CryoDiff: Cryo-EM Synthesis via Biophysics and Cycle-Consistent Diffusion

Genpei Zhang[1,*] Yuntian Yang[2,*] Siqi Wu[1] Ningyan Zhang[1] Seonghui Min[1]
Jie Wu[1] Christopher Braxton Owens[3] Minhao Wu[4] Wanyue Feng[1]
Gus LW Hart[3] Runmin Jiang[1,†] Min Xu[1,†]
[1]Carnegie Mellon University [2]Harvard University
[3]Brigham Young University [4]University of California San Diego
[*]Equal contribution. [†]Corresponding authors.

Abstract

Cryo-electron microscopy (cryo-EM) is a premier technique for near-atomic resolution imaging of macromolecules, yet the development of robust machine learning models for downstream analysis is severely bottlenecked by the scarcity of high-quality, annotated datasets. While synthetic data generation offers a promising alternative, existing simulators struggle to concurrently model the vast structural diversity of biological specimens and the complex, highly structured noise inherent to realistic cryo-EM environments. To bridge this synthetic-to-real gap, we propose ***CryoDiff****, a novel synthesis framework that seamlessly integrates biophysical modeling with conditional cycle-consistent diffusion. Specifically, our biophysical engine first generates noise-free, multi-scale micrographs that capture authentic compositional heterogeneity, cellular context, and physical imaging properties. To faithfully translate these structures into realistic noisy domains without suffering from structural hallucinations, CryoDiff employs a conditional diffusion model. This generation process is regularized by a cycle-consistency constraint to strictly preserve structural fidelity, and is further refined by mask-guided contrastive learning to accurately capture fine-grained, spatially adaptive noise patterns. Extensive evaluations demonstrate that CryoDiff produces highly authentic micrographs that substantially boost performance in critical downstream tasks, outperforming state-of-the-art baselines in both particle picking and pose estimation. We will release our codebase as a flexible resource for generating customized, task-specific datasets, aiming to accelerate algorithm development and model training within the cryo-EM community.*

1. Introduction

In recent years, single-particle cryo-electron microscopy (cryo-EM) has revolutionized structural biology, enabling near-atomic resolution reconstructions of macromolecules in their native states [7, 30]. By vitrifying biological specimens and imaging them with high-energy electron beams, cryo-EM has significantly deepened our understanding of complex protein architectures and molecular mechanisms [40, 48]. However, realizing its full potential remains bottlenecked by several data-centric challenges—namely, the scarcity of diverse datasets, inherently low signal-to-noise ratios (SNR), and the heavy reliance on labor-intensive manual annotations [18, 45]. Acquiring a sufficient number of high-quality micrographs often requires days of expensive instrument time, while expertly curating these datasets can take weeks. These limitations collectively impede the training and generalization of robust models for downstream tasks, including particle picking [3], pose estimation [32], and 3D reconstruction [73, 74].

To fuel the development of learning-based cryo-EM algorithms [16, 63, 73, 77], recent efforts have pivoted towards synthetic data generation via biophysically inspired modeling [11, 55, 60]. Frameworks like InsilicoTEM [60] and LBPN [27] simulate high-fidelity multi-angle projections to approximate realistic imaging conditions, though they incur prohibitive computational costs. Alternatively, VirtualIce [47] improves efficiency by simulating particle behaviors—such as aggregation and preferred orientations—on top of real vitrified backgrounds. Despite these algorithmic advances, existing simulators face two critical hurdles: limited support for structural diversity [37, 73], and an oversimplified assumption of additive Gaussian noise.

In practice, real cryo-EM micrographs exhibit a complex, structured noise distribution comprising detector artifacts, electron scattering, radiation damage, and heterogeneous background signals [33, 50]. While generative approaches like CryoGEM [68] attempt to model this complexity using a physics-informed GAN framework, their lack of bidirectional constraints [75] often leads to mode collapse, structural distortion, and limited controllability over the noise synthesis process. While diffusion models have recently demonstrated unprecedented success in natural image synthesis, their direct application to cryo-EM remains challenging due to the domain's extreme noise levels and unique physical constraints.

To bridge this synthetic-to-real gap, we propose **CryoDiff**, a novel framework integrating biophysical modeling with conditional cycle-consistent diffusion for realistic cryo-EM synthesis. Our method employs a modular biophysical engine to generate structurally diverse components within authentic cellular contexts, capturing true compositional heterogeneity. To translate these noise-free synthetic structures into realistic micrographs, CryoDiff pioneers a conditional diffusion approach tailored for cryo-EM data. Compared to GAN-based alternatives [17, 68], our diffusion formulation guarantees stable, high-fidelity generation without mode collapse. By explicitly conditioning the reverse diffusion process on structural masks, our method ensures that the synthesized noise features remain strictly aligned with the underlying biological macromolecules. Furthermore, the cycle-consistency constraint acts as a robust safeguard against structural hallucinations, a common pitfall in standard generative models. We further introduce mask-guided contrastive learning to enhance the representation of fine-grained features—such as edges, internal textures, and spatial noise patterns—effectively preventing background conflation.

Extensive evaluations across multiple structural classes demonstrate that CryoDiff achieves substantial gains in realism, controllability, and task-specific performance. When utilized for downstream training, our generated data consistently improves particle picking accuracy (AUPRC) and enhances the resolution of pose estimation. Our primary contributions are summarized as follows:

- **Structural Diversity & Control:** We develop a versatile biophysical engine that simulates cryo-EM micrographs with complex compositional heterogeneity, spatial context, and physics-informed imaging properties.
- **Realistic Noise Generation:** We propose a conditional cycle-consistent diffusion model augmented with mask-guided contrastive learning, surpassing conventional noise models and GANs in capturing authentic noise characteristics while preserving structural fidelity.
- **Downstream Utility:** We validate CryoDiff on practical tasks, demonstrating that our synthesized datasets establish new state-of-the-art performance in particle picking and pose estimation baselines.

2. Related Work

Cryo-EM/ET Synthesis. The development of cryo-EM/ET data synthesis has evolved from early physics-based models of electron scattering [10, 60] to more realistic frameworks that account for imaging noise and sample heterogeneity [13, 19, 24, 70, 72]. To balance realism and controllability, hybrid simulators like VirtualIce [47] and cryo-TomoSim [52] generate annotation-ready data with physical priors. More recently, specialized approaches have been developed to handle cryo-ET challenges, such as PolNet [43] for cellular variability and LBPN [27] for the missing wedge problem. However, most methods simulate noise by adding Gaussian perturbations, which fail to capture the complex noise patterns observed in real micrographs. To bridge this gap, CryoETGAN [64] introduces cycle-consistent unpaired translation, while FakET [17] leverages neural style transfer for efficient synthesis. CryoGEM [68] further advances realism by integrating physics-based simulation with mask-guided contrastive learning. In this work, we aim to improve cryo-EM synthesis by introducing biophysical modeling for structural diversity and diffusion models for realistic noise.

Unpaired Image-to-Image Translation. Realistic noise generation in cryo-EM can be cast as an unpaired image-to-image translation task, where clean simulated micrographs are mapped to realistic noisy ones without paired supervision. In vision, this has been widely studied, from early GAN-based frameworks [8, 31, 54, 67, 76] to variants with shared latent spaces [36], attention [28], and contrastive learning [25, 49, 61, 71]. Extending these ideas, GAN-based models such as CryoETGAN [64] and CryoGEM [68] have been applied to cryo-EM/ET, though often limited by instability and mode collapse. In contrast, our method enables stable and structurally faithful noise generation.

