Supplementary Material for: cryoSENSE: Compressive Sensing Enables High-throughput Microscopy with Sparse and Generative Priors on the Protein Cryo-EM Image Manifold

A. Implementation of Sparse Priors

We selected three priors in cryoSENSE: sparse L1 recovery with the DCT, sparse L1 recovery with the wavelet transform, and TV regularization. We provide an overview of these methods here.

Sparse DCT recovery. Sparse DCT recovery assumes that images are sparse in the DCT. We seek to minimize:

$$\widehat{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{y} - \mathcal{A}(\mathbf{x})\|_2^2 + \lambda \, \|\Psi_{\text{DCT}}(\mathbf{x})\|_1 \,,$$

where $\Psi_{\text{DCT}}(\mathbf{x})$ denotes the DCT of $\mathbf{x}$. The L1 norm promotes sparsity in the DCT.

Sparse Wavelet recovery. Similarly, sparse wavelet reconstruction assumes that images are sparse in the wavelet transform. We solve the following optimization problem:

$$\widehat{\mathbf{x}} = \arg\min_{\mathbf{x}} \quad \|\mathbf{y} - \mathcal{A}(\mathbf{x})\|_2^2 + \lambda \, \|\Psi_{\text{WT}}(\mathbf{x})\|_1 \,,$$

where $\Psi_{\text{WT}}(\mathbf{x})$ denotes the wavelet transform of $\mathbf{x}$.

Total Variation regularization. TV regularization promotes piecewise smoothness by penalizing the gradient magnitude across the image. We minimize:

$$\widehat{\mathbf{x}} = \arg\min_{\mathbf{x}} \quad \|\mathbf{y} - \mathcal{A}(\mathbf{x})\|_2^2 + \lambda \, \text{TV}(\mathbf{x}).$$

Here, $\text{TV}(\mathbf{x})$ is the anisotropic total variation operator over a vector $\mathbf{x}$:

$$\text{TV}(\mathbf{x}) = \sum_{i,j} (|x_{i+1,j} - x_{i,j}| + |x_{i,j+1} - x_{i,j}|).$$

To implement the proximal gradient algorithm in sparse DCT reconstruction and sparse wavelet reconstruction, we utilize the Iterative Soft Thresholding Algorithm (ISTA) [4]. At each epoch, we perform a gradient descent step $\mathbf{z}_t = \mathbf{x}_t - \alpha \cdot \nabla \|\mathbf{y} - \mathcal{A}(\mathbf{x}_t)\|_2^2$ on the data fidelity term, then apply the proximal operator by computing the transform $\Psi(\mathbf{z}_t)$, applying soft thresholding with threshold λ, and computing the inverse transform to obtain $\mathbf{x}_{t+1}$. To implement the proximal gradient algorithm in TV regularization, we utilize the $prox\_tv$ package (https://github.com/albarji/proxTV) [1, 2].

For sparse DCT reconstruction, sparse wavelet reconstruction, and TV regularization, we solve each optimization problem by performing proximal gradient descent over 200 epochs. There are two hyperparameters to consider: λ, which governs the strength of the regularizer, and the learning rate (which we set as a constant rate α, such that $\alpha_t = \alpha$). To determine these hyperparameters, for each protein, we perform a grid search of $\lambda = [0.1, 0.01, 0.001, 0.0001]$ and $\alpha = [0.001, 0.01, 0.1, 0.5, 1]$ over two training images.

B. Implementation of Generative Priors

B.1. Nesterov Momentum Acceleration

Algorithm 1 cryoSENSE 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 $t = T, \ldots, 1$ **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 $\mathbf{x}_0$

cryoSENSE integrates an accelerated correction step into the standard DDPM sampling procedure. Rather than directly modifying the reverse SDE, we thus apply measurement consistency guidance at each denoising step using a momentum-based approach inspired by Nesterov acceleration. At each diffusion step t, we first compute the denoised estimate $\widehat{\mathbf{x}}_0$ of the clean underlying image:

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

where $\alpha_t = 1 - \beta_t$ and $\bar{\alpha}_t = \prod_{s=1}^{t} \alpha_s$. Using $\widehat{\mathbf{x}}_0$, we obtain the unconstrained image $\mathbf{x}'_{t-1}$, which represents the next sample generated by the DDPM without guidance:

$$\mathbf{x}'_{t-1} = \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\mathbf{x}_t + \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}\widehat{\mathbf{x}}_0 + \tilde{\sigma}_t \mathbf{z}, \quad (2)$$

where $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. The reverse diffusion variance $\tilde{\sigma}_t^2$ is computed as $\tilde{\sigma}_t^2 = \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t$, as in [8]. Equation (2) is a natural discretization of the reverse process given a fixed number of steps. To guide the denoising trajectory, cryoSENSE introduces the look-ahead point $\mathbf{p}_t$ by extrapolating from $\widehat{\mathbf{x}}_0$ along the current momentum direction:

$$\mathbf{p}_t = \widehat{\mathbf{x}}_0 - \kappa_t \mathbf{m}_t, \quad (3)$$

where $\mathbf{m}_t$ is the accumulated momentum vector and κ_t is a time-dependent extrapolation coefficient. The gradient of $\mathbf{p}_t$, denoted as $\mathbf{g}_t = \nabla_{\mathbf{p}_t} \|\mathbf{y} - \mathcal{A}(\mathbf{p}_t)\|_2^2$, is used to enforce measurement consistency when guiding DDPM toward the corrected image $\mathbf{x}_{t-1}$. At each timestep, we update $\mathbf{m}_t$ to $\mathbf{m}_{t-1}$ via $\mathbf{m}_{t-1} = \kappa_t \mathbf{m}_t + \zeta_t \mathbf{g}_t$, where ζ_t governs the strength of measurement consistency. We then obtain $\mathbf{x}_{t-1}$ as $\mathbf{x}_{t-1} = \mathbf{x}'_{t-1} - \mathbf{m}_{t-1}$.

B.2. DDPM Training Details

The DDPM model [8] was trained within the standard denoising diffusion probabilistic modeling framework using a UNet2D architecture. The network comprises six down-sampling blocks with output channels of $[128, 128, 256, 256, 512, 512]$, mirrored by six up-sampling blocks; self-attention is introduced in the fifth down block (*AttnDownBlock2D*) and the second up block (*AttnUpBlock2D*). Because cryo-EM images are grayscale, both the input and output channel dimensions are set to 1. Optimization is performed with AdamW ($\beta_1 = 0.95$, $\beta_2 = 0.999$, weight decay 1×10^{-6}, Adam epsilon $\epsilon = 1 \times 10^{-8}$) combined with a cosine learning-rate schedule preceded by 500 warm-up steps. Training runs on eight NVIDIA A6000 GPUs in mixed-precision (bfloat16), using a per-GPU batch size shown in the table below and gradient-accumulation of 2. Gradients are clipped to an ℓ_2-norm of 1.0, and an exponential moving average of the model parameters is maintained (inverse gamma = 1.0, power = 0.75, maximum decay = 0.9999). Input images are normalized to $[-1, 1]$ and corrupted with a linear noise (β) schedule across 1000 diffusion timesteps. The model is trained for 100 epochs.

Table S-1 lists the key hyper-parameters for each dataset. The per-GPU batch size was set to the maximum that fits in memory on eight NVIDIA A6000 cards—128 images for 128×128 inputs and 16 images for 256×256 inputs. We used a learning rate of 2×10^{-4} for the 128×128 resolution dataset batches and reduced it by one order of magnitude (2×10^{-5}) for the 256×256 image dataset, following settings similar to those reported in the original DDPM paper [8].

Table S-1. Training details for DDPM models on protein datasets.

Dataset	Resolution	Train Size	LR	Batch
EMPIAR-10076	128×128	105,519	2e-4	128
EMPIAR-11526	128×128	200,260	2e-4	128
EMPIAR-10166	128×128	190,904	2e-4	128
EMPIAR-10786	128×128	230,927	2e-4	128
EMPIAR-10648	256×256	187,964	2e-5	16

B.3. Implementation of Posterior Sampling

For stability and to ensure that generated images remain within a valid range, we clip both the predicted clean image $\widehat{\mathbf{x}}_0$ and the subsequent sample $\mathbf{x}_{t-1}$ to the interval $[-1, 1]$.

We additionally set $\zeta_{\min} = 10^{-10}$ to stabilize optimization in the early stages of sampling, ensuring that the influence of the measurement consistency gradient is initially minimal and gradually increases over time. Empirically, we found that setting $\zeta_{\max} = 1.0$ yields stable reconstructions — measured by LPIPS, SSIM, and PSNR — for kernel sizes 2, 4, 8, and 16 (see Section D for definition of kernel size). For the largest kernel size 32, we observed that $\zeta_{\max} = 10.0$ provided the best reconstruction quality. Intuitively, lower-resolution measurements required stronger guidance from the measurement consistency gradient to achieve accurate reconstructions.

