FROM PHYSICS TO EXPERIMENT: LEARNING RESOLUTION-CONTROLLABLE CRYO-EM DENSITY WITH STRUCTURED DIFFUSION

Anonymous authors
Paper under double-blind review

ABSTRACT

Cryo-electron microscopy (cryo-EM) density maps are central to macromolecular structure determination, yet generating synthetic maps that faithfully approximate experimental reconstructions from an atomic model remains challenging. Analytical renderers provide explicit resolution control but rely on simplified physical approximations, whereas learned methods can capture experimental characteristics but often provide limited controllability. We introduce **CryoNet.Density**, a 3D Brownian Bridge Diffusion Model that formulates cryo-EM density-map generation as *structured physical-to-experimental translation*. Given an atomic model, a differentiable, resolution-matched renderer first produces a physically rendered map; CryoNet.Density then learns the correction toward the paired experimental reference map rather than generating density from noise. Fourier-feature conditioning enables continuous inference from 1.5–10 Å, while a one-transition stochastic sampler requires only two denoiser evaluations. On a held-out 1,923-case test set, CryoNet.Density outperforms analytical renderers and Struc2mapGAN across seven spatial- and Fourier-domain metrics. The generated density maps further improve secondary-structure recognition and residue identification, supporting their structural fidelity. We additionally construct the **CryoNet.Density Database**, comprising 3,640,616 traceable generated density maps from 260,044 eligible PDB entries at 14 controlled resolutions. Together, these results establish structured physical-to-experimental diffusion as a scalable framework for controllable synthetic cryo-EM density-map generation.

1 INTRODUCTION

Cryo-electron microscopy (cryo-EM) density maps are a primary source of structural information, producing a rapidly growing collection of experimentally reconstructed density maps in the Electron Microscopy Data Bank (EMDB) (Lawson et al., 2013). In parallel, the Protein Data Bank (PDB) provides an increasingly comprehensive collection of atomic models describing macromolecular geometry (Burley et al., 2019). Bridging these two representations—from an input atomic model to an experimentally reconstructed density map—is important for developing, training, and evaluating computational methods for cryo-EM structure analysis. Yet, despite the apparent simplicity of this mapping, generating a synthetic density map that is both physically grounded and faithful to experimental reconstructions remains challenging.

A standard solution is to produce a physically rendered map from the atomic model. ChimeraX `molmap` (Pettersen et al., 2021), TEMPy `StructureBlurrer` (Farabella et al., 2015), `phenix.model_map` (Afonine et al., 2018), and EMAN2 `e2pdb2mrc` (Tang et al., 2007) use analytical atomic contributions with resolution-dependent smoothing or filtering. These methods are efficient and controllable, but their idealized maps do not fully reproduce experimental imaging, reconstruction, and heterogeneity effects, limiting their utility for training and evaluation.

Learning-based approaches learn mappings from atomic models to experimental reference maps using paired data. Struc2mapGAN (Zhang et al., 2025), for example, uses adversarial learning to generate experimental-like density maps. However, end-to-end synthesis does not explicitly distinguish features captured by analytical rendering from discrepancies that must be learned from

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

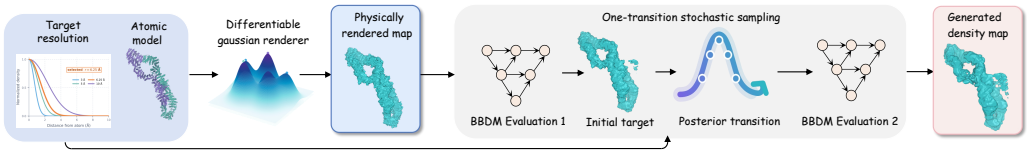

Figure 1: Workflow of CryoNet.Density. Two evaluations of the shared BBDM denoiser surround one stochastic posterior transition.

experimental data. Using a physically rendered map as the starting point allows learning to focus on these discrepancies.

We introduce CryoNet.Density, a resolution-conditioned 3D Brownian Bridge Diffusion Model for *physical-to-experimental map translation*. Starting from a physically rendered map of an atomic model at the requested nominal resolution, CryoNet.Density combines Fourier-FiLM conditioning with one-transition stochastic sampling requiring only two denoiser evaluations. From a curated collection of 19,273 atomic-model–experimental-reference-map pairs, we construct a cluster-disjoint benchmark with a 1,923-case held-out test set. CryoNet.Density outperforms analytical renderers and Struc2mapGAN, supports inference from 1.5–10 Å, and improves frozen downstream structure-analysis performance.

Our contributions are fourfold. (1) We formulate cryo-EM density-map generation as *structured physical-to-experimental translation*, using analytical rendering to anchor generation in atomic geometry and focusing learning on discrepancies from experimental reconstructions. (2) We develop a resolution-conditioned 3D Brownian Bridge Diffusion Model that integrates differentiable rendering, Fourier-FiLM conditioning, and a one-transition stochastic sampler, enabling continuous target-resolution specification with only two denoiser evaluations. (3) We demonstrate improved agreement with experimental reference maps and downstream structure-analysis performance with frozen models, linking map fidelity to structural utility. (4) We construct the CryoNet.Density Database, a 29.5-TB resource containing 3,640,616 traceable generated density maps from 260,044 eligible PDB entries at 14 target resolutions, providing a standardized resource for training and evaluating data-driven cryo-EM methods.

2 Related Work

2.1 Analytical Rendering and Learning-Based Map Generation

Analytical rendering converts an atomic model into a three-dimensional map by combining atomic contributions at a specified nominal resolution. ChimeraX `molmap` (Pettersen et al., 2021) and TEMPy `StructureBlurrer` (Farabella et al., 2015) use Gaussian-based representations, whereas `phenix.model_map` uses atomic scattering factors (Afonine et al., 2018); EMAN2 `e2pdb2mrc` (Tang et al., 2007) uses resolution-dependent Gaussian smoothing. These methods provide efficient, interpretable physically rendered maps, but their resolution parameters are not equivalent to an experimental estimate from independent half-maps and do not capture spatially varying signal quality, heterogeneity, or processing effects. Differentiable renderers, including CryoNet `molmap` (Huang et al., 2026), permit gradient propagation but do not by themselves model these characteristics. Struc2mapGAN (Zhang et al., 2025) instead learns a simulated-to-experimental mapping with a nested U-Net, but its published formulation lacks an explicit continuous target-resolution condition. Our method couples resolution-matched rendering with a conditioned stochastic bridge so that the requested resolution controls both the physically rendered map and the generated density map.

2.2 Diffusion Models and Brownian Bridges

Denoising diffusion models (Ho et al., 2020) learn to reverse a progressive corruption process, providing a flexible framework for generative modeling. For paired translation tasks, Brownian Bridge Diffusion Models (BBDM) (Li et al., 2023) construct stochastic paths between paired observations, making one observation an endpoint of the diffusion process rather than solely an auxiliary condition. This formulation is well suited to learning transformations between physically rendered maps and experimental reference maps that share molecular geometry but differ in local density patterns and frequency content. Conditional generation additionally requires a mechanism for representing

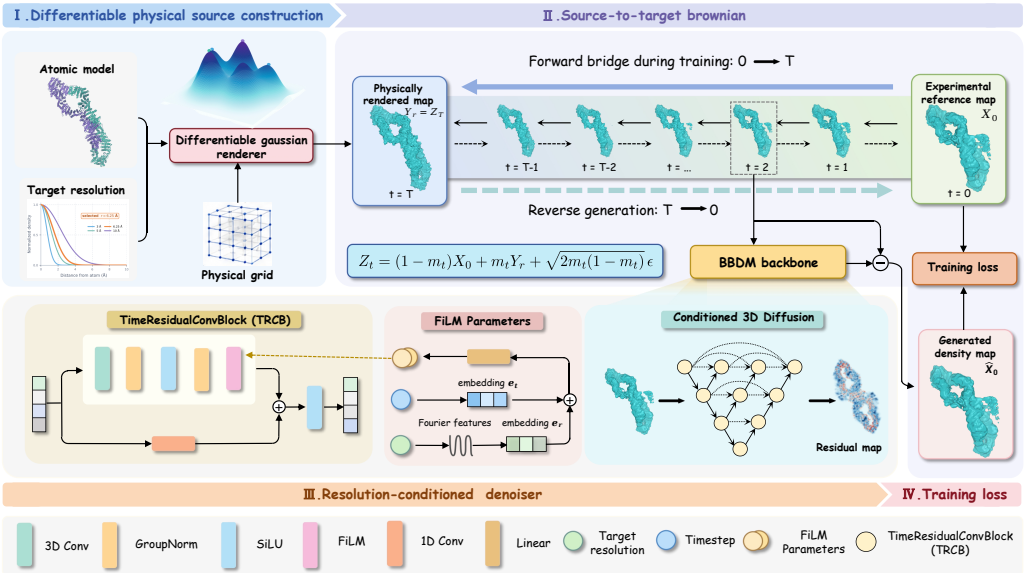

Figure 2: Training of CryoNet.Density. Each example samples one physical-to-experimental BBDM state; supervision is applied to the residual and reconstructed experimental reference map.

the requested output properties: FiLM (Perez et al., 2018) modulates intermediate features through condition-dependent scaling and shifting, which we use with Fourier features to encode continuous target resolution. In structural biology, AlphaFold 3 (Abramson et al., 2024) applies diffusion to atomic-coordinate generation, while CryoNet.Refine (Huang et al., 2026) uses diffusion-based updates for density-guided coordinate refinement. These methods operate on atomic models, whereas our task concerns volumetric map generation conditioned on an atomic model and a target resolution. Sampling efficiency is a further consideration for three-dimensional maps. Prior studies identify discrepancies between training and sampling states as a potential source of error in iterative diffusion sampling (Ning et al., 2023; 2024; Li & van der Schaar, 2024). Consistency models (Song et al., 2023) enable one- or few-step generation by learning mappings along probability-flow trajectories, and consistency diffusion bridge models extend this approach to paired-observation bridges. Within this broader literature, our contribution is a resolution-conditioned bridge for cryo-EM map generation, coupled with a one-transition stochastic sampler requiring two denoiser evaluations.

3 Methods

3.1 Overview of CryoNet.Density

Figures 1 and 2 summarize inference and training, respectively. We denote the input atomic model by P, its physically rendered map at target nominal resolution r by Y_r, the paired experimental reference map by X_0, the bridge state at timestep t by Z_t, and the generated density map by $\widehat{X}_0$. The framework combines differentiable rendering, a Brownian bridge between paired maps, and a resolution-conditioned denoiser trained with a shared map-level objective. Each denoiser evaluation predicts a residual that is subtracted from the current bridge state to estimate the experimental reference map.

Given the non-hydrogen atoms of an atomic model, $P = \{(\mathbf{p}_i, w_i)\}_{i=1}^{N}$, where $\mathbf{p}_i$ and w_i denote atomic coordinates and rendering weights, a target resolution r, and an output grid $\mathcal{G} = (\mathbf{n}, v, \mathbf{o})$ with dimensions $\mathbf{n}$, isotropic voxel spacing v, and origin $\mathbf{o}$, CryoNet.Density computes a physically rendered map using the differentiable renderer from CryoNet.Refine (Huang et al., 2026):

$$Y_r = R(P, r; \mathcal{G}). \tag{1}$$

The denoiser learns the residual $D_t = Z_t - X_0$, with additional supervision at the clean physically rendered map to match initialization at inference. Default sampling uses two evaluations of the same

denoiser, separated by one stochastic bridge transition. Starting from $Z_{\mathrm{src}} = Y_r$,

$$\widetilde{X}_0 = \mathrm{clip}_{[0,1]}\left[Z_{\mathrm{src}} - f_{\bar{\theta}}(Z_{\mathrm{src}}, T-1, Y_r, r)\right], \tag{2}$$

$$Z_{\mathrm{tar}} \sim q_{T-1\to 0}\left(\cdot \mid Z_{\mathrm{src}}, \widetilde{X}_0, Y_r\right), \tag{3}$$

$$\widehat{X}_0 = \mathrm{clip}_{[0,1]}\left[Z_{\mathrm{tar}} - f_{\bar{\theta}}(Z_{\mathrm{tar}}, 0, Y_r, r)\right]. \tag{4}$$

Here $f_{\bar{\theta}}$ denotes the inference denoiser. The first evaluation produces an experimental-reference estimate $\widetilde{X}_0$, which replaces the unavailable experimental reference map in the Gaussian bridge posterior, $q_{T-1\to 0} = \mathcal{N}(\mu_{T-1\to 0}, \widetilde{\delta}_{T-1\to 0} I)$. The second evaluation converts the sampled experimental-reference-side state into the final generated density map. Appendix C specifies the posterior parameters, physically-rendered-map-boundary supervision, and complete algorithms. The sole stochastic operation during default sampling is

$$Z_{\mathrm{tar}} = \mu_{T-1\to 0}\left(Z_{\mathrm{src}}, \widetilde{X}_0; Y_r\right) + \sqrt{\widetilde{\delta}_{T-1\to 0}}\,\xi, \qquad \xi \sim \mathcal{N}(0, I). \tag{5}$$

Both evaluations retain the same clean physically rendered map Y_r as a spatial condition; only the bridge state changes.

3.1.1 Differentiable Atomic Model Rendering

For a voxel index $\mathbf{j}_{\mathrm{xyz}} = (j_x, j_y, j_z)$ in physical (x, y, z) axis order and atom centers $\mathbf{c}_i = (\mathbf{p}_i - \mathbf{o})/v$ in voxel coordinates, the renderer computes

$$Y_r^{\mathrm{raw}}(\mathbf{j}_{\mathrm{xyz}}) = \sum_{i=1}^{N} w_i \exp\left(-\frac{\|\mathbf{j}_{\mathrm{xyz}} - \mathbf{c}_i\|_2^2}{\rho(r,v)^2}\right), \qquad \rho(r,v) = \frac{r}{\pi v}. \tag{6}$$

The physically rendered map encodes atomic geometry with a resolution-dependent Gaussian width. During training, rendering on the experimental-reference-map grid places paired maps in the same physical coordinate frame. Here r controls the renderer's nominal bandwidth rather than specifying an independently measured map resolution. Data processing and variable-size batching are described in Appendices F and B. This construction supplies the geometric starting point for learning discrepancies from experimental reconstructions.

3.1.2 Physical-to-Experimental Brownian Bridge

For each aligned pair (Y_r, X_0), BBDM (Li et al., 2023) samples a timestep $t \in \{0, \ldots, T-1\}$ and a Gaussian noise field to construct a bridge state and its residual target:

$$Z_t = (1 - m_t)X_0 + m_t Y_r + \sqrt{\delta_t}\,\epsilon, \qquad \epsilon \sim \mathcal{N}(0, I), \tag{7}$$

$$D_t = Z_t - X_0 = m_t(Y_r - X_0) + \sqrt{\delta_t}\,\epsilon, \qquad \delta_t = 2(m_t - m_t^2)\delta_{\max}. \tag{8}$$

We use $T = 1000$, $\delta_{\max} = 1$, and a linear schedule from $m_0 = 0.001$ to $m_{T-1} = 0.999$. The denoiser estimates the experimental reference map as $\widehat{X}_0 = Z_t - f_\theta(Z_t, t, Y_r, r)$, with the physically rendered map Y_r retained as a spatial condition. The forward process runs from the experimental-reference-map side toward the physically-rendered-map side, whereas generation starts from the physically rendered map and proceeds in the reverse direction. Because the numerical schedule excludes the exact endpoints, physically-rendered-map initialization is handled by the additional boundary supervision described in Appendix C. Sampling (t, ϵ) exposes the denoiser to both low-variance states near the endpoints and higher-variance states in the bridge interior.

3.1.3 Resolution-Conditioned Denoising with Fourier-FiLM

The residual predictor is a four-scale residual 3D U-Net++ (Zhou et al., 2018) that takes channel-concatenated Z_t and Y_r and returns a residual field of the same spatial dimensions. Multiscale features combine local density patterns with broader molecular context. Sinusoidal timestep embeddings and continuous resolution Fourier features jointly determine FiLM (Perez et al., 2018) scale and shift parameters in every residual block. Resolution thus conditions both rendering and denoising:

$$r \longrightarrow \left\{R(P, r; \mathcal{G}),\ \phi(r) \longrightarrow \mathrm{FiLM}\right\}. \tag{9}$$

These two paths control the rendered-map bandwidth and the learned transformation, respectively. Appendix D details the architecture and FiLM placement. For resolution bounds $r_{\min}$ and $r_{\max}$, the continuous encoding is

$$u(r) = \frac{r - r_{\min}}{r_{\max} - r_{\min}}, \qquad \phi(r) = \left[2u(r) - 1,\ \{\sin(2\pi f u(r)), \cos(2\pi f u(r))\}_{f\in\{1,2,4\}}\right]. \quad (10)$$

The linear component and Fourier features represent resolution at multiple scales without discrete bins. The timestep and resolution embeddings are added before each FiLM projection, jointly conditioning every residual block on bridge position and target resolution.