Diffusion Models. Diffusion models have emerged as a powerful generative framework, with DDPMs defining a forward noise process and learning its reversal [21, 58]. Early efforts focused on sampling efficiency and image quality [12, 46, 53], while recent advances enable conditional generation via classifier guidance and context-aware conditioning [20, 66, 69]. For unpaired image-to-image translation, diffusion models offer greater stability and diversity than GANs. UNIT-DDPM [56] pioneered this direction but lacked structural consistency. CycleDiffusion [62] introduced a unified latent space to enforce cycle consistency, improving fidelity. Despite these advances, diffusion models remain underexplored in cryo-EM synthesis. To fill this gap, we introduce the first conditional, cycle-consistent diffusion framework tailored for this domain.

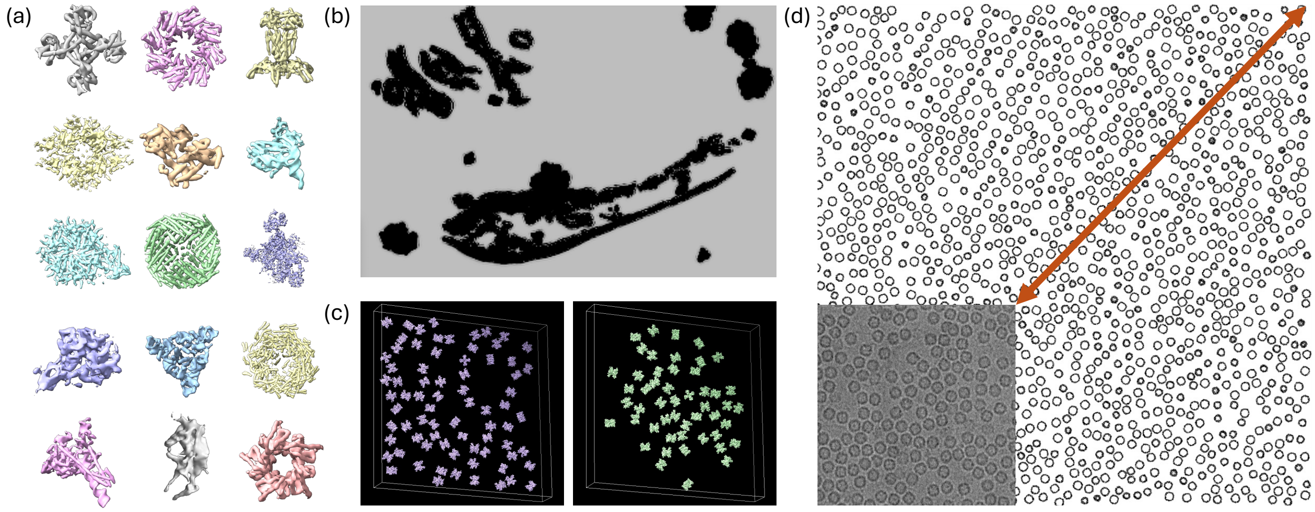

Figure 1. CryoEM simulation results: (a) Generation of multi-scale synthetic data from real CryoEM images according to EMPIAR-10421. (b) Different placement strategies are adopted based on the particles' properties. (c) Simulated images of mitochondria and part of the endoplasmic reticulum. (d) Visualization of the processed particles. (e) Visualization of simulated cryo-EM images.

3. Biophysical Modeling for Diverse Structural Simulation

We introduced a unified biophysical modeling framework that leverages both experimental and generated structures, context-aware organelle embedding, and scale-adaptive placement to synthesize high-fidelity, biologically grounded cryo–EM micrographs. By integrating data-driven and probabilistic placement strategies with realistic imaging physics, our pipeline captures the compositional and spatial heterogeneity of cellular specimens.

Library Construction. To enrich the structural basis for simulation, we build a model library that supports heterogeneity. We retrieved over 10,000 experimentally resolved PDB entries spanning molecular weights from small enzymes (50 kDa) to large viral capsids (¿50,000 kDa), and symmetry classes from high-order tetrahedral assemblies to asymmetric C_1 complexes. These entries contribute well-characterized, functionally diverse macromolecular structures that form the backbone of the simulation.

All models are then standardized through Gaussian smoothing and isosurface extraction, followed by triangulated mesh conversion with care taken to preserve fine-scale features such as active-site geometry. The processed particles are shown in Figure 1(a).

Particle Placement. To achieve biologically plausible positioning and orientation of macromolecules, we combine empirical annotations with probabilistic sampling under controllable fidelity-diversity tradeoffs. Particles are positioned using dual approaches that integrate experimental measurements with synthetic sampling, feeding into the subsequent class-specific distribution modeling:

- **Data-Driven Placement:** We convert RELION-derived 2D picks (x, y) into 3D translations T_{exp} by mapping pixel positions into volume coordinates, and transform Euler angles (α, β, γ) into quaternions q_{exp} via standard attitude conversion conventions.
- **Synthetic Insertion:** To augment sampling beyond mapped particles, we draw translations T_{syn} from density-matched or uniform distributions calibrated to empirical radial functions, and sample orientations q_{syn} uniformly on S^3. Confidence-weighted blending of $(T_{\text{exp}}, q_{\text{exp}})$ and $(T_{\text{syn}}, q_{\text{syn}})$ ensures both fidelity and diversity. These placement pools then interface with the Class-Specific Distribution Modeling module to enforce biologically informed spatial patterns for each particle category.

Class-Specific Distribution Modeling. To recapitulate distinct spatial arrangements of molecular subtypes, we apply particle-type–specific priors inferred from experimental distributions (Figure 1(c)). A dedicated module applies distribution rules extracted from experimental data: soluble enzymes disperse uniformly in the volume; ribosomes cluster according to measured inter-particle distances to mimic polysomes; organelle-associated proteins remain confined inside segmented organelle meshes; and viral capsids follow separation distributions observed on Cryo–EM grids to avoid clashes.

Multi-Scale Volume Modeling. To enable simulation across multiple scales, our simulator adjusts mesh complexity and placement parameters across scales (Figure 1(d)). At larger scales, simplified meshes and broader sampling capture global spatial patterns. While at finer scales, detailed meshes and denser sampling with stricter collision checks preserve sub-nanometer features. This scale-adaptive modeling ensures consistent biological realism from whole-cell panoramas down to molecular interfaces.

Ice-Layer Modeling. To reflect imaging conditions shaped by vitrified ice, we simulate an ice layer with realistic topography and density fluctuations. We generate a vitreous ice slab with thickness drawn from a log-normal distribution and modulate its surface via Perlin-noise to introduce realistic thickness variations. The ice density field incorporates Gaussian fluctuations to simulate beam-induced noise before particle embedding.

4. Conditional Cycle-consistent Diffusion for Realistic Noise Translation

4.1. Overview

In this section, we propose a novel framework for realistic noise generation in Cryo-EM image synthesis that integrates four key components: diffusion models, conditional control, cycle consistency, and mask-guided contrastive learning. The procedure is outlined in Algorithm 1. Figure 2 illustrates the CryoDiff pipeline (we take the synthetic-to-real-to-real stage as an example). The method and corresponding motivations are summarized as follows:

- We first introduce diffusion models for stable cryo-EM noise generation, ensuring consistent and high-quality image synthesis.
- Conditional control through segmentation masks enhances scalability, guiding the network to focus on key structural features.
- A cycle-consistent learning mechanism preserves structural integrity, maintaining alignment of positional information across domains.
- Mask-guided contrastive learning refines the model, enhancing the representation of fine-grained features such as edges, textures, and spatial noise patterns.

