Generative Priors for Cryo-EM Image Reconstruction

Anonymous Authors[1]

Abstract

Single-particle cryo-electron microscopy (cryo-EM) is the premier technique for determining 3D biomolecular structures, yet its reliance on hundreds of thousands of high-fidelity 2D images creates substantial data throughput bottlenecks. To address this, we formulate the recovery of full-resolution cryo-EM images from compressively sampled measurements as an inverse problem, solved via posterior sampling under a learned generative prior. By training a denoising diffusion probabilistic model (DDPM) on EMPIAR datasets, we capture the low-dimensional manifold of protein images and accurately reconstruct them from both spatial and Fourier-domain undersampled data. Our approach successfully recovers 2D images at compression factors up to 2× while strictly preserving the biological signal required for downstream structural analysis, including conformational heterogeneity identification and atomic model building. Ultimately, this work demonstrates that generative diffusion priors can decode highly compressed measurements without sacrificing the high-resolution biological signal necessary for structural biology, offering a robust computational pathway to accelerate cryo-EM workflows.

1. Introduction

Single particle cryogenic electron microscopy (cryo-EM) reconstructs three-dimensional (3D) protein structures from hundreds of thousands of two-dimensional (2D) particle images acquired on a direct electron detector (Yip et al., 2020; Nakane et al., 2020; Nogales, 2016). Modern direct electron detectors generate exceptionally large volumes of raw image data (Datta et al., 2021; Chua et al., 2022; Poger et al., 2023), and reducing it without sacrificing structural information is a natural inverse problem: given a compressed or undersampled measurement of an image, recover the full-resolution image.

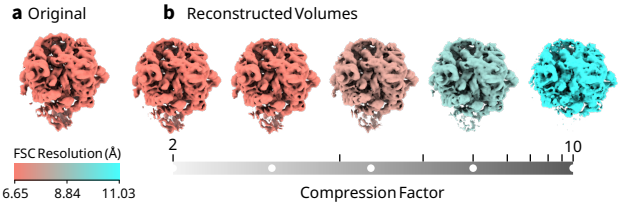

Figure 1. **Diffusion-prior reconstructions preserve 3D structural detail under compression. a**, Original 3D cryo-EM volume. **b**, Reconstructions from compressively acquired images using a learned diffusion prior, across compression factors; volumes are color-coded by FSC resolution (lower is better).

Despite their high dimensionality, raw cryo-EM images are highly structured and inherently reside on a low-dimensional manifold (Evans et al., 2025). The rapid expansion of the Electron Microscopy Public Image Archive (EMPIAR) (Iudin et al., 2023) (now encompassing over 2,800 entries and roughly 8.5 PiB of data) provides the massive scale required to mathematically map this manifold strictly from data. To harness this repository, we leverage denoising diffusion probabilistic models (DDPMs) (Ho et al., 2020; Song et al., 2021; Dhariwal & Nichol, 2021), which excel at establishing generative priors for complex scientific imaging. By coupling a DDPM trained on EMPIAR with advanced posterior sampling techniques (Chung et al., 2023; Song et al., 2022; 2024), we successfully capture the protein image manifold. This learned prior provides the precise structural guidance necessary to robustly reconstruct full-resolution images from compressed measurements (Fig. 2).

To model potential compressive acquisition pipelines, we evaluate two distinct measurement schemes. *Pixel-space masking* downsamples the image and projects it through random binary masks, whereas *Fourier-space masking* subsamples the image's Fourier coefficients via a back-focal-plane binary mask. Treating these physical measurement processes as feasible prerequisites, our work focuses exclusively on the computational reconstruction.

Contributions.

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Submitted to the 2026 Workshop on Generative and Agentic AI for Biology (ICML 2026). Do not distribute.

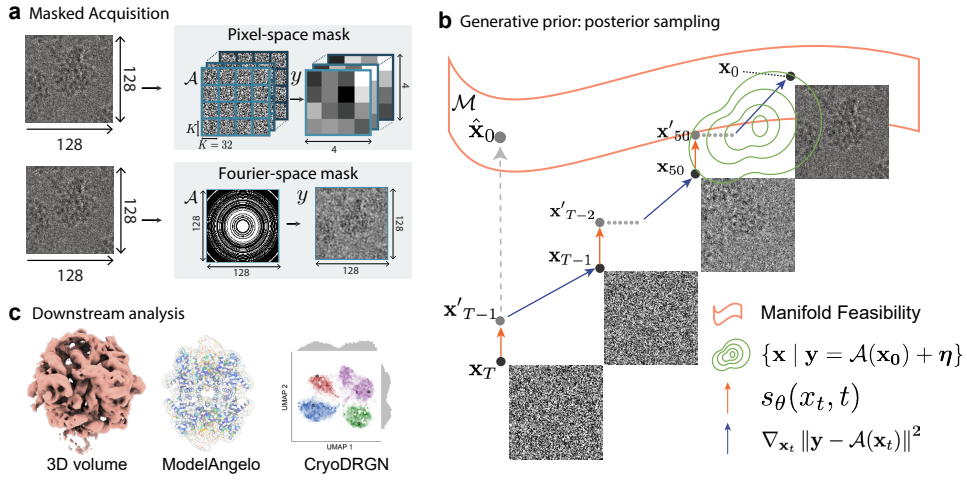

Figure 2. **Overview. a**, We assume compressed cryo-EM measurements obtained via pixel-space or Fourier-space masking. **b**, A denoising diffusion model learns a low-dimensional manifold $\mathcal{M}$ of protein cryo-EM images and guides posterior sampling toward reconstructions consistent with the measurements. **c**, Reconstructed images are evaluated on downstream biological tasks: 3D volume reconstruction, atomic model building, and conformational heterogeneity analysis.