The momentum coefficients κ_t were set linearly between $\kappa_{\min} = 0.1$ and $\kappa_{\max} = 0.9$, following common practice in Nesterov-accelerated gradient methods.

The momentum coefficients, ζ_t and κ_t, vary linearly with timestep t as follows:

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

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

C. Cryo-EM Dataset Details

We provide full details on dataset sizes, preprocessing, and resolution settings used in our experiments.

EMPIAR-10166. Human 26S proteasome bound to Oprozomib [7]. We use particle images and pose metadata provided by the CESPED benchmark [9], with 190,904 training and 23,863 validation particles. Images were originally 284 × 284 pixels and downsampled to 128 × 128 using Fourier cropping via the CryoDRGN downsample utility.

EMPIAR-10076. E. coli large ribosomal subunit assembly intermediates [5]. We used the dataset provided by the CryoDRGN Zenodo repository [13], with 105,519 training particles and 26,380 validation particles. The original images were 320 × 320 pixels and were downsampled to 128 × 128 using Fourier cropping for all experiments.

EMPIAR-10648. PKM2 protein bound to a small-molecule inhibitor [10]. We used particle images and pose metadata from the CESPED benchmark [9], with 187,964 particles for training and 23,496 for validation. Images were provided at 222 $\times$ 222 pixels and were upsampled to 256 $\times$ 256 via bicubic interpolation prior to model input to match the dimensional requirements of our DDPM framework. We verified that this upsampling had no adverse effect on the structural information, as comparable 3D volume resolutions were obtained before and after upsampling.

EMPIAR-10786. Substance P–Neurokinin Receptor G protein complexes [6]. We used the particle images and pose metadata provided by the CESPED benchmark [9], with 230,927 training particles and 28,866 validation particles. Images were downsampled from 184 $\times$ 184 to 128 $\times$ 128 via Fourier cropping.

EMPIAR-11526. Small ribosomal subunit assembly intermediates in E. coli. We used the dataset released by Sun et al. [11] via Zenodo, with 200,260 training particles and 25,032 validation particles. Images were provided at 256 $\times$ 256 pixels and were downsampled to 128 $\times$ 128 via Fourier cropping.

D. Obtaining Linear Measurements

Pixel-space masking. In pixel-space masking, the forward operator $\mathcal{A}$ simulates the cryo-EM measurement process by applying structured masking and downsampling to high-resolution images. Given an input image $\mathbf{x}^* \in \mathbb{R}^n$, $\mathcal{A}$ is defined as follows:

1. Multiple random binary masks $\{B_i\}_{i=0}^{b} \in \mathbb{R}^n$, each drawn independently from a Bernoulli distribution with probability $p = 0.5$, are generated and element-wise multiplied with $\mathbf{x}^*$ to simulate partial observations. Each mask B_i randomly selects a subset of pixels to retain.
2. After masking, a kernel-wise summation pooling operation with kernel size K is applied to the masked images, reducing the spatial resolution. This pooling step aggregates pixel values over non-overlapping $K \times K$ blocks, simulating low-resolution measurements.
3. The final measurements $\mathbf{y} \in \mathbb{R}^m$, with $m = bn/K^2$, are collected across all masks.

Formally, the measurement operator can be expressed as:

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

where $\odot$ denotes element-wise multiplication and Pool_K denotes block-wise summation pooling over kernels of size $K \times K$. Fig. S-1 illustrates the forward operator $\mathcal{A}$ applied with a single random binary mask, demonstrating various pooling kernel sizes $K \times K$ and their corresponding measurement outputs $\mathbf{y}$.

Fourier-space masking. In Fourier-space masking, $\mathcal{A}$ is defined as follows (see Fig. S-2):

1. A single binary mask $B \in \mathbb{R}^n$ is generated and element-wise multiplied with the Fourier transform of $\mathbf{x}^*$, denoted as $\mathcal{F}(\mathbf{x}^*)$, to subsample a total of m Fourier coefficients.
2. The rest of the Fourier coefficients not subsampled are set to zero. The inverse Fourier transform, denoted as $\mathcal{F}^{-1}$, is then applied on the subsampled image to obtain the low-resolution measurement $\mathbf{y}$.

Formally, the measurement operator can be expressed as:

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

In cryoSENSE, we test the following three Fourier masks: (1) uniform, (2) annular ring with low-frequency bias, and (3) radial spoke. These three masks are defined as follows:

- **Uniform.** This mask samples $1/C$ many Fourier coefficients, chosen uniformly at random.
- **Annular ring with low-frequency bias.** This mask partitions the Fourier plane into a 100 concentric, equal-area rings. A subset of $k = 100/C$ rings is then sampled. This sampling is probabilistically weighted, biasing the sampling toward selecting lower-frequency rings (lower-frequency Fourier coefficients are shifted to the center of the image via the `fftshift` command). For each ring, given its mid-radius as r, its weight w is given by:

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

  where $\nu = n/8$.
- **Radial spoke.** This mask partitions the image into 100 radial spokes of equal angular width. A subset of $k = 100/C$ spokes is then selected uniformly at random, without replacement.

Fig. S-3 visualizes how these different masks look and their respective $\mathbf{y}$ measurements.

E. Cryo-EM Dataset Details

We provide full details on dataset sizes, preprocessing, and resolution settings used in our experiments.

EMPIAR-10166. Human 26S proteasome bound to Oprozomib [7]. We use particle images and pose metadata provided by the CESPED benchmark [9], with 190,904 training and 23,863 validation particles. Images were originally 284 $\times$ 284 pixels and downsampled to 128 $\times$ 128 using Fourier cropping via the CryoDRGN downsample utility.

EMPIAR-10076. E. coli large ribosomal subunit assembly intermediates [5]. We used the dataset provided by the CryoDRGN Zenodo repository [13], with 105,519 training particles and 26,380 validation particles. The original images were 320 $\times$ 320 pixels and were downsampled to 128 $\times$ 128 using Fourier cropping for all experiments.

EMPIAR-10648. PKM2 protein bound to a small-molecule inhibitor [10]. We used particle images and pose metadata from the CESPED benchmark [9], with 187,964 particles for training and 23,496 for validation. Images were provided