4.2. Conditional Cycle-consistent Diffusion

To generate realistic micrographs aligned with the physical properties of cryo-EM data, we employ a conditional DDPM [21] which accelerated by DPM-Solver [38] and guided by mask-based conditioning.

We define two generative diffusion models: G_{AB}, which translates synthetic micrographs into the realistic cryo-EM domain, and G_{BA}, which reconstructs synthetic-like images from real micrographs. This bidirectional mapping enables unpaired training while preserving structural fidelity.

Algorithm 1 Training procedure for CryoDiff

```
Input: Synthetic images $\mathbf{x} \in \mathcal{A}$, real images $\mathbf{y} \in \mathcal{B}$, masks $\mathbf{m}$
Output: Trained $G_{AB}$, $G_{BA}$, $D_A$, $D_B$
for each iteration do
    Sample $\mathbf{x} \sim \mathcal{A}$, $\mathbf{y} \sim \mathcal{B}$, $\mathbf{m}$
    Compute noisy $\mathbf{x}_t$, $\mathbf{y}_t$ via forward diffusion
    Generate $\hat{\mathbf{y}} = G_{AB}(\mathbf{x}_t, \mathbf{m})$, $\hat{\mathbf{x}} = G_{BA}(\mathbf{y}_t)$
    Compute cycle: $\mathbf{x}_{\text{cyc}} = G_{BA}(\hat{\mathbf{y}})$, $\mathbf{y}_{\text{cyc}} = G_{AB}(\hat{\mathbf{x}}, \mathbf{m})$
    Compute losses: $\mathcal{L}_{\text{diff}}$, $\mathcal{L}_{\text{GAN}}$, $\mathcal{L}_{\text{cyc}}$, $\mathcal{L}_{\text{NCE}}$
    Update models
end for
```

The forward diffusion process is:

$$q(\mathbf{x}_t \mid \mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t; \sqrt{1-\beta_t}\mathbf{x}_{t-1}, \beta_t \mathbf{I}),$$

and the reverse process is:

$$p_\theta(\mathbf{x}_{t-1} \mid \mathbf{x}_t, \mathbf{m}) = \mathcal{N}(\mathbf{x}_{t-1}; \mu_\theta(\mathbf{x}_t, t, \mathbf{m}), \sigma_t^2 \mathbf{I}),$$

with

$$\mu_\theta(\mathbf{x}_t, t, \mathbf{m}) = \frac{1}{\sqrt{\alpha_t}} \left(\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}} \epsilon_\theta(\mathbf{x}_t, t, \mathbf{m}) \right).$$

Note that the mask conditioning $\mathbf{m}$ is strictly applied to the synthetic-to-real translation G_{AB}. The real-to-synthetic mapping G_{BA} operates unconditionally, where the reverse process simplifies to $p_\theta(\mathbf{x}_{t-1} \mid \mathbf{x}_t)$.

The denoising objective is:

$$\mathcal{L}_{\text{diff}} = \mathbb{E}_{t, \mathbf{x}_0, \epsilon, \mathbf{m}} \left[\| \epsilon - \epsilon_\theta(\mathbf{x}_t, t, \mathbf{m}) \|^2 \right].$$

The segmentation mask $\mathbf{m}$ guides the model to attend to particle regions, enhancing structural preservation. Sampling is accelerated using DPM-Solver:

$$\mathbf{x}_{t-1} = \mathbf{x}_t - \eta \nabla_{\mathbf{x}_t} \log p_\theta(\mathbf{x}_t \mid \mathbf{m}).$$

To ensure consistency across domains without paired supervision, we impose a cycle-consistency constraint that penalizes structural drift between the original and round-trip reconstructed images:

$$\begin{aligned} \mathcal{L}_{\text{cyc}} = {} & \mathbb{E}_{\mathbf{x} \sim \mathcal{A}} \left[\| G_{BA}(G_{AB}(\mathbf{x}, \mathbf{m})) - \mathbf{x} \|_1 \right] \\ & + \mathbb{E}_{\mathbf{y} \sim \mathcal{B}} \left[\| G_{AB}(G_{BA}(\mathbf{y}), \mathbf{m}) - \mathbf{y} \|_1 \right]. \end{aligned}$$

4.3. Mask-Guided Contrastive Learning

To enhance structural discrimination under high noise, we introduce a mask-guided contrastive learning strategy. Unlike standard contrastive sampling that may conflate particles and background, we leverage segmentation masks to

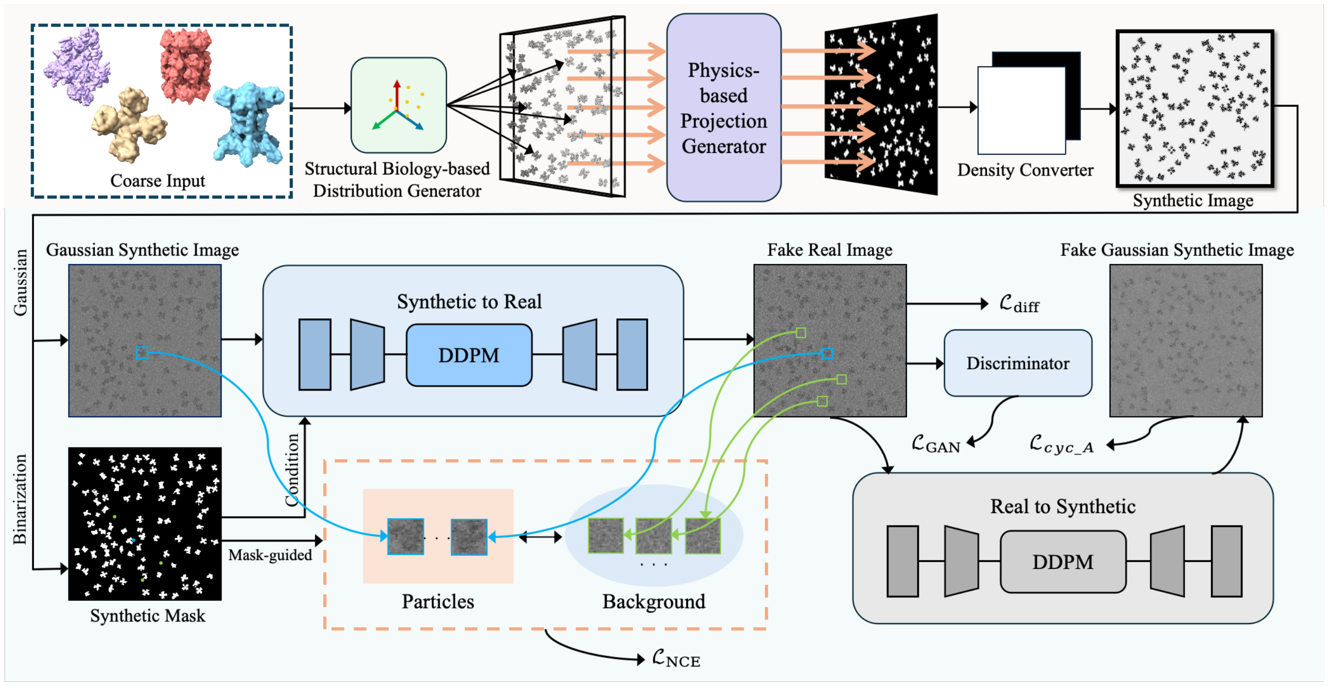

Figure 2. CryoDiff pipeline: (1) The input structures are inserted into a volume through specific placement strategies, which then undergoes physics-based projection and density conversion to generate synthetic images. (2) In the image translation, we use two ddpms, mask-guided contrastive learning, and discriminator to achieve realistic synthetic-to-real image translation.

explicitly define semantically meaningful positives and negatives: particle regions serve as anchors and positives, while background regions are treated as negatives.