- **A generative-prior approach to compressed cryo-EM reconstruction.** We formulate the recovery of full-resolution cryo-EM images from spatial and Fourier-domain undersampled measurements as an inverse problem, solved using a diffusion prior trained over the protein image manifold.
- **Empirical compression limits.** We achieve faithful 3D reconstruction up to $2\times$ compression. Performance degrades sharply at higher sparsity, such as $\geq 3.3\times$ Fourier-space compression or aggressive pixel-space downsampling.
- **Biological signal is preserved.** Reconstructions support downstream structural biology: 87.9% cluster agreement with the original CryoDRGN (Zhong et al., 2021) latent space and atomic model building with backbone RMSD of 2.34 Å in ModelAngelo (Jamali et al., 2024).

Related work. Diffusion priors routinely solve inverse problems in natural and medical imaging (Chung et al., 2023; Song et al., 2022; 2024). While diffusion models are emerging in cryo-EM, none have addressed compressive image reconstruction for single-particle analysis. Unlike handcrafted sparse priors, our approach learns dataset-scale structure directly from EMPIAR (Evans et al., 2025). Methodologically, we build upon diffusion posterior sampling (Chung et al., 2023), extending it with a Nesterov-accelerated reverse process for efficient inference.

2. Method

2.1. Inverse Problem Formulation

We recover a cryo-EM image $\mathbf{x}^* \in \mathbb{R}^n$ from noisy compressed measurements $\mathbf{y} = \mathcal{A}(\mathbf{x}^*) + \boldsymbol{\eta} \in \mathbb{R}^m$, where $\mathcal{A} : \mathbb{R}^n \to \mathbb{R}^m$ is a known linear measurement operator (in pixel or Fourier space) and $\boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$ (Fig. 2a). Because $m \ll n$, the problem is ill-posed and requires a prior. We assume cryo-EM images of proteins live on a low-dimensional manifold $\mathcal{M}$ that we learn directly from EMPIAR data using a DDPM, and we recover $\mathbf{x}^*$ via posterior sampling.

2.2. Image Recovery with a Diffusion Prior

Denoising diffusion probabilistic model. We approximate $\mathcal{M}$ with a DDPM (Ho et al., 2020; Song et al., 2021) that learns to reverse a gradual noising process. Let $\mathbf{x}_0$ be a clean cryo-EM image and $\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ pure Gaussian noise. The forward process is the variance-preserving SDE

$$d\mathbf{x} = -\tfrac{\beta_t}{2}\mathbf{x}\,dt + \sqrt{\beta_t}\,d\mathbf{w}, \tag{1}$$

with noise schedule $\beta_t > 0$ and Wiener process $\mathbf{w}$. Denoising follows the corresponding reverse-time SDE (Anderson, 1982)

$$d\mathbf{x} = \left[-\tfrac{\beta_t}{2}\mathbf{x} - \beta_t \nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t)\right] dt + \sqrt{\beta_t}\,d\bar{\mathbf{w}}, \tag{2}$$

where $\bar{\mathbf{w}}$ is the reverse-time Wiener process. The score $\nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t)$ is approximated by a neural network $\mathbf{s}_\theta(\mathbf{x}_t, t)$ trained via score matching (Vincent, 2011).

Posterior sampling. To reconstruct images consistent with $\mathbf{y}$, we sample from the posterior $p(\mathbf{x}_t \mid \mathbf{y})$ rather than the prior $p(\mathbf{x}_t)$ (Chung et al., 2023) (Fig. 2b). Bayes' rule decomposes the conditional score as

$$\nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t \mid \mathbf{y}) = \nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t) + \nabla_{\mathbf{x}_t} \log p(\mathbf{y} \mid \mathbf{x}_t). \tag{3}$$

Table 1. LPIPS and SSIM for pixel-space masking reconstructions on EMPIAR-10076 across downsampling level K and compression factor C.

Prior	K	2				4				8				16				32			
	C	4	2	1.3	1	16	2.7	1.5	1	10.7	2.6	1.5	1	10.2	2.5	1.4	1	10	2.5	1.4	1
Ours	LPIPS (↓)	0.14	0.12	0.09	0.06	0.27	0.12	0.09	0.06	0.20	0.15	0.11	0.07	0.21	0.17	0.12	0.10	0.23	0.18	0.17	0.16
	SSIM (↑)	0.31	0.50	0.68	0.81	0.18	0.44	0.68	0.84	0.15	0.15	0.43	0.66	0.15	0.44	0.62	0.72	0.17	0.41	0.48	0.53

The first term is the learned score $\mathbf{s}_\theta(\mathbf{x}_t, t)$. The second has no closed form because the likelihood couples to the clean image $\mathbf{x}_0$ rather than the noisy iterate $\mathbf{x}_t$. We resolve this via Tweedie's formula (Efron, 2011; Kim & Ye, 2021), which gives the posterior mean of the clean image as a deterministic function of the score:

$$\widehat{\mathbf{x}}_0 = \mathbf{x}_t + \sigma_t^2 \nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t). \tag{4}$$