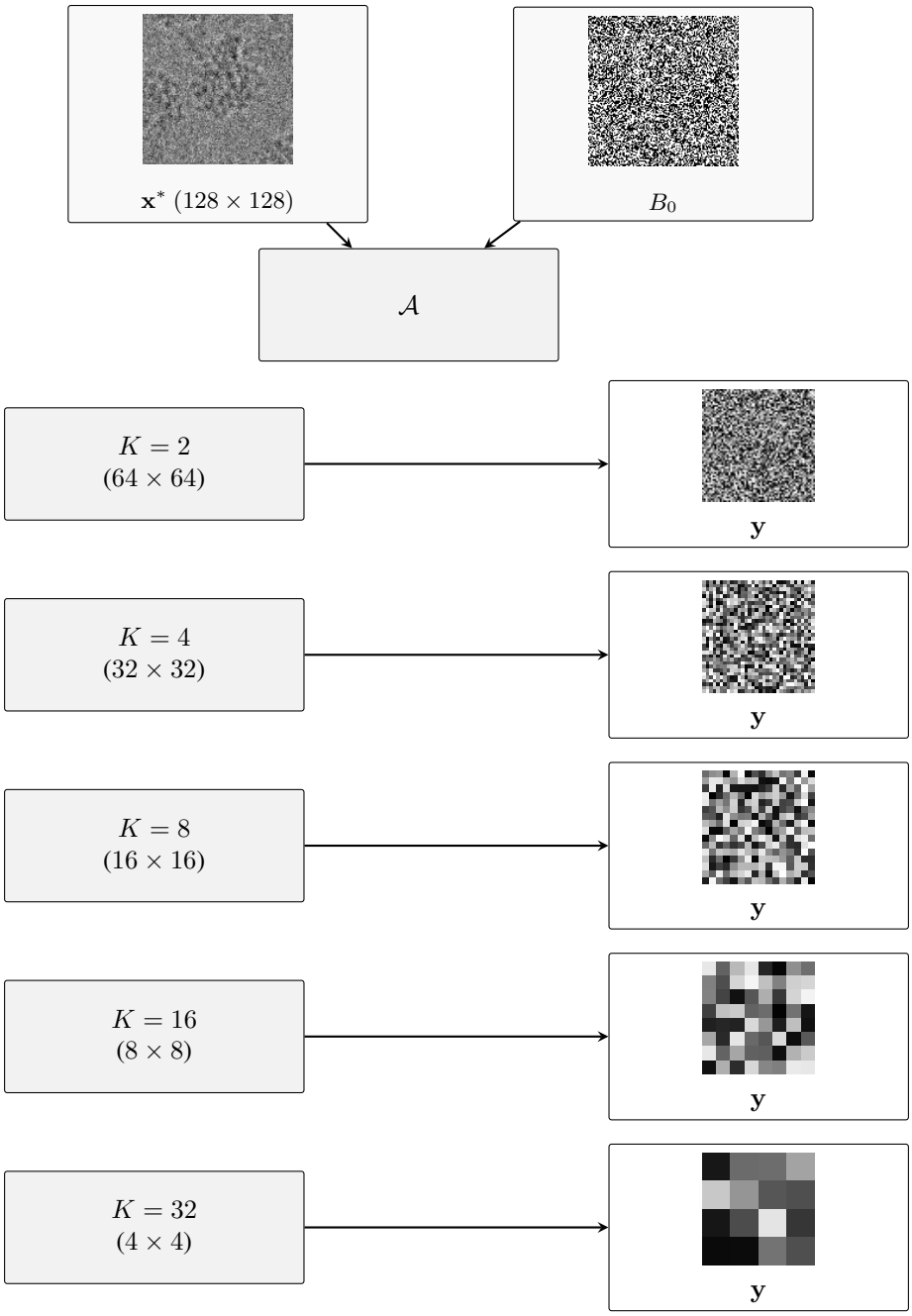

Figure S-1. Visualization of forward operator $\mathcal{A}$ with pixel-space masking over a single mask. The process begins with a 128×128 input image $\mathbf{x}_0$ and a random binary mask B_0. Non-overlapping kernel-wise convolution is applied over $K \times K$ patches, resulting in progressively downsampled measurement resolutions as kernel size increases.

at 222×222 pixels and were upsampled to 256×256 via bicubic interpolation prior to model input to match the dimensional requirements of our DDPM framework. We verified that this upsampling had no adverse effect on the structural information, as comparable 3D volume resolutions were obtained before and after upsampling.

EMPIAR-10786. Substance P–Neurokinin Receptor G protein complexes [6]. We used the particle images and pose metadata provided by the CESPED benchmark [9], with 230,927 training particles and 28,866 validation particles. Images were downsampled from 184×184 to 128×128 via Fourier cropping.

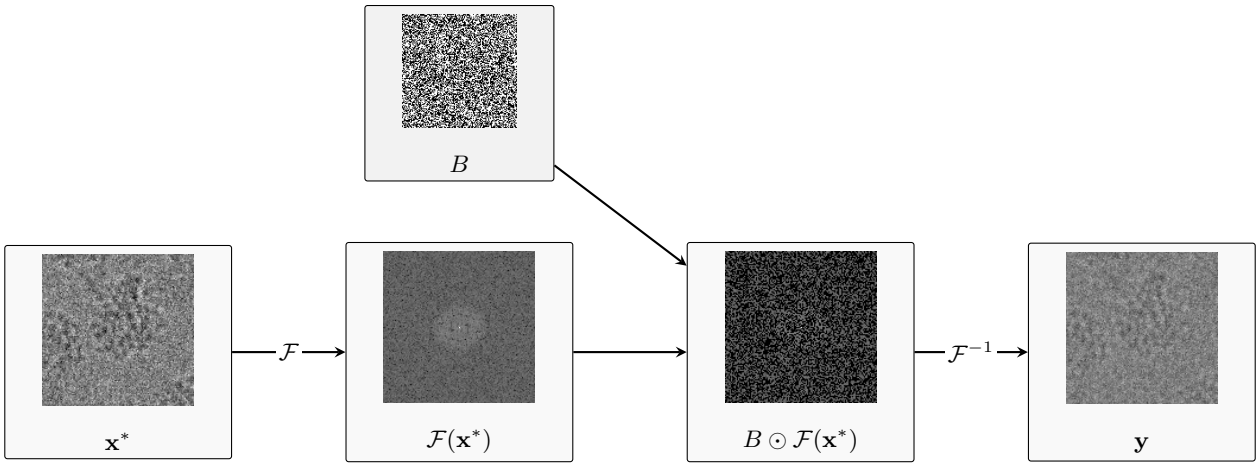

Figure S-2. Visualization of forward operator $\mathcal{A}$ with Fourier-space masking. An input image $\mathbf{x}^*$ is transformed into the Fourier domain, $\mathcal{F}(\mathbf{x}^*)$. A binary mask is then applied to the Fourier domain to subsample a subset of Fourier coefficients. The inverse Fourier transform ($\mathcal{F}^{-1}$) yields the corresponding low-resolution measurement $\mathbf{y}$ in the spatial domain.

EMPIAR-11526. Small ribosomal subunit assembly intermediates in E. coli. We used the dataset released by Sun et al. [11] via Zenodo, with 200,260 training particles and 25,032 validation particles. Images were provided at 256 $\times$ 256 pixels and were downsampled to 128 $\times$ 128 via Fourier cropping.

F. Experimental Setup

All code, project website, model weights, and reproduction instructions are included in the supplementary material accompanying this submission.

F.1. High-fidelity Reconstruction Setup

For our reconstruction experiments, we selected 16 random images from the validation set, which the DDPMs used in cryoSENSE did not see during training, and used that across all runs. For all runs, we compute LPIPS, SSIM, and PSNR and plot the mean and standard deviation.

F.2. CryoDRGN Training Parameters

All CryoDRGN experiments used CryoDRGN v3.4.0 on EMPIAR-10076 validation particles. Models are trained for 100 epochs with an 8-dimensional latent space, encoder and decoder sizes of 1024, and 3 residual layers. Batch size is set to 32. The `ResidLinearMLP` encoder and `FTPositionalDecoder` decoder architectures are used with standard settings. Pose and CTF metadata are provided as input. All training runs use multi-GPU acceleration with AMP mixed-precision.

F.3. ModelAngelo Inference

All atomic models are generated using ModelAngelo v1.0 with the `nucleotides` model bundle. Input cryo-EM density maps are reconstructed from EMPIAR-10648 validation particles for three cases: (1) original high-resolution images and cryoSENSE (2) sparse and (3) generative prior reconstructed images. ModelAngelo inference is performed using default parameters: box size of 64, stride of 16, batch size of 4, and a threshold of 0.05 for Cα prediction. Three rounds of GNN-based model refinement are performed for all datasets, and the output models from the third round are used for all subsequent analyses.

cryoSENSE Parameters. We run cryoSENSE at downsampling level $K = 2$ using $b = 3$ masks (corresponding to $C = 1.33$). This configuration is selected by setting an SSIM threshold of 0.8 and choosing the generative prior setup that requires the fewest measurements.

G. Additional Experiments

In this section, we add additional experiments that did not fit in the main text.

G.1. High-fidelity Reconstruction

Figs. S-4 and S-5 detail reconstruction performance with pixel-space masking, while Figs. S-6 and S-7 detail reconstruction performance with Fourier-space masking. Since the compression factor in pixel-space masking also depends on K, we opt to plot reconstruction performance sweeping across values of $1/C$, which ensures our plots are normalized between 0 and 1. $1/C$ can be interpreted as the ratio of measurements needed to recover the original image, where 1 corresponds to no compression being done. For the tables in the main text, we extract average LPIPS, SSIM, and PSNR values at evenly spaced points of $1/C$.

We observe similar trends to that reported in the main text, demonstrating that cryoSENSE is able to generalize