3.1.4 Training and One-Transition Stochastic Sampling

Training alternates between random bridge states with probability $1 - p_{\text{src}}$ and the clean physically rendered map with probability $p_{\text{src}} = 0.1$. The boundary branch uses Y_r at timestep $T - 1$ with target residual $Y_r - X_0$, matching inference initialization. Both branches share the denoiser and objective. At inference, the first map estimate parameterizes a Gaussian bridge posterior; a second denoiser evaluation converts the sampled state into the generated density map (Figure 1). Thus, the default sampler uses one stochastic transition and two denoiser evaluations without traversing the bridge interior. Longer trajectories use the same denoiser at additional scheduled states conditioned on the preceding map estimate.

3.2 Training Objective

For an input state Z_t, the denoiser predicts $\widehat{D}_t = f_\theta(Z_t, t, Y_r, r)$ and reconstructs the generated density map $\widehat{X}_0 = Z_t - \widehat{D}_t$. Both training branches, indexed by $h \in \{\text{bridge}, \text{boundary}\}$, minimize a weighted residual L_1 loss and a reconstructed-map Pearson correlation coefficient (PCC) loss. Let M be the soft molecular mask from the physically rendered map and V the valid-voxel mask excluding padding. With $W_1 = (0.1 + 0.9M)V$ and $W_{\text{PCC}} = \max(M, 0.01)V$, where the maximum is elementwise,

$$\mathcal{L}_1(h) = \frac{\sum_{\mathbf{u}} W_{1,\mathbf{u}} |\widehat{D}^h_{\mathbf{u}} - D^h_{\mathbf{u}}|}{\sum_{\mathbf{u}} W_{1,\mathbf{u}} + \varepsilon}, \quad (11)$$

$$\mathcal{L}_{\text{PCC}}(h) = 1 - \text{PCC}_{W_{\text{PCC}}}(X_0, \widehat{X}^h_0). \quad (12)$$

Here $\mathbf{u}$ indexes voxels and $\varepsilon > 0$ ensures numerical stability. The combined map-level loss is

$$\mathcal{L}_{\text{map}}(h) = \mathcal{L}_1(h) + \lambda_{\text{PCC}}\mathcal{L}_{\text{PCC}}(h), \qquad \lambda_{\text{PCC}} = 0.1. \quad (13)$$

For a branch indicator $b \sim \text{Bernoulli}(p_{\text{src}})$, the stochastic training loss is

$$\mathcal{L} = (1 - b)\mathcal{L}_{\text{map}}(\text{bridge}) + b\mathcal{L}_{\text{map}}(\text{boundary}). \quad (14)$$

The residual L_1 loss constrains voxelwise map amplitudes, since $\widehat{D}_t - D_t = X_0 - \widehat{X}_0$, while the PCC loss promotes spatial correlation. The weights emphasize molecular regions while retaining background supervision and excluding padded voxels. Appendix E specifies mask construction, weighted correlation, safeguards, and training details. Unlike direct regression from Y_r to X_0, bridge training includes perturbed states and the clean physically rendered map. Both branches target the residual to the experimental reference map: $Z_t - X_0$ for bridge states and $Y_r - X_0$ for the physically rendered map boundary.

4 Experiments

We evaluate fidelity to experimental reference maps, resolution control at fixed atomic coordinates, and structural utility for frozen downstream models. Ablations examine the contributions of resolution-matched rendering, Fourier-FiLM conditioning, bridge training, and sampling trajectory.

4.1 Benchmarks and Baselines

We construct the primary benchmark from 19,273 atomic-model–experimental-reference-map pairs from EMDB and a particle-subsampled reconstruction set (`te2_maps`; Appendix F). For the latter, we sample multiple particle subsets and use cryoSPARC (Punjani et al., 2017) to reconstruct

lower-resolution maps de novo. After filtering at Phenix $CC_{\text{mask}} \geq 0.5$, PDB-ID and RCSB 40% sequence-cluster assignment (RCSB Protein Data Bank, 2026b) yields 13,659 training, 1,936 validation, 1,923 test, and 1,755 homology-excluded cases. We also evaluate the 130-pair Struc2mapGAN benchmark. Baselines are ChimeraX `molmap`, `phenix.model_map`, EMAN2 `e2pdb2mrc`, TEMPy `StructureBlurrer`, Struc2mapGAN, and CryoNet `molmap`, which supplies our physically rendered maps. All methods share test cases, reference grids, and scoring; Appendix G gives implementation details.

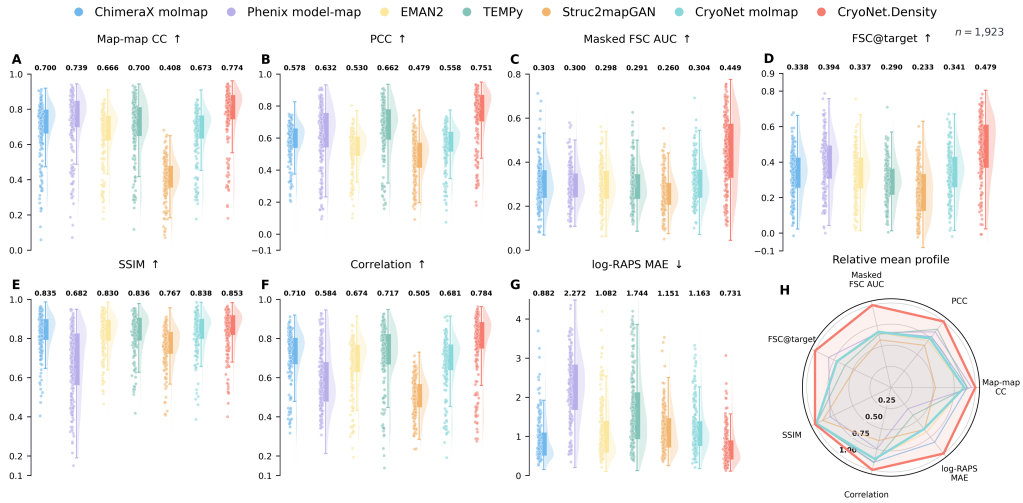

Figure 3: Comparison of generated maps with paired experimental reference maps across seven evaluation metrics. Panels A–G report scores for individual test cases; panel H summarizes relative performance based on mean scores across the test set.

4.2 Agreement with Experimental Reference Maps

We assess map fidelity using seven metrics defined in Appendix H. Map–map CC, Cosine similarity, and PCC quantify real-space agreement; SSIM assesses local structure. Masked FSC AUC summarizes Fourier-shell correlation, FSC@target measures correlation at spatial frequency $1/r$, and log-RAPS MAE compares logarithmic radial power spectra. CryoNet.Density achieves the best mean and median for all seven metrics (Table 1); its mean Masked FSC AUC is 0.449 versus 0.304 for the strongest baseline, while mean PCC increases from 0.662 to 0.751. Appendix I gives stratified and Struc2mapGAN-benchmark results.

Table 1: Map fidelity on the primary benchmark against paired experimental reference maps. Higher is better except for log-RAPS MAE; bold marks the best statistic.

Method	Statistic	Map–map CC	Masked FSC AUC	FSC@target	SSIM	Cosine similarity	log-RAPS MAE	PCC
ChimeraX `molmap`	Mean	0.700	0.303	0.338	0.835	0.710	0.882	0.578
	Median	0.745	0.296	0.344	0.854	0.752	0.701	0.610
`phenix.model_map`	Mean	0.739	0.300	0.394	0.682	0.584	2.272	0.632
	Median	0.786	0.289	0.410	0.718	0.563	2.305	0.665
EMAN2	Mean	0.666	0.298	0.337	0.830	0.674	1.082	0.530
	Median	0.710	0.292	0.343	0.849	0.715	0.839	0.556
TEMPy	Mean	0.700	0.291	0.290	0.836	0.717	1.744	0.662
	Median	0.730	0.282	0.290	0.860	0.745	1.444	0.699
Struc2mapGAN	Mean	0.408	0.260	0.233	0.767	0.505	1.151	0.479
	Median	0.420	0.253	0.230	0.784	0.510	1.038	0.502
CryoNet `molmap`	Mean	0.673	0.304	0.341	0.838	0.681	1.163	0.558
	Median	0.716	0.297	0.348	0.862	0.720	1.043	0.587
CryoNet.Density	Mean	**0.774**	**0.449**	**0.479**	**0.853**	**0.784**	**0.731**	**0.751**
	Median	**0.831**	**0.463**	**0.503**	**0.879**	**0.835**	**0.603**	**0.807**

Per-case comparisons complement the aggregate scores, showing improved agreement with experimental reference maps in both spatial density patterns and frequency content (Figure 3).

Qualitative examples from three resolution intervals illustrate molecular envelopes and internal density features (Figure 4), with complete stratified comparisons in Appendix I. We next assess target-resolution control while holding the input atomic model fixed.

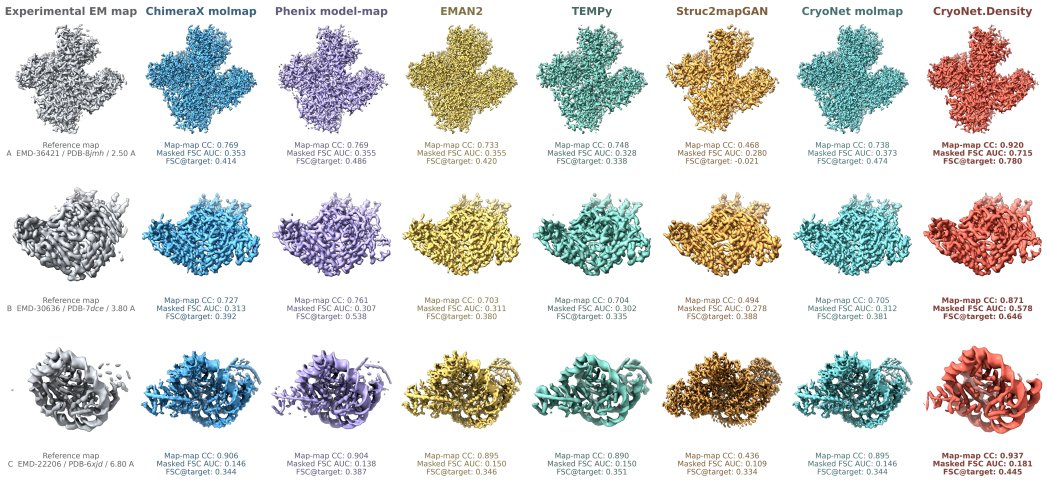

Figure 4: Experimental reference maps and generated density maps across three resolution intervals.

4.3 Resolution-Controlled Map Generation

For the fixed atomic model PDB-*7kod*, CryoNet.Density more closely reproduces the resolution-dependent differences among six deposited experimental reference maps than the comparison methods (Figure 5).

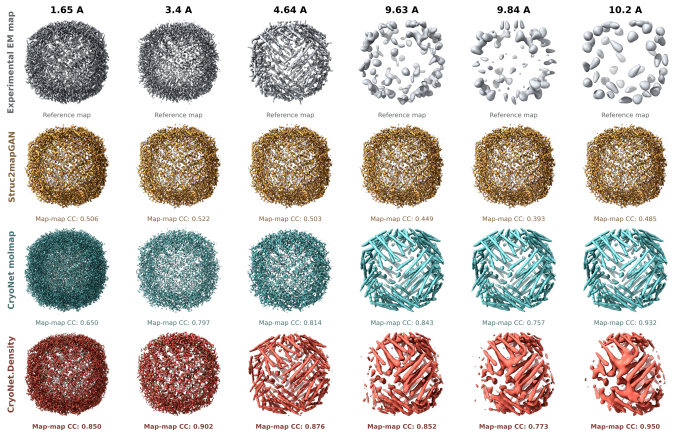

Figure 5: Resolution-controlled generation from atomic model PDB-*7kod*, compared with six experimental reference maps.

Across 96 test atomic models, sweeps over five target resolutions spanning 2–10 Å show consistent monotonic changes in effective spectral bandwidth (Appendix I.3). We estimate bandwidth from each generated density map's radial power spectrum and compute its Spearman correlation with target resolution; a negative correlation is expected because larger resolution values correspond to lower spatial-frequency cutoffs. We also quantify the error in recovering the requested resolution from the estimated bandwidth. Renderer-only and FiLM-only ablations assess the two conditioning paths. We next test downstream structural interpretation.

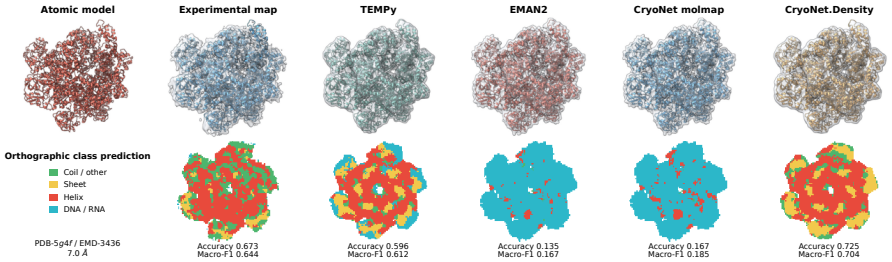

Figure 6: Secondary-structure recognition using Emap2sec+ on EMD-3436/PDB-*5g4f*. Top: each density map is overlaid with the deposited atomic model, with isosurfaces enclosing equal voxel fractions. Bottom: projected predictions for coil/other, sheet, helix, and DNA/RNA after spatial refinement and four-fold probability fusion. Labels report accuracy and macro-F1. Predictions from the experimental reference map and the CryoNet.Density generated density map agree on 67.2% of jointly evaluated voxels.

4.4 Downstream Evaluation of Experimental-Like Map Features

We test whether generated density maps reproduce task-relevant features (such as secondary-structure and residue-type classification) of experimental reference maps using frozen Emap2sec+ (Wang et al., 2021) and ModelAngelo (Jamali et al., 2023). Both models are evaluated without retraining.

For Emap2sec+, we use the 136 complete 5–10 Å cases from the 1,923-case primary test cohort and report case-equal means; DNA/RNA F1 uses its 35 labeled cases. Table 2 gives the results, and Figure 6 shows EMD-3436/PDB-*5g4f* (7.0 Å).

For ModelAngelo, we use the fixed 711-case subset of 1,524 protein-containing 1.88–4.0 Å primary-test cases with nonzero sequence matches for all four inputs. Table 3 reports case-equal means using a 3 Å chain-matching radius. CryoNet.Density leads in all three metrics. Appendix I.4 gives both protocols and cohort derivations.

Table 2: Frozen Emap2sec+ on 136 complete 5–10 Å cases (DNA/RNA: 35 labeled cases). Case-equal means; underlines mark second best.

Metric	Exp.	TEMPy	EMAN2	Renderer	Ours
Accuracy	<u>0.687</u>	0.520	0.517	0.461	**0.698**
Helix F1	**0.766**	0.538	0.590	0.595	<u>0.733</u>
Sheet F1	**0.364**	0.136	0.165	0.078	<u>0.339</u>
Coil/other F1	<u>0.521</u>	0.494	0.486	0.380	**0.596**
DNA/RNA F1	0.713	**0.867**	0.502	0.612	**0.867**

Table 3: Frozen ModelAngelo: case-equal means on the 711-case paired-success cohort.

Input map	Residue accuracy	Sequence match (%)	Sequence score
Experimental	0.687	92.6	0.886
ChimeraX `molmap`	0.522	86.6	0.861
CryoNet `molmap`	0.538	87.6	0.870
CryoNet.Density	**0.698**	**95.7**	**0.949**

Exp.: experimental reference map; Renderer: CryoNet `molmap`; Ours: CryoNet.Density. Residue accuracy uses 20 amino-acid classes.

5 Ablation Study

We isolate resolution-matched rendering, Fourier-FiLM, BBDM training, and transition count on the same test cohort. The full model with the one-transition stochastic sampler is the reference unless stated otherwise.