Approximating $\nabla_{\mathbf{x}_t} \log p(\mathbf{y} \mid \mathbf{x}_t) \simeq \nabla_{\mathbf{x}_t} \log p(\mathbf{y} \mid \widehat{\mathbf{x}}_0)$ (Chung et al., 2023) and assuming Gaussian measurement noise yields the measurement-consistency gradient

$$\nabla_{\mathbf{x}_t} \log p(\mathbf{y} \mid \mathbf{x}_t) \simeq -\tfrac{1}{\sigma^2} \nabla_{\mathbf{x}_t} \|\mathcal{A}(\widehat{\mathbf{x}}_0) - \mathbf{y}\|_2^2. \tag{5}$$

Substituting the learned score and consistency gradient into Eq. (2) yields a guided reverse-time SDE that produces reconstructions on $\mathcal{M}$ consistent with $\mathbf{y}$. We integrate Nesterov-accelerated gradients into the reverse process to make sampling efficient (Wang et al., 2025; Li et al., 2022; Jin, 2025); full sampling pseudocode and DDPM training details are in Appendix A.

3. Experimental Results

3.1. Experimental Setup

Datasets. We evaluate on two EMPIAR datasets providing single-particle images, CTF, and pose metadata: 10076 (Davis et al., 2016) (50S ribosomal complex, 128×128) and 10648 (Saur et al., 2020) (PKM2 protein-ligand complex, 256×256). Full preprocessing details are in Appendix B. **Evaluation metrics.** For 2D image quality we use SSIM (Wang et al., 2004) and LPIPS (Zhang et al., 2018). For 3D volumes we backproject reconstructions using known particle poses, compute FSC resolution at the 0.143 cutoff (Rosenthal & Henderson, 2003). We also use volume correlation (VC) in UCSF ChimeraX (Meng et al., 2023). **Measurement operator.** Compression factor $C = n/m \geq 1$. *Pixel-space masking* projects $\mathbf{x}$ through b binarized random masks followed by non-overlapping $K \times K$ kernel-wise summation, yielding $C = K^2/b$. *Fourier-space masking* subsamples $m = n/C$ Fourier coefficients via uniform, annular ring (low-frequency-biased), or radial spoke patterns; we use uniform for 3D reconstructions. Full operator definitions are in Appendix C.

3.2. High-fidelity Reconstruction and Acquisition Limits

We benchmark on EMPIAR-10076 across pixel-space masking ($K \in \{2, \ldots, 32\}$) and three Fourier-space subsampling strategies (Table 1, 2). Reconstructions from Fourier-sampled data achieve ground-truth fidelity (6.65 Å) at $C = 2$ and maintain high structural accuracy at $C = 2.5$ (6.76 Å). Beyond this threshold, performance degrades sharply; at $C = 3.3$, the reconstruction fails to support accurate pose estimation, with median angular error jumping from 3.5° to 72.7° (Table 3).

For pixel-space masking, high-fidelity recovery is maintained up to $K = 16$ downsampling at $C = 1.3$, yielding a resolution of 6.99 Å. However, aggressive spatial downsampling ($K = 32$) leads to loss in 3D reconstruction resolution (7.91 Å) even with a relatively low compression ($C = 1.4$), marking the resolution-dependent limit of spatial compression. These results establish $C \simeq 2$ as a robust empirical limit for compressive cryo-EM acquisition using our learned diffusion prior.

Table 2. LPIPS and SSIM for Fourier-space masking on EMPIAR-10076 across masking strategies and compression factor C.

Prior	Mask	Uniform				Annular ring				Radial spoke			
	C	10	2.5	1.4	1	10	2.5	1.4	1	10	2.5	1.4	1
Ours	LPIPS (↓)	0.21	0.13	0.05	0.00	0.26	0.13	0.06	0.00	0.20	0.11	0.04	0.00
	SSIM (↑)	0.22	0.63	0.91	1.00	0.14	0.44	0.72	1.00	0.23	0.70	0.92	1.00

Table 3. 3D reconstruction performance on EMPIAR-10076 (6.65 Å ground-truth). Metrics include FSC resolution (Å), volume correlation (VC), and pose estimation accuracy (angular error AE, translational shift TSE).

Pixel-space						Fourier-space				
K	C	Å ↓	VC ↑	AE ↓	TSE ↓	C	Å ↓	VC ↑	AE ↓	TSE ↓
2	4	9.98	0.83	3.1	1.6	10	11.03	0.77	128.8	22.1
4	2	8.38	0.80	1.9	1.0	5	9.11	0.91	116.4	16.6
8	1.5	8.06	0.81	4.3	2.2	3.3	7.62	0.96	72.7	8.8
16	1.3	6.99	0.97	1.7	0.9	2.5	6.76	0.98	3.5	2.5
32	1.4	7.91	0.93	4.6	2.3	2	6.65	0.99	1.8	0.9

3.3. Conformational Heterogeneity via CryoDRGN

To test whether reconstructions preserve biologically meaningful variation across particles, we compare CryoDRGN (Zhong et al., 2021) latent spaces trained on origi-