The contrastive loss is defined as:

$$\mathcal{L}_{\text{NCE}} = -\log \frac{\exp(\text{sim}(q, k^{+})/\tau)}{Z},$$

$$\text{where} \quad Z = \exp(\text{sim}(q, k^{+})/\tau) + \sum_{k^{-}} \exp(\text{sim}(q, k^{-})/\tau).$$

where q is the query feature, k^{+} is its corresponding positive, and k^{-} are sampled negatives. $\text{sim}(\cdot,\cdot)$ denotes cosine similarity, and τ is a temperature parameter.

To further improve local realism, we incorporate a PatchGAN discriminator D_B, trained with the following adversarial loss:

$$\begin{aligned}\mathcal{L}_{\text{GAN}} = {} & \mathbb{E}_{\mathbf{y}\sim\mathcal{B}}[\log D_B(\mathbf{y})] \\ & + \mathbb{E}_{\mathbf{x}\sim\mathcal{A}}[\log(1 - D_B(G_{AB}(\mathbf{x}, \mathbf{m})))].\end{aligned}$$

We also apply physics-informed preprocessing to synthetic inputs, including weight-map normalization $\tilde{\mathbf{x}} = \text{IN}(w \odot \mathbf{x})$ and CTF-based noise simulation, to better match the statistical characteristics of real cryo-EM images.

The overall training objective combines all components:

$$\mathcal{L} = \mathcal{L}_{\text{diff}} + \lambda_{\text{GAN}}\mathcal{L}_{\text{GAN}} + \lambda_{\text{cyc}}\mathcal{L}_{\text{cyc}} + \lambda_{\text{NCE}}\mathcal{L}_{\text{NCE}},$$

where λ are empirically tuned weights.

5. Experiments

Datasets. We trained CryoDiff using synthetic particles derived from 15 publicly available EMPIAR datasets [23]. For visual quality evaluation purposes, we selected *six* datasets: 1) **TRPV1** [34], a tetrameric membrane channel resolved at 3.4 Å; 2) β**-galactosidase** [2], a 2.2 Å map with ~800 solvent peaks; 3) **Rhino/enterovirus** [1], small icosahedral virions requiring symmetry-expansion and high-fidelity difference mapping to resolve the conserved VP1-VP3 pocket; 4) **Innexin-6** [5], nanodisc-embedded hemichannels where lipid-induced pore closure demands focused classification and membrane-signal subtraction; 5) **MLA complex** [42], apo/ATP/ADP states that necessitate heterogeneous 3-D clustering and multi-body refinement to trace lipid transport; and 6) **GroEL** [14], a D7-symmetry chaperonin. Comprehensive descriptions of all datasets are provided in Appendix A.1.

Baselines. We evaluate the performance of our method compared to several traditional noise baselines and deep generative models. We use **Poisson** noise, **Gaussian** noise, and Poisson-Gaussian mixed noise (**Poi-Gau**) as traditional baselines. We choose **CryoGEM** [68], and **CycleDiffusion** [62] as deep generative baselines. In Appendix A.2, we detail their specific settings.

Implementation Details. All experiments are conducted on a server at Brigham Young University equipped with an NVIDIA GeForce RTX A100 GPU (80 GB of mem-

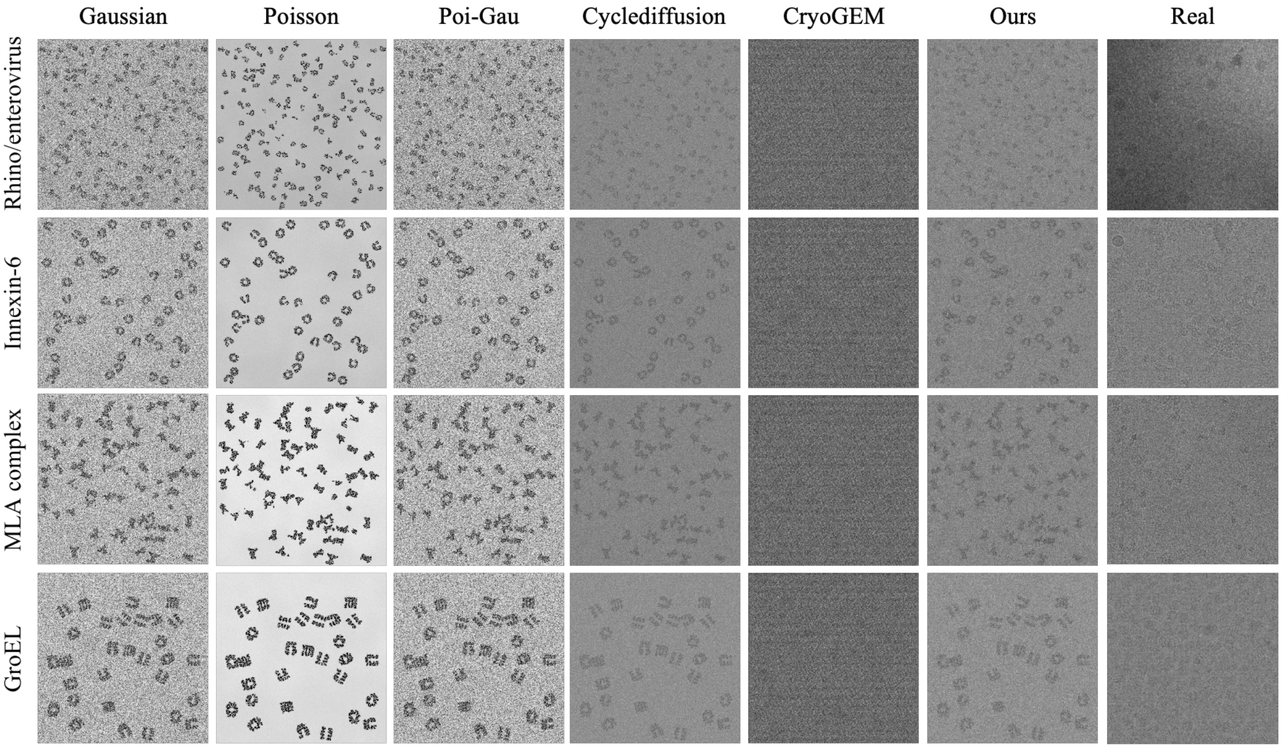

Figure 3. Comparison between real images and generated realistic micrographs. Our method produces more authentic micrographs with superior noise characteristics across all datasets.

ory). The models are trained for 50 epochs using the Adam optimizer with a learning rate of 1×10^{-4} and a batch size of 16. The training dataset, further detailed in Section A.1, consists of 500 synthetic and 500 real images, with all images standardized to a resolution of 1024×1024 pixels.

Our synthetic data generation process begins with biophysical modeling to simulate noise-free cryo-EM micrographs. Given a real cryo-EM image, the corresponding simulated noise-free counterpart, and a structural mask, the CryoDiff framework performs inference to generate realistic noisy images that closely mimic experimental conditions. To explore the effects of different generative sampling strategies, we evaluate four diffusion samplers within the CryoDiff framework: DDIM [58], DDPM [21], and their accelerated variants, DPM-Solver [38] and DPM-Solver++ [39]. These sampling methods are examined to understand the trade-off between synthesis quality and inference speed under various configurations.