Resolution conditioning and bridge training. A0 removes both resolution paths; A1 retains only Fourier-FiLM; and A2 retains only resolution-matched rendering. A1 improves Fourier agreement, A2 favors SSIM and mean PCC, and their combination best balances spatial and spectral fidelity (Appendix J.1). BBDM training also outperforms same-backbone direct regression across all reported metrics. Matched rendering establishes the physically rendered map's bandwidth, whereas Fourier-FiLM selects the request-dependent correction; neither alone reproduces their joint balance. Appendix Tables 17 and 18 compare same-backbone direct regression with 1-NFE BBDM Direct, then fixed-weight 1- versus 2-NFE sampling.

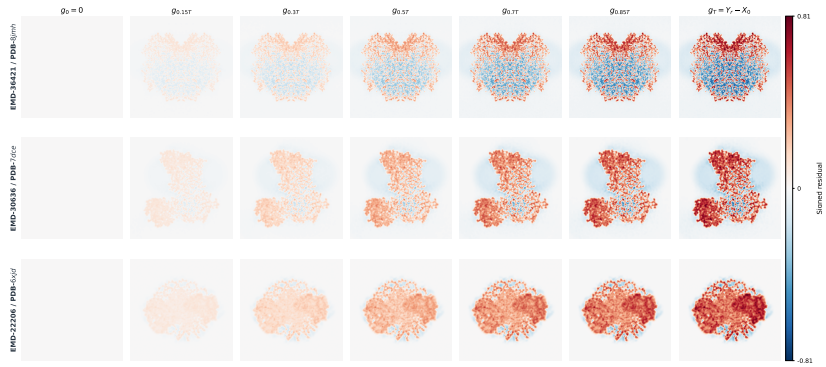

Figure 7: Mean bridge residuals $g_t = \mathbb{E}_\epsilon[Z_t - X_0] = m_t(Y_r - X_0)$ for the cases in Figure 4. Visualization includes $m = 0$ and $m = 1$; training uses $m_t \in [0.001, 0.999]$. The sequence scales a fixed map difference.

Figure 7 makes the BBDM objective concrete: along the mean bridge path, the supervised residual grows continuously from zero at the experimental-reference boundary to $Y_r - X_0$ at the physically-rendered-map boundary.

Sampling trajectory. One transition yields the strongest spatial and Fourier-shell fidelity among the tested schedules, whereas longer trajectories gradually reduce agreement. The trend is consistent with finite prediction error shifting recursively constructed bridge states as the trajectory enters the high-variance bridge interior (Ning et al., 2023; 2024; Li & van der Schaar, 2024). Appendix J.3.1 provides the full sweep, matched-compute intervention, and oracle checks; Appendix J.4 reports conditional diversity.

Source-to-reference residual. The paired correction $C = X_0 - Y_r$ is the density that must be added to the physically rendered map. A useful bookkeeping model is

$$C \approx (H_{\text{eff}} - H_G)O_{\text{model}} + H_{\text{eff}}\Delta O + n_{\text{recon}} + C_{\text{prep}}, \tag{15}$$

where the terms collect transfer-function differences, density absent from the fitted atomic model, reconstruction variation, and deterministic preprocessing. Empirically, C has a colored spectrum, spatial autocorrelation, and energy near molecular support. Phase randomization preserves its spectrum but removes alignment with gradients and curvature of the physically rendered map, identifying a structured map–model difference field (Joseph et al., 2020). The residual can combine imaging, reconstruction, heterogeneity, and atomic-model effects; quantitative controls appear in Appendix J.5.

6 CryoNet.Density Database

Using the frozen inference pipeline over 1.5–10 Å, we generate 3,640,616 density maps (29.542 TB uncompressed) from 260,044 eligible entries in the RCSB PDB snapshot dated September 17, 2026 (RCSB Protein Data Bank, 2026a). The website will provide these maps, traceable metadata, and generation from user-supplied atomic models; Appendix K gives the filtering and resolution distribution.

7 Conclusion

We introduced CryoNet.Density, a resolution-conditioned 3D Brownian Bridge Diffusion Model that translates physically rendered maps of atomic models into experimental-like cryo-EM density maps. On held-out benchmarks, CryoNet.Density outperforms analytical renderers and Struc2mapGAN across spatial- and Fourier-domain metrics. Improved recognition by frozen downstream models supports the compatibility of generated maps with task-relevant experimental density features. The CryoNet.Density Database provides a large-scale synthetic-map resource for developing and evaluating data-driven cryo-EM methods. However, generation depends on input atomic-model accuracy and training-data coverage, and may not reproduce experimental features arising from unmodeled conformations or acquisition conditions underrepresented in training.

Statement on the Use of Artificial Intelligence

We used artificial-intelligence tools to improve the clarity, grammar, and presentation of the manuscript and to assist with research-code development, including implementation, refactoring, debugging, and documentation. All AI-assisted text and code were reviewed, revised, and verified by the authors. The authors made all decisions concerning the research design, experiments, analysis, interpretation, and scientific claims, and take full responsibility for the content of this work.

References

Josh Abramson, Jonas Adler, Jack Dunger, Richard Evans, Tim Green, Alexander Pritzel, Olaf Ronneberger, Lindsay Willmore, Andrew Ballard, Joshua Bambrick, Sebastian Bodenstein, David Evans, Chia-Chun Hung, Michael O'Neill, David Reiman, Kathryn Tunyasuvunakool, Zachary Wu, Akvilė Žemgulytė, Eirini Arvaniti, and John Jumper. Accurate structure prediction of biomolecular interactions with alphafold 3. *Nature*, 630:493–500, 05 2024. doi: 10.1038/s41586-024-07487-w.

Pavel Afonine, Bruno Klaholz, Nigel Moriarty, Billy Poon, Oleg Sobolev, Tom Terwilliger, Paul Adams, and Alexandre Urzhumtsev. New tools for the analysis and validation of cryo-em maps and atomic models. *Acta Crystallographica Section D*, 74:814–840, 09 2018. doi: 10.1107/S2059798318009324.

Alok Bharadwaj and Arjen Jakobi. Electron scattering properties of biological macromolecules and their use for cryo-em map sharpening. *Faraday Discussions*, 240:168–183, 08 2022. doi: 10.1039/D2FD00078D.

Stephen K Burley, Helen M Berman, Charmi Bhikadiya, Chunxiao Bi, Li Chen, Luigi Di Costanzo, Cole Christie, Ken Dalenberg, Jose M Duarte, Shuchismita Dutta, Zukang Feng, Sutapa Ghosh, David S Goodsell, Rachel K Green, Vladimir Guranović, Dmytro Guzenko, Brian P Hudson, Tara Kalro, Yuhe Liang, Robert Lowe, Harry Namkoong, Ezra Peisach, Irina Periskova, Andreas Prlić, Chris Randle, Alexander Rose, Peter Rose, Raul Sala, Monica Sekharan, Chenghua Shao, Lihua Tan, Yi-Ping Tao, Yana Valasatava, Maria Voigt, John Westbrook, Jesse Woo, Huanwang Yang, Jasmine Young, Marina Zhuravleva, and Christine Zardecki. Rcsb protein data bank: biological macromolecular structures enabling research and education in fundamental biology, biomedicine, biotechnology and energy. *Nucleic Acids Research*, 47(D1):D464–D474, 01 2019. ISSN 0305-1048. doi: 10.1093/nar/gky1004. URL https://doi.org/10.1093/nar/gky1004.

Irene Farabella, Daven Vasishtan, Agnel Praveen Joseph, Arun Prasad Pandurangan, Harpal Sahota, and Maya Topf. Tempy: a python library for assessment of three-dimensional electron microscopy density fits. *Journal of Applied Crystallography*, 48:1314 – 1323, 2015. URL https://api.semanticscholar.org/CorpusID:16940678.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 6840–6851. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/4c5bcfec8584af0d967f1ab10179ca4b-Paper.pdf.

Fuyao Huang, Xiaozhu Yu, Kui Xu, and Qiangfeng Zhang. Cryonet.refine: A one-step diffusion model for rapid refinement of structural models with cryo-em density map restraints. In C. Vondrick, B. Hariharan, C. Raffel, L. Pinto, D. Yang, and A. Faust (eds.), *International Conference on Learning Representations*, volume 2026, pp. 1448–1470, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/file/0354c884ec52dcdfd928dca663893298-Paper-Conference.pdf.

Kiarash Jamali, Lukas Käll, Rui Zhang, Alan Brown, Dari Kimanius, and Sjors Scheres. Automated model building and protein identification in cryo-em maps. *bioRxiv : the preprint server for biology*, 05 2023. doi: 10.1101/2023.05.16.541002.

Agnel Praveen Joseph, Ingvar Lagerstedt, Arjen Jakobi, Tom Burnley, Ardan Patwardhan, Maya Topf, and Martyn Winn. Comparing cryo-em reconstructions and validating atomic model fit

using difference maps. *Journal of Chemical Information and Modeling*, 60(5):2552–2560, 05 2020. ISSN 1549-9596. doi: 10.1021/acs.jcim.9b01103. URL https://doi.org/10.1021/acs.jcim.9b01103.

Catherine L. Lawson, Ardan Patwardhan, Matthew L. Baker, Corey F. Hryc, Eduardo Sanz Garcia, Brian P. Hudson, Ingvar C. Lagerstedt, Steven J. Ludtke, Grigore D. Pintilie, Raul Sala, John D. Westbrook, Helen M. Berman, Gerard J. Kleywegt, and Wah Chiu. Emdatabank unified data resource for 3dem. *Nucleic Acids Research*, 44:D396 – D403, 2013. URL https://api.semanticscholar.org/CorpusID:6760338.

Bo Li, Kaitao Xue, Bin Liu, and Yu-Kun Lai. Bbdm: Image-to-image translation with brownian bridge diffusion models. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1952–1961, 2023. doi: 10.1109/CVPR52729.2023.00194.

Yangming Li and Mihaela van der Schaar. On error propagation of diffusion models. In B. Kim, Y. Yue, S. Chaudhuri, K. Fragkiadaki, M. Khan, and Y. Sun (eds.), *International Conference on Learning Representations*, volume 2024, pp. 32791–32807, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/file/8b465dd58ac50e1b0b22894fd581f62f-Paper-Conference.pdf.

Mang Ning, Enver Sangineto, Angelo Porrello, Simone Calderara, and Rita Cucchiara. Input perturbation reduces exposure bias in diffusion models. In *Proceedings of the 40th International Conference on Machine Learning*, ICML'23. JMLR.org, 2023.

Mang Ning, Mingxiao Li, Jianlin Su, Albert Ali Salah, and Itir Onal Ertugrul. Elucidating the exposure bias in diffusion models. In B. Kim, Y. Yue, S. Chaudhuri, K. Fragkiadaki, M. Khan, and Y. Sun (eds.), *International Conference on Learning Representations*, volume 2024, pp. 15167–15189, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/file/4267d84ca2f6fbb4aa5172b76b433aca-Paper-Conference.pdf.

Ethan Perez, Florian Strub, Harm de Vries, Vincent Dumoulin, and Aaron Courville. Film: visual reasoning with a general conditioning layer. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence and Thirtieth Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Educational Advances in Artificial Intelligence*, AAAI'18/IAAI'18/EAAI'18. AAAI Press, 2018. ISBN 978-1-57735-800-8.

Eric F. Pettersen, Thomas D. Goddard, Conrad C. Huang, Elaine C. Meng, Gregory S. Couch, Tristan I. Croll, John H. Morris, and Thomas E. Ferrin. Ucsf chimerax: Structure visualization for researchers, educators, and developers. *Protein Science*, 30(1):70–82, 2021. doi: https://doi.org/10.1002/pro.3943. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/pro.3943.

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. doi: 10.1038/nmeth.4169. URL https://doi.org/10.1038/nmeth.4169.

RCSB Protein Data Bank. File download services: Holdings data. https://www.rcsb.org/docs/programmatic-access/file-download-services, 2026a. Accessed September 17, 2026.

RCSB Protein Data Bank. Sequence-based clustering. https://www.rcsb.org/docs/grouping-structures/sequence-based-clustering, 2026b. Accessed September 23, 2026.

Randy Read, Claudia Millán, Airlie McCoy, and Tom Terwilliger. Likelihood-based signal and noise analysis for docking of models into cryo-em maps. *Acta Crystallographica Section D*, 79:271–280, 03 2023. doi: 10.1107/S2059798323001596.

Yang Song, Prafulla Dhariwal, Mark Chen, and Ilya Sutskever. Consistency models. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett (eds.), *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 32211–32252. PMLR, 23–29 Jul 2023. URL https://proceedings.mlr.press/v202/song23a.html.

Guang Tang, Liwei Peng, Philip R. Baldwin, Deepinder S. Mann, Wen Jiang, Ian Rees, and Steven J. Ludtke. Eman2: An extensible image processing suite for electron microscopy. *Journal of Structural Biology*, 157(1):38–46, 2007. ISSN 1047-8477. doi: https://doi.org/10.1016/j.jsb.2006.05.009. URL https://www.sciencedirect.com/science/article/pii/S1047847706001894. Software tools for macromolecular microscopy.

Alexandre Urzhumtsev, Vladimir Y. Lunin, and A. Thorn. Analytic modeling of inhomogeneous-resolution maps in cryo-electron microscopy and crystallography. *IUCrJ*, 9:728–734, 2022. ISSN 2052-2525. doi: https://doi.org/10.1107/S2052252522008260. URL https://www.sciencedirect.com/science/article/pii/S2052252522010004.

Xiao Wang, Eman Alnabati, Tunde Aderinwale, Sai Raghavendra Maddhuri Venkata Subramaniya, Genki Terashi, and Daisuke Kihara. Detecting protein and dna/rna structures in cryo-em maps of intermediate resolution using deep learning. *Nature Communications*, 12:2302, 04 2021. doi: 10.1038/s41467-021-22577-3.

Chenwei Zhang, Anne Condon, and Khanh Dao Duc. Struc2mapgan: improving synthetic cryogenic electron microscopy density maps with generative adversarial networks. *Bioinformatics Advances*, 5(1):vbaf179, 01 2025. ISSN 2635-0041. doi: 10.1093/bioadv/vbaf179. URL https://doi.org/10.1093/bioadv/vbaf179.

Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: A nested u-net architecture for medical image segmentation. In Danail Stoyanov, Zeike Taylor, Gustavo Carneiro, Tanveer Syeda-Mahmood, Anne Martel, Lena Maier-Hein, João Manuel R.S. Tavares, Andrew Bradley, João Paulo Papa, Vasileios Belagiannis, Jacinto C. Nascimento, Zhi Lu, Sailesh Conjeti, Mehdi Moradi, Hayit Greenspan, and Anant Madabhushi (eds.), *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pp. 3–11, Cham, 2018. Springer International Publishing. ISBN 978-3-030-00889-5.

Jasenko Zivanov, Takanori Nakane, Björn Forsberg, Dari Kimanius, Wim Hagen, Erik Lindahl, and Sjors Scheres. New tools for automated high-resolution cryo-em structure determination in relion-3. *eLife*, 7, 11 2018. doi: 10.7554/eLife.42166.