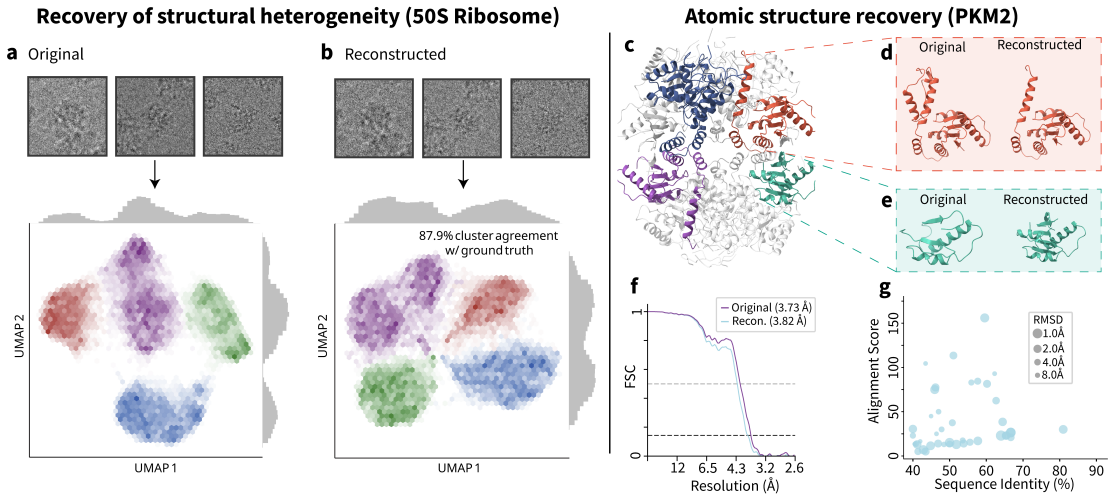

Figure 3. **Reconstructions preserve downstream biological signal. a, b**, Recovery of structural heterogeneity on the 50S ribosome (EMPIAR-10076). UMAP of the CryoDRGN latent space for original (**a**) and reconstructed (**b**) particles ($K = 16$, $C = 1.25$); 87.9% cluster agreement. **c–e**, Atomic structure recovery on PKM2 (EMPIAR-10648, $K = 2$, $C = 1.33$). **c**, ModelAngelo atomic model. **d, e**, Two highlighted regions comparing original and reconstructed atomic detail. **f**, FSC curves (3.73 Å vs 3.82 Å). **g**, Sequence alignment score versus sequence identity for matched chains; marker size encodes backbone RMSD.

nal and reconstructed particle sets (training details in Appendix D.2). We embed EMPIAR-10076 validation particles into an 8-dimensional latent space and visualize via UMAP. Gaussian mixture modeling on the UMAP projection identifies four distinct conformational clusters in the original data (Fig. 3a), corresponding to the four ribosomal assembly intermediates known for this dataset. We then apply pixel-space masking at $K = 16$ and $C = 1.25$, reconstruct, and re-train CryoDRGN on the reconstructed particles (Fig. 3b). To quantify how well reconstructions preserve heterogeneity, we color the reconstructed UMAP by the GMM labels assigned in the original space and compute, per cluster, the fraction of particles whose original and reconstructed assignments match. Per-cluster agreement ranges 82.1–91.4% with an average of 87.9%, indicating that the diffusion prior preserves the fine-grained particle-to-particle variation that drives heterogeneity analysis even at $K = 16$ downsampling, where each acquired pixel summarizes a 16×16 patch of the underlying image.

3.4. Atomic Model Recovery via ModelAngelo

We evaluate *de novo* atomic model building using ModelAngelo (Jamali et al., 2024) on EMPIAR-10648 volumes ($K = 2, C = 1.33$; PDB `6ttf`). Mean per-residue confidence (62.5 vs. 67.3 Original) and FSC resolutions (3.82 Å vs. 3.73 Å) indicate that reconstructed volumes retain the structural detail required for reliable modeling (Fig. 3c, f). Quantitative comparison across 42 matched chain pairs yields a mean backbone RMSD of 2.34 Å and mean alignment score of 38.1 (Fig. 3g), confirming that long polypeptide segments are recovered with high fidelity. Specific alignment parameters and chain-level correspondences are provided in Appendix D.3. Conversely, the DMPlug baseline (Wang et al., 2024) fails to preserve this signal, yielding significantly higher RMSD (2.75 Å) and lower confidence (50.4) while failing to resolve conformational heterogeneity (Appendix E).

4. Discussion

Implications and limitations. A learned diffusion prior over the protein cryo-EM manifold provides a robust framework for solving compressive inverse problems, successfully recovering 2D images and critical downstream structural biology. While we treat the measurement operator as computationally given, leaving the hardware realization of such masking schemes to future instrument design, this work proves the algorithmic viability of highly compressed acquisition in cryo-EM. Natural extensions include scaling the reconstruction pipeline to full micrographs and training the diffusion model across diverse protein families to develop a fully generalized prior.

Impact Statement

This paper advances machine learning methods for compressive cryo-EM reconstruction. By offering a computational pathway to alleviate data bottlenecks, such approaches stand to accelerate structural biology and downstream biomedical research. We foresee no specific ethical concerns or negative societal impacts from this methodological work.

References

Anderson, B. D. Reverse-time diffusion equation models. *Stochastic Processes and their Applications*, 12(3):313–326, 1982. ISSN 0304-4149.

Chua, E. Y., Mendez, J. H., Rapp, M., Ilca, S. L., Tan, Y. Z., Maruthi, K., Kuang, H., Zimanyi, C. M., Cheng, A., Eng, E. T., et al. Better, faster, cheaper: recent advances in cryo–electron microscopy. *Annual review of biochemistry*, 91(1):1–32, 2022.