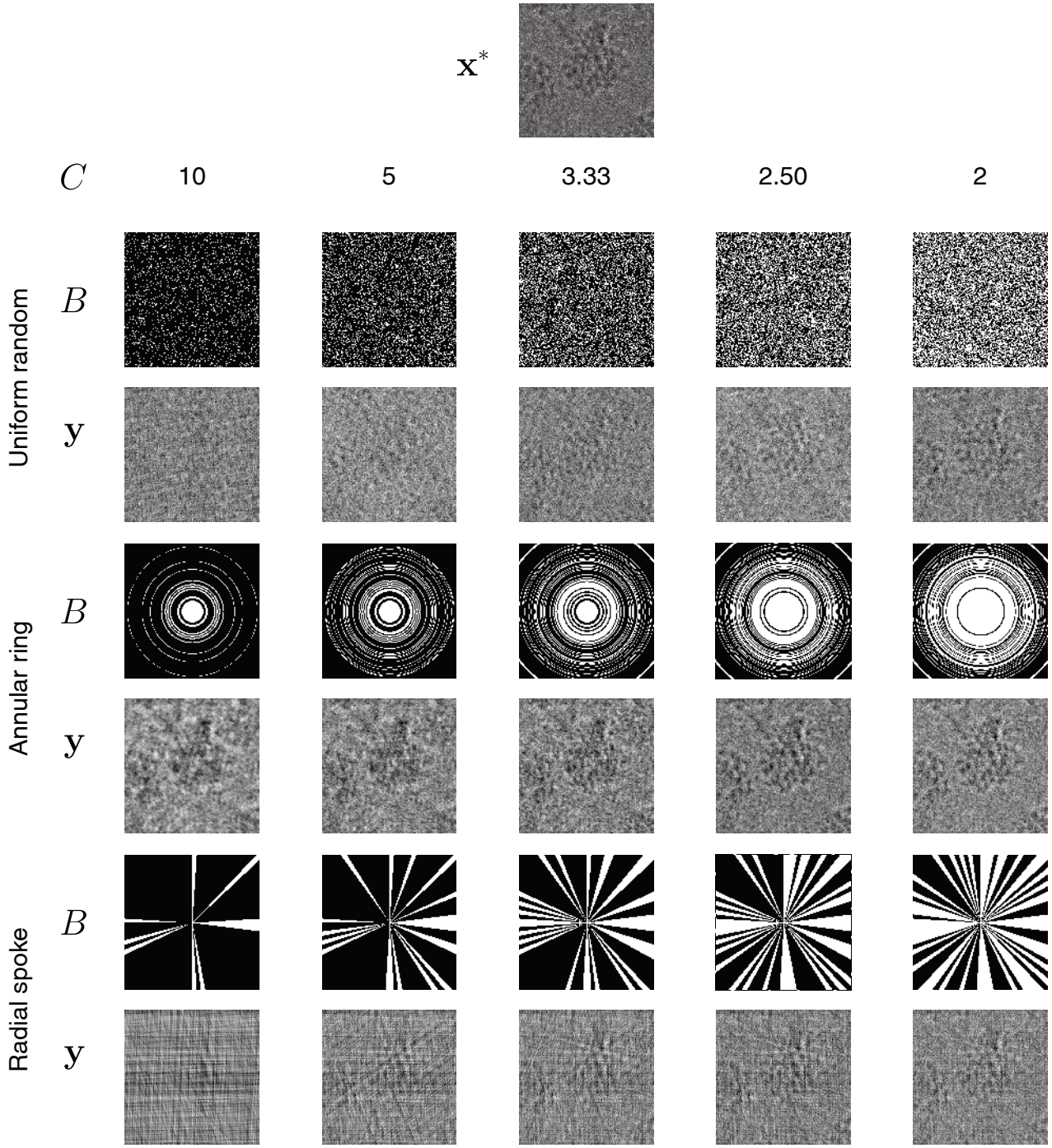

Figure S-3. Visualization of uniform, annular ring, and radial spoke Fourier masks at different levels of C, and their respective outputs $\mathbf{y}$ given $\mathbf{x}^*$.

to other proteins. Across all pixel-space and Fourier-space experiments, we observe that DCT is the best performing sparse prior on average. Therefore, we use DCT for all 3D reconstruction tasks. In Fourier-space masking experiments, we additionally observe that uniform masks exhibit lower LPIPS and higher SSIM scores on average compared to annular ring and radial spoke masks.

G.2. High-fidelity Noisy Reconstruction

We repeat cryoSENSE reconstructions in the noisy case, where $\sigma^2 = 0.01$. Figs. S-8 and S-9 detail reconstruction performance with pixel-space masking, while Figs. S-10 and S-11 detail reconstruction performance with Fourier-space masking. We observe the same trends as those made in the main text, where sparse priors are consistently better than generative priors in Fourier-space masking and under lower levels of compression, whereas generative priors are better than sparse priors in pixel-space masking and under extreme levels of downsampling and compression. This is unsurprising, as sparse and generative priors are known to be robust to moderate levels of noise [3, 4].

G.3. FSC Curves for 3D Reconstruction of EMPIAR-10076

Fig. S-12 presents the FSC curves quantitatively comparing DDPM and DCT reconstructions across both pixel-space and Fourier-space masks, which were summarized in Table 3 of the main text. Similar to our 2D analysis, for pixel-space masking, we note that DDPM outperforms DCT in $K = 16$ and 32 downsampling regimes. For Fourier masking, DCT generally outperforms DDPM.

G.4. Comparison with Baseline Method

Table S-2. Downstream validation: cryoSENSE (DDPM) vs. DMPlug.

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

We compare cryoSENSE (DDPM) against DMPlug [12], a diffusion-based super-resolution (SR) method. While SR methods can produce visually smooth images, they do not enforce consistency with compressed measurements, which leads to structural inaccuracies under aggressive undersampling. We evaluate both methods on the downstream tasks presented in the main text: conformational heterogeneity recovery via CryoDRGN and atomic model building via ModelAngelo. As shown in Table S-2, cryoSENSE outperforms DMPlug across all metrics.

G.5. Noise Model and Robustness Analysis

To evaluate noise robustness, we sweep measurement noise across SNR levels from 30 to -20 dB on EMPIAR-10076 at 128×128 resolution, using 16 validation images. We select one representative configuration for each masking type: pixel-space masking with $K = 4$, $C = 2$ and Fourier-space masking with $C = 2.5$ (uniform subsampling). We define the measurement SNR as $\text{SNR} = 10\log_{10}(\text{Var}(\mathbf{y}_0)/\sigma^2)$, where $\mathbf{y}_0 = \mathcal{A}(\mathbf{x}_{\text{exp}})$ denotes the compressed measurement obtained from the experimentally acquired cryo-EM image before adding Gaussian measurement noise, and $\text{Var}(\mathbf{y}_0)$ is the average variance of these pre-noise measurements across the 16 images. Because the measurement operator $\mathcal{A}$ differs between pixel-space and Fourier-space masking, the variance of $\mathbf{y}_0$ differs substantially between the two configurations: $\text{Var}(\mathbf{y}_0) \approx 2.38$ for pixel-space and $\text{Var}(\mathbf{y}_0) \approx 0.027$ for Fourier-space. To ensure a fair comparison, we calibrate σ for each case such that the same SNR in dB corresponds to the same ratio of signal power to noise power: $\sigma = \sqrt{\text{Var}(\mathbf{y}_0)/10^{\text{SNR}/10}}$.

Fig. S-13 shows LPIPS, SSIM, and PSNR as a function of measurement SNR for both DCT and DDPM priors under pixel-space and Fourier-space masking. The corresponding noise standard deviations range from $\sigma = 0.049$ (30 dB) to $\sigma = 15.4$ (-20 dB) for pixel-space masking and from $\sigma = 0.005$ (30 dB) to $\sigma = 1.65$ (-20 dB) for Fourier-space masking. At high SNR (≥ 20 dB), all four methods converge to similar reconstruction quality. At low SNR (≤ -10 dB), the DDPM prior is more robust than DCT, particularly in pixel-space masking where DCT-Real degrades sharply while DDPM-Real maintains substantially lower LPIPS and higher PSNR. In Fourier-space masking, DDPM also outperforms DCT at low SNR, though both degrade. At intermediate SNR (0–10 dB), Fourier-space masking consistently outperforms pixel-space masking for both priors, and DCT pixel-space performance recovers to match DDPM pixel-space.