A NOTATION

	Framework and Rendering
$P = \{(\mathbf{p}_i, w_i)\}_{i=1}^{N}$	Atomic coordinates $\mathbf{p}_i$ and element amplitudes w_i
r	Requested global resolution
$\mathcal{G} = (\mathbf{n}, v, \mathbf{o})$	Density grid: dimensions, voxel spacing, and origin
$R(P, r; \mathcal{G})$	Differentiable atomic-density renderer
$\rho(r, v), \tau, h$	Gaussian width, cutoff, and support half-width
Y_r^{raw}, Y_r	Raw and normalized physically rendered maps
X^{raw}, X_0	Raw crop and normalized experimental reference map
$\widehat{X}_0$	Generated density map
M, V	Molecular-support and valid-volume masks
	Brownian Bridge and Sampling
t, T	Current and total diffusion timesteps
$m_t, \delta_t, \delta_{\max}$	Bridge interpolation, variance, and maximum variance
$\epsilon, \epsilon_s, \xi$	Gaussian fields for training bridge states, general posterior transitions, and the default stochastic transition
Z_t	Intermediate bridge state
$Z_{\mathrm{src}}, Z_{\mathrm{tar}}$	Boundary states adjacent to Y_r and X_0 in default sampling
$D_t = Z_t - X_0, \widehat{D}_t$	Reference and predicted residual fields
$f_\theta, f_{\bar{\theta}}$	Denoiser and its exponential-moving-average copy
p_{src}, b	Probability and indicator of boundary supervision at Y_r
$Z_{T-1}^{\mathrm{src}}, D_{T-1}^{\mathrm{src}}$	Physically rendered map boundary state and residual
$J, \{t_j\}$	Posterior-transition count and inference timestep schedule; NFE is $J + 1$
$s, \eta, \widetilde{\delta}_{t\to s}$	Next scheduled timestep, stochasticity, and posterior variance
$a_{t\to s}, \mu_{t\to s}$	Posterior residual coefficient and mean
$[q]_+$	Nonnegative projection $\max(q, 0)$
	FiLM-Conditioned BBDM Denoiser
$H_t = [Z_t, Y_r]$	Channel-concatenated denoiser input
C_{base}, C_i	Base width and channels at scale i
$X^{i,j}$	Multiscale denoiser node at scale i and nested depth j
$\mathcal{D}_i, \mathcal{U}, \mathcal{B}_{i,j}$	Downsampling, upsampling, and residual blocks
$\phi(r), \mathcal{F}$	Fourier resolution features and frequencies
e_t, e_r, e	Timestep, resolution, and combined conditioning embeddings
γ, β	FiLM channel-wise scale and shift
	Training Objective
W_1, W_{PCC}	Spatial weights for the L_1 and PCC terms
μ_W, C_W, E_W	Weighted mean, cross-product, and energy
PCC_W	Weighted Pearson correlation
ε	Numerical-stability constant
$\mathcal{L}_1(h), \mathcal{L}_{\mathrm{PCC}}(h)$	Weighted residual and reconstructed-map PCC losses
$\mathcal{L}_{\mathrm{map}}(h)$	Map objective for training branch h
$\mathcal{L}, \lambda_{\mathrm{PCC}}$	Total objective and PCC-loss weight

B ATOMIC MODEL RENDERING AND MAP PREPROCESSING

B.1 ATOMIC MODEL PARSING AND GRID CONSTRUCTION

The parser retains heavy atoms from `ATOM` and `HETATM` records, accepts blank or `A` alternate locations, and discards nonfinite coordinates. The PDB parser processes every qualifying coordinate record in the file; for mmCIF input, model 1 is used when a model identifier is present. The atomic

model contains non-hydrogen atoms, including ligands, ions, and other heteroatoms. The tensor address corresponding to $\mathbf{j}_{\mathrm{xyz}}$ follows the canonical MRC array order (j_z, j_y, j_x). For isotropic voxel spacing v, the physical origin is $\mathbf{o} = \mathbf{o}_{\mathrm{MRC}} + v(n_{x\mathrm{start}}, n_{y\mathrm{start}}, n_{z\mathrm{start}})$.

For paired training, the experimental reference map grid supplies the initial voxel spacing and origin. The smallest cubic atomic bounding region is expanded by 12 Å, rounded up to a multiple of eight voxels, and shifted inside the available map extent. A candidate is rejected if the requested cube is larger than the shortest map dimension or if its output edge is below 32 voxels. The target crop is read before normalization. A cube larger than 192 voxels per side is trilinearly resized to 192 voxels; if its original edge is L and output edge is L', the new spacing is $v' = vL/L'$ and the origin is shifted by $\frac{1}{2}(v' - v)$ after applying the crop offset. This keeps output voxel centers in the same physical coordinate system.

B.2 Gaussian Rendering

The physically rendered map generator follows the parameter-free differentiable density formulation of CryoNet.Refine (Huang et al., 2026) on a fixed physical grid with finite Gaussian support. Let a grid point with physical-order integer index $\mathbf{j}_{\mathrm{xyz}} = (j_x, j_y, j_z)$ have position $\mathbf{x}(\mathbf{j}_{\mathrm{xyz}}) = (o_x + vj_x, o_y + vj_y, o_z + vj_z)$. Each non-hydrogen atom contributes an element-weighted Gaussian:

$$R(P, r; \mathcal{G})_{\mathbf{j}_{\mathrm{xyz}}} = \sum_{i=1}^{N} w_i \exp\left[-\frac{\|\mathbf{j}_{\mathrm{xyz}} - \mathbf{c}_i\|_2^2}{\rho(r, v)^2}\right], \qquad \mathbf{c}_i = \frac{\mathbf{p}_i - \mathbf{o}}{v}, \qquad \rho(r, v) = \frac{r}{\pi v}. \tag{16}$$

The finite support uses cutoff $\tau = 10^{-5}$ and an axis-aligned cubic support with half-width $h = \lceil \sqrt{-\log \tau}\, \rho \rceil$ voxels along each axis. Gaussian values inside this support remain differentiable with respect to coordinates and r; discrete support construction makes the complete renderer piecewise differentiable.

The amplitudes use fixed element-dependent atomic weights: $w_{\mathrm{C}} = 12$, $w_{\mathrm{N}} = 14$, $w_{\mathrm{O}} = 16$, $w_{\mathrm{P}} = 31$, $w_{\mathrm{S}} = 32$, $w_{\mathrm{Mg}} = 24$, $w_{\mathrm{Fe}} = 56$, and $w_{\mathrm{Zn}} = 65$; other heavy elements use 14.

B.3 Map Normalization and Support Masks

Online physical rendering and normalization. Let Y_r^{raw} be the rendered map. Its normalization scale is the positive 99.9th percentile,

$$s_Y = Q_{0.999}\left(\{Y_{r,\mathbf{u}}^{\mathrm{raw}} : Y_{r,\mathbf{u}}^{\mathrm{raw}} > 0\}\right), \qquad Y_r = \mathrm{clip}_{[0,1]}(Y_r^{\mathrm{raw}}/s_Y). \tag{17}$$

A binary support is obtained at 10^{-5} of the physically rendered map maximum, dilated with a 5^3 max-pooling kernel, and smoothed with a 5^3 average-pooling kernel to form the soft mask M.

Target normalization. The experimental reference map is normalized online on the complete cropped volume. Background intensity is the median over voxels with $M < 0.05$. After subtraction, the scale is the positive 99.9th percentile over voxels with $M > 0.5$; if fewer than 100 such positive voxels are available, all positive voxels are used. For an empty positive set, s_X is the maximum absolute centered intensity, lower-bounded by 10^{-8}. The normalized experimental reference map is

$$X_0 = \mathrm{clip}_{[0,1]}\left(\frac{X^{\mathrm{raw}} - b_X}{s_X}\right). \tag{18}$$

Each cropped case uses its own background and scale statistics.

Processing order. For each case, the data loader reads the coordinate file and experimental crop and applies the physical resize when required. After these data are transferred to the GPU, the training-step preparation renders Y_r^{raw} on the resulting grid, normalizes the physically rendered map, constructs M, normalizes the complete experimental reference map with the same case-level mask, and finally pads cases for size-bucketed batching. Thus, physically rendered maps are computed online during training-step preparation, while the experimental reference maps remain complete case-level crops.

B.4 Variable-Size Batching

Cases are bucketed by output edge length before batching. Within a minibatch, tensors are padded to the largest depth, height, and width present in that batch. The original region is marked by $V = 1$ and all padding by $V = 0$. The mask-weighted objectives in Equations 42–49 therefore remain invariant to the amount of padding introduced by neighboring samples.

C Brownian Bridge Training and Sampling

C.1 Forward Bridge and Residual Prediction

Drawing on the paired-map formulation of Brownian Bridge Diffusion Models (BBDM) (Li et al., 2023), the forward process interpolates between the experimental reference map X_0 and the matched physically rendered map Y_r. For a discrete timestep $t \in \{0, \dots, T-1\}$,

$$Z_t = (1 - m_t)X_0 + m_t Y_r + \sqrt{\delta_t}\,\epsilon, \qquad \epsilon \sim \mathcal{N}(0, I), \tag{19}$$

$$D_t = Z_t - X_0 = m_t(Y_r - X_0) + \sqrt{\delta_t}\,\epsilon. \tag{20}$$

For the linear schedule, m_t increases from 0.001 to 0.999 and

$$\delta_t = 2\left(m_t - m_t^2\right)\delta_{\max}. \tag{21}$$

The denoiser predicts the residual $D_t = Z_t - X_0$ relative to the experimental reference map, which gives the generated density map

$$\widehat{X}_0 = Z_t - f_\theta(Z_t, t, Y_r, r). \tag{22}$$

Conditioning on the physically rendered map supplies Y_r as a separate spatial input and preserves the atomic-model signal throughout the bridge trajectory.

Figure 8 visualizes stochastic states sampled from this forward process for three additional paired cases with complementary morphologies.

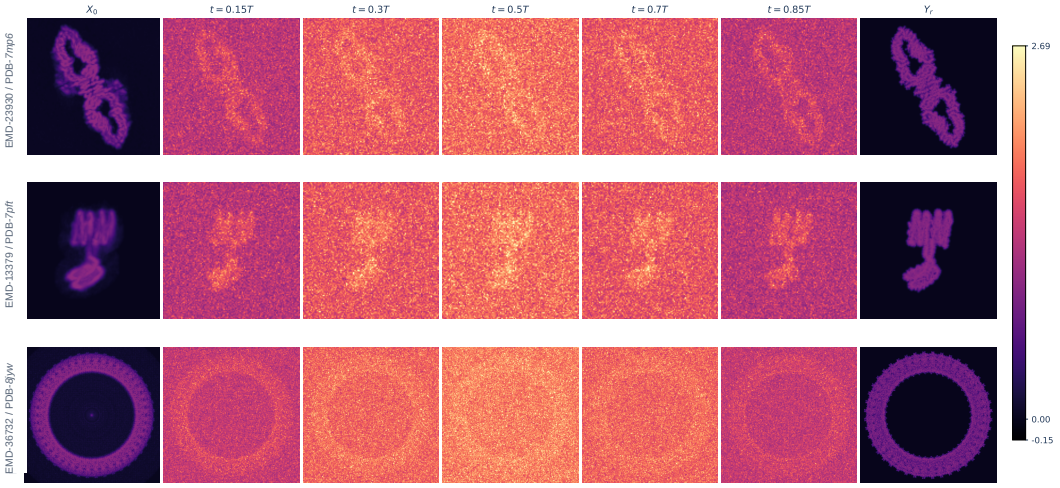

Figure 8: Visualization of stochastic Brownian-bridge states for three paired cases. Within each row, the columns reuse one Gaussian field to isolate how the closed-form marginal changes with t. Actual training independently draws one (t, ϵ) pair for each example and supervises that single state. The intermediate states become most noise-dominated near the middle of the bridge, consistent with the variance schedule in Equation 21.

Removing the stochastic term gives the corresponding mean-path residuals in Figure 9. The matched main-text view in Figure 7 instead reuses the three cases from Figure 4.

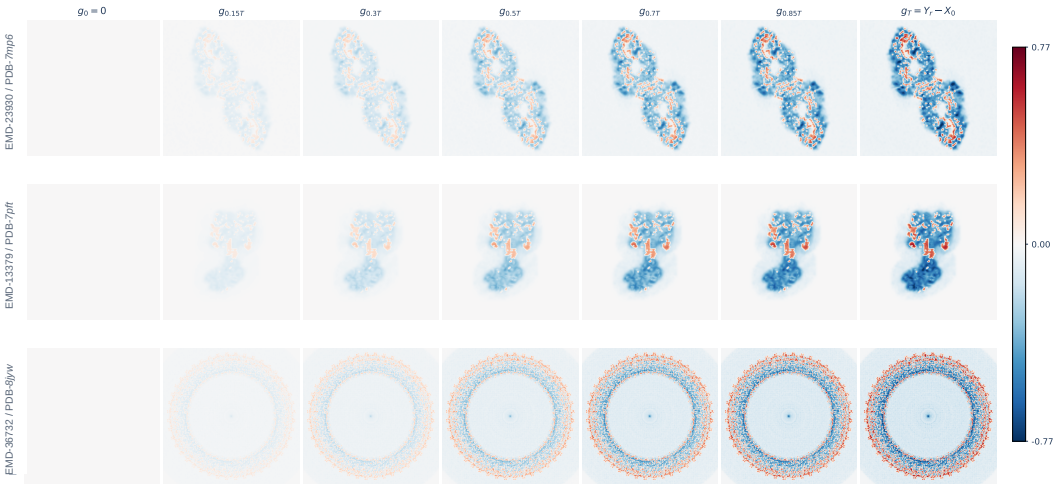

Figure 9: Mean forward-bridge residuals for the three additional cases in Figure 8: EMD-23930/PDB-*7mp6*, EMD-13379/PDB-*7pft*, and EMD-36732/PDB-*8jyw*.

C.2 Source-Boundary Supervision

Default one-transition stochastic sampling begins from the exact physically rendered map. Training includes the corresponding boundary state

$$Z_{T-1}^{\mathrm{src}} = Y_r, \tag{23}$$

$$D_{T-1}^{\mathrm{src}} = Y_r - X_0, \tag{24}$$

$$\widehat{D}_{T-1}^{\mathrm{src}} = f_\theta(Y_r, T-1, Y_r, r). \tag{25}$$

The boundary reconstruction is $\widehat{X}_0^{\mathrm{src}} = Y_r - \widehat{D}_{T-1}^{\mathrm{src}}$. For either training state, $\widehat{X}_0 = Z_t - f_\theta(Z_t, t, Y_r, r)$ is a generated density map estimate from one network evaluation. This boundary supervision exposes the first state used by the sampler.

C.3 Posterior Transitions

For transition-count analysis, let $J \geq 1$ be the number of posterior bridge-state transitions. The sampler uses a decreasing schedule $T-1 = t_J > \cdots > t_0 = 0$, containing $J+1$ evaluations of the shared denoiser and J posterior transitions. The default uses $J = 1$ and the schedule $(T-1, 0)$. The transition-count experiment evaluates $J \in \{1, 2, 3, 4, 5, 10, 15, 20\}$, corresponding to NFE $\in \{2, 3, 4, 5, 6, 11, 16, 21\}$. In code, `sample_steps` is set to $J+1$; throughout the paper, sampler length is reported by both J and NFE. For the linear schedule, $m_{T-1} = 0.999$ and $m_0 = 0.001$, so the default $T-1 \to 0$ transition has variance 0.001998 and standard deviation 0.0447. Given the current bridge state Z_t and $\widehat{X}_0 = \mathrm{clip}_{[0,1]}[Z_t - f_{\bar{\theta}}(Z_t, t, Y_r, r)]$, let $s < t$ be the next scheduled timestep and define

$$\widetilde{\delta}_{t\to s} = \left[\left(\delta_t - \delta_s \frac{(1-m_t)^2}{(1-m_s)^2}\right)\frac{\delta_s}{\delta_t}\right]_+, \tag{26}$$

The transition is

$$a_{t\to s} = \sqrt{\frac{[\delta_s - \eta^2 \widetilde{\delta}_{t\to s}]_+}{\delta_t}}, \tag{27}$$

$$\mu_{t\to s} = (1-m_s)\widehat{X}_0 + m_s Y_r + a_{t\to s}\left[Z_t - (1-m_t)\widehat{X}_0 - m_t Y_r\right], \tag{28}$$

$$Z_s = \mu_{t\to s} + \eta\sqrt{\widetilde{\delta}_{t\to s}}\,\epsilon_s, \qquad \epsilon_s \sim \mathcal{N}(0, I). \tag{29}$$

where $[q]_+ = \max(q, 0)$. The coefficient $\eta \in [0, 1]$ jointly controls the innovation variance and the residual coefficient $a_{t\to s}$; $\eta = 1$ gives the full BBDM posterior used by default inference and every

transition-count ablation. Independent samples reuse the same physically rendered map and draw a new Gaussian field at each transition.

C.4 One-Transition Sampling

Given P, r, and an output grid $\mathcal{G}$, inference renders Y_r and evaluates the EMA denoiser at $T-1$ to obtain an initial experimental reference map estimate. The BBDM posterior draws one state at $t=0$ using its full prescribed Gaussian variance, after which the shared denoiser is evaluated a second time. Algorithm 1 summarizes this data flow.