Finally, we validate the utility of our generated data in two downstream tasks: particle picking (Section B.1) and pose estimation (Section B.2), which demonstrates the effectiveness and generalization of the proposed framework.

5.1. Visual Quality

We compared CryoDiff against various baselines through visual assessment of generated images of evaluation datasets. Figure 3 illustrates that our method produces significantly more authentic micrographs with superior noise characteristics while preserving structural details such as particle boundaries, internal textures, and contrast between particles and background. Conventional approaches (Gaussian, Poisson, and Poisson-Gaussian mixed noise models) consistently failed to capture the complex noise patterns inherent in real cryo-EM data, often resulting in unnatural global intensity distributions and oversmoothed textures. Among deep generative models, CycleDiffusion often produced images where the particle boundaries were blurred and indistinguishable from the background. CryoGEM tended to collapse when handling our composite synthetic datasets, failing to generate recognizable particles. Quantitatively, we employed Frechet Inception Distance (**FID**) to measure generation quality. Table 1 confirms that our method achieves substantially lower FID scores across all tested datasets. Specifically, CryoDiff achieves an average FID of 146.65, demonstrating an improvement of 34% compared to the strongest baseline, CycleDiffusion (222.12). This performance leap is particularly evident on structurally heterogeneous datasets like the MLA complex, where our method reaches a remarkably low FID of 100.63, nearly halving the error of competing deep generative models.

Table 1. Quantitative comparison of FID scores of different noise methods on six Cryo-EM datasets (lower is better).

Dataset	**TRPV1**	**β-galactosidase**	**Rhino/enterovirus**	**Innexin-6**	**MLA complex**	**GroEL**	**Avg.**
Gaussian	287.04	351.27	292.56	328.17	225.20	364.28	308.09
Poisson	396.59	515.91	395.59	467.19	388.26	410.38	428.99
Poi–Gau	290.68	356.96	301.77	272.24	239.43	369.95	305.17
CycleDiffusion [62]	199.47	247.09	210.22	208.23	197.15	270.58	222.12
CryoGEM [68]	192.28	283.28	224.18	223.02	189.46	279.20	231.90
Ours	**147.75**	**150.26**	**155.37**	**148.45**	**100.63**	**177.44**	**146.65**

Table 2. **Quantitative comparison of particle picking.** Our approach consistently achieves the best in AUPRC and Precision metrics.

Metric	**AUPRC↑**					**Precision↑**				
Dataset	Proteasome	Integrin	PhageMS2	HumanBAF	Avg.	Proteasome	Integrin	PhageMS2	HumanBAF	Avg.
Gaussian	0.463	0.233	0.575	0.449	0.430	0.412	0.204	0.538	0.397	0.388
Poisson	0.458	0.195	0.408	0.392	0.363	0.386	0.168	0.365	0.341	0.315
Poi–Gau	0.438	0.244	0.601	0.376	0.415	0.397	0.227	0.519	0.347	0.373
CycleDiffusion [62]	0.357	0.233	0.420	0.369	0.345	0.321	0.196	0.347	0.326	0.298
CryoGEM [68]	0.257	0.208	0.159	0.178	0.201	0.221	0.175	0.148	0.159	0.176
Topaz	0.301	0.510	0.317	0.487	0.404	0.279	0.452	0.286	0.462	0.370
Ours	**0.471**	**0.518**	**0.712**	**0.521**	**0.555**	**0.423**	**0.475**	**0.612**	**0.531**	**0.510**

5.2. Particle Picking

To validate the realism of the generated noise characteristics, we evaluated its impact on particle picking using several publicly available EMPIAR datasets with annotations. Realistic micrographs generated by CryoDiff and multiple baselines were incorporated into the Topaz [3] particle-picking pipeline. Picked particles were subsequently subjected to a standard single-particle reconstruction workflow using cryoSPARC [51].

Table 2 summarizes both the Area Under the Precision-Recall Curve (AUPRC) and overall precision for each method (detailed metric definitions are provided in Appendix B). Overall, CryoDiff exhibits consistent improvements over classical noise models (Gaussian, Poisson, Hybrid) as well as recent learning-based approaches. These gains manifest across both high-contrast specimens (Proteasome, PhaseMS2) and more challenging low-contrast cases (Integrin, HumanBAF), suggesting that our model balances sensitivity and specificity more effectively.

By selecting a fixed confidence cutoff, we observe a uniform reduction in false positives. In terms of AUPRC, our method achieves the best results on all four datasets. While traditional models fail to register meaningful precision at the fixed cutoff, our method maintains valid detection outputs and achieves superior localization accuracy, particularly on Integrin and HumanBAF where noise and low contrast typically pose significant challenges. To further assess the structural integrity recovered from the picked particles, we calculate the Fourier shell correlation (FSC) to determine the resolution of the reconstructions. As illustrated in Figure 4, the filtered back-projection (FBP) reconstructions utilizing particles identified through our method consistently reach higher resolutions at the gold-standard 0.143 threshold compared to other noise baselines.

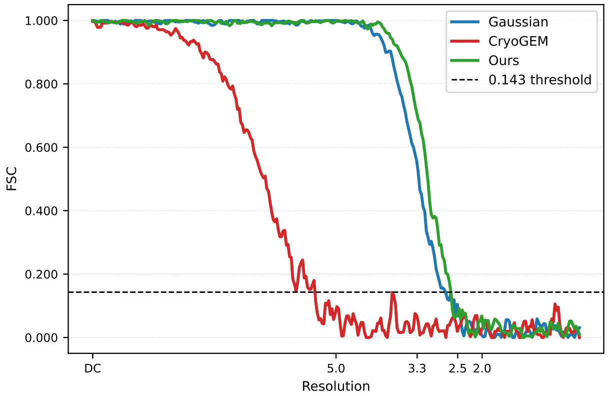

Figure 4. Filtered back-projection reconstruction FSC curves at the 0.143 threshold. Our method consistently achieves higher resolution compared to the baselines.

5.3. Pose Estimation

Accurate particle orientation determination represents a critical bottleneck in achieving high-resolution cryo-EM reconstructions. We evaluated CryoDiff's capacity to enhance pose estimation by applying our noisy synthetic datasets to train state-of-the-art ab-initio reconstruction methods like CryoFIRE [32]. The synthetic particles, complete with precise orientation annotations, provided an ideal training cor-

Table 3. **Quantitative comparison of pose estimation.** Our approach achieves the best performance in both Res(px) and Rot.(rad) metrics across all datasets.

Metric	Res(px)↓					Rot.(rad)↓				
Dataset	Proteasome	Integrin	PhaseMS2	HumanBAF	Avg.	Proteasome	Integrin	PhaseMS2	HumanBAF	Avg.
Gaussian	3.03	7.01	5.96	7.08	5.77	0.51	1.49	0.99	1.58	1.14
Poisson	3.12	7.14	6.09	9.33	6.42	1.09	1.66	1.03	1.60	1.35
Poi–Gau	3.15	7.99	5.98	9.09	6.55	1.32	1.57	0.73	1.51	1.28
CycleDiffusion [62]	3.67	9.04	6.33	9.02	7.01	0.47	1.41	0.97	1.67	1.13
CryoGEM [68]	7.99	9.24	15.38	10.79	10.85	1.72	1.94	1.20	1.78	1.66
CryoFIRE	6.02	13.27	18.03	7.17	11.12	1.54	0.97	0.75	1.49	1.19
Ours	**2.95**	**6.25**	**5.62**	**7.04**	**5.46**	**0.45**	**0.94**	**0.61**	**1.48**	**0.87**

pus for this supervised learning paradigm. We subsequently assessed performance on real micrographs using Filtered Back-Projection (FBP) as our reconstruction algorithm. Detailed mathematical definitions of the pose loss and evaluation metrics are provided in Appendix B.