Chung, H., Kim, J., McCann, M. T., Klasky, M. L., and Ye, J. C. Diffusion posterior sampling for general noisy inverse problems. In *International Conference on Learning Representations*, 2023.

Datta, A., Ng, K. F., Balakrishnan, D., Ding, M., Chee, S. W., Ban, Y., Shi, J., and Loh, N. D. A data reduction and compression description for high throughput time-resolved electron microscopy. *Nature communications*, 12(1):664, 2021.

Davis, J. H., Tan, Y. Z., Carragher, B., Potter, C. S., Lyumkis, D., and Williamson, J. R. Modular assembly of the bacterial large ribosomal subunit. *Cell*, 167(6):1610–1622, 2016.

Dhariwal, P. and Nichol, A. Diffusion models beat GANs on image synthesis. In *Advances in Neural Information Processing Systems*, volume 34, pp. 8780–8794, 2021.

Efron, B. Tweedie's formula and selection bias. *Journal of the American Statistical Association*, 106(496):1602–1614, 2011.

Evans, L., Murad, O.-V., Dingeldein, L., Cossio, P., Covino, R., and Meila, M. Cryo-EM images are intrinsically low dimensional. *PRX Life*, 3(3):033025, 2025.

Ho, J., Jain, A., and Abbeel, P. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, volume 33, pp. 6840–6851, 2020.

Iudin, A., Korir, P. K., Somasundharam, S., Weyand, S., Cattavitello, C., Fonseca, N., Salih, O., Kleywegt, G. J., and Patwardhan, A. EMPIAR: the electron microscopy public image archive. *Nucleic Acids Research*, 51(D1): D1503–D1511, 2023.

Jamali, K., Käll, L., Zhang, R., Brown, A., Kimanius, D., and Scheres, S. H. Automated model building and protein identification in cryo-EM maps. *Nature*, 628(8007):450–457, 2024.

Jin, Q. Adaptive nesterov momentum method for solving ill-posed inverse problems. *Inverse Problems*, 41(2), February 2025. ISSN 0266-5611. Publisher Copyright: © 2025 The Author(s). Published by IOP Publishing Ltd.

Kim, K. and Ye, J. C. Noise2Score: Tweedie's approach to self-supervised image denoising without clean images. In *Advances in Neural Information Processing Systems*, volume 34, pp. 864–874, 2021.

Li, R., Zha, H., and Tao, M. Hessian-free high-resolution nesterov acceleration for sampling, 2022.

Meng, E. C., Goddard, T. D., Pettersen, E. F., Couch, G. S., Pearson, Z. J., Morris, J. H., and Ferrin, T. E. UCSF ChimeraX: Tools for structure building and analysis. *Protein Science*, 32(11):e4792, 2023.

Nakane, T., Kotecha, A., Sente, A., McMullan, G., Masiulis, S., Brown, P. M., Grigoras, I. T., Malinauskaite, L., Malinauskas, T., Miehling, J., et al. Single-particle cryo-EM at atomic resolution. *Nature*, 587(7832):152–156, 2020.

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

Poger, D., Yen, L., and Braet, F. Big data in contemporary electron microscopy: challenges and opportunities in data transfer, compute and management. *Histochemistry and cell biology*, 160(3):169–192, 2023.

Rosenthal, P. B. and Henderson, R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *Journal of Molecular Biology*, 333(4):721–745, 2003.

Sanchez-Garcia, R., Saur, M., Vargas, J., Poelking, C., and Deane, C. M. CESPED: A benchmark for supervised particle pose estimation in cryo-em. *Phys. Rev. Res.*, 6: 023245, Jun 2024.

Saur, M., Hartshorn, M. J., Dong, J., Reeks, J., Bunkoczi, G., Jhoti, H., and Williams, P. A. Fragment-based drug discovery using cryo-EM. *Drug Discovery Today*, 25(3): 485–490, 2020.

Song, B., Kwon, S. M., Zhang, Z., Hu, X., Qu, Q., and Shen, L. Solving inverse problems with latent diffusion models via hard data consistency. In *International Conference on Learning Representations*, 2024.

Song, Y., Sohl-Dickstein, J., Kingma, D. P., Kumar, A., Ermon, S., and Poole, B. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations (ICLR)*, 2021.

Song, Y., Shen, L., Xing, L., and Ermon, S. Solving inverse problems in medical imaging with score-based generative models. In *International Conference on Learning Representations (ICLR)*, 2022.

Vincent, P. A connection between score matching and denoising autoencoders. *Neural Computation*, 23(7):1661–1674, 2011.

Wang, G., Cai, Y., Li, L., Peng, W., and Su, S. PFDiff: Training-free acceleration of diffusion models combining past and future scores, 2025.

Wang, H., Zhang, X., Li, T., Wan, Y., Chen, T., and Sun, J. DMPlug: A plug-in method for solving inverse problems with diffusion models. In *Advances in Neural Information Processing Systems*, 2024.

Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.

Yip, K. M., Fischer, N., Paknia, E., Chari, A., and Stark, H. Atomic-resolution protein structure determination by cryo-EM. *Nature*, 587(7832):157–161, 2020.

Zhang, R., Isola, P., Efros, A. A., Shechtman, E., and Wang, O. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 586–595, 2018.

