Simulating Cryo-EM: Cycle-Consistent Predictor–Corrector Diffusion with Biophysical Modeling

Anonymous authors
Paper under double-blind review

Abstract

Single-particle cryo-electron microscopy (cryo-EM) has become a cornerstone of structural biology, enabling near-atomic resolution analysis of macromolecules through advanced computational methods. However, the development of cryo-EM processing tools is constrained by the scarcity of high-quality annotated datasets. Synthetic data generation offers a promising alternative, but existing approaches lack thorough biophysical modeling of heterogeneity and fail to reproduce the complex noise observed in real imaging. To address these limitations, we present CryoCCD, a comprehensive simulator for cryo-EM, unifying versatile biophysical modeling with the first diffusion model for realistic noise generation. The biophysical engine provides multi-functional generation capabilities to capture authentic biological organization, and the diffusion model is enhanced with cycle consistency and predictor–corrector sampling to improve realism and structural fidelity. Extensive experiments demonstrate that CryoCCD generates structurally faithful micrographs, enhances particle picking and pose estimation, as well as achieves superior performance over state-of-the-art baselines, while also generalizing effectively to held-out protein families.

1 Introduction

In recent years, single-particle cryo-electron microscopy (cryo-EM) has emerged as an essential technique in structural biology, enabling near-atomic resolution reconstructions of macromolecules in their native states (Cheng, 2015; Kühlbrandt, 2014). By vitrifying biological specimens and imaging them with high-energy electron beams, cryo-EM has significantly deepened our understanding of protein architectures and molecular mechanisms, thereby accelerating drug discovery and expanding biological insight (Nogales, 2016; Lyumkis, 2019). However, its full potential is constrained by several data-centric limitations, including the scarcity of datasets, extremely low signal-to-noise ratios (SNR), and reliance on labor-intensive annotation processes (Nakane et al., 2020; Henderson, 2013). These challenges collectively impede the development of robust models for downstream tasks such as particle picking (Bepler et al., 2019), pose estimation (Levy et al., 2022), and 3D reconstruction (Zhong et al., 2021a;b).

To support the development of learning-based cryo-EM algorithms (Zhu et al., 2017b; Gupta et al., 2021; Zhong et al., 2021a; Wu et al., 2021; Jiang et al., 2025), recent efforts have focused on generating synthetic data (Rullgård et al., 2011; Vulović et al., 2013; Dhakal et al., 2023) through biophysically inspired modeling. Several cryo-EM–specific frameworks, including InsilicoTEM (Vulović et al., 2013) and LBPN (Kiewisz et al., 2025), simulate high-fidelity multi-angle projection processes to approximate realistic imaging conditions. While effective, these methods often incur substantial computational costs. VirtualIce (Noble, 2024) improves efficiency by simulating particle behaviors such as aggregation, overlap, and preferred orientations on top of real vitrified backgrounds. However, it relies on simplified Gaussian noise injection, which fails to capture the complex and structured noise patterns observed in real cryo-EM micrographs (Li et al., 2022).

Despite these advances, existing biophysical simulation approaches (Vulović et al., 2013; Rullgård et al., 2011) still face two major limitations. First, they offer limited support for modeling structural diversity and spatial variability, essential for generating datasets tailored to specific biological contexts (Zhong et al., 2021a; Liu et al., 2023). Second, they typically assume additive Gaussian noise, whereas real cryo-EM data contain a mixture of detector noise, electron scattering artifacts, radiation damage, and heterogeneous background

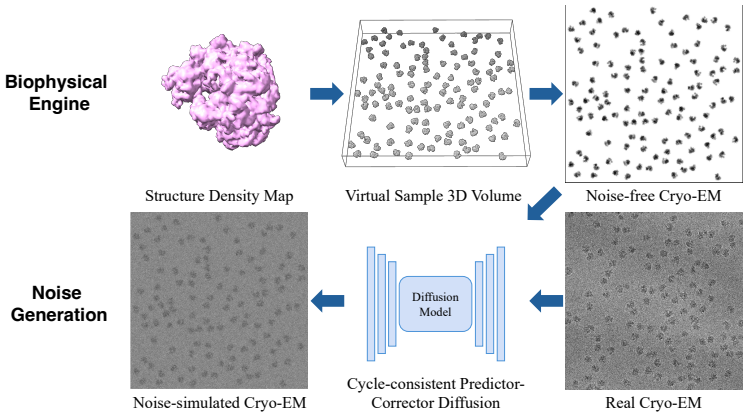

Figure 1: A schematic overview of simulating cryo-EM data via CryoCCD. The biophysical engine simulates noise-free cryo-EM from structure density maps, while a diffusion model synthesizes realistic noise distributions learned from real cryo-EM data (example: EMPIAR-10028; Plasmodium falciparum 80S ribosome).

signals (Li et al., 2022; Parkhurst et al., 2024). To mitigate this gap, CryoGEM (Zhang et al., 2024) introduces a physics-informed generative framework that better models noise distributions. Nevertheless, its GAN-based formulation lacks bidirectional constraints (Zhu et al., 2017a), making it prone to structure distortion and incapable of controllable noise synthesis. A schematic overview is provided in Fig. 1.

To address these limitations, we propose **CryoCCD**, a **C**ycle-consistent predictor–**C**orrector **D**iffusion framework with biophysical modeling for **Cryo**-EM synthesis. We design a modular biophysical engine for multi-scale cryo-EM synthesis, using structurally diverse particles from the PDB and AlphaFold3 (wwPDB consortium, 2018; Abramson et al., 2024) to simulate virtual cellular specimens and modeling imaging physics to generate micrographs that faithfully capture heterogeneity. To the best of our knowledge, CryoCCD is the first to apply diffusion models for realistic noise generation in cryo-EM simulation. Compared to GAN-based methods (Zhang et al., 2024; Harar et al., 2025), which are prone to mode collapse, conditional diffusion models provide more stable generation. To preserve structural fidelity, we introduce a cycle-consistent approach that aligns positional information during domain translation. Furthermore, mask-guided contrastive learning enhances the representation of fine-grained features such as edges, textures, and spatial noise patterns, and predictor–corrector sampling improves noise realism by better recovering particle-level and background-level statistical characteristics.

We validate CryoCCD on six diverse cryo-EM datasets, showing consistently superior scores on both FID and CMMD compared to traditional noise models and recent generative baselines. Beyond visual realism, our framework also improves downstream tasks: particle picking achieves higher AUPRC, and pose estimation attains lower angular error. Moreover, CryoCCD generalizes to held-out datasets spanning both unseen protein families and unseen imaging conditions, highlighting its transferability. The core code—the data-generation pipeline, the model, and training scripts—will be released upon acceptance to facilitate further research.

Our contributions are summarized as follows:

- **Comprehensive Biophysical Engine:** We develop a unified biophysical modeling engine that encodes structural and biophysical priors, enabling multi-functional cryo-EM synthesis with realistic compositional and conformational heterogeneity.
- **Realistic Noise Generation:** We introduce the first diffusion-based framework for cryo-EM noise synthesis, where cycle consistency preserves structural integrity, mask-guided contrastive learning enhances fine-grained representation, and predictor–corrector sampling improves fine-scale noise fidelity.
- **Robust Generalization:** We validate CryoCCD on six public cryo-EM datasets, achieving state-of-the-art (**SOTA**) performance across the evaluated visual-quality benchmarks and downstream tasks

(particle picking and pose estimation), and demonstrating strong generalization to held-out datasets spanning unseen protein families and imaging conditions.

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 (Cowley & Moodie, 1957; Vulović et al., 2013) to more realistic frameworks that account for imaging noise and sample heterogeneity (Himes & Grigorieff, 2021; Zhang et al., 2020; Dsouza et al., 2023; Zheng et al., 2023; Joosten et al., 2024). To balance realism and controllability, hybrid simulators like VirtualIce (Noble, 2024) and cryo-TomoSim (Purnell et al., 2023) generate annotation-ready data with physical priors. Recently, specialized approaches have been developed to handle cryo-ET challenges, such as PolNet (Martinez-Sanchez et al., 2024) for cellular variability and LBPN (Kiewisz et al., 2025) 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 (Wu et al., 2022) applies GAN-based noise synthesis, FakET (Harar et al., 2025) leverages neural style transfer, and CryoGEM (Zhang et al., 2024) incorporates physics-guided contrastive learning. Here, we enhance realism and diversity via biophysical modeling and diffusion-based noise generation.

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 (Zhu et al., 2017a; Yi et al., 2017; Lee et al., 2018; Royer et al., 2020; Choi et al., 2018) to variants with shared latent spaces (Liu et al., 2017), attention (Kim et al., 2020), and contrastive learning (Park et al., 2020; Jung et al., 2022; Wang et al., 2021; Zheng et al., 2021). Extending these ideas, GAN-based models such as CryoETGAN (Wu et al., 2022) and CryoGEM (Zhang et al., 2024) 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 (Ho et al., 2020; Song et al., 2020). Early efforts focused on sampling efficiency and image quality (Nichol & Dhariwal, 2021; Dhariwal & Nichol, 2021; Rombach et al., 2022), while recent advances enable conditional generation via classifier guidance and context-aware conditioning (Ho & Salimans, 2022; Yang et al., 2024; Zhang et al., 2023; Min et al., 2024). For unpaired image-to-image translation, diffusion models offer greater stability and diversity than GANs. CycleDiffusion (Wu & De la Torre, 2022) introduced a unified latent space to enforce cycle consistency, improving fidelity. Currently, CycleNet (Xu et al., 2023) introduces a lightweight cycle consistency regularizer for text-guided diffusion. Despite these advances, diffusion models remain underexplored in cryo-EM synthesis. To fill this gap, we integrate these ideas—cycle consistency, contrastive learning, conditional diffusion, and predictor–corrector sampling—into the first diffusion-based framework tailored to cryo-EM noise synthesis.

3 Biophysical Modeling for Diverse Structural Simulation

We introduce a unified biophysical modeling framework that combines both experimental and generated structures with 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. Representative outputs of its core components are illustrated in Fig. 3, the overall pipeline follows the stages depicted in Fig. 4, and a step-wise visualization is provided in Fig. 6. Full specifications are provided in Appendix A.

Library Construction. To enrich the structural basis for simulation, we construct a heterogeneity-supporting model library by combining a broad collection of experimentally resolved PDB structures, spanning molecular weights from small enzymes ($\sim$50 kDa) to large viral capsids (>50,000 kDa) and symmetry classes from high-order tetrahedral assemblies to asymmetric C_1 complexes, and combining AlphaFold predictions on curated proteomes to fill gaps in experimental structures (see Fig. 2), thereby expanding structural coverage to improve the robustness and generalizability. All models are then standardized through Gaussian

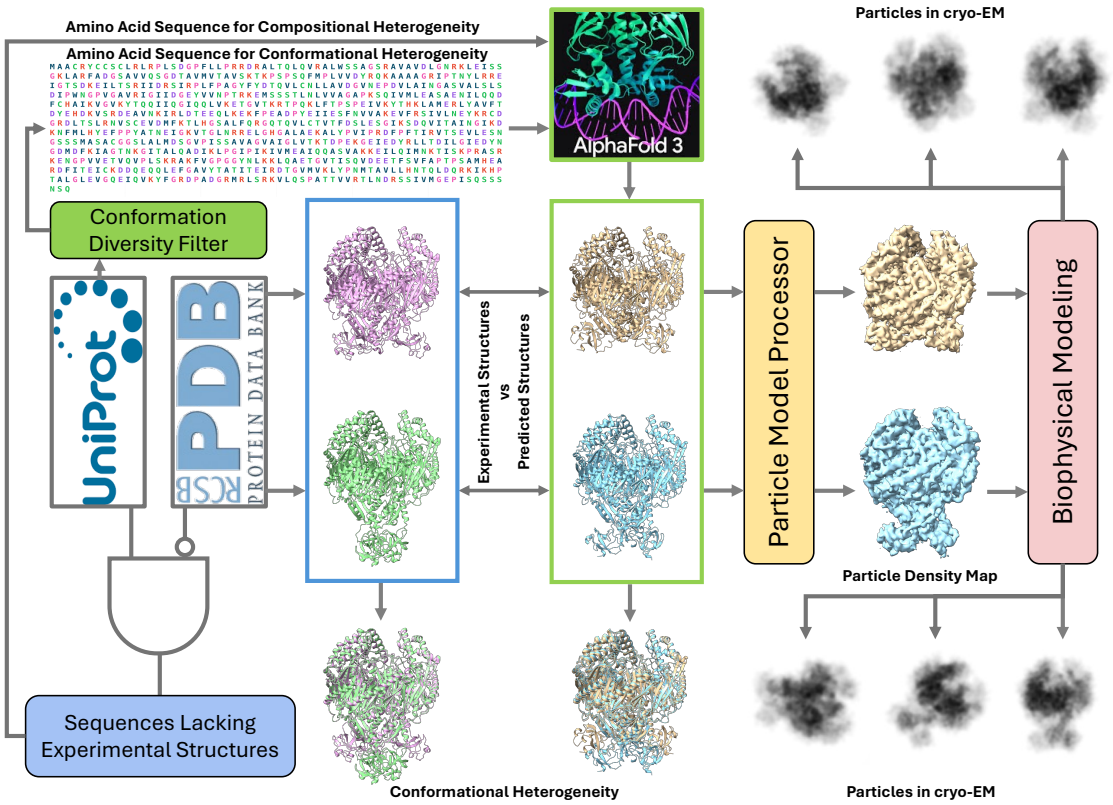

Figure 2: **AlphaFold3-based pipeline for enhancing compositional and conformational heterogeneity of the structure library.** Sequences lacking experimental structures are selected from the UniProt (Consortium, 2019) database to increase compositional heterogeneity, and entries with potential conformational diversity are selected to enhance conformational heterogeneity. The green frame shows the atomic models of two human PNPase (Q8TCS8) conformations processed based on AlphaFold3; in the blue frame, for comparison, are the open formation (9KJR) and closed formation (9KJT) of human PNPase from the Protein Data Bank. Below the frames are illustrations of the differences between these conformations. The generated particles are then converted into density maps and undergo biophysical modeling to generate simulated cryo-EM images.

smoothing and isosurface extraction, preserving fine-scale features. The processed particles are shown in Fig. 3 (a).