Table 3 reports the reconstruction resolution (Res(px)) and angular precision (Rot(rad)) for each baseline and our method. Our pose estimator consistently produces sharper reconstructions—reflected in significantly lower Res(px) values—on both static (Proteasome, HumanBAF) and dynamic (Integrin, PhaseMS2) specimens, indicating more accurate orientation estimates. Similarly, angular precision improvements demonstrate that a greater proportion of predicted rotations fall within acceptable error bounds, which is key for high-fidelity 3D volume recovery.

Compared to the CryoFIRE baseline trained without our synthesized noise, our method improves resolution by 52.9% on Integrin (6.25 vs. 13.27 px) and reduces rotation error by 18.6% on PhaseMS2 (0.61 vs. 0.75 rad). Classical noise baselines (e.g., Gaussian, Poisson) exhibit large errors across all datasets due to their limited noise realism, while CryoGEM suffers from adversarial distortions that degrade both resolution and angular accuracy. Our model maintains low errors on both rigid (e.g., 2.95 px, 0.45 rad on Proteasome) and flexible structures, confirming the robust generalization capacity of our framework across diverse structural classes.

5.4. Ablation Study

We validate the effectiveness of different *sampling configurations* on the *Ribosome* dataset. Table 4 summarises two complementary studies: **Sampling steps.** We vary the DDPM step budget from 5 to 100. An FID drop is observed when increasing the budget from 5 to 10 steps. Interestingly, further enlarging the trajectory to 20 and 50 steps leads to *non-monotonic* behaviour (20.44 and 20.06), before a new minimum emerges at 100 steps (18.31). This suggests that moderate-length trajectories may accumulate numerical error or suffer from over-smoothing effects—where the denoising network is repeatedly applied to low-noise inputs, potentially washing out fine structural details. **Sampler algorithm.** With the step budget fixed to 10, we compare four popular samplers. DDPM attains the best FID (19.08), closely followed by DPM-Solver++ and DPM-Solver; DDIM lags behind at 20.33. DDIM underperforms likely because its deterministic nature lacks the stochasticity that may help preserve sample diversity and reduce structural artifacts when the step count is small.

Table 4. Quantitative comparison of FID scores on the *Ribosome*: varying DDPM steps vs. different samplers.

Configuration	Steps	Sampler Algorithm	FID↓
5 Steps	5	DDPM	21.97
10 Steps	10	DDPM	**19.08**
20 Steps	20	DDPM	20.44
50 Steps	50	DDPM	20.06
100 Steps	100	DDPM	18.31
DDIM	10	DDIM	20.33
DDPM	10	**DDPM**	**19.08**
DPM-Solver	10	DPM-Solver	19.27
DPM-Solver++	10	DPM-Solver++	19.13

6. Conclusion

Broader Impacts. We introduce CryoDiff, a synthesis framework that combines biophysical modeling with conditional cycle-consistent diffusion to generate structurally diverse and noise-realistic cryo-EM micrographs. Our method enhances structural fidelity and spatial noise modeling, leading to improved performance on downstream tasks. **Limitation Discussion.** A limitation of our current approach is its dependence on masks as conditioning inputs. Future work may address this limitation by incorporating richer conditioning information, such as density maps or spatial priors.

References

[1] Rana Abdelnabi, James A Geraets, Yipeng Ma, Carmen Mirabelli, Justin W Flatt, Aušra Domanska, Leen Delang, Dirk Jochmans, Timiri Ajay Kumar, Venkatesan Jayaprakash, et al. A novel druggable interprotomer pocket in the capsid of rhino-and enteroviruses. *PLoS Biology*, 17 (6):e3000281, 2019. 5, 12

[2] Alberto Bartesaghi, Alan Merk, Soojay Banerjee, Doreen Matthies, Xiongwu Wu, Jacqueline LS Milne, and Sriram Subramaniam. 2.2 å resolution cryo-em structure of β-galactosidase in complex with a cell-permeant inhibitor. *Science*, 348(6239):1147–1151, 2015. 5, 12

[3] Tristan Bepler, Andrew Morin, Micah Rapp, Julia Brasch, Lawrence Shapiro, Alex J. Noble, and Bonnie Berger. Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nature Methods*, 16(11):1153–1160, 2019. 1, 7, 13

[4] Raquel Bromberg, Yirui Guo, Dominika Borek, and Zbyszek Otwinowski. High-resolution cryo-em reconstructions in the presence of substantial aberrations. *IUCrJ*, 7(3):445–452, 2020. 12

[5] Batuujin Burendei, Ruriko Shinozaki, Masakatsu Watanabe, Tohru Terada, Kazutoshi Tani, Yoshinori Fujiyoshi, and Atsunori Oshima. Cryo-em structures of undocked innexin-6 hemichannels in phospholipids. *Science advances*, 6(7): eaax3157, 2020. 5, 12

[6] Melody G. Campbell, David Veesler, Alan Cheng, Clinton S. Potter, and Bridget Carragher. 2.8 Å resolution reconstruction of the thermoplasma acidophilum 20s proteasome using cryo-electron microscopy. *eLife*, 4:e06380, 2015. 12

[7] Yifan Cheng. Single-particle cryo-em at crystallographic resolution. *Cell*, 161(3):450–457, 2015. 1

[8] Yunjey Choi, Minje Choi, Munyoung Kim, Jung-Woo Ha, Sunghun Kim, and Jaegul Choo. Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8789–8797, 2018. 2

[9] Anthony Cormier, Melody G. Campbell, Saburo Ito, Shenping Wu, Jianlong Lou, James Marks, Jody L. Baron, Stephen L. Nishimura, and Yifan Cheng. Cryo-EM structure of the $\alpha_v\beta_8$ integrin reveals a mechanism for stabilizing integrin extension. *Nature Structural & Molecular Biology*, 25(8):698–704, 2018. 12

[10] John M Cowley and Alexander Forbes Moodie. The scattering of electrons by atoms and crystals. i. a new theoretical approach. *Acta Crystallographica*, 10(10):609–619, 1957. 2

[11] Ashwin Dhakal, Rajan Gyawali, Liguo Wang, and Jianlin Cheng. A large expert-curated cryo-em image dataset for machine learning protein particle picking. *Scientific Data*, 10(1):392, 2023. 1

[12] Prafulla Dhariwal and Alexander Quinn Nichol. Diffusion models beat GANs on image synthesis. In *Advances in Neural Information Processing Systems*, 2021. 2

[13] Raison Dsouza, Ghoncheh Mashayekhi, Roshanak Etemadpour, Peter Schwander, and Abbas Ourmazd. Energy landscapes from cryo-em snapshots: a benchmarking study. *Scientific reports*, 13(1):1372, 2023. 2

[14] Jack Godek, Jared Sivinski, Edmond R Watson, Felicidad Lebario, Wenli Xu, Mckayla Stevens, Christopher J Zerio, Andrew J Ambrose, Xiaoyi Zhu, Carlee A Trindl, et al. Bis-sulfonamido-2-phenylbenzoxazoles validate the groes/el chaperone system as a viable antibiotic target. *Journal of the American Chemical Society*, 146(30):20845–20856, 2024. 5, 13