Zhong, E. D., Bepler, T., Berger, B., and Davis, J. H. CryoDRGN: reconstruction of heterogeneous cryo-EM structures using neural networks. *Nature Methods*, 18(2): 176–185, 2021.

A. Implementation of Generative Priors

A.1. Nesterov Momentum Acceleration

Algorithm 1 Sampling with Nesterov momentum.

Require: $\mathbf{y}$, $\mathcal{A}$, $s_\theta(\mathbf{x}_t, t)$, $\{\alpha_t\}_{t=1}^T$, $\{\kappa_t\}_{t=1}^T$, $\{\zeta_t\}_{t=1}^T$
$\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, $\mathbf{m}_T \leftarrow \mathbf{0}$
for $t = T, \dots, 1$ **do**
  $\mathbf{s} \leftarrow s_\theta(\mathbf{x}_t, t)$
  $\widehat{\mathbf{x}}_0 \leftarrow \frac{1}{\sqrt{\bar{\alpha}_t}}\big(\mathbf{x}_t + (1-\bar{\alpha}_t)\mathbf{s}\big)$
  $\mathbf{x}'_{t-1} \leftarrow \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\mathbf{x}_t + \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}\widehat{\mathbf{x}}_0 + \tilde{\sigma}_t\mathbf{z}$
  $\mathbf{p}_t \leftarrow \widehat{\mathbf{x}}_0 - \kappa_t\mathbf{m}_t$
  $\mathbf{g}_t \leftarrow \nabla_{\mathbf{p}_t}\big\|\mathbf{y} - \mathcal{A}(\mathbf{p}_t)\big\|_2^2$
  $\mathbf{m}_{t-1} \leftarrow \kappa_t\mathbf{m}_t + \zeta_t\mathbf{g}_t$
  $\mathbf{x}_{t-1} \leftarrow \mathbf{x}'_{t-1} - \mathbf{m}_{t-1}$
end for
return $\mathbf{x}_0$

Our method enhances the conventional DDPM sampling framework by introducing an accelerated correction phase. Rather than altering the underlying reverse SDE, we enforce measurement consistency during every denoising iteration utilizing a momentum strategy based on Nesterov acceleration. Specifically, at each timestep t, we initially calculate the clean image prediction $\widehat{\mathbf{x}}_0$:

$$\widehat{\mathbf{x}}_0 = \frac{1}{\sqrt{\bar{\alpha}_t}}\left(\mathbf{x}_t + (1-\bar{\alpha}_t)s_\theta(\mathbf{x}_t, t)\right), \tag{6}$$

with $\alpha_t = 1 - \beta_t$ and $\bar{\alpha}_t = \prod_{s=1}^t \alpha_s$. From this clean estimate $\widehat{\mathbf{x}}_0$, we compute the intermediate state $\mathbf{x}'_{t-1}$, corresponding to the standard reverse-diffusion step prior to the inclusion of any measurement conditioning:

$$\mathbf{x}'_{t-1} = \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\mathbf{x}_t + \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}\widehat{\mathbf{x}}_0 + \tilde{\sigma}_t\mathbf{z}, \tag{7}$$

where $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. Consistent with (Ho et al., 2020), the variance of the reverse process $\tilde{\sigma}_t^2$ is evaluated as $\tilde{\sigma}_t^2 = \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t$. The expression in Equation (7) provides a standard discrete realization of the backward dynamics over a set schedule of timesteps. To properly steer the sampling trajectory, we define a look-ahead coordinate $\mathbf{p}_t$, which is constructed by projecting forward from the clean prediction $\widehat{\mathbf{x}}_0$ using the active momentum vector:

$$\mathbf{p}_t = \widehat{\mathbf{x}}_0 - \kappa_t\mathbf{m}_t, \tag{8}$$

Here, $\mathbf{m}_t$ represents the accumulated momentum state, while κ_t acts as a schedule-dependent extrapolation parameter. To guarantee that the generation remains faithful to the actual observations, we evaluate the gradient at this look-ahead coordinate, defined as $\mathbf{g}_t = \nabla_{\mathbf{p}_t}\|\mathbf{y} - \mathcal{A}(\mathbf{p}_t)\|_2^2$. This gradient supplies the necessary data-consistency guidance to steer the DDPM toward the proper target $\mathbf{x}_{t-1}$. During each iteration, the momentum is refreshed via $\mathbf{m}_{t-1} = \kappa_t\mathbf{m}_t + \zeta_t\mathbf{g}_t$, with the coefficient ζ_t modulating the intensity of the measurement constraint. Ultimately, the conditioned sample is yielded by subtracting this momentum adjustment: $\mathbf{x}_{t-1} = \mathbf{x}'_{t-1} - \mathbf{m}_{t-1}$.

A.2. DDPM Training Details

We optimized the generative prior utilizing the conventional denoising diffusion framework (Ho et al., 2020), implemented via a 2D U-Net backbone. This architecture is structured with six progressive down-sampling stages featuring channel widths of $[128, 128, 256, 256, 512, 512]$, accompanied by a symmetric arrangement of six up-sampling stages. Spatial self-attention mechanisms are integrated specifically at the fifth down-stage (*AttnDownBlock2D*) and the corresponding second up-stage (*AttnUpBlock2D*). To accommodate the single-channel nature of grayscale cryo-EM micrographs, the input and target dimensions are both restricted to 1. For network optimization, we employ the AdamW solver (hyperparameters: $\beta_1 = 0.95$, $\beta_2 = 0.999$, $\epsilon = 1 \times 10^{-8}$, weight decay $= 1 \times 10^{-6}$). The learning rate follows a cosine decay trajectory after an initial 500-step linear warm-up phase.