Virtual Sample Modeling. To assemble in-silico specimens for imaging, we build virtual samples that balance fidelity and diversity. The steps are as follows:

- **Particle Placement.** To achieve biologically plausible positioning and orientation of macromolecules, we combine empirical annotations with probabilistic sampling under tunable fidelity-diversity tradeoffs. RELION-derived (Kimanius et al., 2021) picks and angles from real data are mapped from pixel position to volume coordinates to yield translations T_{exp} and quaternion orientations q_{exp}. To augment sampling beyond mapped particles, we sample translations T_{syn} from density-matched or uniform distributions calibrated to empirical radial functions and 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 fidelity and diversity.

- **Class-Specific Distribution Modeling.** To recapitulate distinct spatial arrangements of molecular subtypes, we apply particle-type–specific priors inferred from experimental distributions, shown in Fig. 3 (b). 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

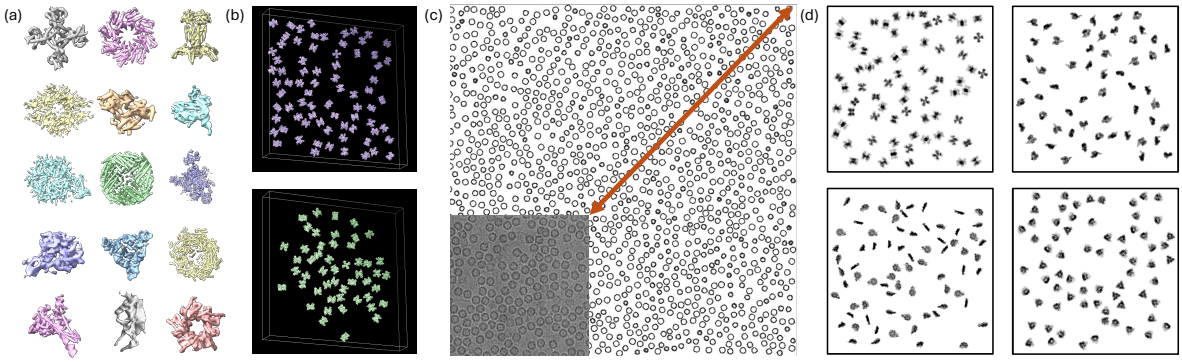

Figure 3: Results from the CryoCCD Biophysical Modeling: (a) Visualization of the particles density maps processed from PDB data. (b) Different placement strategies are adopted based on the particles' properties. (c) Generation of multi-scale synthetic data from real cryo-EM images according to EMPIAR-10421. (d) Visualization of simulated cryo-EM images.

mimic polysomes, 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, as detailed in Fig. 3 (c). 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 simulates 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 without real-sample input. The ice density field incorporates Gaussian fluctuations to simulate noise before particle embedding.

Projection and Electrostatic Potential Assembly. To produce high-fidelity micrographs from the virtual sample, we compute the total electrostatic potential from all embedded components as

$$\Phi(\mathbf{r}) = \sum_{i=1}^{L} \rho_i\big(R(q_i)^{-1}(\mathbf{r} - T_i)\big), \tag{1}$$

and project it under the weak-phase approximation:

$$P(x, y) = \int_{-z_0/2}^{z_0/2} \Phi(x, y, z)\,\mathrm{d}z. \tag{2}$$

The generated projections are shown in Fig. 3 (d). We then apply the instrument contrast transfer function in Fourier space and perform an inverse Fourier transform to obtain a noise-free micrograph for subsequent diffusion-based noise synthesis (see Appendix A.1).

4 Conditional Cycle-consistent Diffusion for Realistic Noise Translation

Realistic cryo-EM noise generation can be posed as unpaired image-to-image translation from clean simulations to realistic noisy micrographs. Unlike Gaussian or Poisson injection, this enables learning structured background and detector-specific noise directly from real data. However, existing GAN-based approaches often suffer from mode collapse and lack explicit constraints to preserve structural fidelity. To overcome these limitations, we introduce CryoCCD, which incorporates cycle-consistent predictor–corrector diffusion as described in Section 4.1 and mask-guided contrastive learning as described in Section 4.2. The overall module is illustrated in Fig. 4.

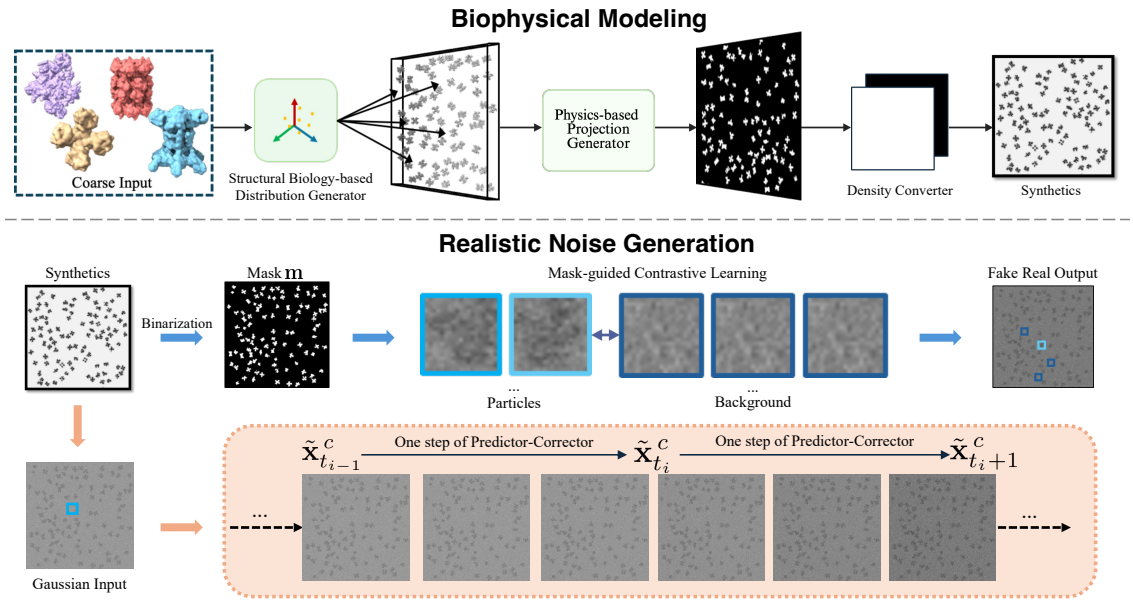

Figure 4: CryoCCD pipeline: (1) Biophysical modeling. Input structures are placed into a 3D volume via particle placement strategies, then passed through physics-based projection and density conversion to synthesize cryo-EM micrographs. (2) Realistic noise generation. Synthetic images are translated into realistic noisy micrographs using Cycle-Consistent Predictor–Corrector Diffusion, mask-guided contrastive learning; the "One Step of Predictor–Corrector" inset shows that, within one diffusion step, the predictor and corrector sequentially advance the state from t_{i-1} to the corrected state at t_i.

4.1 Cycle-Consistent Predictor-Corrector Diffusion

Let $\mathcal{A}$ denote the synthetic domain, where each clean simulation $\mathbf{x} \in \mathcal{A}$ is accompanied by a segmentation mask $\mathbf{m}$. Let $\mathcal{B}$ denote the real micrograph domain, where masks are unavailable. Accordingly, we use a mask-conditioned generator $G_{AB}(\cdot, \mathbf{m})$ for $\mathcal{A} \to \mathcal{B}$ and an unconditional generator $G_{BA}(\cdot)$ for $\mathcal{B} \to \mathcal{A}$.

Forward and reverse processes. Let $\{\beta_t\}_{t=1}^{T}$ be the variance schedule, $\alpha_t = 1 - \beta_t$, and $\bar{\alpha}_t = \prod_{s=1}^{t} \alpha_s$. The forward process is

$$q(\mathbf{x}_t \mid \mathbf{x}_{t-1}) = \mathcal{N}\Big(\mathbf{x}_t; \sqrt{1-\beta_t}\,\mathbf{x}_{t-1},\, \beta_t \mathbf{I}\Big)\,. \tag{3}$$

We parameterize the reverse process with a noise predictor ϵ_θ for G_{AB} and ϵ_ϕ for G_{BA}. For G_{AB}, ϵ_θ is conditioned on the source image $\mathbf{x}$ and its mask $\mathbf{m}$; for G_{BA}, on the source image alone.

Mask-conditioned Predictor-Corrector Sampler. To accelerate sampling, we adopt UniPC (Zhao et al., 2023) in the half log-SNR domain. we define the signal scaling factor $\tilde{\alpha}_t = \sqrt{\bar{\alpha}_t}$, the noise scaling factor $\tilde{\sigma}_t = \sqrt{1-\bar{\alpha}_t}$, and the half log-SNR $\lambda_t = \log(\tilde{\alpha}_t/\tilde{\sigma}_t)$. We choose an inference schedule $t_0 > t_1 > \cdots > t_M$ with step size $h_i = \lambda_{t_i} - \lambda_{t_{i-1}}$. Let $t_\lambda(\cdot)$ denote the inverse mapping from λ to the diffusion time index.

Given intermediate collocation nodes $0 < r_1 < \cdots < r_p = 1$ and their corresponding time steps $s_m = t_\lambda(r_m h_i + \lambda_{t_{i-1}})$, we define the noise differences used in the high-order correction:

$$D_m = \epsilon_\vartheta(\tilde{\mathbf{x}}_{s_m}, s_m, \mathbf{m}_\vartheta) - \epsilon_\vartheta(\tilde{\mathbf{x}}^c_{t_{i-1}}, t_{i-1}, \mathbf{m}_\vartheta), \tag{4}$$

where the superscript c denotes the previously corrected state. Here, $\vartheta \in \{\theta, \phi\}$ indexes the generation direction, and the condition mask is set as $\mathbf{m}_\theta = \mathbf{m}$ for mask-conditioned generation (G_{AB}) and $\mathbf{m}_\phi = \varnothing$ for unconditional generation (G_{BA}).

Using a step-size function $B(h) = \mathcal{O}(h)$ (e.g., $B(h) = h$ or $B(h) = e^h - 1$), the Exponential-Euler base term is formulated as:

$$\mathcal{E}_i(\tilde{\mathbf{x}}^c_{t_{i-1}}; \mathbf{m}_\vartheta) = \frac{\tilde{\alpha}_{t_i}}{\tilde{\alpha}_{t_{i-1}}} \tilde{\mathbf{x}}^c_{t_{i-1}} - \tilde{\sigma}_{t_i}\left(e^{h_i} - 1\right) \epsilon_\vartheta(\tilde{\mathbf{x}}^c_{t_{i-1}}, t_{i-1}, \mathbf{m}_\vartheta). \tag{5}$$

A complete UniPC step consists of a **Predictor** (UniP-p) to estimate the initial state and a **Corrector** (UniC-p) to refine it:

$$\tilde{\mathbf{x}}_{t_i} = \mathcal{E}_i\big(\tilde{\mathbf{x}}^c_{t_{i-1}}; \mathbf{m}_\vartheta\big) - \tilde{\sigma}_{t_i}\, B(h_i) \sum_{m=1}^{p-1} \frac{a^{\mathrm{P}}_m}{r_m}\, D_m, \tag{6}$$

$$\tilde{\mathbf{x}}^c_{t_i} = \mathcal{E}_i\big(\tilde{\mathbf{x}}_{t_i}; \mathbf{m}_\vartheta\big) - \tilde{\sigma}_{t_i}\, B(h_i) \sum_{m=1}^{p} \frac{a^{\mathrm{C}}_m}{r_m}\, D_m, \tag{7}$$

where a^{P}_m and a^{C}_m are UniPC solver coefficients, and D_p evaluates the predictor output via $\epsilon_\vartheta(\tilde{\mathbf{x}}_{t_i}, t_i, \mathbf{m}_\vartheta)$.