[15] Basil J Greber, Juan M Perez-Bertoldi, Kif Lim, Anthony T Iavarone, Daniel B Toso, and Eva Nogales. The cryoelectron microscopy structure of the human cdk-activating kinase. *Proceedings of the National Academy of Sciences*, 117 (37):22849–22857, 2020. 13

[16] Harshit Gupta, Michael T McCann, Laurene Donati, and Michael Unser. Cryogan: A new reconstruction paradigm for single-particle cryo-em via deep adversarial learning. *IEEE Transactions on Computational Imaging*, 7:759–774, 2021. 1

[17] Pavol Harar, Lukas Herrmann, Philipp Grohs, and David Haselbach. Faket: Simulating cryo-electron tomograms with neural style transfer. *Structure*, 33(4):820–827, 2025. 2

[18] Richard Henderson. Avoiding the pitfalls of single particle cryo-electron microscopy: Einstein from noise. *Proceedings of the National Academy of Sciences*, 110(45):18037–18041, 2013. 1

[19] Benjamin Himes and Nikolaus Grigorieff. Cryo-tem simulations of amorphous radiation-sensitive samples using multislice wave propagation. *IUCrJ*, 8(Pt 6):943–953, 2021. 2

[20] Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance. *arXiv preprint arXiv:2207.12598*, 2022. 2

[21] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020. 2, 4, 6

[22] Corey F Hryc, Dong-Hua Chen, Pavel V Afonine, Joanita Jakana, Zhao Wang, Cameron Haase-Pettingell, Wen Jiang, Paul D Adams, Jonathan A King, Michael F Schmid, et al. Accurate model annotation of a near-atomic resolution cryo-em map. *Proceedings of the National Academy of Sciences*, 114(12):3103–3108, 2017. 12

[23] Andrii Iudin, Paul K Korir, Sriram Somasundharam, Simone Weyand, Cesare Cattavitello, Neli Fonseca, Osman Salih, Gerard J Kleywegt, and Ardan Patwardhan. Empiar: the electron microscopy public image archive. *Nucleic Acids Research*, 51(D1):D1503–D1511, 2023. 5

[24] Maarten Joosten, Joel Greer, James Parkhurst, Tom Burnley, and Arjen J. Jakobi. Roodmus: a toolkit for benchmarking heterogeneous electron cryo-microscopy reconstructions. *IUCrJ*, 11(Pt 6):951–965, 2024. Open Access. 2

[25] Chanyong Jung, Gihyun Kwon, and Jong Chul Ye. Exploring patch-wise semantic relation for contrastive learning in image-to-image translation tasks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 18260–18269, 2022. 2

[26] David Kalbermatter, Neeta Shrestha, Flavio M Gall, Marianne Wyss, Rainer Riedl, Philippe Plattet, and Dimitrios Fotiadis. Cryo-em structure of the prefusion state of canine

distemper virus fusion protein ectodomain. *Journal of structural biology: X*, 4:100021, 2020. 12

[27] Robert Kiewisz, Gabriel Meyer-Lee, and Tristan Bepler. A latent back-projection network for novel projection synthesis for improved cryo-ET. In *ICLR 2025 Workshop on Generative and Experimental Perspectives for Biomolecular Design*, 2025. 1, 2

[28] Junho Kim, Minjae Kim, Hyeonwoo Kang, and Kwang Hee Lee. U-gat-it: Unsupervised generative attentional networks with adaptive layer-instance normalization for image-to-image translation. In *International Conference on Learning Representations*, 2020. 2

[29] R. I. Koning, J. Gomez-Blanco, I. Akopjana, J. Vargas, A. Kazaks, K. Tars, J. M. Carazo, and A. J. Koster. Asymmetric cryo-em reconstruction of phage ms2 reveals genome structure in situ. *Nature Communications*, 7:12524, 2016. 12

[30] Werner Kühlbrandt. The resolution revolution. *Science*, 343 (6178):1443–1444, 2014. 1

[31] Hsin-Ying Lee, Hung-Yu Tseng, Jia-Bin Huang, Maneesh Singh, and Ming-Hsuan Yang. Diverse image-to-image translation via disentangled representations. In *Proceedings of the European conference on computer vision (ECCV)*, pages 35–51, 2018. 2

[32] Axel Levy, Gordon Wetzstein, Julien NP Martel, Frederic Poitevin, and Ellen Zhong. Amortized inference for heterogeneous reconstruction in cryo-em. *Advances in neural information processing systems*, 35:13038–13049, 2022. 1, 7

[33] Hongjia Li, Hui Zhang, Xiaohua Wan, Zhidong Yang, Chengmin Li, Jintao Li, Renmin Han, Ping Zhu, and Fa Zhang. Noise-transfer2clean: denoising cryo-em images based on noise modeling and transfer. *Bioinformatics*, 38 (7):2022–2029, 2022. 1

[34] Maofu Liao, Erhu Cao, David Julius, and Yifan Cheng. Structure of the trpv1 ion channel determined by electron cryo-microscopy. *Nature*, 504(7478):107–112, 2013. 5, 12

[35] Ting-Yu Lin, Leon Kleemann, Jakub Jezowski, Dominika Dobosz, Michal Rawski, Paulina Indyka, Grzegorz Wazny, Rahul Mehta, Andrzej Chramiec-Glabik, Lukasz Koziej, et al. The molecular basis of trna selectivity by human pseudouridine synthase 3. *Molecular Cell*, 84(13):2472–2489, 2024. 13

[36] Ming-Yu Liu, Thomas Breuel, and Jan Kautz. Unsupervised image-to-image translation networks. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, page 700–708, Red Hook, NY, USA, 2017. Curran Associates Inc. 2

[37] Xinhang Liu, Yan Zeng, Yifan Qin, Hao Li, Jiakai Zhang, Lan Xu, and Jingyi Yu. Cryoformer: Continuous heterogeneous cryo-em reconstruction using transformer-based neural representations. *arXiv preprint arXiv:2303.16254*, 2023. 1

[38] Cheng Lu, Yuhao Zhou, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu. Dpm-solver: A fast ode solver for diffusion probabilistic model sampling in around 10 steps. *Advances in Neural Information Processing Systems*, 35:5775–5787, 2022. 4, 6

[39] Cheng Lu, Yuhao Zhou, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu. Dpm-solver++: Fast solver for guided sampling of diffusion probabilistic models. *arXiv preprint arXiv:2211.01095*, 2022. 6

[40] Dmitry Lyumkis. Challenges and opportunities in cryo-em single-particle analysis. *Journal of Biological Chemistry*, 294(13):5181–5197, 2019. 1

[41] Brandon Malone, James Chen, Qi Wang, Eliza Llewellyn, Young Joo Choi, Paul Dominic B Olinares, Xinyun Cao, Carolina Hernandez, Edward T Eng, Brian T Chait, et al. Structural basis for backtracking by the sars-cov-2 replication–transcription complex. *Proceedings of the National Academy of Sciences*, 118(19):e2102516118, 2021. 13

[42] Daniel Mann, Junping Fan, Kamolrat Somboon, Daniel P Farrell, Andrew Muenks, Svetomir B Tzokov, Frank DiMaio, Syma Khalid, Samuel I Miller, and Julien RC Bergeron. Structure and lipid dynamics in the maintenance of lipid asymmetry inner membrane complex of a. baumannii. *Communications biology*, 4(1):817, 2021. 5, 12

[43] Antonio Martinez-Sanchez, Lorenz Lamm, Marion Jasnin, and Harold Phelippeau. Simulating the cellular context in synthetic datasets for cryo-electron tomography. *IEEE Transactions on Medical Imaging*, 43(11):3742–3754, 2024. 2