All models are trained in a bfloat16 mixed-precision environment distributed across eight NVIDIA A6000 GPUs. We apply a gradient accumulation factor of 2, with the specific per-device batch capacities detailed subsequently. To ensure training stability, gradients are constrained to a maximum ℓ_2-norm of 1.0. Furthermore, an exponential moving average (EMA) of the network weights is tracked throughout training (decay limits bounded at 0.9999, power $= 0.75$, inverse gamma $= 1.0$). Prior to injection into the network, pixel intensities are scaled to the $[-1, 1]$ range. The forward corruption process utilizes a linear β schedule spanning 1000 timesteps, and the total optimization duration is fixed at 100 epochs.

Principal dataset-specific configurations are summarized in Table 4. We maximize hardware utilization by saturating the VRAM on the A6000 units, which permits a batch size of 128 per GPU for the 128×128 resolution data, and limits the batch size to 16 per GPU for the larger 256×256 inputs. Drawing on standard practices established in the foundational DDPM literature (Ho et al., 2020), the base learning rate is set to 2×10^{-4} for the lower-resolution models. For the 256×256 experiments, this rate is scaled down by a factor of ten to 2×10^{-5}.

Table 4. Training details for DDPM models on protein datasets.

Dataset	Resolution	Train Size	LR	Batch
EMPIAR-10076	128×128	105,519	2e-4	128
EMPIAR-10648	256×256	187,964	2e-5	16

A.3. Implementation of Posterior Sampling

To maintain numerical stability and enforce valid pixel intensities, the clean prediction $\widehat{\mathbf{x}}_0$ and the updated state $\mathbf{x}_{t-1}$ are both clamped to the $[-1, 1]$ range. The measurement consistency weight originates at $\zeta_{\min} = 10^{-10}$. This small initial value prevents instability during early generation phases, allowing the conditioning gradient's impact to scale up progressively. Through empirical evaluation using PSNR, SSIM, and LPIPS metrics, an upper bound of $\zeta_{\max} = 1.0$ proved optimal for spatial kernels of 2, 4, 8, and 16 (kernel definitions in Section C). For the extreme 32 kernel, reconstruction quality peaked at $\zeta_{\max} = 10.0$. This aligns with the expectation that heavily compressed observations necessitate more aggressive data guidance.

Following standard conventions for Nesterov acceleration, the momentum term κ_t is bounded between $\kappa_{\min} = 0.1$ and $\kappa_{\max} = 0.9$. Both scheduling parameters, ζ_t and κ_t, advance linearly across timesteps as defined below:

$$\zeta_t = \zeta_{\min} + \frac{t-1}{T-1}\left(\zeta_{\max} - \zeta_{\min}\right),$$

$$\kappa_t = \kappa_{\min} + \frac{t-1}{T-1}\left(\kappa_{\max} - \kappa_{\min}\right).$$

B. Cryo-EM Dataset Details

This section outlines the specific dataset splits, image resolutions, and preprocessing pipelines utilized for all empirical evaluations.

EMPIAR-10076. This dataset captures assembly intermediates of the *E. coli* large ribosomal subunit (Davis et al., 2016). Particle stacks were acquired directly from the CryoDRGN Zenodo archive (Zhong et al., 2021), yielding a split of 105,519 training samples alongside 26,380 validation samples. To meet our experimental constraints, the native 320×320 pixel images were uniformly downsampled to 128×128 dimensions via Fourier cropping.

EMPIAR-10648. This dataset contains the PKM2 protein complexed with a small-molecule inhibitor (Saur et al., 2020). We sourced both the particle images and the corresponding pose metadata through the CESPED benchmark framework (Sanchez-Garcia et al., 2024). The partition consists of 187,964 particles allocated for training and 23,496 held out for validation. Because the base resolution of these images is 222×222 pixels, they were resized to 256×256 using bicubic interpolation before being passed to the network, ensuring compatibility with our DDPM architecture. Control reconstructions confirmed that this interpolation step did not degrade the underlying structural fidelity, yielding equivalent 3D resolution metrics.

C. Obtaining Linear Measurements

Pixel-space masking. The spatial masking strategy models the acquisition process by subjecting high-resolution targets to a combination of binary masking and spatial downsampling. For a given clean image $\mathbf{x}^* \in \mathbb{R}^n$, the forward measurement operator $\mathcal{A}$ operates through the following steps:

1. A set of b independent binary masks $\{B_i\}_{i=0}^{b} \in \mathbb{R}^n$ is instantiated via a Bernoulli distribution ($p = 0.5$). To emulate incomplete sensing, these masks are applied element-wise to $\mathbf{x}^*$, where each B_i dictates which specific pixels are preserved.
2. Following the masking stage, spatial resolution is degraded using a block-wise summation pooling operation parametrized by a kernel size K. By aggregating intensities within disjoint $K \times K$ pixel patches, this step mimics a lower-resolution detector.
3. The resulting undersampled observations across all b masks are concatenated to form the final measurement vector $\mathbf{y} \in \mathbb{R}^m$, where $m = bn/K^2$.