Coefficient construction and accuracy. We derive the UniPC coefficient vectors $\mathbf{a}^{\mathrm{P}} = [a^{\mathrm{P}}_1, \ldots, a^{\mathrm{P}}_{p-1}]^\top$ and $\mathbf{a}^{\mathrm{C}} = [a^{\mathrm{C}}_1, \ldots, a^{\mathrm{C}}_p]^\top$ via moment matching of the exponential integrator. We define the Vandermonde matrix $\mathbf{R}_q(h)$ and the corresponding moment vector $\boldsymbol{\phi}_q(h)$ as follows:

$$\mathbf{R}_q(h) = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ r_1 h & r_2 h & \cdots & r_q h \\ \vdots & \vdots & \ddots & \vdots \\ (r_1 h)^{q-1} & (r_2 h)^{q-1} & \cdots & (r_q h)^{q-1} \end{pmatrix}, \tag{8}$$

$$\boldsymbol{\phi}_q(h) = \big(h \cdot 1!\, \varphi_2(h),\, \ldots,\, h^q q!\, \varphi_{q+1}(h)\big)^\top,$$

where $\varphi_0(h) = e^h$ and $\varphi_{n+1}(h) = (\varphi_n(h) - 1/n!)/h$. The non-learned solver weights are then computed as:

$$\mathbf{a}^{\mathrm{P}} = \mathbf{R}_{p-1}(h_i)^{-1}\, \boldsymbol{\phi}_{p-1}(h_i) \,/\, B(h_i), \qquad \mathbf{a}^{\mathrm{C}} = \mathbf{R}_p(h_i)^{-1}\, \boldsymbol{\phi}_p(h_i) \,/\, B(h_i), \tag{9}$$

These vectors provide the elements $\{a^{\mathrm{P}}_m\}$ and $\{a^{\mathrm{C}}_m\}$ used to linearly combine the noise differences $\{D_m\}$ in Eqs. equation 6 and equation 7. This moment-matching formulation accurately matches the exponential-integrator expansion of the exact reverse ODE, yielding a p-th order predictor and a $(p{+}1)$-th order corrector (Zhao et al., 2023). We apply the same coefficient construction to both directions (G_{AB} and G_{BA}); mask conditioning is handled through $\mathbf{m}_\vartheta$ with $\mathbf{m}_\theta = \mathbf{m}$ and $\mathbf{m}_\phi = \varnothing$.

Cycle consistency. Without paired supervision, we enforce that translating across domains and back should approximately recover the original input. Concretely, we define:

$$\begin{aligned} \mathcal{L}_{\mathrm{cyc}} = \mathbb{E}_{(\mathbf{x},\mathbf{m})\sim p_{\mathcal{A}}} \big[\big\| G_{BA}\big(G_{AB}(\mathbf{x}, \mathbf{m})\big) - \mathbf{x} \big\|_1 \big] \\ + \mathbb{E}_{\mathbf{y}\sim p_{\mathcal{B}}} \big[\big\| G_{AB}\big(G_{BA}(\mathbf{y}), \varnothing\big) - \mathbf{y} \big\|_1 \big]. \end{aligned} \tag{10}$$

where the first term penalizes reconstruction error along the synthetic-real-synthetic cycle and the second term along the real-synthetic-real cycle. Since real micrographs carry no mask, G_{AB} uses an empty mask ($\mathbf{m}_\phi = \varnothing$) when closing this cycle, and the intermediate sampling within each cycle is run deterministically so the reconstruction is well defined.

4.2 Mask-Guided Contrastive Learning

Standard contrastive sampling often conflates particle and background regions, weakening structural discrimination under high noise. To overcome this limitation, we introduce mask-guided contrastive learning, where segmentation masks explicitly define particles as positives and background as negatives, improving fine-grained feature representation in edges, textures, and spatial noise patterns. The contrastive loss is defined as:

$$\mathcal{L}_{\mathrm{NCE}} = -\log \frac{\exp\left(\cos(q, k^+)/\tau\right)}{\exp\left(\cos(q, k^+)/\tau\right) + \sum_{k^-} \exp\left(\cos(q, k^-)/\tau\right)}, \tag{11}$$

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

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

The overall training objective combines all components where γ are empirically tuned weights:

$$\mathcal{L} = \gamma_{\mathrm{cyc}}\mathcal{L}_{\mathrm{cyc}} + \gamma_{\mathrm{NCE}}\mathcal{L}_{\mathrm{NCE}}. \tag{12}$$

Because the translation is unpaired, no ground-truth noisy target exists per clean input; both generators are therefore supervised via the cycle-consistency and mask-guided contrastive objectives rather than a per-image denoising loss, as is standard for diffusion-based unpaired translation.

5 Experiments

Datasets. We trained CryoCCD using synthetic particles derived from 15 publicly available EMPIAR datasets (Iudin et al., 2023) and the details are provided in Appendix B.1. For visual quality evaluation purposes, we selected *six* datasets: 1) **TRPV1** (Liao et al., 2013), a tetrameric membrane channel resolved at 3.4 Å; 2) **β-galactosidase** (Bartesaghi et al., 2015), a 2.2 Å map with ∼800 solvent peaks; 3) **Rhino/enterovirus** (Abdelnabi et al., 2019), small icosahedral virions requiring symmetry-expansion and high-fidelity difference mapping to resolve the conserved VP1-VP3 pocket; 4) **Innexin-6** (Burendei et al., 2020), nanodisc-embedded hemichannels where lipid-induced pore closure demands focused classification and membrane-signal subtraction; 5) **MLA complex** (Mann et al., 2021), apo/ATP/ADP states that necessitate heterogeneous 3-D clustering and multi-body refinement to trace lipid transport; and 6) **GroEL** (Godek et al., 2024), a D7-symmetry chaperonin. Comprehensive descriptions of all datasets are provided in Appendix B.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** (Zhang et al., 2024), **CycleDiffusion** (Wu & De la Torre, 2022) and **CycleNet** (Xu et al., 2023) as deep generative baselines. In Appendix B.2, we detail their specific settings.

Implementation Details. All experiments are conducted on a NVIDIA A100 GPU and trained for 50 epochs. The training dataset consists of 500 synthetic and 500 real images, with all images standardized to a resolution of 1024×1024 pixels. For the six datasets also used for visual-quality evaluation, we use an 80/20 micrograph-level split (the disjoint 20% is used only for Table 1; Appendix B.1), so training and evaluation micrographs never overlap. In addition to this mixed-dataset configuration, we also train CryoGEM and CryoCCD *separately* on each dataset without mixing EMPIAR sources, in order to assess their performance in the per-dataset setting. Further implementation details of sampler setups are provided in Appendix B.3.

5.1 Visual Quality

We compared CryoCCD against various baselines through visual assessment on the evaluation datasets, as shown in Table 1. As shown in Fig. 5, CryoCCD produces more faithful micrographs with superior noise characteristics while preserving structural details such as particle boundaries and particle–background contrast. In contrast, conventional approaches (Gaussian, Poisson, and Poisson-Gaussian) fail to capture the complex noise patterns in real data. Among deep generative models, CycleDiffusion often blurs particle boundaries, while CryoGEM and CycleNet collapse on our composite synthetic datasets, failing to generate recognizable particles. In the separately trained setting, the reduced complexity alleviates the collapse issue for CryoGEM and noticeably improves its results. However, CryoCCD consistently remains the top-performing method.

Quantitatively, we report both Fréchet Inception Distance (FID) and Class-Mean Maximum Discrepancy with learned features (CMMD) (Jayasumana et al., 2024) to complement FID. Table 1 compares these two metrics across six cryo-EM datasets and shows that CryoCCD achieves the lowest scores on both metrics compared with traditional methods, the recent GAN-based baseline (CryoGEM), and diffusion-based baselines (CycleDiffusion, CycleNet).

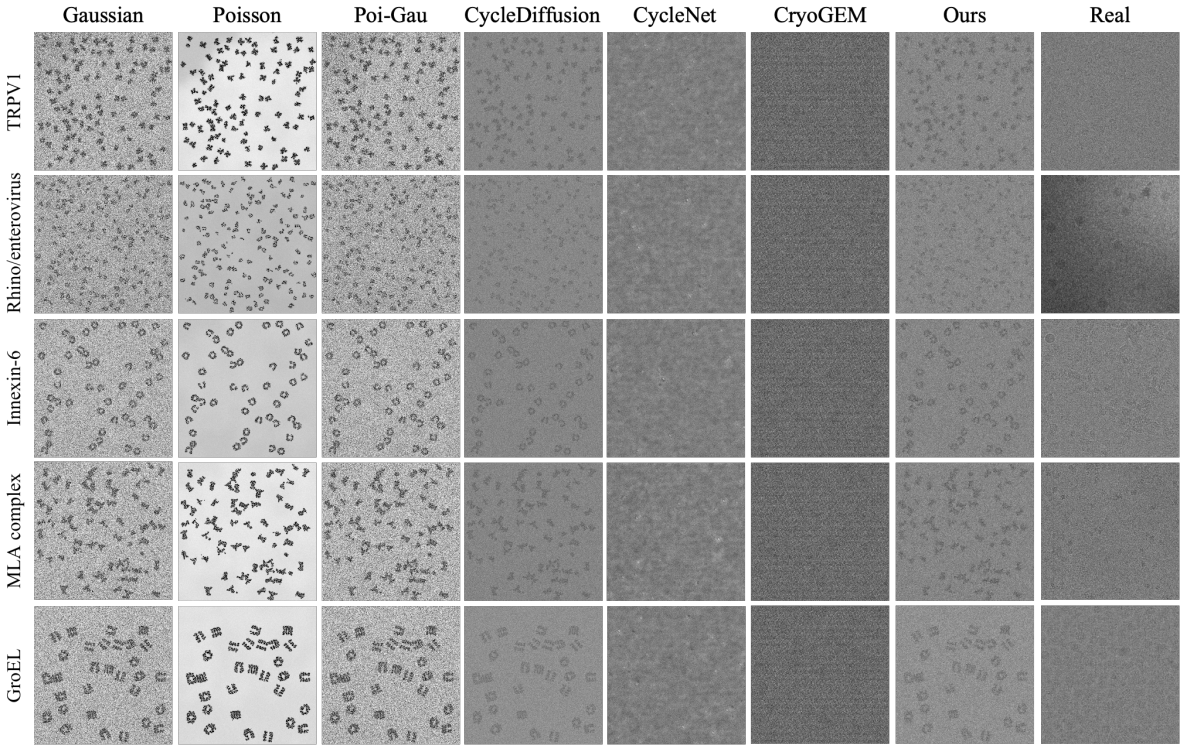

Figure 5: Comparison between real images and generated images. Our method produces more authentic micrographs with superior noise characteristics across all datasets.

Step-wise visualizations of the CryoCCD pipeline are provided in Fig. 6. Extracted structures are embedded into 3D virtual volumes, which are projected to noise-free micrographs with masks. Noise is then generated on the micrographs to produce the final simulated cryo-EM that, shown alongside real micrographs, closely match their overall appearance and noise characteristics.

Table 1: Quantitative comparison of FID and CMMD on six Cryo-EM datasets (lower is better, **ST** means separately trained).