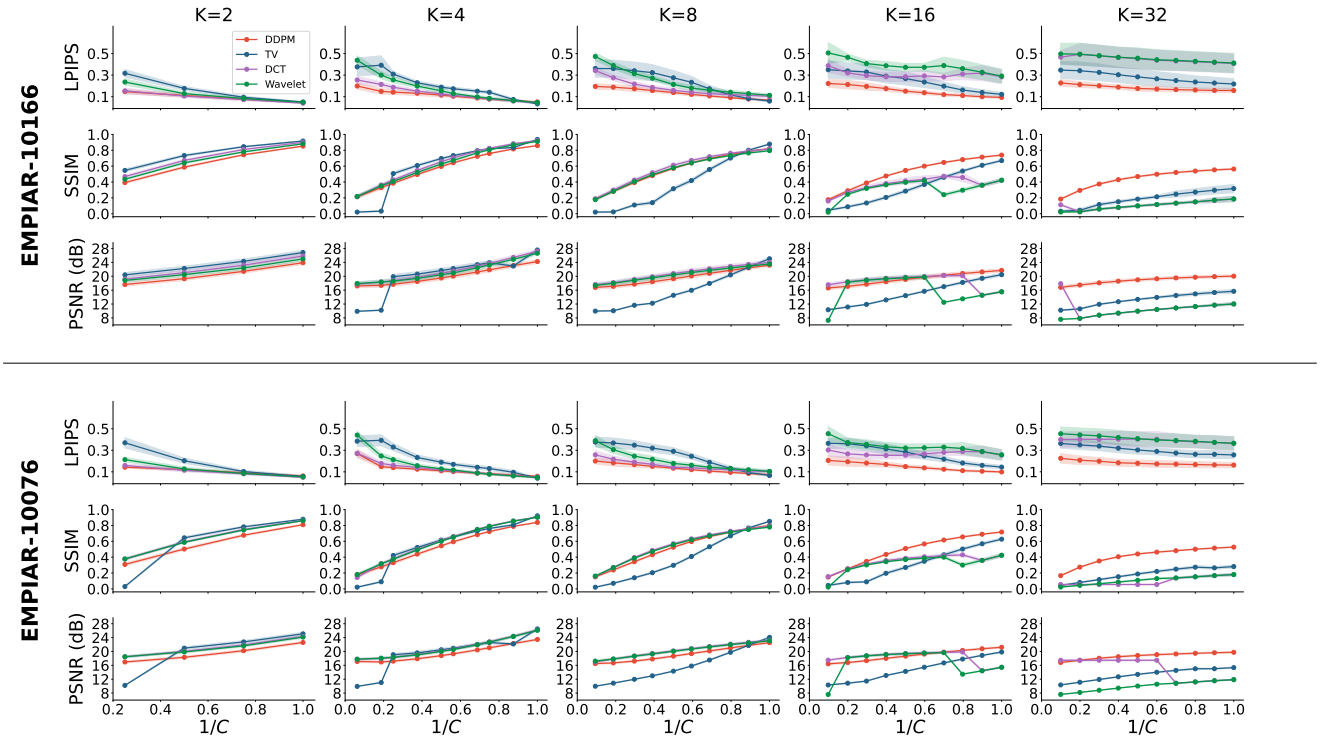

Figure S-4. LPIPS, SSIM, and PSNR scores for cryoSENSE reconstructions with pixel-space masks across various levels of K and $1/C$ for EMPIAR-10076 and EMPIAR-10166 (no noise $\sigma = 0.0$).

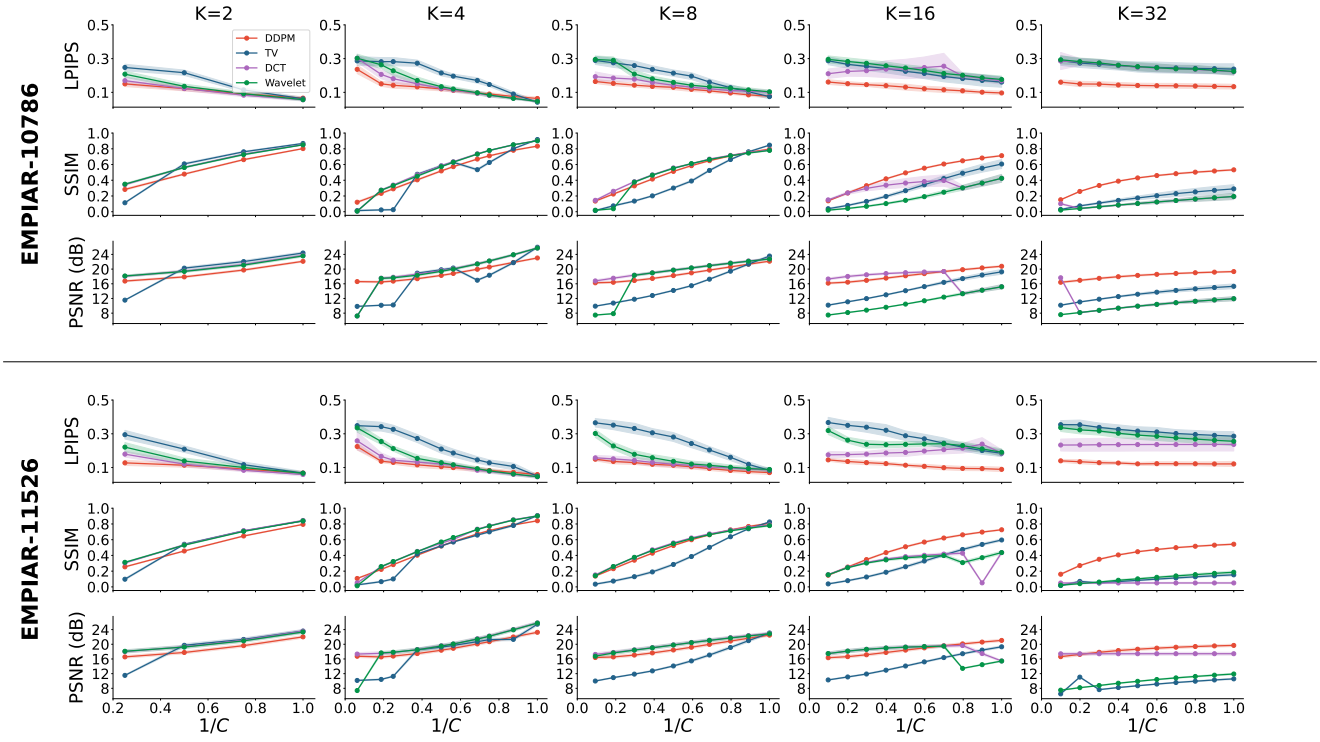

Figure S-5. LPIPS, SSIM, and PSNR scores for cryoSENSE reconstructions with pixel-space masks across various levels of K and $1/C$ for EMPIAR-10786 and EMPIAR-11526 (no noise $\sigma = 0.0$).

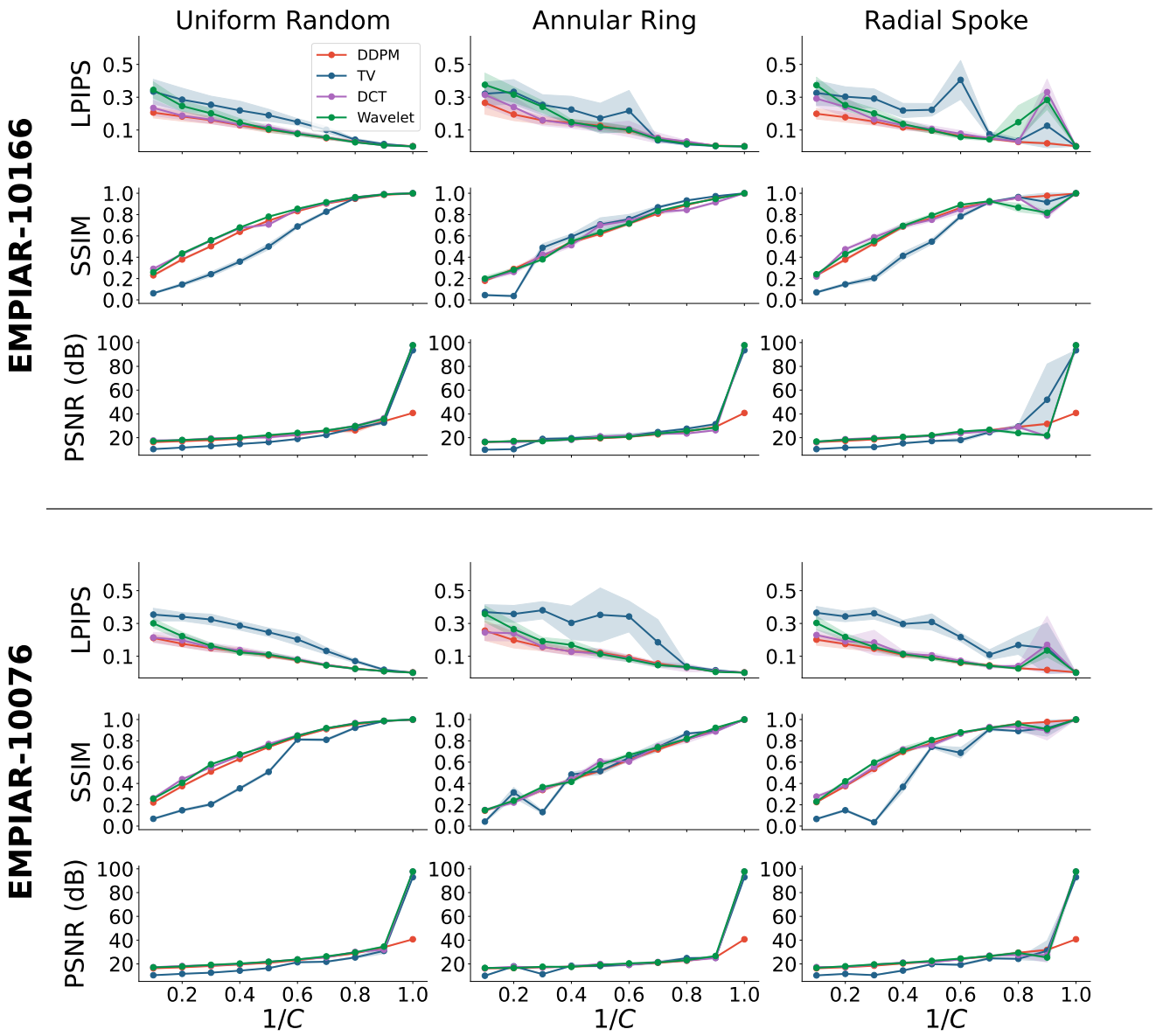

Figure S-6. LPIPS, SSIM, and PSNR scores for cryoSENSE reconstructions across various Fourier masks and $1/C$ for EMPIAR-10076 and EMPIAR-10166 (no noise $\sigma = 0.0$).