Mathematically, this process is formulated as:

$$\mathcal{A}(\mathbf{x}^*) = \left\{\text{Pool}_K\left(B_i \odot \mathbf{x}^*\right)\right\}_{i=1}^{b},$$

with $\odot$ representing the Hadamard product and Pool_K indicating the $K \times K$ block summation.

Fourier-space masking. Alternatively, the frequency-domain measurement operator $\mathcal{A}$ is constructed as follows:

1. We define a singular binary mask $B \in \mathbb{R}^n$ and apply it via element-wise multiplication to the frequency representation of the image, $\mathcal{F}(\mathbf{x}^*)$, thereby isolating exactly m Fourier coefficients.
2. All unselected frequencies are zeroed out. The final low-resolution observation $\mathbf{y}$ is then recovered by applying the inverse Fourier transform, $\mathcal{F}^{-1}$, to this sparsified signal.

This frequency-based subsampling is analytically expressed as:

$$\mathcal{A}(\mathbf{x}^*) = \mathcal{F}^{-1}(B \odot \mathcal{F}(\mathbf{x}^*)).$$

We evaluate reconstruction performance across three distinct Fourier masking patterns: (1) uniform, (2) low-frequency biased annular rings, and (3) radial spokes. Their construction logic is as follows:

- **Uniform.** Exactly $1/C$ of the available Fourier coefficients are retained using a uniformly random selection process.

- **Annular ring with low-frequency bias.** The frequency domain is divided into 100 concentric rings of equal area, from which $k = 100/C$ rings are drawn. The selection probability is intentionally skewed to favor low-frequency information (which is centralized using the `fftshift` operation). The specific sampling weight w for a ring with a midpoint radius of r is calculated as:

$$w = e^{-\frac{r}{2\nu^2}},$$

  setting $\nu = n/8$.

- **Radial spoke.** The Fourier space is segmented angularly into 100 identical spokes. From this set, $k = 100/C$ individual spokes are sampled uniformly at random and without replacement.

D. Extended Experimental Details

D.1. High-fidelity Reconstruction Setup

To evaluate reconstruction performance, a fixed subset of 16 images was randomly sampled from the held-out validation split, ensuring no overlap with the DDPM training data. This identical cohort was utilized for every experimental trial. Throughout all evaluations, we calculate the PSNR, SSIM, and LPIPS metrics, reporting both their mean values and standard deviations.

D.2. CryoDRGN Training Parameters

For the heterogeneity analysis, we deployed CryoDRGN v3.4.0, training specifically on the validation subset of EMPIAR-10076. The optimization protocol ran for 100 epochs using a batch size of 32. We configured the network with an 8-dimensional latent bottleneck, supported by 3 residual layers and uniform encoder and decoder widths of 1024. Architecturally, we utilized the default `ResidLinearMLP` for the encoder and the `FTPositionalDecoder` for the decoder. Explicit CTF parameters and pose orientations were supplied as model inputs. Furthermore, all training was distributed across multiple GPUs employing AMP mixed-precision capabilities.

D.3. ModelAngelo Inference and Alignment Details

De novo atomic modeling was executed via ModelAngelo v1.0 utilizing the `nucleotides` weights bundle. Inference relied on the default configuration: a prediction threshold of 0.05 for Cα atoms, a batch size of 4, a spatial stride of 16, and a box size of 64. Every dataset underwent three iterations of GNN-driven structural refinement, with the tertiary output models serving as the basis for all comparative analyses. For our generative approach, measurements were extracted using $b = 3$ masks at a spatial reduction of $K = 2$ (yielding $C = 1.33$). This specific operating point was chosen to minimize the required measurement fraction while maintaining a baseline SSIM of at least 0.8.

For chain-to-chain validation, we isolated continuous polypeptide segments containing ≥ 20 residues. Global pairwise sequence alignments were then computed using the `pairwise2` utility within Biopython (scoring weights: gap extend: -0.5, gap open: -2, mismatch: -1, match: 2). Only paired chains demonstrating a sequence identity of $\geq 40\%$ were kept for spatial comparison. Following structural superposition, the backbone RMSD was calculated strictly across the (N, Cα, C, O) atomic coordinates.

Highlighted examples of this correspondence feature the Original chain `Aa` matching with generated chains `Ah` and `Ak` (yielding alignment scores of 156.0 and 113.5, with sequence identities of 0.60 and 0.51, respectively). Another instance pairs `AT` with `Af` (score 104.5; 0.43 identity). Such matches emphasize that extensive structural motifs remain highly accurate despite the compressive nature of the simulated acquisition.

E. Comparison with Baseline Method

Table 5. Downstream biological validation comparing our generative prior against the DMPlug baseline.

Method	Heterogeneity Acc. ↑	Chains ↑	RMSD (Å) ↓	Confidence ↑
Ours	**82.1–91.4%**	**42**	**2.34**	**62.5**
DMPlug	Random	28	2.75	50.4

To provide a comparative baseline, we evaluate our approach against DMPlug (Wang et al., 2024), a super-resolution (SR) technique grounded in diffusion models. Although SR algorithms typically generate visually coherent textures, they fail to explicitly constrain the output to the underlying compressed measurements. This limitation introduces severe structural artifacts during high-ratio undersampling. Both models were assessed using the core downstream biological tasks introduced previously: atomic model construction through ModelAngelo and conformational landscape mapping via CryoDRGN. As detailed in Table 5, our framework achieves superior performance across the entire suite of quantitative metrics.