Algorithm 1: One-transition stochastic density generation in CryoNet.Density**.**

Require: atomic model P, requested resolution r, output grid $\mathcal{G}$, EMA denoiser $f_{\bar{\theta}}$

1: $Y_r \leftarrow R(P, r; \mathcal{G})$ and normalize the physically rendered map.
2: $\widehat{D}_{T-1} \leftarrow f_{\bar{\theta}}(Y_r, T-1, Y_r, r)$ and $\widetilde{X}_0 \leftarrow \mathrm{clip}_{[0,1]}(Y_r - \widehat{D}_{T-1})$.
3: Draw $\xi \sim \mathcal{N}(0, I)$ and set $Z_{\mathrm{tar}} \leftarrow \mu_{T-1\to 0}(Y_r, \widetilde{X}_0; Y_r) + \sqrt{\widetilde{\delta}_{T-1\to 0}}\,\xi$. **(one stochastic transition)**
4: $\widehat{D}_0 \leftarrow f_{\bar{\theta}}(Z_{\mathrm{tar}}, 0, Y_r, r)$.
5: $\widehat{X}_0 \leftarrow \mathrm{clip}_{[0,1]}(Z_{\mathrm{tar}} - \widehat{D}_0)$.
6: **return** $\widehat{X}_0$ on $\mathcal{G}$.

Zero-transition controls. The appendix also evaluates two one-evaluation controls. *Direct* evaluates the denoiser on the exact physically rendered map boundary state and returns its clipped experimental reference map estimate. *Noisy Direct* adds a boundary-scale Gaussian perturbation, $Z_{T-1} = Y_r + \sqrt{\delta_{T-1}}\,\xi$, while retaining the physically rendered map Y_r as the spatial condition, then applies the same single evaluation. Both controls have zero posterior transitions and one network evaluation; comparison with default one-transition stochastic sampling quantifies the changes introduced by the posterior draw and second denoiser evaluation.

C.5 Training and Sampling Algorithms

Minibatch training algorithm. Each minibatch samples the physically rendered map boundary branch once. For bridge minibatches, every case receives an independent timestep and Gaussian noise. Physical rendering and experimental reference map normalization precede padding, and per-case losses are averaged across the minibatch.

Algorithm 2: One training iteration.

Require: paired cases $(P_i, r_i, \mathcal{G}_i, X_i^{\mathrm{raw}})$, model f_θ, boundary probability p_{src}

1: Render $Y_{r_i}^{\mathrm{raw}} \leftarrow R(P_i, r_i; \mathcal{G}_i)$ for every case.
2: Normalize Y_{r_i}; derive soft support M_i; background-center and normalize $X_{0,i}$.
3: Pad tensors to the largest batch shape and construct valid mask V_i.
4: Draw one $b \sim \mathrm{Bernoulli}(p_{\mathrm{src}})$ for the minibatch.
5: If $b=1$: set $t_i = T-1$, $Z_{t_i} = Y_{r_i}$, and $D_{t_i} = Y_{r_i} - X_{0,i}$.
6: Else: draw $t_i \sim \mathrm{Unif}\{0, \ldots, T-1\}$ and $\epsilon_i \sim \mathcal{N}(0, I)$, then form (Z_{t_i}, D_{t_i}) by Equations 19–20.
7: Predict $\widehat{D}_{t_i} = f_\theta(Z_{t_i}, t_i, Y_{r_i}, r_i)$ and reconstruct $\widehat{X}_{0,i} = Z_{t_i} - \widehat{D}_{t_i}$.
8: Compute the per-case losses using $W_1 = (0.1 + 0.9 M_i) V_i$ and $W_{\mathrm{PCC}} = \max(M_i, 0.01) V_i$, then average across cases.
9: Backpropagate, clip the global gradient norm, update AdamW, then update EMA parameters.

D Resolution-Conditioned Denoiser

D.1 Network Architecture

Input and four-scale feature hierarchy. The network concatenates the current bridge state and physically rendered map along the channel dimension:

$$H_t = \mathrm{concat}_{\mathrm{ch}}(Z_t, Y_r) \in \mathbb{R}^{B\times 2\times N_z\times N_y\times N_x}. \tag{30}$$

Let $C_i = 2^i C_{\mathrm{base}}$ for levels $i = 0, 1, 2, 3$, with $C_{\mathrm{base}} = 16$ in the main configuration. Downsampling $\mathcal{D}_i$ is a stride-two $3 \times 3 \times 3$ convolution that preserves C_i channels; the subsequent residual block changes the width to C_{i+1}. Upsampling $\mathcal{U}$ uses nearest-neighbor interpolation to the spatial dimensions of the receiving feature.

Nested multiscale graph. We denote the node at spatial level i and nested depth j by $X^{i,j}$. The encoder column is

$$X^{0,0} = \mathcal{B}_{0,0}(H_t, e), \qquad X^{i,0} = \mathcal{B}_{i,0}(\mathcal{D}_{i-1}X^{i-1,0}, e), \quad i = 1, 2, 3, \tag{31}$$

and each nested decoder node is

$$X^{i,j} = \mathcal{B}_{i,j}\left(\mathrm{concat}_{\mathrm{ch}}\left[X^{i,0}, \ldots, X^{i,j-1}, \mathcal{U}(X^{i+1,j-1})\right], e\right). \tag{32}$$

The instantiated graph contains ten residual blocks and the following exact pathways:

Table 4: Exact multiscale denoiser nodes. The main model uses base width $C_{\mathrm{base}} = 16$; "upsampled" denotes interpolation to the receiving node's shape.

Node	Concatenated input	Operation	Output channels
$X^{0,0}$	$[Z_t, Y_r]$	residual block	C_{base}
$X^{1,0}$	$\mathcal{D}_0 X^{0,0}$	residual block	$2C_{\mathrm{base}}$
$X^{2,0}$	$\mathcal{D}_1 X^{1,0}$	residual block	$4C_{\mathrm{base}}$
$X^{3,0}$	$\mathcal{D}_2 X^{2,0}$	residual block	$8C_{\mathrm{base}}$
$X^{0,1}$	$[X^{0,0},$ upsampled $X^{1,0}]$	residual block	C_{base}
$X^{1,1}$	$[X^{1,0},$ upsampled $X^{2,0}]$	residual block	$2C_{\mathrm{base}}$
$X^{2,1}$	$[X^{2,0},$ upsampled $X^{3,0}]$	residual block	$4C_{\mathrm{base}}$
$X^{0,2}$	$[X^{0,0}, X^{0,1},$ upsampled $X^{1,1}]$	residual block	C_{base}
$X^{1,2}$	$[X^{1,0}, X^{1,1},$ upsampled $X^{2,1}]$	residual block	$2C_{\mathrm{base}}$
$X^{0,3}$	$[X^{0,0}, X^{0,1}, X^{0,2},$ upsampled $X^{1,2}]$	residual block	C_{base}

Operationally, the encoder applies three stride-two convolutions to build four progressively coarser feature scales. At every decoder node, the coarser feature is first upsampled to the receiving grid, concatenated with all earlier features at that spatial scale, and processed by a residual block. The sum of the timestep and requested-resolution embeddings modulates each of the ten residual blocks through FiLM as defined in Equation 40. The final full-resolution node $X^{0,3}$ therefore combines local detail with all four scales and predicts the residual through a single head, $\widehat{D}_t = \mathrm{Conv}_{1\times1\times1}(X^{0,3})$. Table 4 lists every concatenation and output width needed to reproduce this computation.

D.2 Timestep and Resolution Embeddings

Time embedding. For timestep t, we construct a sinusoidal vector of width C_{base} with exponentially spaced frequencies,

$$\omega_k = \exp\left[-k\frac{\log 10000}{\max(C_{\mathrm{base}}/2 - 1, 1)}\right], \tag{33}$$

$$s_t = [\sin(t\omega_0), \ldots, \sin(t\omega_{C_{\mathrm{base}}/2-1}), \cos(t\omega_0), \ldots, \cos(t\omega_{C_{\mathrm{base}}/2-1})]. \tag{34}$$

The MLP $C_{\mathrm{base}} \to 128 \to 128$ with a SiLU nonlinearity produces e_t.

Continuous resolution encoding. For a configured physical interval $[r_{\min}, r_{\max}]$, the requested resolution is normalized as

$$u(r) = \mathrm{clip}_{[0,1]}\left(\frac{r - r_{\min}}{r_{\max} - r_{\min}}\right). \tag{35}$$

For the main model, $[r_{\min}, r_{\max}] = [1.5, 15.0]$ Å and $\mathcal{F} = \{1, 2, 4\}$. The resulting seven-dimensional feature vector is

$$\phi(r) = [2u(r) - 1, \{\sin(2\pi f u(r)), \cos(2\pi f u(r))\}_{f\in\mathcal{F}}]. \tag{36}$$

The condition mapping is $e_r = W_r\,\mathrm{SiLU}(\phi(r)) + b_r \in \mathbb{R}^{128}$; the shared block condition is $e = e_t + e_r$. Therefore the single affine projection at each block is jointly conditioned on timestep and resolution.

D.3 FiLM Residual Blocks

Residual block and FiLM location. Each of the ten nodes in Table 4 is a FiLM-modulated residual block. For input A and the shared condition e, it computes

$$H_1 = \mathrm{SiLU}(\mathrm{GN}_1(\mathrm{Conv}_1^3(A))), \tag{37}$$

$$H_2 = \mathrm{GN}_2(\mathrm{Conv}_2^3(H_1)), \tag{38}$$

$$[\gamma, \beta] = \mathrm{Linear}_{128\to 2C_{\mathrm{out}}}(e), \tag{39}$$

$$\mathcal{B}(A, e) = \mathrm{SiLU}((1 + \gamma) \odot H_2 + \beta + \mathcal{S}(A)), \tag{40}$$

where both convolutions use 3^3 kernels with unit padding, GroupNorm uses the largest divisor in $\{8, 4, 2, 1\}$, and $\mathcal{S}$ is identity or a 1^3 projection when the channel count changes. The scale and shift are broadcast over $N_z \times N_y \times N_x$. FiLM is therefore applied after the second GroupNorm and before residual addition and the final SiLU, at every encoder and nested decoder node.

The fixed encoding gives resolution the same physical semantics during training and inference. Resolution affects the method through two coordinated paths: the Gaussian width in Equation 16 and FiLM throughout the BBDM denoiser.

E Training Objective and Implementation Details

E.1 Residual and Correlation Losses

For a random-time bridge sample, Equations 19 and 20 define $Z^{\mathrm{br}} = Z_t$, $D^{\mathrm{br}} = Z_t - X_0$, and $t^{\mathrm{br}} = t$. Physically rendered map boundary supervision uses $Z^{\mathrm{src}} = Y_r$, $D^{\mathrm{src}} = Y_r - X_0$, and $t^{\mathrm{src}} = T - 1$, matching the boundary state at the start of inference. For either branch $h \in \{\mathrm{br}, \mathrm{src}\}$, the shared denoiser predicts

$$\widehat{D}^h = f_\theta(Z^h, t^h, Y_r, r), \qquad \widehat{X}_0^h = Z^h - \widehat{D}^h, \tag{41}$$

with $W_1 = (0.1 + 0.9M)V$. Its weighted residual loss is

$$\mathcal{L}_1(h) = \frac{1}{B}\sum_{i=1}^{B} \frac{\sum_{\mathbf{u}} W_{1,i,\mathbf{u}} \left|\widehat{D}^h_{i,\mathbf{u}} - D^h_{i,\mathbf{u}}\right|}{\sum_{\mathbf{u}} W_{1,i,\mathbf{u}} + \varepsilon}. \tag{42}$$

Together with the PCC loss below, the state-specific objective is

$$\mathcal{L}_{\mathrm{map}}(h) = \mathcal{L}_1(h) + 0.1\mathcal{L}_{\mathrm{PCC}}(h). \tag{43}$$

With $b \sim \mathrm{Bernoulli}(p_{\mathrm{src}})$ and $p_{\mathrm{src}} = 0.1$, one minibatch objective is

$$\mathcal{L} = (1 - b)\mathcal{L}_{\mathrm{map}}(\mathrm{br}) + b\mathcal{L}_{\mathrm{map}}(\mathrm{src}). \tag{44}$$

Thus both state types have the same prediction semantics and optimize the same denoiser parameters.

Weighted PCC and numerical stability. The mask derived from the physically rendered map concentrates both objectives on molecular support while retaining weak background supervision. Let $W_{\mathrm{PCC}} = \max(M, 0.01)V$ and

$$\mu_W(A) = \frac{\sum_{\mathbf{u}} W_{\mathrm{PCC},\mathbf{u}} A_{\mathbf{u}}}{\sum_{\mathbf{u}} W_{\mathrm{PCC},\mathbf{u}}}, \tag{45}$$

$$C_W(A,B) = \sum_{\mathbf{u}} W_{\mathrm{PCC},\mathbf{u}} [A_{\mathbf{u}} - \mu_W(A)][B_{\mathbf{u}} - \mu_W(B)], \tag{46}$$

$$E_W(A) = \sum_{\mathbf{u}} W_{\mathrm{PCC},\mathbf{u}} [A_{\mathbf{u}} - \mu_W(A)]^2. \tag{47}$$

We use

$$\mathrm{PCC}_W(A,B) = \begin{cases} C_W(A,B)/\sqrt{E_W(A)E_W(B)}, & E_W(A)E_W(B) > \varepsilon, \\ 0, & \text{otherwise}, \end{cases} \tag{48}$$

$$\mathcal{L}_{\mathrm{PCC}}(h) = \frac{1}{B}\sum_{i=1}^{B} \left[1 - \mathrm{PCC}_W(X_{0,i}, \widehat{X}^h_{0,i})\right]. \tag{49}$$

The L_1 weight retains coefficient 0.1 in valid background; the PCC weight retains coefficient 0.01. Padding receives zero weight through V. Weighted means and energies are accumulated in FP32 with $\varepsilon = 10^{-8}$.

E.2 Optimization and Model Selection

The main model is trained for 100 epochs on eight NVIDIA H20 GPUs. We use AdamW with an initial learning rate of 10^{-4}. The validation objective controls a reduce-on-plateau schedule that halves the learning rate after two unimproved validation epochs, with a relative improvement threshold of 0.002 and a minimum learning rate of 10^{-6}. Training uses bfloat16 autocast, clips the global gradient norm to 1.0, and updates an exponential moving average of the denoiser with decay 0.999. The reported checkpoint is selected by the lowest validation objective.

Validation and model selection. Validation uses the EMA denoiser and a fixed state assignment for each record. One tenth of the records use the physically rendered map boundary state, while the others use record-specific timesteps and Gaussian fields that remain constant across epochs. The mean validation objective in Equation 13 controls the learning-rate schedule and model selection.

F Datasets and Benchmark Construction

F.1 Data Sources and Particle Subsampling

Pair collection. We construct one full-case crop for each experimental reference map–atomic model pair. The first source is the EMDB archive (Lawson et al., 2013) and its associated PDB/mmCIF coordinates. The second is our particle-subsampled reconstruction set (`te2_maps`). Starting from 220 high-resolution EMDB maps with raw data in EMPIAR, we reproduced the deposited processing workflows by downloading particle images or picking them from the corresponding micrographs. This yielded 109 high-resolution-map/particle-set pairs, with 14,262–730,118 particles per set. For each pair, we sampled 6–435 particle subsets and reconstructed maps de novo with cryoSPARC (Punjani et al., 2017), producing 15,679 intermediate- and low-resolution experimental reference maps labelled with the atomic model paired to the original EMDB map. Each EMDB–PDB pair defines one record: multiple maps for one PDB entry and multiple deposited atomic models for one map remain separate pairs. Across both sources, the archive contributes 7,857 candidate pairs and `te2_maps` contributes 15,679, for 23,536 candidates in total; 20,862 pass the validity checks below.

F.2 Quality Control

Input Validation. For every candidate, we require a reported resolution in $[1.5, 15.0]$ Å, at least 100 and at most 500,000 non-hydrogen atoms, standard MRC axis order, an isotropic positive voxel

size no larger than 4 Å, and a reported resolution no finer than twice the voxel spacing. The deposited map grid must contain at least 95% of the parsed atomic coordinates. The useful density region is the smallest cubic bounding box of the atomic coordinates plus a 12 Å physical margin. The physical crop size is rounded to a multiple of eight and may exceed 192 voxels; it is then resampled by trilinear interpolation so that the network input lies between 32 and 192 voxels, with voxel spacing and origin updated to preserve physical coordinates. Cases whose estimated differentiable-renderer cost exceeds 4×10^8 atom-voxel interactions are removed. These checks remove 2,674 candidates and retain 20,862 of 23,536.