Dataset	**TRPV1**		**β-galactosidase**		**Rhino/enterovirus**		**Innexin-6**		**MLA complex**		**GroEL**	
	FID↓	**CMMD↓**	**FID↓**	**CMMD↓**	**FID↓**	**CMMD↓**	**FID↓**	**CMMD↓**	**FID↓**	**CMMD↓**	**FID↓**	**CMMD↓**
Gaussian	287.04	2.81	351.27	2.96	292.56	2.74	328.17	2.38	225.20	1.80	364.28	1.87
Poisson	396.59	3.19	515.91	2.89	395.59	3.58	467.19	2.66	388.26	2.57	410.38	2.33
Poi–Gau	290.68	2.86	356.96	3.01	301.77	2.77	272.24	2.28	239.43	1.80	369.95	1.87
CycleDiffusion	199.47	2.20	247.09	2.04	210.22	1.98	208.23	1.95	197.15	2.00	270.58	2.62
CryoGEM	192.28	2.17	283.28	2.73	224.18	3.41	223.02	3.76	189.46	2.93	279.20	2.77
CryoGEM (ST)	61.70	0.74	70.11	0.79	68.87	0.73	82.22	0.82	54.05	0.69	87.50	0.85
CycleNet	211.30	2.33	235.02	2.09	228.79	3.10	215.66	2.47	203.50	2.18	290.41	2.86
Ours	**121.32**	**1.53**	**137.57**	**1.66**	**130.89**	**1.63**	**130.52**	**1.46**	**88.55**	**0.94**	**147.41**	**1.32**
Ours (ST)	**57.42**	**0.71**	**60.13**	**0.70**	**51.88**	**0.67**	**70.57**	**0.75**	**45.09**	**0.67**	**82.52**	**0.77**

5.2 Particle Picking

To validate the realism of the noise characteristics, we evaluated its impact on particle picking. Specifically, we selected several publicly available EMPIAR datasets with particle annotations to conduct this evaluation. Fake real micrographs generated by CryoCCD and multiple baselines were incorporated into Topaz (Bepler et al., 2019). Qualitatively, particle-picking visualizations demonstrated that Topaz models finetuned with CryoCCD-generated noisy synthetic data provided superior discrimination between true particles and background noise. Improvements were confirmed through the Area Under the Precision-Recall Curve (AUPRC) metric (Bepler et al., 2019; Wagner et al., 2019), where our method consistently outperformed all other baselines; a standard CryoSPARC (Punjani et al., 2017) single-particle workflow further

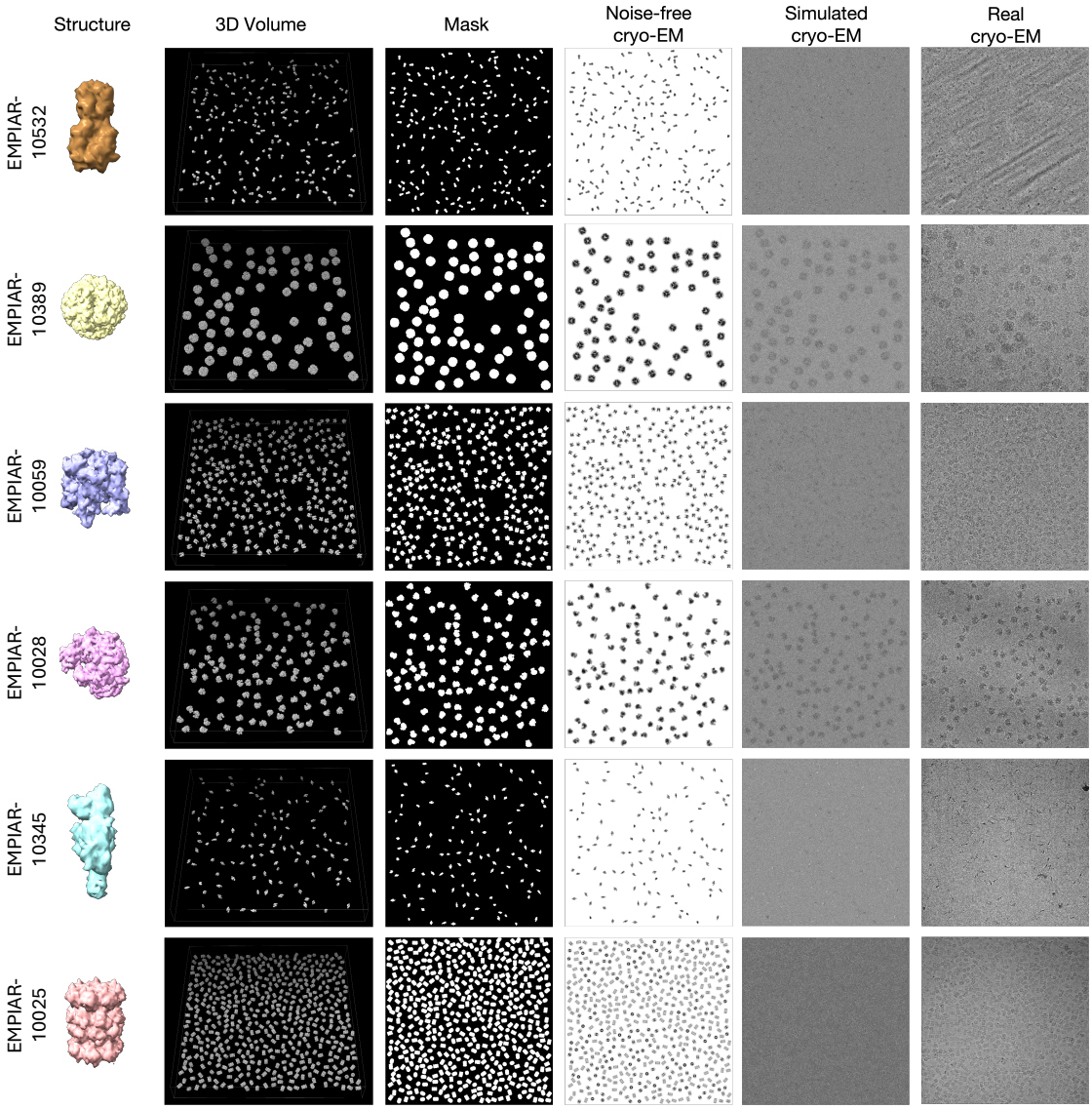

Figure 6: Visual examples from the CryoCCD biophysical simulation pipeline.

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

Method	**AUPRC↑**				**Precision↑**			
	Proteasome	Integrin	PhageMS2	HumanBAF	Proteasome	Integrin	PhageMS2	HumanBAF
Gaussian	0.463	0.233	0.575	0.449	0.412	0.204	0.538	0.397
Poisson	0.458	0.195	0.408	0.392	0.386	0.168	0.365	0.341
Poi–Gau	0.438	0.244	0.601	0.376	0.397	0.227	0.519	0.347
CycleDiffusion	0.357	0.233	0.420	0.369	0.321	0.196	0.347	0.326
CryoGEM	0.257	0.208	0.159	0.178	0.221	0.175	0.148	0.159
CycleNet	0.343	0.251	0.399	0.297	0.307	0.234	0.340	0.303
Topaz	0.301	0.510	0.317	0.487	0.279	0.452	0.286	0.462
Ours	**0.479**	**0.532**	**0.797**	**0.588**	**0.437**	**0.497**	**0.736**	**0.618**

confirms improved reconstruction resolutions (Appendix C.1). Fig. 7 shows our visualization results, and metrics are documented in Table 2 and Appendix C.1.

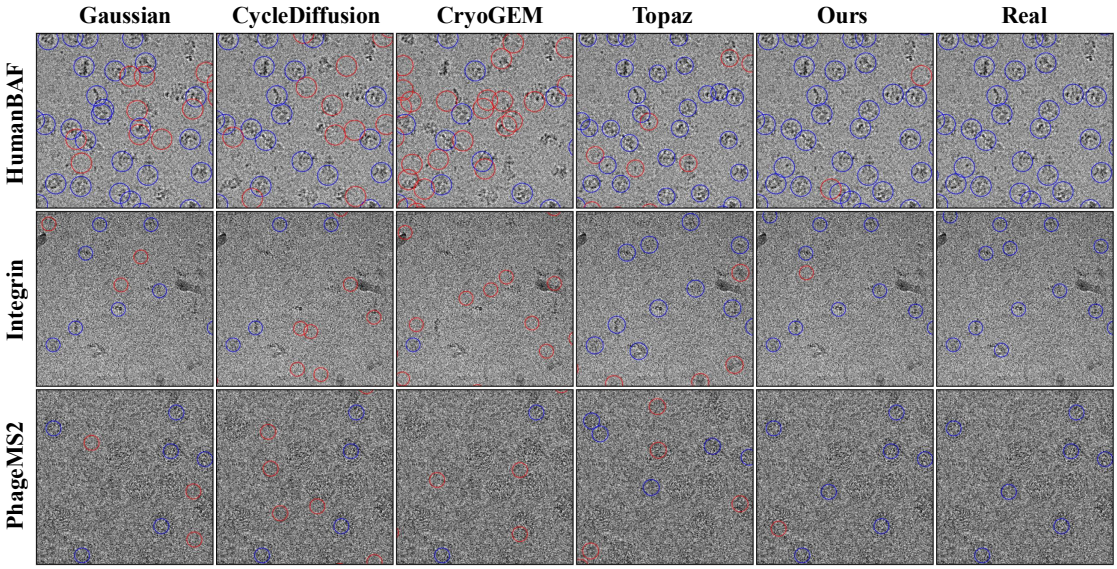

Figure 7: Qualitative comparison of particle picking. Blue circles indicate correct picks and red circles denote misses or false picks. Finetuning on CryoCCD-generated cryo-EM images the highest picking accuracy.

5.3 Pose Estimation

Accurate particle orientation determination represents a critical bottleneck in achieving high-resolution cryo-EM reconstructions. We evaluated CryoCCD's capacity to enhance pose estimation by applying our noisy synthetic datasets to train the ab-initio reconstruction methods CryoFIRE (Levy et al., 2022). The synthetic particles, complete with precise orientation annotations, provided an ideal training corpus for this supervised learning paradigm. We subsequently assessed performance on real micrographs using Filter Back-Projection (FBP) as our reconstruction algorithm. Our method demonstrated improvements over other baselines, achieving higher rotation accuracy and enhanced resolution quality across several datasets. Table 3 presents the pose estimation results, while Fig. 8 in Appendix C.2 provides the corresponding FSC curves.

Table 3: Quantitative comparison of pose estimation. Our approach achieves the best in Res(px)↓ and Rot.(rad)↓ metrics.

Method	**Res(px)↓**				**Rot.(rad)↓**			
	Proteasome	Integrin	PhageMS2	HumanBAF	Proteasome	Integrin	PhageMS2	HumanBAF
Gaussian	3.03	7.01	5.96	7.08	0.51	1.49	0.99	1.58
Poisson	3.12	7.14	6.09	9.33	1.09	1.66	1.03	1.60
Poi–Gau	3.15	7.99	5.98	9.09	1.32	1.57	0.73	1.51
CycleDiffusion	3.67	9.04	6.33	9.02	0.47	1.41	0.97	1.67
CryoGEM	7.99	9.24	15.38	10.79	1.72	1.94	1.20	1.78
CycleNet	4.58	8.31	9.04	8.91	1.03	1.50	1.19	1.69
CryoFIRE	6.02	13.27	18.03	7.17	1.54	0.97	0.75	1.49
Ours	**2.87**	**5.89**	**5.37**	**7.01**	**0.44**	**0.90**	**0.51**	**1.47**

5.4 Generalization to Held-Out Protein Families

To demonstrate generalization, we evaluate CryoCCD on four EMPIAR datasets not used during training. Three of them—EMPIAR-10028 (Wong et al., 2014), EMPIAR-10389 (Righetto et al., 2020), and EMPIAR-10532 (Tan & Rubinstein, 2020)—have no protein-family overlap with training and are thus held-out protein families; EMPIAR-10059 (Gao et al., 2016) (TRPV1) is instead a held-out imaging condition, since the training set already includes a TRPV1 dataset (Appendix B.1). As shown in Table 4, CryoCCD consistently achieves the lowest FID and CMMD against Gaussian, Poisson, Gaussian–Poisson mixture, CryoGEM, and CycleNet, indicating strong generalization to unseen protein families and imaging conditions.

Table 4: Generalization to held-out protein families: FID↓ and CMMD↓ on four EMPIAR datasets not used during training.

Dataset	EMPIAR-10028		EMPIAR-10059		EMPIAR-10389		EMPIAR-10532	
	FID↓	CMMD↓	FID↓	CMMD↓	FID↓	CMMD↓	FID↓	CMMD↓
Gaussian	290.34	2.85	301.07	2.90	275.46	2.62	299.01	2.87
Poisson	443.13	3.43	389.78	3.31	391.27	3.36	410.98	3.50
Poi–Gau	286.46	2.77	300.98	2.87	277.04	2.64	307.87	2.91
CryoGEM	232.69	2.01	247.58	2.08	244.33	2.05	280.61	2.74
CycleNet	210.47	1.94	227.48	2.03	230.09	1.81	222.56	2.10
Ours	**135.77**	**1.67**	**140.54**	**1.79**	**131.90**	**1.60**	**158.75**	**1.84**

5.5 Ablation Study

Impact of Diffusion model components. We further ablate learning modules of the reconstruction model: *cycle consistency*, *mask-guided contrastive learning*, and *predictor-corrector sampler*. Table 5 reports FID on six datasets. Removing cycle consistency substantially worsens FID and disrupts particle localization after noise augmentation; removing mask-guided contrastive learning also degrades FID and yields blurred boundaries with poor foreground–background separation. We include visual comparisons in the supplement to illustrate these effects.

Impact of Biophysical modeling. To assess the impact of our physics-informed imaging components, we performed a quantitative ablation study focusing on the *ice layer* model and the *CTF* parameterization. We found that both elements are crucial for high-fidelity image simulation. We show the results in Table 5 and detail the analysis in Appendix D.1.

Impact of Samplers and Sampling Steps. We further analyze the effect of different diffusion sampling steps and samplers. Increasing the step budget from 5 to 50 yields a clear FID improvement. The complete quantitative results are reported in Table 7 in Appendix D.2.

Table 5: **Ablation on diffusion and biophysical components.** Results (FID↓) across six datasets, where CC denotes the Cycle Consistency component, MGCL denotes Mask-guided Contrastive Learning, PCS denotes Predictor-Corrector Sampler, GL-Ice refers to the global-gradient ice-layer model, CTF-S refers to the small-defocus CTF condition, and CTF-L refers to the large-defocus CTF condition.

