSTRUCTION METHODS IN OUR COMPARISONS

Homogeneous reconstruction. Our comparisons span voxel-based, implicit neural, and explicit Gaussian representations. CryoSPARC reconstructs voxelized maps, with non-uniform refinement adapting regularization to spatially varying disorder through cross-validation (Punjani et al., 2017; 2020). The homogeneous CryoDRGN model uses a coordinate-based network in Fourier space (Zhong et al., 2021), while CryoNERF uses a real-space neural field with multiresolution hash encoding (Qu et al., 2025). CryoSplat develops normalized orthographic Gaussian projection with CTF modulation and CUDA rasterization for homogeneous reconstruction from randomly initialized anisotropic Gaussians (Chen & Ling, 2026). CryoGaussian emphasizes data-derived control of representation detail and capacity: image sampling constrains Gaussian widths, while sampling and estimated spatial support determine the Gaussian count. Relocation inspired by 3DGS-MCMC redistributes Gaussians during optimization while preserving this count (Kheradmand et al., 2024). Our CUDA rasterizer organizes gradient accumulation per Gaussian and schedules additive projection for batched and distributed training, supporting homogeneous and heterogeneous reconstruction within one framework (Appendix G). Experimental configurations and score sources are detailed in Appendix C.4.

Heterogeneous reconstruction. Our CryoBench comparisons use the same supplied ground-truth poses as fixed inputs (Jeon et al., 2024). The compared methods represent structural variability through discrete states, continuous subspaces, deformation fields, or neural representations. 3D classification estimates a finite set of volumes (Punjani et al., 2017; Jeon et al., 2024), whereas 3DVA learns a linear subspace (Punjani & Fleet, 2021). RECOVAR combines regularized covariance estimation with adaptive kernel regression for per-state reconstruction (Gilles & Singer, 2025); SOLVAR directly optimizes a low-rank covariance representation and uses RECOVAR's reconstruction procedure with its estimated principal components (Yadgar et al., 2026). 3DFlex learns latent-conditioned deformation fields that transport a canonical density (Punjani & Fleet, 2023). CryoDRGN learns image embeddings with an encoder and a latent-conditioned neural field (Zhong et al., 2021), while CryoDRGN-AI-fixed directly optimizes particle-image latents (Levy et al., 2025; Jeon et al., 2024). OPUS-DSD uses a convolutional encoder–decoder with a structural disentanglement prior (Luo et al., 2023). CryoHype uses a transformer hypernetwork to condition implicit neural representation weights on each particle image (Gu et al., 2026), and CryoKRAQEN combines quantized embeddings, kernel-guided assignment, and a triplane representation (Gao et al., 2026). Among Gaussian approaches, CryoSPIRE jointly optimizes particle-image latents and Gaussian features in a model initialized from a reconstructed density. Feature clustering identifies structural parts and anchors for a hierarchical representation of part-level motion, local deformation, and occupancy variation (Shekarforoush et al., 2025). CryoGaussian instead models compositional and conformational heterogeneity directly through latent-conditioned Gaussian attributes, starting from randomly initialized Gaussians. Particle-image and per-Gaussian latents jointly condition this representation; the inferred particle-image latents are then held fixed while a newly initialized Gaussian representation is fitted for per-state reconstruction.

H.2 FUTURE INTEGRATION WITH AGENTIC STRUCTURAL WORKFLOWS

CryoGaussian could serve as a reconstruction component in agentic workflows that repeatedly select particles, reconstruct structural states, and evaluate their agreement with experimental observations. Scientific agents such as Coscientist demonstrate how language models can coordinate specialized tools and use feedback to plan experiments (Boiko et al., 2023). In cryo-EM, such a workflow could invoke CryoGaussian to compare alternative particle selections and reconstruct candidate structural subpopulations. Accelerating projection and gradient computation with our CUDA rasterizer and distributed training could reduce the cost of these repeated analyses, allowing more structural hypotheses to be examined within a computational budget. Where local map quality supports atomic interpretation, reconstructed states could be passed to automated model-building tools such as ModelAngelo (Jamali et al., 2024), connecting heterogeneity analysis to molecular models for downstream study.

A further direction is to connect experimentally resolved structural states with molecular design. ProteinZero improves inverse-folding models through online reinforcement learning with computational feedback (Wang et al., 2026), while PocketXMol uses atomic prompts to unify structure prediction and the design of pocket-interacting small molecules and peptides (Peng et al., 2026). These capabilities motivate workflows in which reconstructed conformations inform structural targets or binding environments, and subsequent experiments test the resulting designs. CryoGaussian could provide the reconstruction component as new particle images become available, enabling comparisons across experimental conditions and structural states. Evaluating this integration would establish whether accelerated reconstruction can make iterative structure-guided design more practical.