Target-quality curation. We compute the Phenix map-model CC_{mask} (Afonine et al., 2018) between each experimental reference map and its paired atomic model after translating the coordinates into the map-local frame. Of the 20,862 geometrically valid cases, 20,831 have a finite value. We retain cases with $CC_{\mathrm{mask}} \geq 0.5$. The resulting primary dataset contains 19,273 cases from 19,005 distinct PDB IDs. The retained subset has CC_{mask} values from 0.5000 to 0.9683, with mean 0.7710. Thus, 31 geometrically valid cases lack a finite score and 1,558 finite-score cases fall below the cutoff: $20{,}862 - 31 - 1{,}558 = 19{,}273$. This cutoff selects paired targets with consistent deposited structures and map registrations. It is computed from each deposited pair before partitioning, training, and evaluation. The curated split is then frozen, and every primary-benchmark method is evaluated on the same 1,923 test cases.

F.3 Training, Validation, and Test Splits

Primary Benchmark. We use the RCSB 40% sequence-identity entity clusters (Burley et al., 2019; RCSB Protein Data Bank, 2026b) and bind all cases sharing a PDB ID to one assignment. Each PDB is represented by RCSB polymer entity 1, with the smallest available entity identifier used when entity 1 is absent. Of 19,005 source PDB IDs, 18,888 map to a cluster and 117 unmatched PDBs each form their own group. This partition enforces disjointness of PDB IDs and primary-entity sequence families across training, validation, and test.

The allocation assigns each indivisible PDB group and primary-entity cluster line wholly to one partition. The 121 Struc2mapGAN protocol pairs present in the primary dataset are assigned to test. Source PDBs that share a primary-entity cluster with these test PDBs form the homology-excluded set. Training, model selection, and primary evaluation use train, validation, and test, respectively. Test and validation are constructed toward 1,928 cases each, allowing at most 24 additional cases to keep each selected PDB group intact. Once a cluster line or PDB is assigned, its primary-entity line is reserved and other source PDBs on that line enter the homology-excluded set. All remaining non-conflicting groups form training. Table 5 reports the resulting counts; the supplementary material provides the sample-ID and PDB-ID lists for every assignment together with the RCSB cluster file used for the split.

Table 5: Primary benchmark accounting. Percentages are computed over the 17,518 assigned train, validation, and test cases; the excluded cases remain part of the 19,273-case curated dataset.

Split	Cases	PDB IDs	Proportion (%)	Resolution range (Å)	Median (Å)	Mean (Å)
Train	13,659	13,438	77.97	1.52–15.00	3.30	3.574
Validation	1,936	1,894	11.05	1.88–14.00	3.39	3.532
Test	1,923	1,923	10.98	1.88–13.00	3.40	3.665
Excluded	1,755	1,750	–	1.51–14.10	3.51	3.775

The three model-development and evaluation partitions contain $13{,}659 + 1{,}936 + 1{,}923 = 17{,}518$ cases. A further 1,755 quality-passing cases are excluded because their PDB IDs or primary-entity 40% sequence clusters conflict with a reserved validation or test assignment. Consequently, $17{,}518 + 1{,}755 = 19{,}273$ cases, matching the curated total above. The PDB-ID counts reconcile in the same way: $13{,}438 + 1{,}894 + 1{,}923 + 1{,}750 = 19{,}005$ distinct IDs. "Training," "validation," and "test" elsewhere in the paper always refer to the first three rows; the homology-excluded cases are used in none of these stages.

Train, validation, and test contain disjoint PDB IDs and mutually exclusive primary-entity RCSB 40% clusters; their map IDs are also disjoint. The 19,273 sample IDs are unique, and all 121 protocol

pairs present in the dataset belong to test. Using the resolution intervals $r < 3.5$, $3.5 \leq r \leq 6$, and $r > 6$ Å, test contains 1,039, 771, and 113 cases in these three intervals, respectively.

F.4 Published Struc2mapGAN Benchmark

To preserve comparability with prior PDB-to-experimental-map generation, we reconstruct the 130 EMDB-PDB pairs in the validation protocol of Struc2mapGAN (Zhang et al., 2025). Every EMDB and PDB identifier is unique. Deposited maps and structures are brought onto a per-case grid shared by every compared method, preserving the experimental reference map's physical frame, voxel spacing, and crop. ChimeraX `molmap`, differentiable CryoNet `molmap`, and TEMPy use the 2 Å source resolution specified by the Struc2mapGAN comparison protocol; EMAN2 uses $\max(2\text{ Å}, v)$, setting the requested resolution lower bound to the output voxel spacing v. `phenix.model_map` uses the reported target resolution because this is the resolution argument of its native map-model calculation, and Struc2mapGAN uses its canonical preprocessing and inference path. Thus, the common-source methods measure agreement of the protocol's nominal 2 Å source with the experimental reference map, while `phenix.model_map` is retained as the protocol's resolution-matched physically rendered map reference. All generated density maps are evaluated with the same per-case Map–map masks and normalization used for our method. The benchmark ranges from 3.0 to 7.9 Å (median 4.2 Å): 21, 91, and 18 cases fall in $r < 3.5$, $3.5 \leq r \leq 6$, and $r > 6$ Å, respectively.

The primary dataset contains 121 matching EMDB-PDB pairs, all assigned to test, while the remaining nine protocol pairs are external to the primary dataset. The training and benchmark PDBs belong to separate primary-entity RCSB 40% clusters. Headline results use the 1,923-case primary test set.

Table 6: Resolution coverage of the two evaluation benchmarks. Bins are defined as $r < 3.5$, $3.5 \leq r \leq 6$, and $r > 6$ Å.

Benchmark	Cases	< 3.5 Å	3.5–6 Å	> 6 Å	Median	Range
Primary test	1,923	1,039	771	113	3.40	1.88–13.00
Struc2mapGAN protocol	130	21	91	18	4.20	3.00–7.90

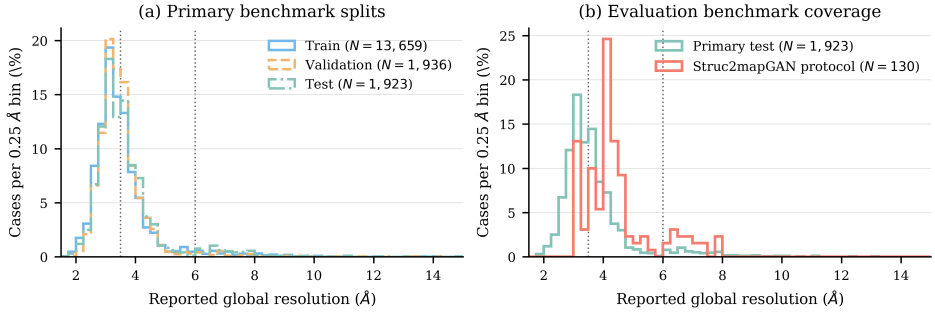

Figure 10: Reported global-resolution distributions in normalized 0.25 Å bins. (a) The step histograms visualize coverage of the three primary-benchmark splits. (b) The primary test set covers a broader range than the 130-case Struc2mapGAN protocol benchmark. Each step histogram is normalized within its split or benchmark to compare distributional shape.

G Baseline Implementations

Shared coordinate and grid geometry. All six baselines receive the same deposited PDB or mmCIF coordinates as CryoNet.Density. For the primary benchmark, direct physical renderers use the reported global resolution r. The experimental reference map defines the output array shape, isotropic voxel spacing, and physical origin. These geometric fields define the common output lattice for baseline generation. Programs that first render on their native grids are subsequently resampled onto the experimental reference grid. Physically rendered maps with arbitrary intensity units are min–max scaled to $[0, 1]$ according to their native inference pipelines. Map–map CC and FSC are then computed directly from the aligned arrays with the shared evaluation mask.

ChimeraX `molmap`. We use UCSF ChimeraX (Pettersen et al., 2021). The `molmap` command represents each atom by a Gaussian whose amplitude is proportional to atomic number and whose standard deviation is $0.225r$ under the default `sigmaFactor`. We invoke `molmap` at the requested resolution, resample the physically rendered map onto the experimental reference map grid with `volume resample ... onGrid`, and min-max normalize the saved MRC volume. This baseline is the standard Gaussian renderer used to construct the input of Struc2mapGAN.

CryoNet `molmap`. This baseline is the differentiable physical renderer R inherited from CryoNet.Refine (Huang et al., 2026) and defined in Equation 16. It uses the same element-dependent amplitudes, finite atom-centered Gaussian support, experimental reference map grid, and case-level normalization as the rendering branch of CryoNet.Density. Its prediction is exactly the physically rendered map Y_r. Comparing this baseline with CryoNet.Density measures the change introduced by the learned physical-to-experimental transformation while holding the physically rendered map fixed.

`phenix.model_map`. We generate an electron-scattering physically rendered map with `phenix.model_map` (Afonine et al., 2018) using the reported target resolution, electron scattering factors, atomic coordinates, elements, occupancies, and deposited atomic displacement parameters. The experimental reference MRC supplies the unit cell and sampling grid. For PDB input, fixed-column preprocessing removes nonphysical lone-pair pseudoatoms; when the installed electron table lacks a charged atom label, the charge annotation is cleared while the element, coordinate, occupancy, and displacement parameter are retained. The resulting model-calculated map is written on the experimental reference map array shape. We convert the CCP4 axis labels to canonical MRC order and, when its zero-origin periodic box differs from the experimental reference map grid origin, test the unique origin-derived periodic roll against the coordinate-only CryoNet `molmap`. This geometry check selects the roll with the higher PCC between the two coordinate-derived maps. The final volume shares the experimental reference map grid's shape, voxel spacing, and physical origin.

EMAN2 `e2pdb2mrc`. We use EMAN2 and `e2pdb2mrc.py` (Tang et al., 2007). The command receives the experimental reference map voxel spacing, array dimensions, physical center, requested resolution, and the `--het` option. mmCIF inputs are converted to PDB before invocation. EMAN2 constrains the nominal resolution to $r_{\text{render}} \geq v$, so these cases use $r_{\text{render}} = \max(r, v)$; the effective rendering resolution is retained with each result. The physically rendered map is placed on the experimental reference map grid and scaled to $[0, 1]$.

TEMPy `StructureBlurrer`. We use BioTEMPy through its TEMPy API (Farabella et al., 2015). PDB and mmCIF files are read with hetero atoms enabled. `StructureBlurrer.gaussian_blur_real_space` places the atoms on a volume initialized from the experimental reference map grid and applies real-space Gaussian blurring with $\sigma = 0.356r$. Internal normalization is disabled, after which the physically rendered map is min-max scaled by the common baseline pipeline.

Struc2mapGAN reproduction protocol. We run the released Struc2mapGAN (Zhang et al., 2025) code and `struc2mapGAN.ckpt` in evaluation mode. ChimeraX `molmap` renders the fixed 2 Å source, which is cubically resampled to isotropic 1 Å voxels, min–max normalized, and divided into overlapping 32^3 cubes. Inference uses batch size 32; the central 20^3 regions are assembled and normalized before a final resampling to the paired experimental reference map grid. The resulting evaluations cover 1,923 primary-test cases and 130 protocol-benchmark cases on their common experimental reference map grids.

Resolution settings for the two benchmarks. On the primary benchmark, ChimeraX `molmap`, `phenix.model_map`, EMAN2, TEMPy, and CryoNet `molmap` render at each case's reported resolution, following the EMAN2 voxel-spacing constraint above; Struc2mapGAN uses its fixed 2 Å physically rendered map. On the Struc2mapGAN protocol benchmark, ChimeraX `molmap` and TEMPy use the protocol's 2 Å source resolution, EMAN2 uses $\max(2\text{ Å}, v)$, `phenix.model_map` uses the reported target resolution, Struc2mapGAN follows its released preprocessing, and CryoNet `molmap` uses the protocol's 2 Å source resolution. We report the setting with every benchmark result and compute their statistics separately. The nominal 2 Å setting reproduces the common

physically rendered map used by Struc2mapGAN; `phenix.model_map` defines its map at the experimental reference map's reported resolution.

H Evaluation Metrics

All methods are evaluated after placing their generated density maps on the experimental reference map grid. Let X denote the min–max-normalized experimental reference map and Y the generated density map. A soft evaluation mask M is obtained by dilating and smoothing the support of the matched physically rendered map; masked Fourier metrics use MX and MY. We report the arithmetic mean over cases.

Map–map CC. We run `phenix.map_comparison` (Afonine et al., 2018) on the aligned experimental reference map and generated density map and report its "CC, input maps" value as the primary real-space Map–map CC.

Masked FSC AUC and FSC@target. For Fourier shell S_k, the masked Fourier shell correlation is

$$\mathrm{FSC}(k) = \frac{\sum_{\mathbf{q}\in S_k} \widehat{MX}(\mathbf{q})\widehat{MY}(\mathbf{q})^*}{\sqrt{\sum_{\mathbf{q}\in S_k} |\widehat{MX}(\mathbf{q})|^2 \sum_{\mathbf{q}\in S_k} |\widehat{MY}(\mathbf{q})|^2}}. \tag{50}$$

Shells contain at least eight Fourier voxels. Masked FSC AUC is the trapezoidal integral of this curve divided by its sampled frequency span. For a target resolution r, FSC@target is the FSC curve linearly interpolated at spatial frequency $1/r$.

log-RAPS MAE. To compare spectral energy independently of direction, we radially average the power spectrum of each masked map. For shell S_k, let

$$P_Z(k) = \frac{1}{|S_k|} \sum_{\mathbf{q}\in S_k} |\widehat{MZ}(\mathbf{q})|^2. \tag{51}$$

Over the B shells up to the Nyquist frequency, we report

$$\text{log-RAPS MAE}(X, Y) = \frac{1}{B} \sum_{k=1}^{B} \left|\log(P_X(k) + 10^{-12}) - \log(P_Y(k) + 10^{-12})\right|. \tag{52}$$

Lower values indicate closer agreement between the radial spectral profiles.

Struc2mapGAN protocol metrics. We additionally reproduce three metrics from the Struc2mapGAN protocol (Zhang et al., 2025): SSIM, Cosine similarity, and PCC. Let X denote the experimental reference map and Y the generated density map after both maps have been placed on the same physical grid. MRC axes are first converted to the same Cartesian order, and the two arrays must have identical shapes. The Cosine similarity uses the complete aligned arrays. SSIM and PCC instead apply the percentile operation in the released Struc2mapGAN evaluator. For a map Z, define

$$q_{90}(Z) = \mathrm{percentile}_{90}\{Z_i : Z_i \neq 0\}, \tag{53}$$

$$\mathcal{T}_{90}(Z)_i = Z_i \, \mathbf{1}[Z_i \geq q_{90}(Z)], \tag{54}$$

$$\mathcal{N}(A) = \frac{A - \min(A)}{\max(A) - \min(A)}, \qquad Z^{90} = \mathcal{N}(\mathcal{T}_{90}(Z)). \tag{55}$$

The threshold is estimated independently for X and Y; thus each map retains its own highest 10% of nonzero density values, while all lower values become zero. Min-max normalization is also performed independently and maps the retained range to $[0, 1]$. Empty or zero-range post-threshold maps are treated as undefined. All 1,923 cases in Table 1 pass these validity checks.

Three-dimensional SSIM. We evaluate $\mathrm{SSIM}_{\mathrm{S2M}}(X, Y) = \mathrm{SSIM}(X^{90}, Y^{90})$. For local means μ_X, μ_Y, variances σ_X^2, σ_Y^2, and covariance σ_{XY}, the local structural-similarity term is

$$\mathrm{SSIM}(X, Y) = \frac{(2\mu_X\mu_Y + c_1)(2\sigma_{XY} + c_2)}{(\mu_X^2 + \mu_Y^2 + c_1)(\sigma_X^2 + \sigma_Y^2 + c_2)}, \tag{56}$$

where the normalized data range is one. We call `structural_similarity` from `skimage.metrics` with `data_range=1` and otherwise retain the library defaults used in our evaluator (scikit-image 0.25.2): a uniform $7 \times 7 \times 7$ window, sample covariance, $c_1 = (0.01)^2$, and $c_2 = (0.03)^2$. The final score is the mean of the local SSIM field after excluding the three-voxel border required by the window. This metric therefore measures local agreement of the strongest density patterns, including their local luminance, contrast, and covariance.