Setting	TRPV1	β-gal.	Rhino	Innexin-6	MLA	GroEL
w/o CC	243.38	254.97	270.01	248.53	199.79	300.35
w/o MGCL	155.95	163.31	159.44	169.01	122.57	191.49
w/o PCS	147.75	150.26	155.37	148.45	100.63	177.44
GL-Ice	199.84	270.32	212.09	228.65	197.52	284.43
CTF-S	149.33	151.79	163.31	147.28	104.01	180.43
CTF-L	148.47	156.90	166.44	150.80	107.09	184.47
Ours	**121.32**	**137.57**	**130.89**	**130.52**	**88.55**	**147.41**

6 Conclusion

In this paper, we introduce CryoCCD, a synthesis framework that integrates biophysical modeling with conditional cycle-consistent predictor-corrector diffusion to generate structurally diverse and noise-realistic cryo-EM micrographs. Our method achieves state-of-the-art performance across the evaluated visual-quality benchmarks and two downstream tasks on six public datasets. Moreover, CryoCCD demonstrates generalization to held-out datasets spanning unseen protein families and imaging conditions. This framework provides a high-fidelity simulation tool that may reduce the manual annotation burden by amplifying limited real annotations into a larger, more diverse synthetic corpus, accelerating the development of computational tools for the cryo-EM community.

References

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.

Josh Abramson, Jonas Adler, Jack Dunger, Richard Evans, Tim Green, Alexander Pritzel, Olaf Ronneberger, Lindsay Willmore, Andrew J Ballard, Joshua Bambrick, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, 630(8016):493–500, 2024.

Xiao-Chen Bai, Greg McMullan, and Sjors HW Scheres. How cryo-EM is revolutionizing structural biology. *Trends in biochemical sciences*, 40(1):49–57, 2015.

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.

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, November 2019. ISSN 1548-7105. doi: 10.1038/s41592-019-0575-8. URL `https://doi.org/10.1038/s41592-019-0575-8`.

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.

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.

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. doi: 10.7554/eLife.06380. URL `https://elifesciences.org/articles/06380`.

Yifan Cheng. Single-particle cryo-EM at crystallographic resolution. *Cell*, 161(3):450–457, 2015.

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

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*, pp. 8789–8797, 2018.

UniProt Consortium. UniProt: a worldwide hub of protein knowledge. *Nucleic acids research*, 47(D1): D506–D515, 2019.

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. doi: 10.1038/s41594-018-0093-x. URL `https://pubmed.ncbi.nlm.nih.gov/30061598/`.

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.

DJ DeRosier and PB Moore. Reconstruction of three-dimensional images from electron micrographs of structures with helical symmetry. *Journal of molecular biology*, 52(2):355–369, 1970.

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.

Prafulla Dhariwal and Alexander Nichol. Diffusion models beat GANs on image synthesis. *Advances in neural information processing systems*, 34:8780–8794, 2021.

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.

Jacques Dubochet, Marc Adrian, Jiin-Ju Chang, Jean-Claude Homo, Jean Lepault, Alasdair W McDowall, and Patrick Schultz. Cryo-electron microscopy of vitrified specimens. *Quarterly reviews of biophysics*, 21 (2):129–228, 1988.

Yuan Gao, Erhu Cao, David Julius, and Yifan Cheng. TRPV1 structures in nanodiscs reveal mechanisms of ligand and lipid action. *Nature*, 534(7607):347–351, 2016.

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.

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.

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.

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

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.

Richard Henderson. Overview and future of single particle electron cryomicroscopy. *Archives of biochemistry and biophysics*, 581:19–24, 2015.

Benjamin Himes and Nikolaus Grigorieff. Cryo-TEM simulations of amorphous radiation-sensitive samples using multislice wave propagation. *IUCrJ*, 8(Pt 6):943–953, 2021. doi: 10.1107/S2052252521008538. URL `https://doi.org/10.1107/S2052252521008538`.

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

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

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.

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.

Sadeep Jayasumana, Srikumar Ramalingam, Andreas Veit, Daniel Glasner, Ayan Chakrabarti, and Sanjiv Kumar. Rethinking fid: Towards a better evaluation metric for image generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9307–9315, 2024.

Runmin Jiang, Jackson Daggett, Shriya Pingulkar, Yizhou Zhao, Priyanshu Dhingra, Daniel Brown, Qifeng Wu, Xiangrui Zeng, Xingjian Li, and Min Xu. BOE-ViT: Boosting Orientation Estimation with Equivariance in Self-Supervised 3D Subtomogram Alignment. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 29352–29362, 2025.

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. ISSN 2052-2525. doi: 10.1107/S2052252524009321. URL `https://doi.org/10.1107/S2052252524009321`. Open Access.

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*, pp. 18260–18269, 2022.

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.

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.

Junho Kim, Minjae Kim, Hyeonwoo Kang, and Kwanghee Lee. U-GAT-IT: Unsupervised Generative Attentional Networks with Adaptive Layer-Instance Normalization for Image-to-Image Translation, 2020. URL `https://arxiv.org/abs/1907.10830`.

Dari Kimanius, Liyi Dong, Grigory Sharov, Takanori Nakane, and Sjors HW Scheres. New tools for automated cryo-EM single-particle analysis in RELION-4.0. *Biochemical Journal*, 478(24):4169–4185, 2021.

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. doi: 10.1038/ncomms12524. URL `https://www.nature.com/articles/ncomms12524`.

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

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)*, pp. 35–51, 2018.

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.

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.

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.

Ting-Yu Lin, Leon Kleemann, Jakub Jeżowski, Dominika Dobosz, Michał Rawski, Paulina Indyka, Grzegorz Ważny, Rahul Mehta, Andrzej Chramiec-Głąbik, Łukasz Koziej, et al. The molecular basis of tRNA selectivity by human pseudouridine synthase 3. *Molecular cell*, 84(13):2472–2489, 2024.

Luping Liu, Yi Ren, Zhijie Lin, and Zhou Zhao. Pseudo numerical methods for diffusion models on manifolds. *arXiv preprint arXiv:2202.09778*, 2022.

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*, NIPS'17, pp. 700–708, Red Hook, NY, USA, 2017. Curran Associates Inc. ISBN 9781510860964.

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.

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.

Cheng Lu, Yuhao Zhou, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu. Dpm-solver++: Fast solver for guided sampling of diffusion probabilistic models. *Machine Intelligence Research*, pp. 1–22, 2025.

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

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.

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.

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. doi: 10.1109/TMI.2024.3398401.

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. doi: 10.1016/j.cell.2020.09.051. URL `https://www.cell.com/cell/fulltext/S0092-8674(20)31248-4`.

Seonghui Min, Hyun-Jic Oh, and Won-Ki Jeong. Co-synthesis of histopathology nuclei image-label pairs using a context-conditioned joint diffusion model. In *European Conference on Computer Vision*, pp. 146–162. Springer, 2024.

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.

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

Alex J Noble. Virtualice: Half-synthetic cryoem micrograph generator. *bioRxiv*, pp. 2024–09, 2024.

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

Mateusz Olek, Kevin Cowtan, Donovan Webb, Yuriy Chaban, and Peijun Zhang. IceBreaker: Software for high-resolution single-particle cryo-EM with non-uniform ice. *Structure*, 30(4):522–531, 2022.

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*, pp. 319–345. Springer, 2020.

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.

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.

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.

Riccardo D. Righetto, Lucien Anton, Ricardo Adaixo, Roman P. Jakob, Jasenko Zivanov, Mohammad A. Mahi, Philippe Ringler, Torsten Schwede, Timm Maier, and Henning Stahlberg. High-resolution cryo-EM structure of urease from the pathogen Yersinia enterocolitica. *Nature Communications*, 11:5101, 2020. doi: 10.1038/s41467-020-18870-2. URL `https://doi.org/10.1038/s41467-020-18870-2`.

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*, pp. 10684–10695, 2022.

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*, pp. 33–49. Springer, 2020.

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.

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.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*, 2020.

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

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.

Thorsten Wagner, Felipe Merino, Markus Stabrin, Toshio Moriya, Claudia Antoni, Amir Apelbaum, Philine Hagel, Oleg Sitsel, Tobias Raisch, Daniel Prumbaum, et al. SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. *Communications biology*, 2(1):218, 2019.

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*, pp. 14020–14029, 2021.

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

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.

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.

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. ISSN 1664-042X. doi: 10.3389/fphys.2022.760404. URL `https://www.frontiersin.org/journals/physiology/articles/10.3389/fphys.2022.760404`.

wwPDB consortium. Protein Data Bank: the single global archive for 3D macromolecular structure data. *Nucleic Acids Research*, 47(D1):D520–D528, 10 2018. ISSN 0305-1048. doi: 10.1093/nar/gky949. URL https://doi.org/10.1093/nar/gky949.

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.

Sihan Xu, Ziqiao Ma, Yidong Huang, Honglak Lee, and Joyce Chai. Cyclenet: Rethinking cycle consistency in text-guided diffusion for image manipulation. *Advances in Neural Information Processing Systems*, 36: 10359–10384, 2023.

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.

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*, pp. 2849–2857, 2017.

Jiakai Zhang, Qihe Chen, Yan Zeng, Wenyuan Gao, Xuming He, Zhijie Liu, and Jingyi Yu. CryoGEM: Physics-Informed Generative Cryo-Electron Microscopy. *Advances in Neural Information Processing Systems*, 37:63222–63249, 2024.

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*, pp. 3836–3847, 2023.

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. ISSN 1549-9596. doi: 10.1021/acs.jcim.9b01176. URL https://doi.org/10.1021/acs.jcim.9b01176. Published on 2020/05/26.

Wenliang Zhao, Lujia Bai, Yongming Rao, Jie Zhou, and Jiwen Lu. Unipc: A unified predictor-corrector framework for fast sampling of diffusion models. *Advances in Neural Information Processing Systems*, 36: 49842–49869, 2023.

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*, pp. 16407–16417, 2021.

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.

Ellen D Zhong, Tristan Bepler, Bonnie Berger, and Joseph H Davis. CryoDRGN: reconstruction of heterogeneous cryo-EM structures using neural networks. *Nature methods*, 18(2):176–185, 2021a.

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*, pp. 4066–4075, 2021b.

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*, pp. 2223–2232, 2017a.

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, 2017b.

A Details of Biophysical Modeling

A.1 Implementation Details

Structure Library Construction. The structure-based mining targets cryo-EM-derived entries from the Protein Data Bank using a resolution cutoff <4.0 Å with molecular weight and symmetry classification spanning biological assemblies across diverse size scales and point group symmetries. AlphaFold3 integration addresses structural coverage gaps and increases heterogeneity by generating predicted models for sequences that lack experimental structures, while expanding conformational heterogeneity by predicting and modeling structures from sequences with potential conformational diversity. The structures are processed through confidence-stratified sampling: very high confidence regions (pLDDT>90) remain static, confident regions (70-90) undergo constrained perturbations, low confidence regions (50-70) experience enhanced sampling, and very low confidence regions (<50) are treated as fully flexible. Interdomain flexibility employs PAE-guided rigid-body sampling for regions with PAE>10 Å, generating structural ensembles that capture prediction uncertainty while preserving fold topology. The pipeline for enhancing compositional and conformational heterogeneity is illustrated in Fig. 2.

Atomic Model Conversion. The atomic model to volume conversion employs element-specific van der Waals radii based on established crystallographic databases, with different radii assigned to reflect the electronic structure and coordination environments of various atomic species. For unrecognized elements, a conservative default radius is applied based on typical organic atom dimensions. The Gaussian density kernel utilizes width $\sigma = R_{vdW}/(2 \cdot \text{spacing})$ where voxel spacing follows spacing = resolution/2.0 to ensure Nyquist-compliant sampling. Box margin calculation applies margin $= \max(3R_{max}, 2 \cdot \text{resolution})$ with automatic expansion by $4R_{max}$ when boundary violations exceed 10% of total atoms. Post-processing includes Gaussian smoothing with $\sigma_{smooth} = \text{resolution}/(2.0 \cdot \text{spacing})$ and background thresholding at 0.005 relative intensity to eliminate noise while preserving structural features.

Scale-Adaptive Parameterization. The automatic scale detection employs size-based classification with specific thresholds, particles <10 Å receive scale factor $s = 1.0$, 10-50 Å receive $s = 0.8$, 50-200 Å receive $s = 0.6$, 200-1000 Å receive $s = 0.4$, and >1000 Å receive $s = 0.2$. The composite scale factor follows $s = 0.7s_{size} + 0.3s_{density}$ where density component reflects local crowding through $s_{density} = \min(1.0, \text{particle_size}^3/(\text{volume_size}/1000))$. Scale-dependent parameter relationships include overlap threshold $= 0.4 - 0.3s$, placement density $= 0.7 + 0.5s$, collision strictness $= 0.5 + 0.5s$, and mesh reduction factor $= 0.7 - 0.7s$. These scaling laws ensure appropriate geometric constraints while optimizing computational resources across biological size ranges. Fig. 6 illustrates the generated 3D volumes rendered at varying scales.