[44] Nikita Mashtalir, Daniel R. D'Avino, Yashar Michel, Benjamin Luo, Charles Zullow, Benjamin McKenzie, and Benjamin K. Kadoch. A structural model of the endogenous human baf complex informs disease mechanisms. *Cell*, 183(3): 802–817.e24, 2020. 13

[45] Takanori Nakane, Abhay Kotecha, Andrija Sente, Greg McMullan, Simonas Masiulis, Patricia MGE Brown, Ioana T Grigoras, Lina Malinauskaite, Tomas Malinauskas, Jonas Miehling, et al. Single-particle cryo-em at atomic resolution. *Nature*, 587(7832):152–156, 2020. 1

[46] Alexander Quinn Nichol and Prafulla Dhariwal. Improved denoising diffusion probabilistic models. In *International conference on machine learning*, pages 8162–8171. PMLR, 2021. 2

[47] Alex J Noble et al. Virtualice: Half-synthetic cryoem micrograph generator. *bioRxiv*, 2023. 1, 2

[48] Eva Nogales. The development of cryo-em into a mainstream structural biology technique. *Nature methods*, 13(1):24–27, 2016. 1

[49] Taesung Park, Alexei A Efros, Richard Zhang, and Jun-Yan Zhu. Contrastive learning for unpaired image-to-image translation. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part IX 16*, pages 319–345. Springer, 2020. 2

[50] James M Parkhurst, Anna Cavalleri, Maud Dumoux, Mark Basham, Daniel Clare, C Alistair Siebert, Gwyndaf Evans, James H Naismith, Angus Kirkland, and Jonathan W Essex. Computational models of amorphous ice for accurate simulation of cryo-em images of biological samples. *Ultramicroscopy*, 256:113882, 2024. 1

[51] 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. 7

[52] Carson Purnell, Jessica Heebner, Michael T Swulius, Ryan Hylton, Seth Kabonick, Michael Grillo, Sergei Grigoryev,

Fred Heberle, M Neal Waxham, and Matthew T Swulius. Rapid synthesis of cryo-et data for training deep learning models. *bioRxiv*, 2023. 2

[53] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10684–10695, 2022. 2

[54] Amélie Royer, Konstantinos Bousmalis, Stephan Gouws, Fred Bertsch, Inbar Mosseri, Forrester Cole, and Kevin Murphy. Xgan: Unsupervised image-to-image translation for many-to-many mappings. In *Domain Adaptation for Visual Understanding*, pages 33–49. Springer, 2020. 2

[55] Hans Rullgård, L-G Öfverstedt, Sergey Masich, Bertil Daneholt, and Ozan Öktem. Simulation of transmission electron microscope images of biological specimens. *Journal of microscopy*, 243(3):234–256, 2011. 1

[56] Hiroshi Sasaki, Chris G Willcocks, and Toby P Breckon. Unit-ddpm: Unpaired image translation with denoising diffusion probabilistic models. *arXiv preprint arXiv:2104.05358*, 2021. 2

[57] Katherine A Senn, Karli A Lipinski, Natalie J Zeps, Amory F Griffin, Max E Wilkinson, and Aaron A Hoskins. Control of $3'$ splice site selection by the yeast splicing factor fyv6. *Elife*, 13:RP100449, 2024. 13

[58] Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models, 2022. 2, 6

[59] Yong Zi Tan and John L Rubinstein. Through-grid wicking enables high-speed cryoem specimen preparation. *Biological Crystallography*, 76(11):1092–1103, 2020. 12

[60] Miloš Vulović, Raimond BG Ravelli, Lucas J van Vliet, Abraham J Koster, Ivan Lazić, Uwe Lücken, Hans Rullgård, Ozan Öktem, and Bernd Rieger. Image formation modeling in cryo-electron microscopy. *Journal of structural biology*, 183(1):19–32, 2013. 1, 2, 13

[61] Weilun Wang, Wengang Zhou, Jianmin Bao, Dong Chen, and Houqiang Li. Instance-wise hard negative example generation for contrastive learning in unpaired image-to-image translation. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 14020–14029, 2021. 2

[62] Chen Henry Wu and Fernando De la Torre. Unifying diffusion models' latent space, with applications to cyclediffusion and guidance. *arXiv preprint arXiv:2210.05559*, 2022. 2, 5, 7, 8, 13

[63] Jia-Geng Wu, Yang Yan, Dong-Xu Zhang, Bo-Wen Liu, Qing-Bing Zheng, Xiao-Liang Xie, Shi-Qi Liu, Sheng-Xiang Ge, Zeng-Guang Hou, and Ning-Shao Xia. Machine learning for structure determination in single-particle cryo-electron microscopy: A systematic review. *IEEE Transactions on Neural Networks and Learning Systems*, 33(2):452–472, 2021. 1

[64] Xindi Wu, Chengkun Li, Xiangrui Zeng, Haocheng Wei, Hong-Wen Deng, Jing Zhang, and Min Xu. Cryoetgan: Cryo-electron tomography image synthesis via unpaired image translation. *Frontiers in Physiology*, Volume 13 - 2022, 2022. 2

[65] Qing Xie, Craig K Yoshioka, and Michael S Chapman. Adeno-associated virus (aav-dj)—cryo-em structure at 1.56 å resolution. *Viruses*, 12(10):1194, 2020. 13

[66] Ling Yang, Zhilong Zhang, Zhaochen Yu, Jingwei Liu, Minkai Xu, Stefano Ermon, and Bin Cui. Contextualized diffusion models for text-guided image and video generation, 2024. 2

[67] Zili Yi, Hao Zhang, Ping Tan, and Minglun Gong. Dualgan: Unsupervised dual learning for image-to-image translation. In *Proceedings of the IEEE international conference on computer vision*, pages 2849–2857, 2017. 2

[68] Jiakai Zhang, Qihe Chen, Yan Zeng, Wenyuan Gao, Xuming He, Zhijie Liu, and Jingyi Yu. CryoGEM: Physics-informed generative cryo-electron microscopy. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024. 2, 5, 7, 8, 13

[69] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. Adding conditional control to text-to-image diffusion models. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 3836–3847, 2023. 2

[70] Y. Zhang, R. Tammaro, P. J. Peters, and R. B. G. Ravelli. Could egg white lysozyme be solved by single particle cryo-em? *Journal of Chemical Information and Modeling*, 60(5): 2605–2613, 2020. Published on 2020/05/26. 2

[71] Chuanxia Zheng, Tat-Jen Cham, and Jianfei Cai. The spatially-correlative loss for various image translation tasks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 16407–16417, 2021. 2

[72] Liming Zheng, Nan Liu, Xiaoyin Gao, Wenqing Zhu, Kun Liu, Cang Wu, Rui Yan, Jincan Zhang, Xin Gao, Yating Yao, et al. Uniform thin ice on ultraflat graphene for high-resolution cryo-em. *Nature Methods*, 20(1):123–130, 2023. 2

[73] 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. 1

[74] Ellen D Zhong, Adam Lerer, Joseph H Davis, and Bonnie Berger. Cryodrgn2: Ab initio neural reconstruction of 3d protein structures from real cryo-em images. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4066–4075, 2021. 1

[75] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 2242–2251, 2017. 2

[76] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017. 2

[77] Yanan Zhu, Qi Ouyang, and Youdong Mao. A deep convolutional neural network approach to single-particle recognition in cryo-electron microscopy. *BMC bioinformatics*, 18:1–10, 2017. 1