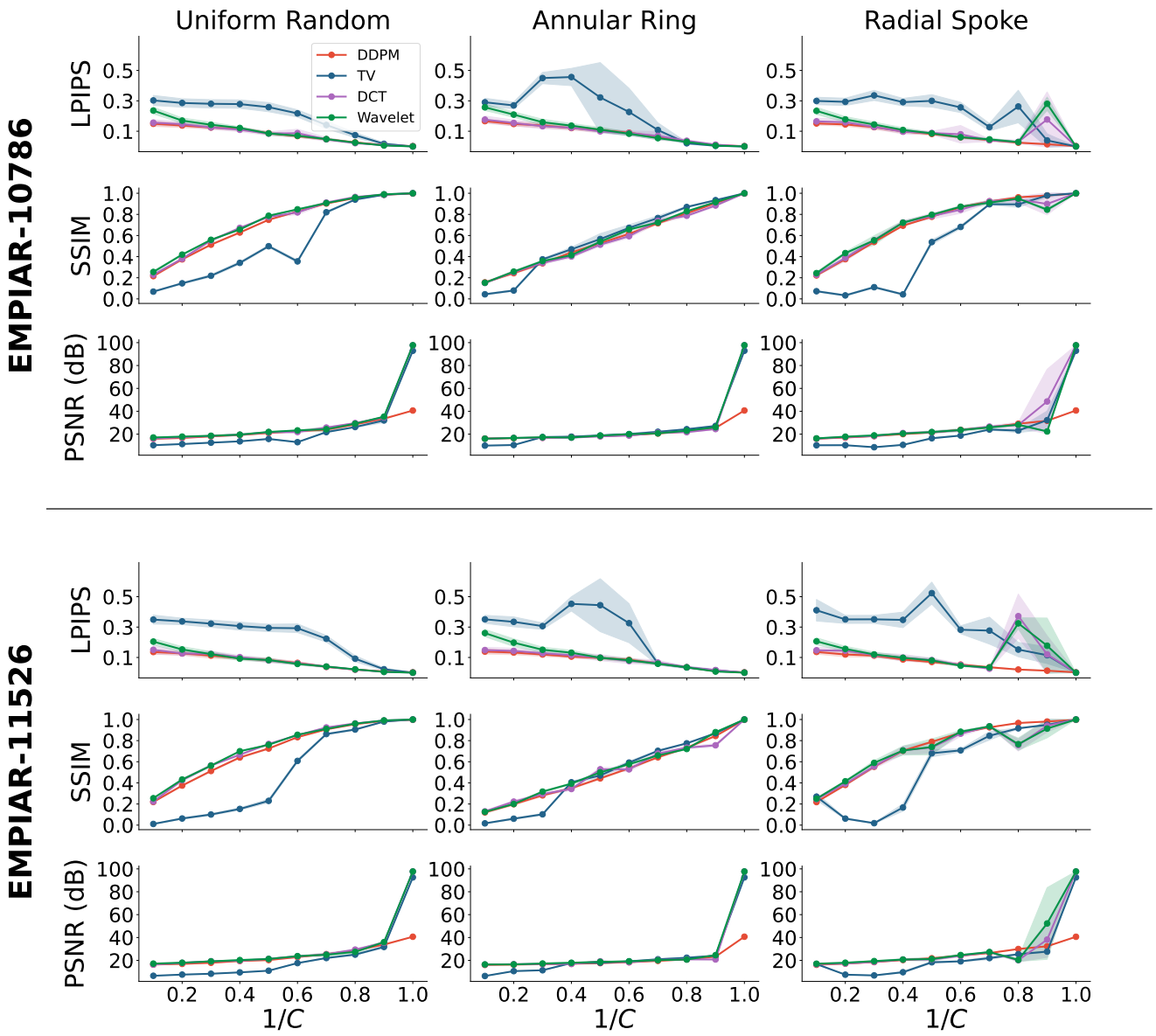

Figure S-7. LPIPS, SSIM, and PSNR scores for cryoSENSE reconstructions across various Fourier masks and $1/C$ for EMPIAR-10786 and EMPIAR-11526 (no noise $\sigma = 0.0$).

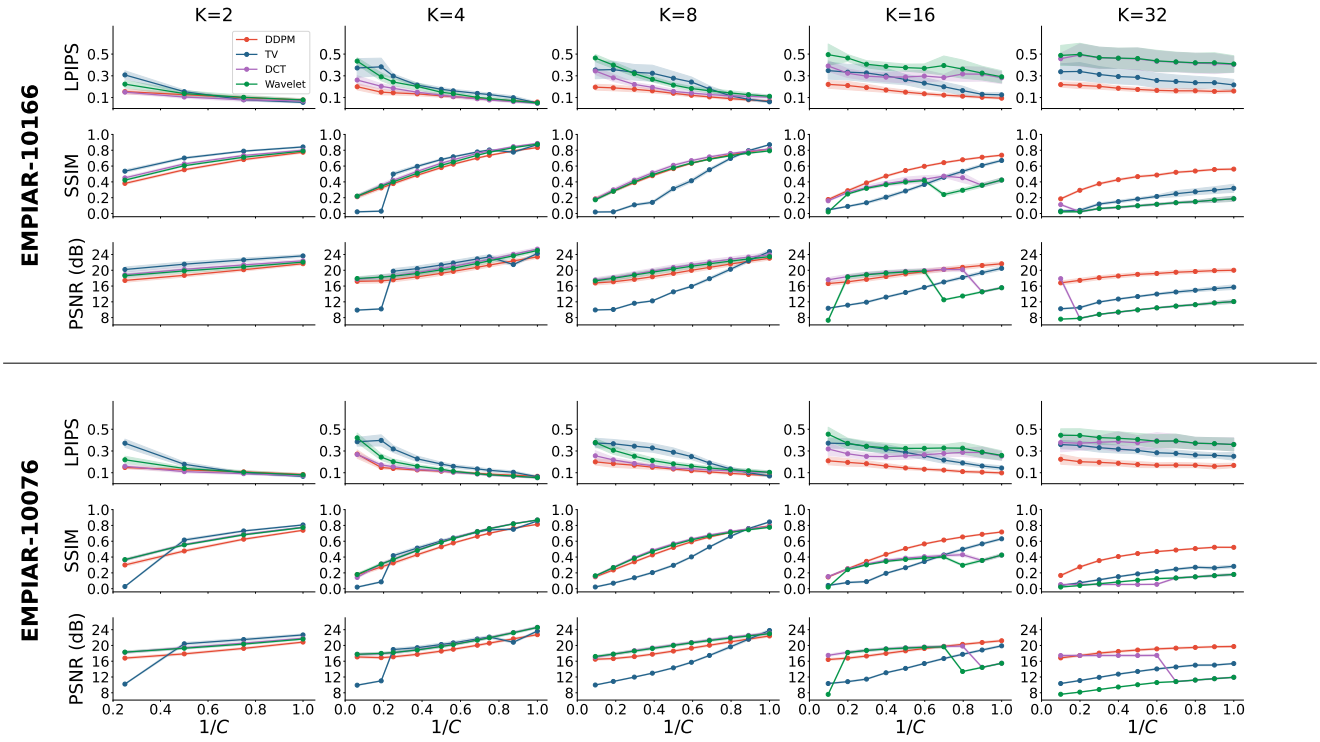

Figure S-8. LPIPS, SSIM, and PSNR scores for cryoSENSE reconstructions with pixel-space masks across various levels of K and $1/C$ for EMPIAR-10076 and EMPIAR-10166 in the presence of noise ($\sigma = 0.01$).