Cosine similarity. The released protocol obtains this value through ChimeraX `measure correlation` (Pettersen et al., 2021). Our batched evaluator computes the same quantity directly on the complete aligned arrays before percentile processing:

$$\mathrm{CosineSimilarity}(X, Y) = \frac{\langle X, Y \rangle}{\|X\|_2 \|Y\|_2}. \tag{57}$$

The inner product and norms include every voxel in the common box. Cosine similarity is uncentered and therefore measures agreement relative to zero density; a shared positive offset contributes to its numerator.

Thresholded PCC. PCC uses the independently thresholded and normalized maps in Equation 55. After flattening X^{90} and Y^{90}, including their zero background voxels, the reported value is

$$\mathrm{PCC}_{\mathrm{S2M}}(X, Y) = \frac{\sum_i (X_i^{90} - \overline{X^{90}})(Y_i^{90} - \overline{Y^{90}})}{\sqrt{\sum_i (X_i^{90} - \overline{X^{90}})^2 \sum_i (Y_i^{90} - \overline{Y^{90}})^2}}. \tag{58}$$

This score emphasizes whether the two maps place their strongest density in the same voxels and largely suppresses weaker density. Its independent 90th-percentile thresholding makes PCC a measure of the centered high-density pattern; full-box Cosine similarity measures the complete aligned arrays.

Case aggregation and interpretation. Each metric is first computed for each experimental reference map–generated density map pair. Table 1 reports the arithmetic mean and median over the fixed common cohort of 1,923 test cases, assigning one score and equal weight to every case. The reported metrics are complementary: SSIM evaluates local high-density structure, Cosine similarity evaluates full-box agreement relative to zero, PCC evaluates the centered high-density pattern, and log-RAPS MAE measures disagreement between radial spectral-energy profiles. The high-density region is selected independently for each map and background voxels remain in the arrays. Joint reporting with Map–map CC and Masked FSC AUC presents their complementary behavior.

I Additional Experimental Results

I.1 Map Fidelity Across Resolution Ranges

We report the $r < 3.5$, $3.5 \leq r \leq 6$, and $r > 6$ Å strata to separate high-, medium-, and low-resolution behavior. The CryoNet.Density rows use the default one-transition stochastic sampler with two shared-denoiser evaluations (2 NFE). Bold marks the best unrounded mean or median; lower is better only for log-RAPS MAE.

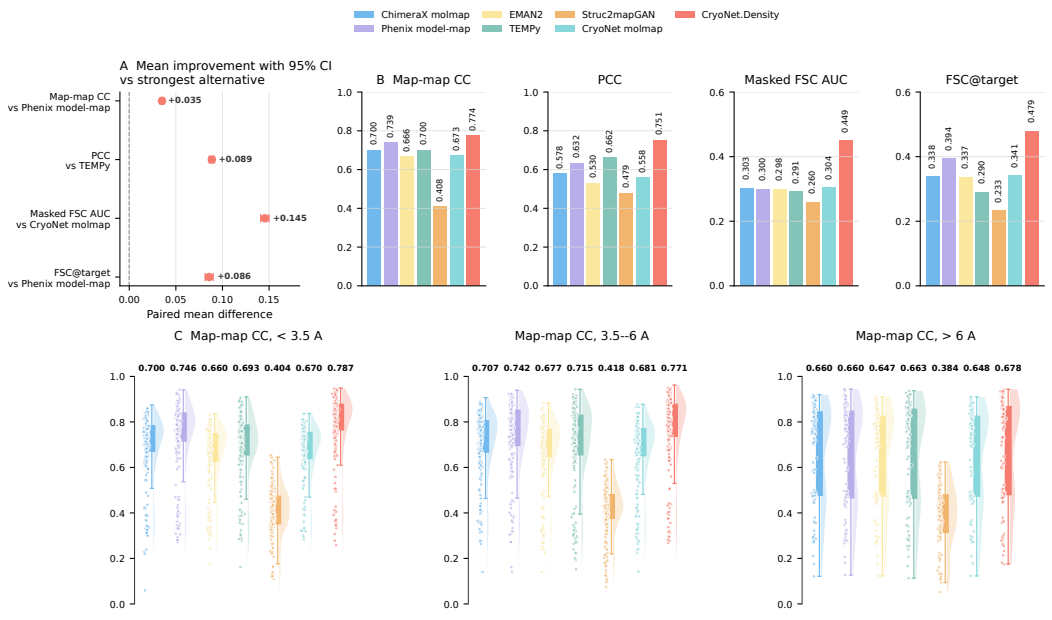

Figure 11: Complementary comparisons across methods and resolution intervals on the primary benchmark.

Table 7: Primary-test results for $r < 3.5$ Å.

Method	Statistic	Map–map CC	Masked FSC AUC	FSC@target	SSIM	Cosine similarity	log-RAPS MAE ↓	PCC
ChimeraX `molmap`	Mean	0.700	0.345	0.353	0.832	0.708	0.845	0.595
	Median	0.743	0.332	0.359	0.846	0.749	0.699	0.626
`phenix.model_map`	Mean	0.746	0.335	0.414	0.692	0.577	2.158	0.647
	Median	0.788	0.319	0.430	0.721	0.556	2.187	0.678
EMAN2	Mean	0.660	0.337	0.350	0.827	0.665	1.002	0.543
	Median	0.706	0.327	0.360	0.841	0.709	0.806	0.571
TEMPy	Mean	0.693	0.324	0.284	0.838	0.710	1.268	0.674
	Median	0.722	0.315	0.286	0.856	0.737	1.183	0.707
Struc2mapGAN	Mean	0.404	0.287	0.173	0.770	0.497	1.201	0.510
	Median	0.412	0.278	0.164	0.786	0.502	1.206	0.530
CryoNet `molmap`	Mean	0.670	0.345	0.357	0.838	0.676	1.023	0.578
	Median	0.713	0.333	0.363	0.857	0.715	0.959	0.609
CryoNet.Density	Mean	**0.787**	**0.532**	**0.534**	**0.861**	**0.792**	**0.573**	**0.784**
	Median	**0.835**	**0.550**	**0.560**	**0.883**	**0.837**	**0.498**	**0.833**

Table 8: Primary-test results for $3.5 \leq r \leq 6$ Å.

Method	Statistic	Map–map CC	Masked FSC AUC	FSC@target	SSIM	Cosine similarity	log-RAPS MAE ↓	PCC
ChimeraX `molmap`	Mean	0.706	0.268	0.326	0.840	0.718	0.878	0.564
	Median	0.749	0.256	0.326	0.866	0.758	**0.674**	0.586
`phenix.model_map`	Mean	0.741	0.270	0.381	0.679	0.588	2.404	0.624
	Median	0.784	0.259	0.394	0.721	0.565	2.456	0.649
EMAN2	Mean	0.677	0.265	0.326	0.836	0.687	1.065	0.517
	Median	0.716	0.254	0.328	0.861	0.724	0.835	0.538
TEMPy	Mean	0.714	0.263	0.299	0.837	0.731	1.938	0.659
	Median	0.752	0.249	0.297	0.866	0.767	1.739	0.689
Struc2mapGAN	Mean	0.417	0.241	0.307	0.767	0.516	0.962	0.459
	Median	0.429	0.228	0.310	0.784	0.523	0.838	0.474
CryoNet `molmap`	Mean	0.680	0.268	0.328	0.839	0.690	1.291	0.539
	Median	0.719	0.254	0.330	0.867	0.726	1.116	0.562
CryoNet.Density	Mean	**0.770**	**0.379**	**0.431**	**0.848**	**0.784**	**0.858**	**0.729**
	Median	**0.824**	**0.363**	**0.437**	**0.877**	**0.831**	0.738	**0.774**

Table 9: Primary-test results for $r > 6$ Å.

Method	Statistic	Map–map CC	Masked FSC AUC	FSC@target	SSIM	Cosine similarity	log-RAPS MAE ↓	PCC
ChimeraX `molmap`	Mean	0.664	0.160	0.289	**0.828**	0.682	**1.244**	0.526
	Median	0.753	0.155	0.294	0.854	0.768	**1.098**	0.561
`phenix.model_map`	Mean	0.664	**0.190**	0.298	0.614	0.613	2.416	0.554
	Median	0.761	**0.181**	0.319	0.654	0.578	2.514	0.539
EMAN2	Mean	0.651	0.164	0.288	0.826	0.668	1.929	0.499
	Median	0.745	0.159	0.291	0.852	0.755	1.763	0.530
TEMPy	Mean	0.667	0.178	0.285	0.809	0.691	4.797	0.580
	Median	0.771	0.171	0.291	0.854	0.792	5.342	0.565
Struc2mapGAN	Mean	0.388	0.136	0.273	0.745	0.512	1.986	0.335
	Median	0.405	0.133	0.285	0.765	0.506	1.934	0.336
CryoNet `molmap`	Mean	0.652	0.164	0.290	0.826	0.668	1.576	0.509
	Median	0.745	0.159	0.300	0.858	0.755	1.591	0.537
CryoNet.Density	Mean	**0.682**	0.173	**0.307**	0.815	**0.712**	1.326	**0.598**
	Median	**0.797**	0.170	**0.331**	**0.867**	**0.812**	1.282	**0.603**

I.2 Results on the Struc2mapGAN Benchmark

We additionally evaluate the 130 paired cases in the Struc2mapGAN protocol using the same three resolution intervals. The CryoNet.Density rows use the default one-transition stochastic sampler with 2 NFE, and bold follows the unrounded values.

Table 10: Protocol-benchmark results for $r < 3.5$ Å.

Method	Statistic	Map–map CC	Masked FSC AUC	FSC@target	SSIM	Cosine similarity	log-RAPS MAE ↓	PCC
ChimeraX `molmap`	Mean	0.623	0.336	0.405	0.805	0.626	1.684	0.502
	Median	0.617	0.335	0.415	0.847	0.625	1.621	0.506
`phenix.model_map`	Mean	0.801	0.323	0.486	0.717	0.593	2.213	0.675
	Median	0.801	0.318	0.484	0.776	0.561	2.509	0.679
EMAN2	Mean	0.478	0.299	0.359	0.798	0.481	2.416	0.402
	Median	0.538	0.295	0.403	0.840	0.543	1.838	0.452
TEMPy	Mean	0.709	0.321	0.345	0.802	0.734	0.758	0.691
	Median	0.712	0.321	0.354	0.833	0.740	0.606	0.699
Struc2mapGAN	Mean	0.434	0.294	0.270	0.772	0.518	1.124	0.576
	Median	0.425	0.286	0.281	0.800	0.512	1.288	0.580
CryoNet `molmap`	Mean	0.554	0.333	0.402	0.817	0.556	1.475	0.491
	Median	0.548	0.334	0.412	0.855	0.554	1.347	0.485
CryoNet.Density	Mean	**0.851**	**0.586**	**0.606**	**0.857**	**0.853**	**0.465**	**0.843**
	Median	**0.858**	**0.579**	**0.598**	**0.886**	**0.866**	**0.438**	**0.852**

Table 11: Protocol-benchmark results for $3.5 \leq r \leq 6$ Å.

Method	Statistic	Map–map CC	Masked FSC AUC	FSC@target	SSIM	Cosine similarity	log-RAPS MAE ↓	PCC
ChimeraX `molmap`	Mean	0.491	0.260	0.347	0.803	0.499	2.321	0.362
	Median	0.510	0.264	0.367	0.828	0.513	2.252	0.374
`phenix.model_map`	Mean	0.757	0.271	**0.413**	0.674	0.597	2.238	0.616
	Median	0.776	0.267	**0.436**	0.737	0.593	2.300	0.641
EMAN2	Mean	0.305	0.211	0.261	0.800	0.310	3.359	0.248
	Median	0.344	0.202	0.201	0.827	0.338	3.289	0.242
TEMPy	Mean	0.612	0.256	0.328	0.743	0.651	1.637	0.577
	Median	0.633	0.258	0.340	0.737	0.673	1.496	0.604
Struc2mapGAN	Mean	0.446	0.248	0.346	0.759	0.539	0.923	0.466
	Median	0.460	0.244	0.364	0.778	0.549	0.833	0.484
CryoNet `molmap`	Mean	0.425	0.258	0.343	0.808	0.432	2.276	0.360
	Median	0.435	0.263	0.361	0.836	0.441	2.181	0.372
CryoNet.Density	Mean	**0.791**	**0.357**	0.402	**0.837**	**0.806**	**0.923**	**0.728**
	Median	**0.833**	**0.343**	0.402	**0.867**	**0.846**	**0.781**	**0.773**

Table 12: Protocol-benchmark results for $r > 6$ Å.

Method	Statistic	Map–map CC	Masked FSC AUC	FSC@target	SSIM	Cosine similarity	log-RAPS MAE ↓	PCC
ChimeraX `molmap`	Mean	0.431	0.190	0.305	0.833	0.442	3.183	0.287
	Median	0.445	0.178	0.290	0.862	0.459	3.000	0.291
`phenix.model_map`	Mean	0.791	**0.272**	**0.348**	0.711	0.636	1.768	0.613
	Median	0.832	**0.248**	**0.348**	0.713	0.578	1.715	0.618
EMAN2	Mean	0.224	0.142	0.179	0.833	0.230	4.024	0.176
	Median	0.158	0.141	0.150	0.859	0.158	4.168	0.154
TEMPy	Mean	0.565	0.194	0.310	0.730	0.610	2.884	0.510
	Median	0.586	0.178	0.306	0.698	0.624	2.813	0.511
Struc2mapGAN	Mean	0.467	0.200	0.327	0.781	0.560	1.975	0.419
	Median	0.464	0.183	0.326	0.797	0.562	2.065	0.410
CryoNet `molmap`	Mean	0.367	0.186	0.295	0.833	0.377	3.344	0.290
	Median	0.380	0.178	0.286	0.866	0.393	3.250	0.287
CryoNet.Density	Mean	**0.822**	0.210	0.332	**0.856**	**0.840**	**1.204**	**0.676**
	Median	**0.857**	0.197	0.331	**0.876**	**0.873**	**1.261**	**0.685**

I.3 Resolution Control

We extend the single-atomic-model PDB-*7kod* example to a 96-atomic-model test cohort, with 32 atomic models in each resolution interval and balanced coverage of four within-interval volume-size quantile bins.

For every atomic model, we keep its atomic coordinates and physical grid fixed and request 2–10 Å in 1 Å increments. The full path changes both matched-resolution rendering and the Fourier-FiLM condition. The renderer-only and FiLM-only sweeps change one path while holding the other at 6 Å. The quantitative experiment uses the final main-model weights and the default one-transition stochastic sampler (2 NFE). Within each atomic model, the same Gaussian field is reused across resolutions and paths, so the paired sweep directly measures the control response.

Metrics and interpretation. For each atomic model and path, let $c(r)$ be the spectral centroid of the generated density map $\widehat{X}_{0,r}$ requested at $r \in \{2, \ldots, 10\}$ Å. We compute $\rho = \mathrm{Spearman}(r, c(r))$ per atomic model. Because a larger Å value should remove high-frequency content, a more negative ρ is better; $\rho < 0$ and $\rho \leq -0.7$ report the fractions of atomic models with the expected direction and a strong monotonic response. Sweep span is the voxelwise $\mathrm{MAE}(\widehat{X}_{0,2}, \widehat{X}_{0,10})$, summarized by its median across atomic models; larger values indicate a wider 2–10 Å control response. For retrieval, each generated density map is matched by PCC to the nine physically rendered maps at 2–10 Å; if r^* is the best-matching resolution, Retrieval MAE averages $|r^* - r|$ over requests and atomic models, so lower is better. Bold values are best within each cohort.

Table 13: Multi-structure resolution control. Arrows indicate the preferred direction; best values within each cohort are bold.