Distribution Modeling. The placement implementation combines data-driven coordinates from RELION with synthetic insertion algorithms. Data-driven placement converts 2D picks (x, y) to 3D translations T_{exp} through coordinate system transformation, while Euler angles (α, β, γ) undergo conversion to quaternions q_{exp} using standard attitude conversion conventions. Synthetic insertion employs multiple distribution strategies, uniform placement uses rejection sampling with collision avoidance radius $2R_{particle}(1 - \text{overlap_threshold})$ and maximum iteration limits; Gaussian clustering utilizes standard deviations $\sigma_{primary} = \min(\text{volume_shape})/6$ for concentrated regions and $\sigma_{secondary} = 0.7\sigma_{primary}$ for peripheral zones; grid placement applies spacing $d = 2R_{particle}(1 - \text{overlap_threshold}/2)$ with positional jitter $\pm d/4$; interface placement constrains particles within distance tolerance from target surfaces. Class-specific rules implement distinct organizational patterns. For example, soluble enzymes follow uniform dispersion, ribosomes cluster according to measured inter-particle distances to mimic polysomes, and viral capsids follow separation distributions to avoid unrealistic aggregation.

Orientational Sampling. Particle orientation sampling employs multiple strategies reflecting different biophysical constraints. Uniform sampling draws from rotation group $SO(3)$ using quaternion parameterization for isotropic coverage. Preferred axis alignment utilizes von Mises-Fisher distributions $f(\mathbf{x}; \boldsymbol{\mu}, \kappa) = \kappa \exp(\kappa \boldsymbol{\mu}^T \mathbf{x})/(4\pi \sinh(\kappa))$ with concentration parameter $\kappa = 10$ providing moderate orienta-

tional bias suitable for air-water interface effects. Limited tilt sampling constrains orientations within $\theta_{max} = \pi/6$ from vertical using truncated normal distributions, reflecting adsorption-induced preferences observed in cryo-EM. Numerical stability ensures quaternion normalization within appropriate tolerance with iterative renormalization when necessary.

Mesh Generation and Geometric Processing. Surface mesh construction employs marching cubes algorithm with adaptive step sizing, unit steps for high-resolution simulations ($s > 0.8$) and increased step sizes for coarse applications. Triangle mesh quality maintenance includes aspect ratio constraints, minimum angle thresholds, and edge length limits to ensure geometric fidelity. Smoothing operations apply $N_{iterations} = \max(1, 5s)$ iterations with relaxation factor 0.2. Mesh decimation utilizes topology-preserving edge collapse when reduction factor $0.6(1-s)$ exceeds threshold values. Surface normal computation employs area-weighted vertex averaging to maintain geometric accuracy. Collision detection uses octree structures with adaptive depth and leaf capacity based on scale factor, enabling efficient proximity queries during placement operations.

Ice-Layer Modeling. Vitrified ice simulation incorporates experimentally-derived parameters for realistic specimen conditions. Thickness distribution follows log-normal parameterization with $\mu = \ln(100)$ and $\sigma = 0.2$, yielding mean thickness 100 nm with variance matching experimental observations. Surface topography employs multi-octave Perlin noise with amplitude scaling $A_i = 5.0/2^i$ nm and fundamental wavelength $L = 10$ pixels, capturing fractal roughness characteristics observed in vitrified specimens. Density fluctuations apply spatially correlated Gaussian noise with amplitude $\sigma = 0.05\rho_{ice}$ and correlation length appropriate for amorphous ice structure. The ice density $\rho_{ice} = 0.92$ g/cm^3 represents vitrified water at cryogenic temperatures, with surface height constraints maintaining reasonable thickness ranges to prevent unrealistic topographic extremes.

Electrostatic Potential Assembly and Projection. The electrostatic potential calculation follows $\Phi(\mathbf{r}) = \sum_{i=1}^{L} \rho_i(R(q_i)^{-1}(\mathbf{r} - T_i))$ with atomic scattering factors incorporating appropriate approximations and relativistic corrections for high-voltage applications. Weak-phase projection $P(x, y) = \int_{-z_0/2}^{z_0/2} \Phi(x, y, z)\, dz$ employs numerical integration with adaptive step sizing. The contrast transfer function uses defocus values optimized for different resolution targets; in the Fourier domain, the CTF is $H(s, a) = -\big(\sqrt{1 - w^2}\, \sin\gamma(s, a) - w\, \cos\gamma(s, a)\big) \exp\big(-\frac{Bs^2}{4}\big)$ with $\gamma(s, a) = \frac{\pi}{2}(2\lambda\, \Delta Z\, s^2 + \lambda^3 C_s\, s^4) - \phi$, $s = \sqrt{k_x^2 + k_y^2}$, and $a = \mathrm{atan2}(k_y, k_x)$. Parameter specifications include spherical aberration $C_s = 2.7$ mm for conventional systems, amplitude contrast $w = 0.07$ for biological specimens, and B-factors reflecting radiation sensitivity and molecular flexibility. The phase function incorporates electron wavelength $\lambda = 12.2643247/\sqrt{kV(1 + 0.978466 \times 10^{-6} \cdot kV)}$ with relativistic corrections for high-voltage applications. An inverse Fourier transform of the filtered spectrum yields a noise-free micrograph used as the clean input for diffusion-based noise synthesis.

B Experiment Settings

B.1 Dataset Details

This section provides detailed information about EMPIAR datasets utilized in our experiments. These datasets represent a diverse range of molecular structures with various characteristics that present unique challenges for cryo-EM reconstruction. Among these datasets, 15 datasets serve as training data (EMPIAR-10005, 10061, 10083, 10199, 10289, 10363, 10366, 10421, 10425, 10431, 10551, 12003, 12330, 12598, 12667), which we merge into a single, large-scale corpus so that the model is exposed to diverse imaging conditions and can learn features that generalize beyond any one specimen. Six of these datasets also serve for visual-quality evaluation: TRPV1 (EMPIAR-10005), β-galactosidase (EMPIAR-10061), Rhino/enterovirus (EMPIAR-10199), Innexin-6 (EMPIAR-10289), MLA complex (EMPIAR-10425), and GroEL (EMPIAR-12667). For each of these six, we apply an 80/20 split at the *micrograph* level: the 80% partition is included in the training corpus, while the disjoint 20% is held out and used only for the visual-quality results in Table 1 and the corresponding figures; training and evaluation micrographs therefore never overlap. Additionally,

four publicly available cryo-EM datasets (EMPIAR-10025, 10075, 10345, and 10590) are used to validate our framework on the downstream tasks of particle picking and pose estimation. Finally, EMPIAR-10028 (*P. falciparum* 80S ribosome), EMPIAR-10389 (*Yersinia* urease), and EMPIAR-10532 (influenza HA) are used as held-out protein families, whereas EMPIAR-10059 (TRPV1) is a held-out imaging condition—an unseen sample of a protein family (TRPV1) already present in training via EMPIAR-10005—together demonstrating the generalization ability of CryoCCD.

EMPIAR-10005. Rat TRPV1 ion channel (Liao et al., 2013) (EMPIAR-10005) achieved a high-resolution result at 3.4 Å, breaking the side-chain resolution barrier for membrane proteins without crystallization. The structure exhibits four-fold symmetry around a central ion pathway formed by transmembrane segments 5-6 and the intervening pore loop, with a wide extracellular mouth and a short selectivity filter.

EMPIAR-10025. The Thermoplasma acidophilum 20S proteasome (Campbell et al., 2015) (EMPIAR-10025) dataset comprises 196 real micrographs and 49,954 manually curated particles, enabling reconstruction at a high resolution of 2.8 Å. The T20S proteasome is a 700 kDa complex composed of 14 α-subunits and 14 β-subunits arranged with D7 symmetry. The micrographs contain particles exhibiting high symmetry, random in-plane orientations, and frequent mutual occlusion, posing challenges for particle picking.

EMPIAR-10028. *Plasmodium falciparum* 80S ribosome bound to the anti-protozoan drug emetine (Wong et al., 2014) (EMPIAR-10028) was determined by cryo-EM at a resolution of 3.2 Å. The structure revealed drug–ribosome interactions in malaria parasites, with related PDB entries 3j79 and 3j7a and EMDB entry EMD-2660. The dataset contains 1081 multi-frame micrographs (4096×4096 pixels, 16 frames each) and 105,247 processed particles, totaling 1.2 TB. These data enabled detailed structural analysis of the malaria ribosome and its inhibition by small molecules.

EMPIAR-10059. TRPV1 ion channel in complex with DkTx and RTX, embedded in lipid nanodiscs (Gao et al., 2016) (EMPIAR-10059) was determined by cryo-EM at a resolution of 2.95 Å. The structure (PDB 5irx; EMDB EMD-8117) revealed mechanisms of ligand and lipid modulation of TRPV1. The dataset contains 1200 motion-corrected micrographs (3838×3710 pixels, 30-frame exposures) and 218,805 raw particle images (192×192 pixels), totaling 93.8 GB.

EMPIAR-10061. *E. coli* β-galactosidase with PETG inhibitor (Bartesaghi et al., 2015) was determined by cryo-EM at an average resolution of 2.2 Å. The map contains identifiable densities for approximately 800 water molecules and for magnesium and sodium ions, demonstrating that specimen preparation quality and protein flexibility are now the major bottlenecks to routinely achieving near 2 Å resolutions.

EMPIAR-10075. The MS2 bacteriophage(Koning et al., 2016) (EMPIAR-10075) dataset contains 300 real micrographs and 12,682 manually curated particles, enabling a reconstruction at 8.7 Å resolution. The virus comprises 178 copies of the coat protein, a single A-protein, and an encapsulated RNA genome. Due to its large size and structural variability, PhageMS2 presents a significant challenge for particle detection and classification.

EMPIAR-10083. P22 bacteriophage capsid (Hryc et al., 2017) presents a 3.3-Å map of this large and complex macromolecular assembly. The dataset demonstrates a computational procedure to derive an atomic model with annotated metadata, identifying previously undescribed molecular interactions between capsid subunits that maintain stability without cementing proteins or cross-linking found in other bacteriophages.

EMPIAR-10199. Rhino- and enterovirus capsid (Abdelnabi et al., 2019) revealed a previously unknown druggable pocket formed by viral proteins VP1 and VP3, conserved across entero-/rhinovirus species. This discovery led to the identification of inhibitor analogues with broad-spectrum activity against multiple virus groups, providing novel insights into viral entry biology.

EMPIAR-10289. *C. elegans* innexin-6 gap junction proteins (Burendei et al., 2020) shows the structures in an undocked hemichannel form. In the nanodisc-reconstituted structure of the wild-type INX-6 hemichannel, flat double-layer densities obstruct the channel pore, revealing insights into lipid-mediated amino-terminal rearrangement and pore obstruction upon nanodisc reconstitution.

EMPIAR-10345. The asymmetric αVβ8 integrin (Cormier et al., 2018) (EMPIAR-10345) dataset consists of 1,644 micrographs and 84,266 manually selected particles, yielding a 3.3 Å resolution structure. This com-

plex features the human $\alpha V\beta 8$ ectodomain bound to porcine L-TGF-$\beta 1$, and is characterized by significant conformational flexibility, particularly in the leg domain, as observed in the micrographs.

EMPIAR-10363. HemQ from *Geobacillus* (Bromberg et al., 2020) provides 2.3 Å cryo-EM reconstructions obtained under substantial optical aberrations. The dataset was collected on a 200 kV microscope and contains unaligned multiframe micrographs (258 images, 100 frames each) in MRCS format, with a pixel spacing of 0.36 Å. It demonstrates that near-atomic resolution can still be achieved in the presence of significant coma and astigmatism, serving as a valuable benchmark for evaluating the robustness of image reconstruction algorithms under imperfect imaging conditions.

EMPIAR-10366. Canine distemper virus F protein (Kalbermatter et al., 2020) presents the prefusion state of the CDV F protein ectodomain at 4.3 Å resolution. Stabilization was achieved by fusing the GCNt trimerization sequence at the C-terminal region and purifying with a morbilliviral fusion inhibitor, with the 3D map showing clear density for the ligand at the protein binding site.

EMPIAR-10389. Urease from the pathogen *Yersinia enterocolitica* (Righetto et al., 2020) (EMPIAR-10389) was determined by cryo-EM at 1.98 Å resolution (PDB 6yl3; EMDB EMD-10835). The high-resolution structure revealed detailed features of the multimeric urease complex and provided insights into its pathogenic function. The dataset contains motion-corrected micrographs, totaling 839.9 GB.

EMPIAR-10421. Human apoferritin light chain vitrified using the "Back-it-up" (BIU) method (Tan & Rubinstein, 2020) demonstrates a high-speed specimen preparation technique where solution sprayed onto one side of a holey grid is wicked through by a glass-fiber filter on the opposite side, enabling vitrification within tens of milliseconds. The dataset contains 3,168 Falcon IV movies (30 frames each, pixel size 0.816 Å) of apoferritin prepared using BIU on a 200 kV microscope, yielding a 2.0 Å reconstruction (EMD-21951). It provides both raw and dose-weighted aligned micrographs, as well as a final polished particle stack with Euler angles and shifts, serving as a benchmark for evaluating high-throughput cryo-EM data processing and motion correction algorithms.