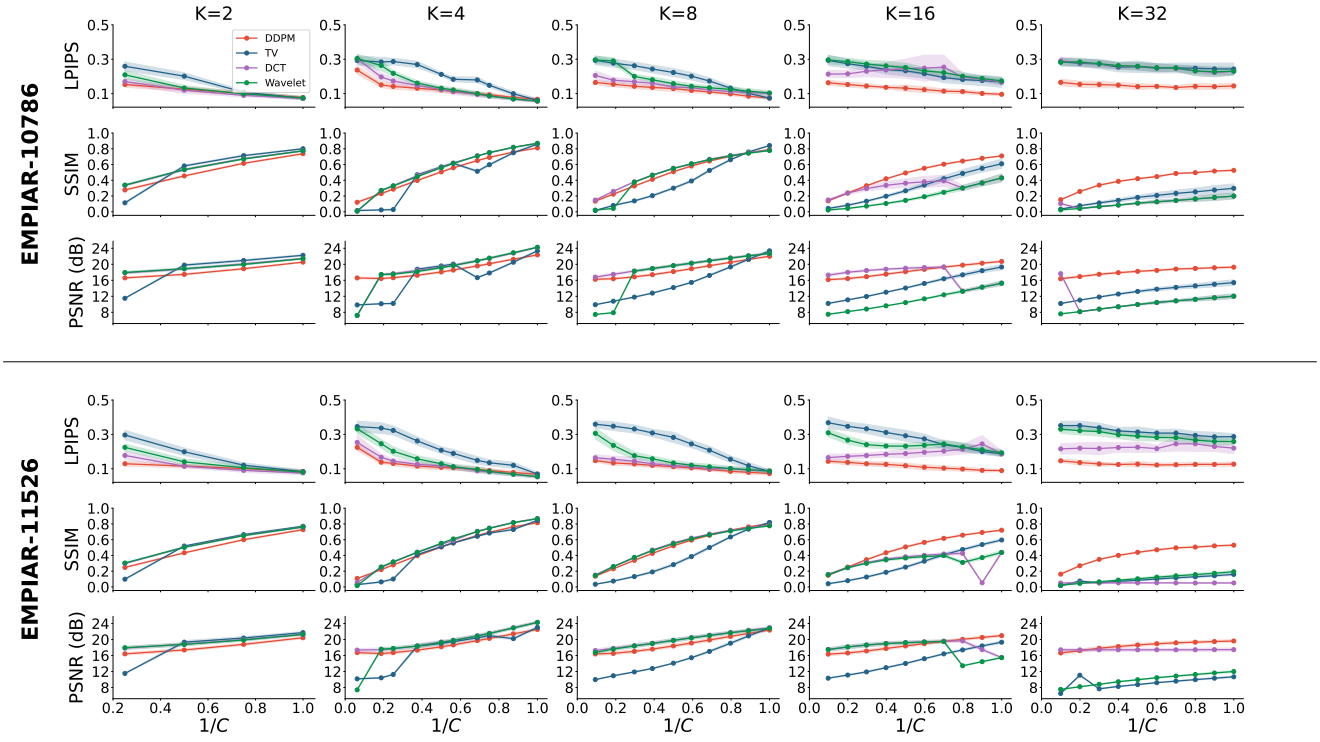

Figure S-9. LPIPS, SSIM, and PSNR scores for cryoSENSE reconstructions with pixel-space masks across various levels of K and $1/C$ for EMPIAR-10786 and EMPIAR-11526 in the presence of noise ($\sigma = 0.01$).

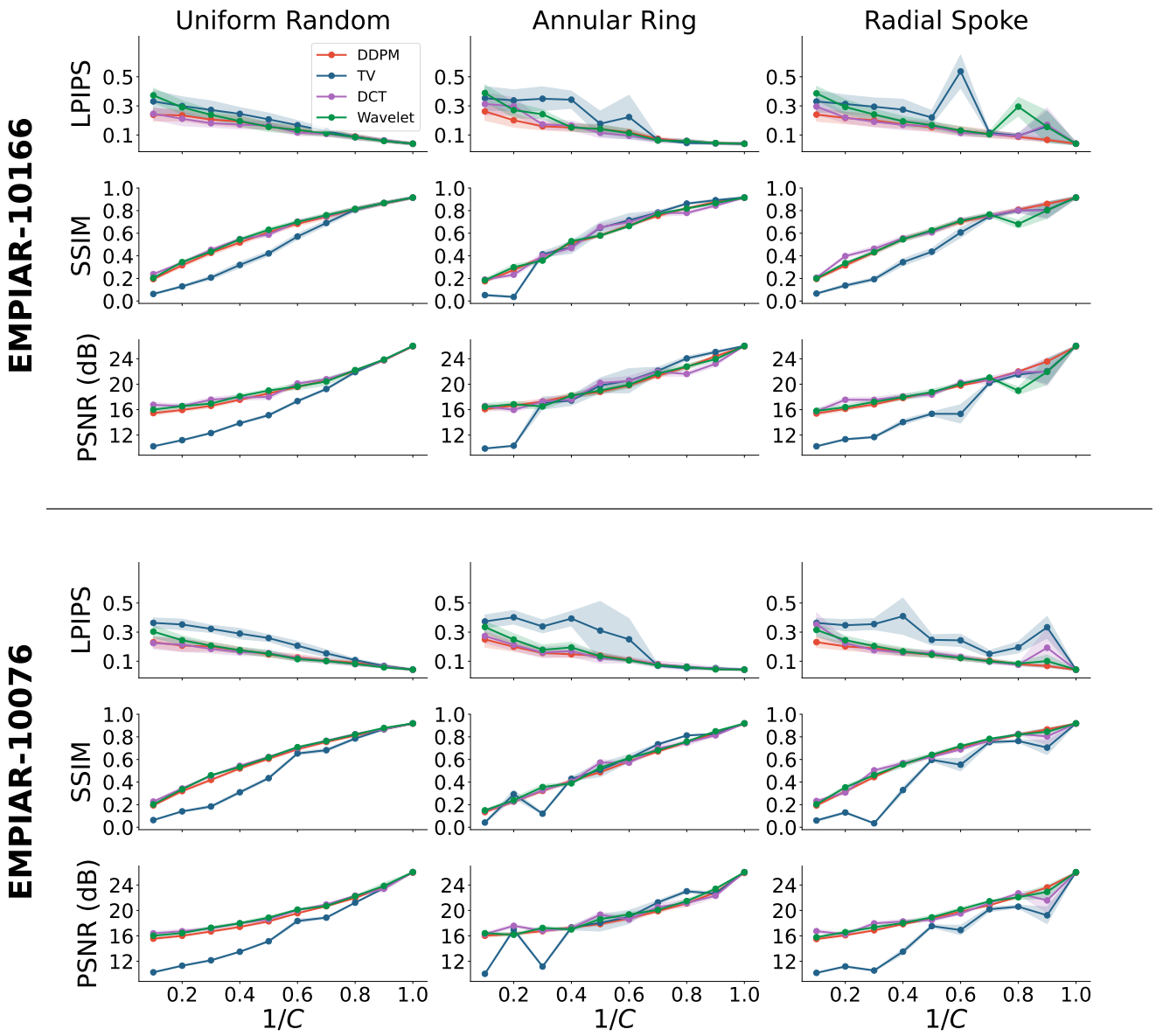

Figure S-10. LPIPS, SSIM, and PSNR scores for cryoSENSE reconstructions across various Fourier masks and $1/C$ for EMPIAR-10076 and EMPIAR-10166 in the presence of noise ($\sigma = 0.01$).

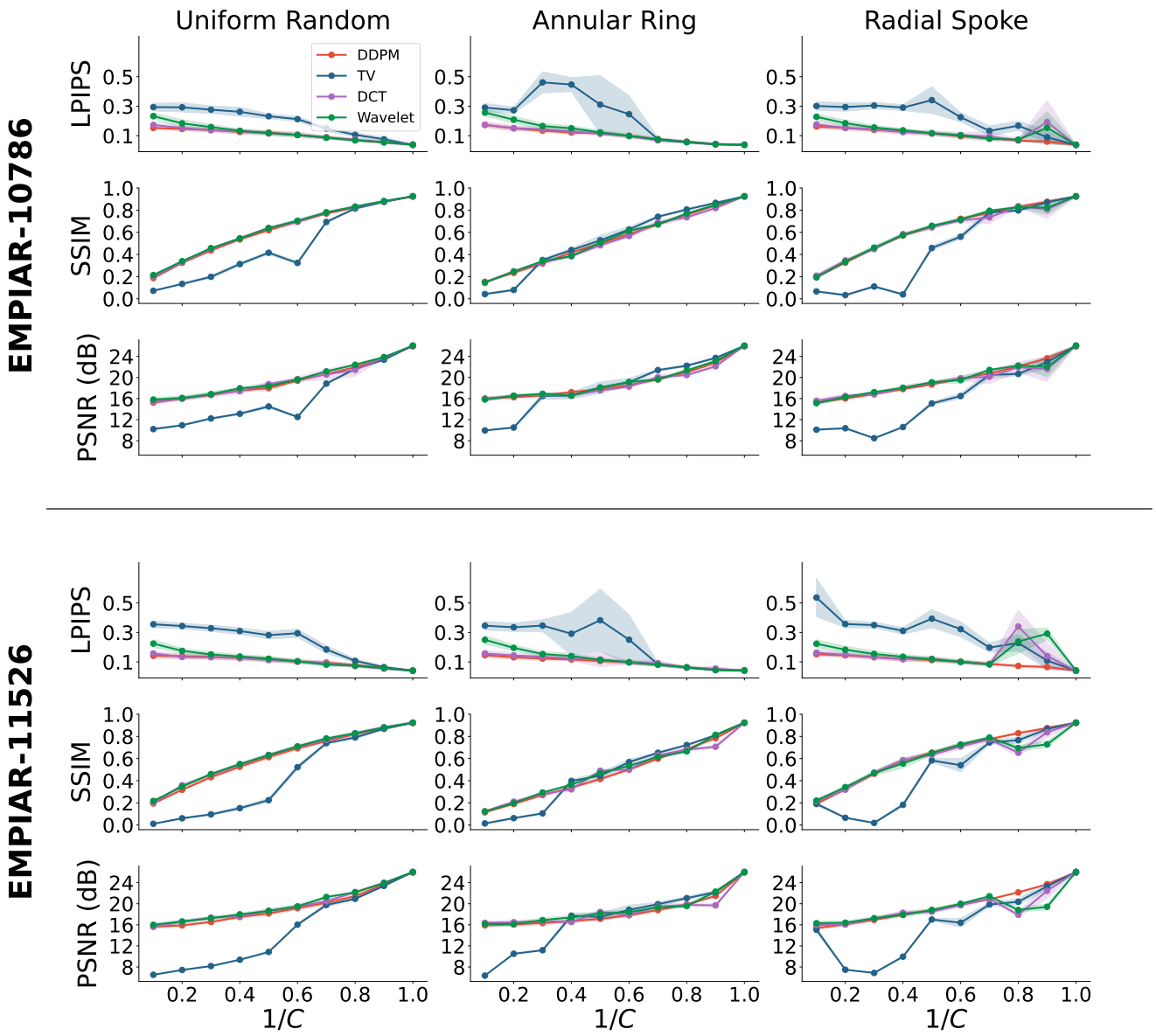

Figure S-11. LPIPS, SSIM, and PSNR scores for cryoSENSE reconstructions across various Fourier masks and $1/C$ for EMPIAR-10786 and EMPIAR-11526 in the presence of noise ($\sigma = 0.01$).

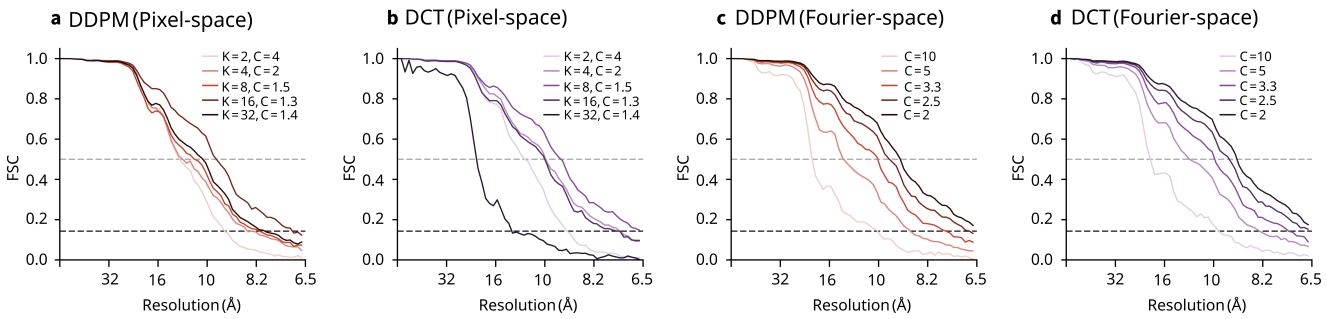

Figure S-12. FSC curves for cryoSENSE reconstructions on EMPIAR-10076 with pixel-space for **a,** DDPM and **b,** DCT and Fourier-space masking for **c,** DDPM and **d,** DCT.

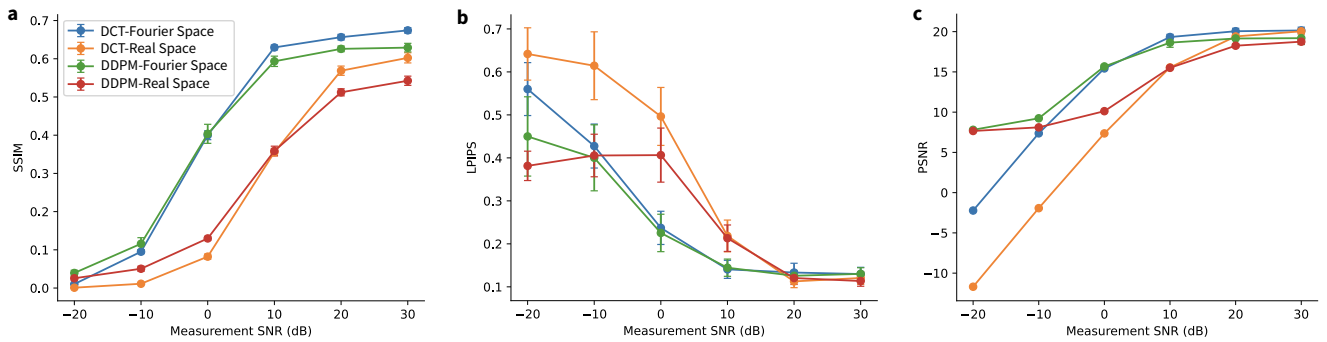

Figure S-13. Noise robustness analysis on EMPIAR-10076. **a**, SSIM, **b**, LPIPS, and **c**, PSNR as a function of measurement SNR for DCT and DDPM priors under pixel-space masking ($K = 4$, $b = 8$, $C = 2$) and Fourier-space masking ($C = 2.5$, uniform subsampling). Error bars indicate standard deviation across 16 validation images. Noise levels are calibrated to the variance of the noiseless measurements for each masking type to ensure comparable SNR across configurations.

References

[1] Alvaro Barbero and Suvrit Sra. Fast Newton-type methods for total variation regularization. In *International Conference on Machine Learning*, pages 313–320, 2011. 1

[2] Alvaro Barbero and Suvrit Sra. Modular proximal optimization for multidimensional total-variation regularization. *Journal of Machine Learning Research*, 19(56):1–82, 2018. 1

[3] Hyungjin Chung, Jeongsol Kim, Michael T. McCann, Marc L. Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. In *International Conference on Learning Representations*, 2023. 7

[4] Ingrid Daubechies, Michel Defrise, and Christine De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communications on Pure and Applied Mathematics*, 57(11):1413–1457, 2004. 1, 7

[5] Joseph H. Davis, Yong Zi Tan, Bridget Carragher, Clinton S. Potter, Dmitry Lyumkis, and James R. Williamson. Modular assembly of the bacterial large ribosomal subunit. *Cell*, 167 (6):1610–1622, 2016. 2, 3

[6] Julian A. Harris, Bryan Faust, Arisbel B. Gondin, Marc André Dämgen, Carl-Mikael Suomivuori, Nicholas A. Veldhuis, Yifan Cheng, Ron O. Dror, David M. Thal, and Aashish Manglik. Selective G protein signaling driven by substance P–neurokinin receptor dynamics. *Nature Chemical Biology*, 18(1):109–115, 2022. 3, 4

[7] David Haselbach, Jil Schrader, Felix Lambrecht, Fabian Henneberg, Ashwin Chari, and Holger Stark. Long-range allosteric regulation of the human 26S proteasome by 20S proteasome-targeting cancer drugs. *Nature Communications*, 8(1):15578, 2017. 2, 3

[8] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, pages 6840–6851, 2020. 2

[9] Ruben Sanchez-Garcia, Michael Saur, Javier Vargas, Carl Poelking, and Charlotte M. Deane. CESPED: A benchmark for supervised particle pose estimation in cryo-em. *Phys. Rev. Res.*, 6:023245, 2024. 2, 3, 4

[10] Michael Saur, Michael J. Hartshorn, Jing Dong, Judith Reeks, Gabor Bunkoczi, Harren Jhoti, and Pamela A. Williams. Fragment-based drug discovery using cryo-EM. *Drug Discovery Today*, 25(3):485–490, 2020. 3

[11] Jingyu Sun, Laurel F. Kinman, Dushyant Jahagirdar, Joaquin Ortega, and Joseph H. Davis. KsgA facilitates ribosomal small subunit maturation by proofreading a key structural lesion. *Nature Structural & Molecular Biology*, 30(10):1468–1480, 2023. 3, 5

[12] Hengkang Wang, Xu Zhang, Taihui Li, Yuxiang Wan, Tiancong Chen, and Ju Sun. DMPlug: A plug-in method for solving inverse problems with diffusion models. In *Advances in Neural Information Processing Systems*, 2024. 7

[13] Ellen D. Zhong, Tristan Bepler, Bonnie Berger, and Joseph H. Davis. CryoDRGN: reconstruction of heterogeneous cryo-EM structures using neural networks. *Nature Methods*, 18(2): 176–185, 2021. 2, 3