EMPIAR-10425. *A. baumannii* MLA complex (Mann et al., 2021) presents cryo-EM maps of the core MlaBDEF complex in apo-, ATP- and ADP-bound states. The maps reveal multiple lipid binding sites in both cytosolic and periplasmic sides of the complex, with molecular dynamics simulations suggesting potential lipid trajectories across the membrane.

EMPIAR-10431. Human CDK-activating kinase (CAK) (Greber et al., 2020) contains the three-dimensional structure of this critical regulator of transcription initiation and the cell cycle. The dataset includes both the catalytic module structure and CAK in complex with covalently bound inhibitor THZ1, providing insights into assembly, CDK7 activation, and inhibitor binding for therapeutic compound design.

EMPIAR-10532. Influenza hemagglutinin (HA) trimer vitrified using the Back-it-up (BIU) device (Tan & Rubinstein, 2020) (EMPIAR-10532) was determined by cryo-EM at 2.9 Å resolution (PDB 6wxb; EMDB EMD-21954). The study demonstrated that through-grid wicking enables high-speed cryo-EM specimen preparation. The dataset contains 1,556 raw Falcon IV movies (4096×4096 pixels, 30 frames each), the corresponding aligned micrographs, and 128,305 refined particle images (256×256 pixels) with assigned Euler angles and shifts, totaling 1.2 TB.

EMPIAR-10551. Adeno-associated virus AAV-DJ (Xie et al., 2020) was determined to 1.56 Å resolution, nearly matching the highest resolutions ever attained through X-ray diffraction of virus crystals. At this exceptional resolution, most hydrogens are clearly visible, improving atomic refinement accuracy and revealing that hydrogen bond networks are quite different from those inferred at lower resolutions.

EMPIAR-10590. The endogenous human BAF complex(Mashtalir et al., 2020) (EMPIAR-10590) includes 300 micrographs and 62,493 manually picked particles, achieving a reconstruction resolution of 7.8 Å. This complex is known for its compositional and conformational heterogeneity, which increases the difficulty of automated particle picking.

EMPIAR-12003. Human pseudouridine synthase 3 (Lin et al., 2024) presents structures in apo form and bound to three tRNAs, showing how the symmetric PUS3 homodimer recognizes tRNAs and positions the target uridine. Combined with structure-guided mutations and transcriptome-wide Ψ site mapping,

this dataset provides the molecular basis for PUS3-mediated tRNA modification and its link to intellectual disabilities.

EMPIAR-12330. Yeast spliceosome with Fyv6 (Senn et al., 2024) contains a high-resolution (2.3 Å) structure of a product complex spliceosome. The structure reveals Fyv6 as the only second step factor contacting the Prp22 ATPase, with Fyv6 binding mutually exclusive with first step factor Yju2, supporting a model where their exchange facilitates exon ligation.

EMPIAR-12598. SARS-CoV-2 RdRp backtracking (Malone et al., 2021) uses cryo-EM, RNA-protein cross-linking, and molecular dynamics simulations to characterize viral RNA polymerase behavior. The results establish the mechanism by which the product RNA extrudes through the RdRp nucleoside triphosphate entry tunnel during backtracking, a process that may aid in proofreading and contribute to viral resistance against antivirals.

EMPIAR-12667. GroEL chaperonin with inhibitors (Godek et al., 2024) employed cryo-EM to establish the binding site of bis-sulfonamido-2-phenylbenzoxazoles at the GroEL ring-ring interface. The biochemical characterization showed potent inhibition of Gram-negative chaperonins but lower potency against Gram-positive organisms, validating this chaperone system as an antibiotic target against bacteria including ESKAPE pathogens.

B.2 Baseline Details

For all learning-based baselines (CycleDiffusion, CryoGEM, CycleNet), we use their official public implementations with default training and inference configurations, changing only the input data to our datasets. In the downstream evaluations, every noise-generation method is compared under an identical setup—the same downstream architecture, initialization, number of synthetic samples, optimization budget, threshold-selection procedure, and real test split—so that differences reflect the generated data rather than the training protocol.

Traditional Noise Simulations. These baselines include Gaussian, Poisson, and Poisson-Gaussian (Poi–Gau) noise simulations, representing commonly used traditional strategies in cryo-EM synthetic data generation. Following Vulovic et al. (Vulović et al., 2013), the Gaussian noise baseline introduces zero-mean additive Gaussian noise to simulate a mixture of readout noise, dark current, and background structural variations. The Poisson noise baseline captures the stochastic nature of electron detection under low-dose imaging conditions, reflecting quantum noise characteristics. The Poi–Gau baseline combines both noise types to model complex imaging artifacts. For fair comparison, the Signal-to-Noise Ratio (SNR) is uniformly set to 0.1 across all three methods.

CycleDiffusion (Wu & De la Torre, 2022). This method employs two diffusion models trained independently on synthetic and real cryo-EM domains, respectively. Each DDIM is trained on 500 images of resolution 512×512 over 10,000 steps with a batch size of 8. The translation from synthetic to real is achieved by running a cycle-consistent sampling loop without paired supervision. We use the publicly available implementation and default inference settings.

CryoGEM (Zhang et al., 2024). CryoGEM is a physics-informed generative model designed to produce synthetic cryo-EM images with realistic noise characteristics. To introduce authentic noise, CryoGEM employs an unpaired noise translation technique that leverages contrastive learning with a novel mask-guided sampling strategy. This approach effectively transforms the simulated images into ones with realistic noise patterns. All experiments are performed with the method's default settings.

CycleNet (Xu et al., 2023). CycleNet enables unpaired image-to-image translation with pre-trained diffusion models by enforcing cycle-consistency regularization. Built upon ControlNet (Zhang et al., 2023) and Stable Diffusion (Rombach et al., 2022), it performs a forward translation from source to target domain and a backward reconstruction to ensure semantic preservation. Compared to mask- or attention-based approaches, CycleNet achieves higher consistency and translation faithfulness, and demonstrates strong zero-shot generalization with limited training data.

Table 6: **Efficiency on a single NVIDIA A100.** Training time (s/epoch), inference time (s/img), and peak memory (MB).

Method	Train	Infer	Memory
CryoGEM	209	0.2	4150
CycleDiffusion	1267	3.1	54377
Ours	805	1.9	68341

B.3 Implementation Details

All experiments are conducted on a single NVIDIA A100 GPU with 80 GB of memory. The models are trained for 50 epochs using the Adam optimizer with a learning rate of 1×10^{-4} and a batch size of 4. To explore the effects of different generative sampling strategies, we evaluate four diffusion samplers within the CryoCCD framework: DDIM (Song et al., 2020), DDPM (Ho et al., 2020), and their accelerated variants, DPM-Solver (Lu et al., 2022), DPM-Solver++ (Lu et al., 2025), UniPC (Zhao et al., 2023), and linear multistep (LMS) sampler (Liu et al., 2022). These sampling methods are examined to understand the trade-off between synthesis quality and inference speed under various configurations.

Our synthetic data generation process begins with biophysical modeling to simulate noise-free cryo-EM micrographs. Training is fully unpaired: the real micrographs and the simulated noise-free images are drawn independently and are aligned only in distribution through cycle consistency, with no one-to-one correspondence between them. At inference, given a simulated noise-free image and its structural mask, the CryoCCD framework generates a realistic noisy image that closely mimics experimental conditions. This design enables controllable and interpretable simulation of complex imaging artifacts.

The training dataset is detailed in Section B.1, and we validate the utility of our generated data in two downstream tasks—particle picking (Section C.1) and pose estimation (Section C.2)—demonstrating the effectiveness and generalizability of the proposed framework.

B.4 Efficiency and Computational Complexity

We benchmark training time, inference latency, and peak memory on a single NVIDIA A100 GPU to clarify computational feasibility in Table 6. As expected for a lightweight GAN, CryoGEM attains the fastest speed and lowest memory, but suffers from instability and eventual collapse in our mixed datasets (see the CryoGEM column in Fig. 5). Compared with CycleDiffusion, CryoCCD reduces training time (805 s vs. 1267 s per epoch) and inference latency (1.9 s vs. 3.1 s per image) while yielding better visual and downstream performance. The efficiency gains stem from mask-guided contrastive learning and a lightweight discriminator, which provide strong structural guidance so fewer diffusion steps suffice. Our peak memory is higher due to these additional modules.

C Details of Downstream Task

C.1 Particle Picking

Topaz (Bepler et al., 2019) is a cryo-EM particle picking framework that assigns each pixel a probability of belonging to a particle. A user-defined threshold is then applied to retain high-confidence detections. We adopt a pre-trained Topaz model as our baseline without further fine-tuning. In Table 2, the "Topaz" row is this pre-trained model without fine-tuning, whereas every other row is Topaz fine-tuned on data generated by the corresponding method under an identical setup.

Evaluation via AUPRC. To compare different picking methods, we rank detection candidates by their confidence scores and select the top N predictions for analysis. Varying a threshold set $\{\tau_i\}_{i=1}^{n-1}$ partitions the confidence range into n levels, allowing computation of precision $\mathrm{Pr}(k)$ and recall $\Re(k)$ at each level.

The area under the resulting precision–recall curve (AUPRC) is defined as

$$\text{AUPRC} = \sum_{k=1}^{n} \Pr(k)\big(\Re(k) - \Re(k-1)\big), \tag{13}$$

where the precision $\Pr(k)$ is the fraction of true positives among predictions with probability $\geq \tau_k$, and the recall $\Re(k)$ is the fraction of true positives with probability $\geq \tau_k$ relative to the total number of true positives.

Evaluation via Precision. Beyond AUPRC, we also report the overall precision at a fixed confidence cutoff. Precision is computed as

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{14}$$

where TP and FP denote true and false positives, respectively, at the chosen threshold. This metric directly reflects the accuracy of detected particles without regard to recall.

Result Analysis. Table 2 summarizes both AUPRC and overall precision for each method. Overall, our detector exhibits consistent improvements over classical noise-model baselines (Gaussian, Poisson, Hybrid) as well as recent learning-based approaches (CycleDiffusion (Wu & De la Torre, 2022), CryoGEM (Zhang et al., 2024), Topaz (Bepler et al., 2019)). These gains manifest across both high-contrast specimens (Proteasome, PhageMS2) 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, directly benefiting downstream reconstruction workflows.

In terms of AUPRC, our method achieves the best results on all four datasets, notably improving over Topaz by +17.8% on PhageMS2 and +10.1% on HumanBAF. While traditional models fail to register meaningful precision at the fixed cutoff, our method maintains valid detection outputs across all test cases, as reflected in the non-zero precision scores. Although all baselines yield near-zero precision under the same conditions, our method preserves confident particle proposals and achieves superior localization accuracy, particularly on Integrin and HumanBAF, where noise and low contrast typically pose significant challenges.

As shown in Fig. 7, particle picking accuracy improves significantly when Topaz is fine-tuned using synthetic data generated by CryoCCD. Compared to other data synthesis methods such as Gaussian noise, CycleDiffusion, and CryoGEM, our approach produces more realistic cryo-EM micrographs that better reflect the structural and noise characteristics of experimental data. As a result, the fine-tuned Topaz model yields more accurate particle localization across both static (e.g., HumanBAF) and dynamic (e.g., Integrin, PhageMS2) specimens, closely matching the annotations in real datasets.

Beyond picking accuracy, we assess the downstream reconstruction quality of the CryoCCD-picked particles using the standard CryoSPARC single-particle workflow. The resulting gold-standard FSC (0.143) resolutions are 2.63 Å (Proteasome), 5.50 Å (Integrin), 7.12 Å (PhageMS2), and 7.69 Å (HumanBAF), confirming that the improved picks translate into high-quality 3D reconstructions. A fully matched cross-method reconstruction comparison under identical particle counts, symmetry, masks, and post-processing is left to future work.

C.2 Pose Estimation

In Table 3, the "CryoFIRE" row is CryoFIRE run in its native ab-initio manner on real data, whereas "Ours" is the CryoFIRE architecture trained on CryoCCD-generated data with ground-truth poses; all noise-generation baselines use identical downstream settings and their official implementations. Our approach predicts particle orientations and translations via direct supervised learning on synthetic data. Each training batch contains images with known ground-truth rotations $\{R_i^{\text{gt}}\}$ and translations $\{T_i^{\text{gt}}\}$. We minimize the pose loss:

$$\mathcal{L}_{\text{pose}} = \frac{1}{B}\sum_{i=1}^{B}\left[\frac{1}{9}\|R_i^{\text{gt}} - R_i^{\text{pred}}\|_F^2 + \frac{1}{2}\|T_i^{\text{gt}} - T_i^{\text{pred}}\|_1\right], \tag{15}$$

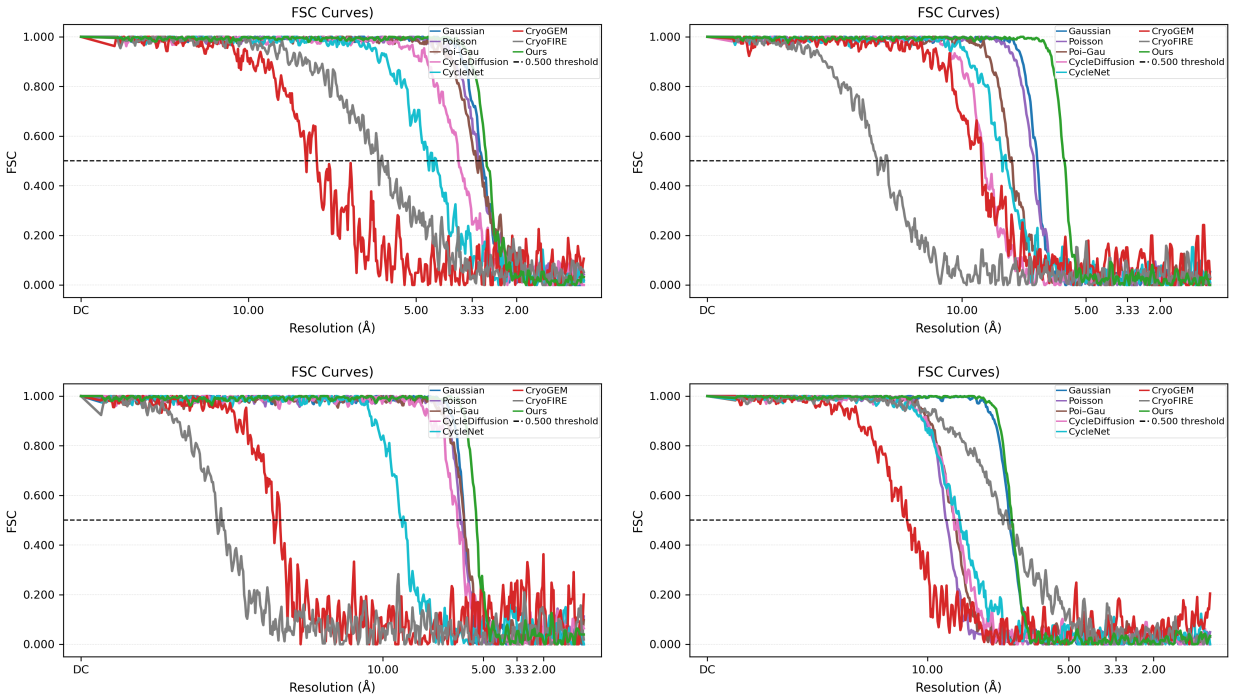

Figure 8: Filter back-projection reconstruction FSC curves at the 0.500 threshold.

where B is the batch size, $\|\cdot\|_F$ denotes the Frobenius norm, and $\|\cdot\|_1$ is the L_1 norm.

Evaluation via Reconstruction Resolution (Res(px)). Predicted poses reconstruct two independent half-volumes using filter back-projection (FBP). Resolution is determined by the Fourier shell correlation (FSC) at threshold 0.5:

$$FSC(r) = \frac{\sum_{r_i \in r} F_1(r_i) \cdot F_2(r_i)^*}{\sqrt{\sum_{r_i \in r} \|F_1(r_i)\|^2 \cdot \sum_{r_i \in r} \|F_2(r_i)\|^2}}, \tag{16}$$

where F_1 and F_2 are Fourier transforms of the two half-volumes. The resulting Res(px) measures pose accuracy without 2D classification bias.

Evaluation via Angular Error (Rot(rad)). We further compute the mean angular error in radians:

$$\text{Rot(rad)} = \frac{180}{\pi\, n_{\text{rots}}} \sum_{i=1}^{n_{\text{rots}}} \arccos\Big(\frac{R_i^{\text{gt}} v}{\|R_i^{\text{gt}} v\|} \cdot \frac{R_i^{\text{pred}} v}{\|R_i^{\text{pred}} v\|} \Big), \tag{17}$$

using unit vector $v = (0, 0, 1)$.

Result Analysis. Table 3 reports reconstruction resolution (Res(px)) and angular precision (Rot(rad)) for each baseline and our method. Our pose estimator consistently produces sharper reconstructions—reflected in lower Res(px) values—on both static (Proteasome, HumanBAF) and dynamic (Integrin, PhageMS2) 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. These overall trends highlight the robustness of our supervised training framework and its capacity to generalize across diverse structural classes.

Compared to CryoFIRE, our method yields consistent gains across all datasets in both resolution and rotation. In particular, we improve Integrin resolution by 55.6% (5.89 vs. 13.27 px) and reduce PhageMS2 rotation error by 32.0% (0.51 vs. 0.75 rad). We also see large margins on Proteasome—52.3% better

Table 7: **Quantitative comparison of FID scores on the *Ribosome*.** Varying DDPM steps vs. different samplers.

Configuration	Steps	Sampler	FID↓
5 Steps	5	DDPM	37.97
10 Steps	10	DDPM	29.08
20 Steps	20	DDPM	24.44
50 Steps	50	DDPM	20.06
DDIM	20	DDIM	20.33
DDPM	20	DDPM	24.44
DPM-Solver	20	DPM-Solver	19.27
DPM-Solver++	20	DPM-Solver++	19.13
LMS	20	LMS	18.68
UniPC (Ours)	20	UniPC	18.54

resolution (2.87 vs. 6.02 px) and 71.4% lower rotation error (0.44 vs. 1.54 rad)—and a smaller but positive gain on HumanBAF (resolution 7.01 vs. 7.17 px, rotation 1.47 vs. 1.49 rad). On average over the four datasets, our approach reduces resolution error by 45.1% and rotation error by 28.0% relative to CryoFIRE.

Classical noise baselines (Gaussian, Poisson, and their mixture) still underperform across datasets: despite occasionally competitive numbers (e.g., Gaussian resolution on HumanBAF), their average errors remain higher (Res: 5.77–6.55 px; Rot.: 1.14–1.35 rad) than ours (Res: 5.29 px; Rot.: 0.83 rad), reflecting limited noise realism. CryoGEM exhibits the largest errors overall (e.g., Res: 7.99–15.38 px; Rot.: 1.72–1.94 rad), suggesting adversarial distortions that harm both angular accuracy and resolution.

Notably, our method achieves the best performance across all evaluated datasets and metrics in Table 3. The margins vary with structural heterogeneity: they are largest on PhageMS2 (complex, flexible; Res: 70.2% better than CryoFIRE) and Proteasome (rigid; Rot.: 71.4% improvement), and narrower on HumanBAF, where we still edge out the next best baseline on both resolution and rotation. These results indicate that our model maintains low errors on both rigid and flexible structures, confirming robustness across structural heterogeneity. Consistent with Fig. 8, our advantage in resolution at the 0.500 FSC threshold on Proteasome persists relative to all competing baselines.

D Ablation Study

D.1 Biophysical modeling.

Our biophysical engine is implemented as an integrated pipeline; thus only the physics-informed imaging components can be isolated for quantitative ablations. Following this principle, we probe two factors: the *ice layer* model and the *CTF* parameterization. For the ice layer, we replace our physically-based simulation with a global-gradient scheme that randomly selects IceBreaker-estimated thickness maps (Olek et al., 2022). For CTF, we compare (i) **small defocus** (8,000–15,000 Å, 300–800 Å astigmatism), (ii) **large defocus** (10,000–30,000 Å, 500–2,500 Å), and (iii) **ours** (15,000–20,000 Å, 100–500 Å), which targets high-quality cryo-EM imaging. As Table 5 shows, replacing the ice layer with the global-gradient approach severely degrades FID across all datasets, confirming the effectiveness of our ice modeling. Across realistic CTF ranges, FID shifts are within $\leq 7\%$, and our setting is best or on par on five of six datasets (Innexin-6 shows a marginal 0.8% gain under small defocus).

D.2 Samplers and Sampling Steps.

We validate the effectiveness of different *sampling configurations* on the *Ribosome* dataset. Table 7 presents two complementary studies:

Sampling steps. We vary the DDPM step budget from 5 to 50. As expected, the FID steadily decreases as the number of steps increases: from 37.97 (5 steps) to 29.08 (10 steps), 24.44 (20 steps), and finally 20.06

(50 steps). This indicates that longer trajectories allow the denoising process to better approximate the data distribution, although with diminishing returns as the step budget grows.

Sampler algorithm. With the step budget fixed to 20, we compare several popular samplers. While DDPM achieves an FID of 24.44, advanced solvers bring consistent improvements: DPM-Solver and DPM-Solver++ reduce FID to 19.27 and 19.13, respectively, while LMS and UniPC achieve the best performance at 18.68 and 18.54. In contrast, DDIM (20.33) underperforms, likely due to its deterministic formulation, which limits stochastic exploration and reduces sample diversity.

E Background of Cryo-EM

E.1 Principles of Cryo-EM

Cryo-EM is a cornerstone of contemporary structural biology, enabling macromolecular structures to be determined at near-atomic resolution in a close-to-native environment (Dubochet et al., 1988; Nogales, 2016). By eliminating the need for crystallization or harsh chemical fixation, cryo-EM allows direct visualization of protein assemblies, viruses, and cellular organelles.

Cryo-EM enables structural imaging of radiation-sensitive biological specimens by operating a transmission electron microscope (TEM) under cryogenic conditions. Electrons, typically accelerated to 300kV, possess a de Broglie wavelength of approximately 0.02Å, vastly shorter than that of visible light (4000–7000Å), thus allowing for atomic-resolution imaging (Henderson, 2015). To mitigate the severe radiation damage caused by electron scattering, cryo-EM employs (i) vitrification via plunge-freezing into liquid ethane to form amorphous ice that preserves the native ultrastructure (Dubochet et al., 1988), and (ii) low-dose imaging strategies, where multiple noisy projections are later computationally averaged.

Modern TEM consists of a high-coherence electron source, condenser lenses that control beam illumination, an objective lens that focuses the electron wavefront after interacting with the specimen, and image magnification lenses. The specimen is maintained below $-170°$C to prevent devitrification during imaging. Transmitted electrons are collected by a direct electron detector, producing a set of 2D projection micrographs. According to the central projection theorem (DeRosier & Moore, 1970), the Fourier transform of each 2D image corresponds to a central slice of the object's 3D Fourier space. By acquiring thousands of such projections at varying orientations, it becomes possible to reconstruct a 3D structure via algorithms such as iterative back-projection or maximum-likelihood refinement (Cheng, 2018).

E.2 Major Applications

Single-particle cryo-EM merges tens of thousands of noisy projections into 3D density maps, often reaching 2–3 Å resolution for favourable specimens (Bai et al., 2015). Three computational stages are pivotal:

Particle picking. Automatic discrimination of particles from ice, carbon, and contaminants is essential for downstream accuracy.

Pose estimation. Assignment of particle orientations is typically achieved via common-line methods or projection matching, both of which are sensitive to noise and heterogeneity.

Demonstrated targets include membrane channels (TRPV1 (Liao et al., 2013; Gao et al., 2016)), soluble enzymes (*E. coli* β-galactosidase (Bartesaghi et al., 2015)), viral capsids, and chaperonins—each posing unique challenges related to symmetry, flexibility, and low SNR.

F Limitations and Broader Impact

Limitations. We evaluate visual realism with FID and CMMD, which are built on natural-image representations: we report FID for comparability with prior work and add CMMD as a complementary metric, but neither directly measures cryo-EM-specific properties such as CTF behavior, correlated detector noise, or local SNR. We therefore treat downstream task performance (particle picking and pose estimation) as the primary, application-relevant evidence of noise realism. Cryo-EM-specific analyses—for example, radial

and 2D power spectra in particle-free and particle-containing regions, background spatial autocorrelation, and CTF-parameter recovery—would further characterize the generated noise and are a natural direction for future work. In the downstream evaluation, running full pipelines (fine-tuning plus reconstruction) across all methods, datasets, and multiple random seeds is computationally prohibitive; we instead rely on the consistency of the improvements across four heterogeneous specimens and two distinct tasks, and note per-condition seed uncertainty as a limitation. Likewise, a fully matched cross-method reconstruction comparison under identical particle counts, symmetry, masks, and post-processing is left to future work.

Broader impact. CryoCCD is a simulation tool intended to reduce the manual annotation burden and accelerate the development of cryo-EM computational methods. Its particle placement incorporates RELION-derived picks and angles from real data, blended in a confidence-weighted manner with purely synthetic sampling; in this sense the method amplifies limited real annotations into a larger and more diverse synthetic corpus rather than generating fully independently of real data, and it still relies on existing annotations as priors. Because small differences in cryo-EM noise statistics can affect downstream structural interpretation, synthetic data produced by CryoCCD is intended to support, not replace, real experimental validation, and users should be aware that downstream models may learn simulator-specific biases.