Cohort	Path	N	Median $\rho \downarrow$	$\rho < 0 \uparrow$	$\rho \leq -0.7 \uparrow$	Sweep span $\uparrow$	Retrieval MAE $\downarrow$ (Å)
All	Full	96	**-1.000**	**100.0%**	**100.0%**	**0.0383**	**1.678**
All	Renderer only	96	-0.783	94.8%	59.4%	0.0325	2.307
All	FiLM only	96	0.717	18.8%	15.6%	0.0174	3.462
< 3.5 Å	Full	32	**-1.000**	**100.0%**	**100.0%**	**0.0447**	**1.618**
< 3.5 Å	Renderer only	32	-0.875	**100.0%**	90.6%	0.0344	1.826
< 3.5 Å	FiLM only	32	0.750	0.0%	0.0%	0.0214	3.615
$3.5 \leq r \leq 6$ Å	Full	32	**-1.000**	**100.0%**	**100.0%**	**0.0400**	**1.622**
$3.5 \leq r \leq 6$ Å	Renderer only	32	-0.817	**100.0%**	71.9%	0.0342	2.229
$3.5 \leq r \leq 6$ Å	FiLM only	32	0.792	0.0%	0.0%	0.0173	3.632
> 6 Å	Full	32	**-0.950**	**100.0%**	**100.0%**	**0.0179**	**1.795**
> 6 Å	Renderer only	32	-0.400	84.4%	15.6%	0.0134	2.865
> 6 Å	FiLM only	32	-0.450	56.3%	46.9%	0.0153	3.139

Under one-transition stochastic sampling, the full rendering-FiLM path gives a monotonic bandwidth response for every structure, with median $\rho = -1.000$ and all 96 structures satisfying $\rho \leq -0.7$. Renderer-only retains the main negative trend. FiLM-only has a positive median ρ in the high- and medium-resolution cohorts and is heterogeneous for $r > 6$ Å; its response varies with the spatial scale and bandwidth of the fixed physically rendered map. Matched-resolution rendering and Fourier-FiLM together produce the most consistent control.

The following panels visualize the same cohort with the default one-transition stochastic sampler used in Table 13.

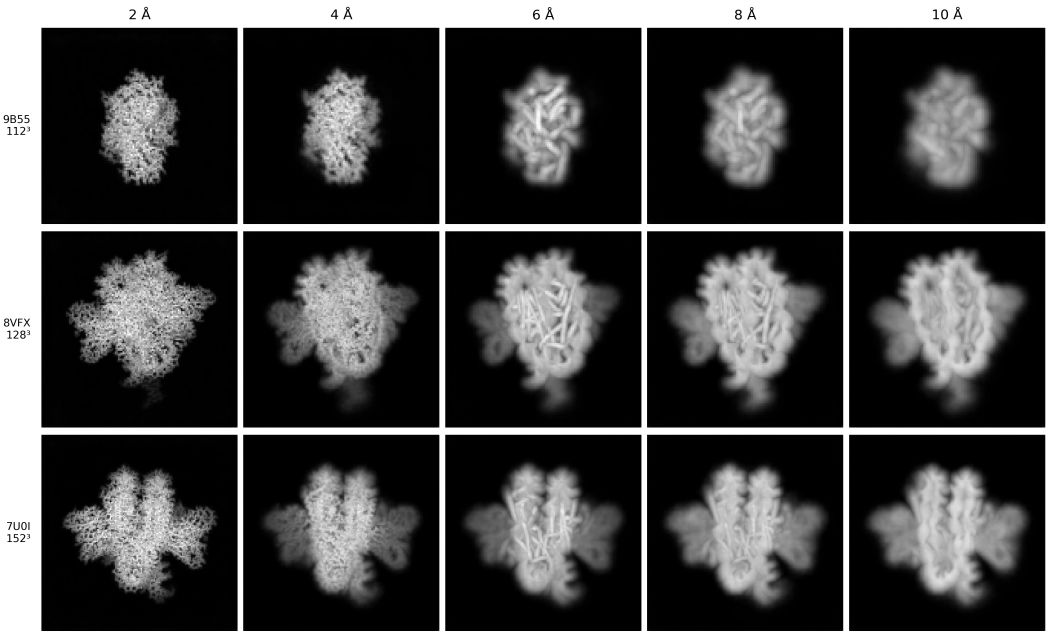

Figure 12: One-transition stochastic examples from the high-resolution test cohort under the full 2–10 Å sweep. Coordinates, output grid, and Gaussian draw remain fixed within each row.

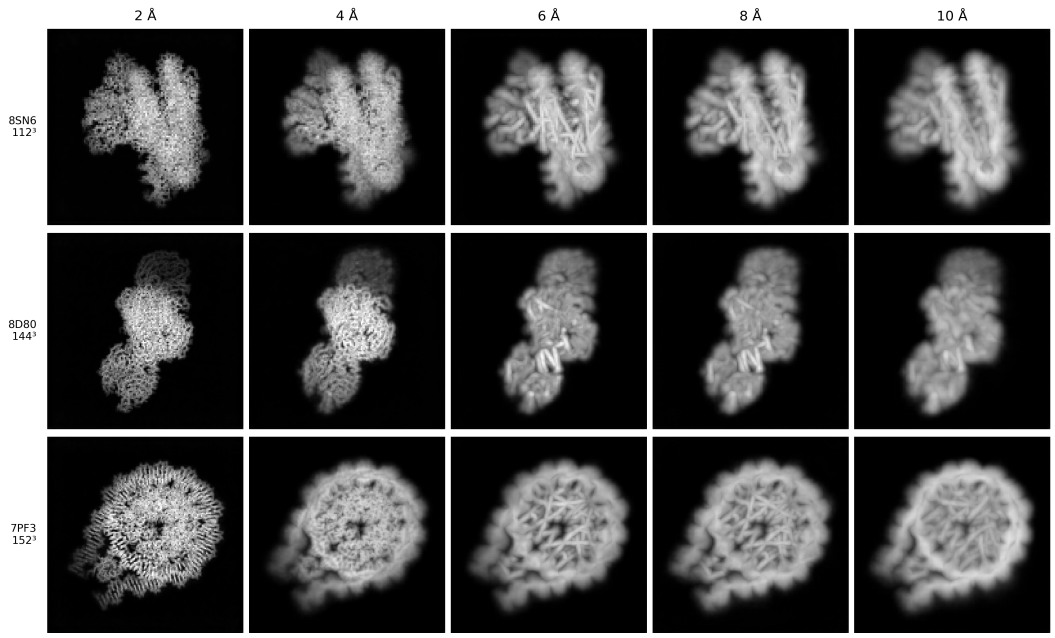

Figure 13: One-transition stochastic examples from the medium-resolution test cohort under the full 2–10 Å sweep. Coordinates, output grid, and Gaussian draw remain fixed within each row.

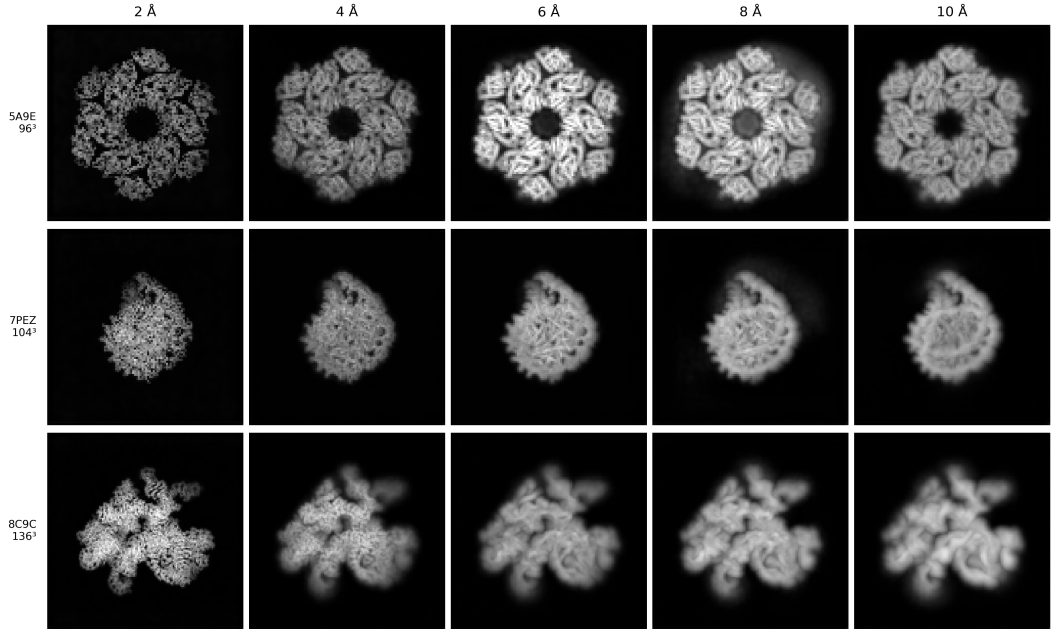

Figure 14: One-transition stochastic examples from the low-resolution test cohort under the full 2–10 Å sweep. Coordinates, output grid, and Gaussian draw remain fixed within each row.

I.4 Downstream Recognition

Emap2sec+. Starting from the 1,923-case primary held-out test cohort, the inclusive filter $5 \leq r \leq 10$ Å and the requirement for complete experimental-reference, TEMPy, EMAN2, CryoNet `molmap`, and CryoNet.Density inputs yield the 136 cases reported in Table 2; no further case sampling is applied. Each input is placed in the same physical frame. Its contour is chosen to match

the occupied-voxel fraction of the coordinate-derived reference so that the five inputs are evaluated on comparable support. The deposited atomic model supplies voxel labels through the released Emap2sec+ preparation pipeline, including STRIDE secondary-structure assignments.

We run the released frozen full cascade independently for each density: the coil, sheet, helix, DNA/RNA, and joint four-class first-stage networks feed `Combine_Modelall.pkl`, followed by the released spatial refinement. Probabilities are fused across the four released folds before taking the final class. Accuracy and per-class F1 are computed over labeled voxels and then averaged equally over cases; when a voxel admits multiple labels, a prediction matching any admissible label is counted as correct. Helix, sheet, and coil/other F1 use all 136 cases. DNA/RNA F1 is averaged only over the 35 cases with at least one DNA/RNA-labeled voxel. Displayed values are rounded only after aggregation.

ModelAngelo. From the primary held-out test cohort, requiring protein content and $r \leq 4.0$ Å yields 1,524 cases; their observed resolution range is 1.88–4.0 Å. We run the same released model on four aligned inputs per case: the experimental reference map, ChimeraX `molmap`, CryoNet `molmap`, and the CryoNet.Density generated density map. Inference comprises frozen Cα detection followed by 20-class amino-acid assignment. Residue accuracy is the fraction of assigned residues with the correct amino-acid class. Predicted chains are compared with the deposited atomic model using `phenix.chain_comparison` with a 3 Å matching radius, which supplies sequence match and sequence score.

Because chain correspondence can fail entirely, Table 3 uses a paired success cohort: all four inputs must have a nonzero sequence match. This criterion retains 711 of the 1,524 eligible cases and excludes the remaining 813. All three reported metrics are then averaged equally over exactly these same 711 cases, so differences cannot arise from method-specific case inclusion. Values are rounded only for display; the table therefore describes performance conditional on successful chain correspondence.

I.5 Runtime and Memory Usage

We profile the unit cost of one denoiser evaluation, decomposing measured inference into tensor preparation and batch assembly, the denoiser forward pass and output materialization, and Python orchestration and synchronization. Profiling uses the final main-model checkpoint on one NVIDIA H20 with batch size one and bfloat16 autocast over all 1,923 variable-size test cases. We perform three warm-up passes and explicitly synchronize CUDA at every timing boundary so that each measured interval covers all previously queued GPU work.

Timing starts after the data loader returns a sample and includes host-to-device tensor preparation, the denoiser evaluation, and output materialization. The measured scope focuses on in-memory model inference. The table reports per-case means over the same 1,923 cases, with component shares computed from those measurements.

Table 14: Runtime profile of one denoiser evaluation on one NVIDIA H20.

Component	Mean time per case (ms)	Share (%)
Tensor preparation and batch assembly	16.97	9.6
Denoiser forward and output materialization	160.01	90.4
Python orchestration and synchronization	0.00	0.0
Total per evaluation	176.97	100.0

The denoiser forward pass and output materialization account for 90.4% of the measured unit cost, whereas tensor preparation and batch assembly account for 9.6%. The default one-transition stochastic sampler uses 2 NFE, so its runtime is dominated by two evaluations of the shared denoiser; each additional transition adds another evaluation. Table 19 reports the full-test transition-count ablation and lists the corresponding NFE.

J Ablation Studies and Residual Analysis

J.1 Resolution Conditioning

These experiments use the same data and backbone to separate the two resolution-conditioning paths and to compare bridge learning with same-backbone direct regression.

A0 fixes the physically rendered map at 2 Å and uses a resolution-invariant denoising path. A1 adds Fourier-FiLM to the fixed physically rendered map, allowing the requested resolution to modulate the BBDM denoiser. A2 uses matched-resolution rendering while retaining a resolution-invariant denoising path. These ablations isolate complementary mechanisms: matched rendering establishes the physically rendered map bandwidth, and Fourier-FiLM supplies request-dependent conditioning to the BBDM.

Table 15: Overall conditioning-path ablation. Every method uses one-transition stochastic sampling; bold marks the best unrounded mean or median.

Method	Statistic	Map–map CC	Masked FSC AUC	FSC@target	SSIM	Cosine similarity	log-RAPS MAE ↓	PCC
CryoNet.Density	Mean	**0.774**	**0.449**	**0.479**	0.853	**0.784**	**0.731**	**0.751**
	Median	**0.831**	**0.463**	**0.503**	0.879	**0.835**	**0.603**	**0.807**
A0	Mean	0.748	0.384	0.439	0.854	0.759	0.879	0.720
	Median	0.798	0.399	0.463	0.880	0.805	0.735	0.770
A1	Mean	0.758	0.414	0.459	0.857	0.770	0.762	0.741
	Median	0.814	0.421	0.480	0.886	0.820	0.674	0.797
A2	Mean	0.761	0.402	0.456	**0.859**	0.772	0.782	0.742
	Median	0.817	0.408	0.484	**0.889**	0.822	0.693	0.797

Table 16: Resolution-stratified conditioning-path ablation. Every method uses one-transition stochastic sampling; bold marks the best unrounded mean or median within each stratum.

Method	Resolution	Statistic	Map–map CC	Masked FSC AUC	FSC@target	SSIM	Cosine similarity	log-RAPS MAE ↓	PCC
CryoNet.Density	< 3.5 Å	Mean	**0.787**	**0.532**	**0.534**	0.861	**0.792**	**0.573**	**0.784**
		Median	**0.835**	**0.550**	**0.560**	0.883	**0.837**	**0.498**	**0.833**
A0	< 3.5 Å	Mean	0.770	0.445	0.478	0.858	0.778	0.781	0.756
		Median	0.816	0.453	0.507	0.879	0.821	0.706	0.798
A1	< 3.5 Å	Mean	0.769	0.496	0.516	0.863	0.777	0.623	0.773
		Median	0.820	0.505	0.544	0.887	0.823	0.571	0.822
A2	< 3.5 Å	Mean	0.774	0.476	0.512	**0.864**	0.780	0.643	0.772
		Median	0.822	0.479	0.540	**0.890**	0.825	0.586	0.820
CryoNet.Density	3.5–6 Å	Mean	**0.770**	**0.379**	**0.431**	0.848	**0.784**	**0.858**	**0.729**
		Median	**0.824**	**0.363**	**0.437**	0.877	**0.831**	0.738	**0.774**
A0	3.5–6 Å	Mean	0.737	0.333	0.406	0.854	0.750	0.871	0.699
		Median	0.781	0.332	0.415	0.886	0.790	**0.707**	0.745
A1	3.5–6 Å	Mean	0.755	0.340	0.407	0.854	0.770	0.892	0.719
		Median	0.806	0.334	0.416	0.888	0.817	0.814	0.767
A2	3.5–6 Å	Mean	0.757	0.336	0.404	**0.858**	0.771	0.916	0.721
		Median	0.807	0.324	0.409	**0.891**	0.817	0.855	0.763
CryoNet.Density	> 6 Å	Mean	**0.682**	**0.173**	0.307	0.815	**0.712**	1.326	0.598
		Median	**0.797**	**0.170**	**0.331**	0.867	**0.812**	1.282	0.603
A0	> 6 Å	Mean	0.621	0.166	**0.310**	**0.821**	0.650	1.838	0.545
		Median	0.683	0.167	0.331	0.860	0.707	1.670	0.558
A1	> 6 Å	Mean	0.677	0.163	0.289	0.818	0.706	1.157	0.588
		Median	0.777	0.165	0.307	**0.870**	0.799	**0.991**	0.590
A2	> 6 Å	Mean	0.679	0.164	0.288	0.817	0.707	**1.153**	**0.599**
		Median	0.784	0.160	0.305	0.868	0.802	1.138	**0.605**

J.2 Bridge Training versus Direct Regression

The direct-regression variant uses the matched-resolution physically rendered map, Fourier-FiLM conditioning, the same backbone and loss weights, and a regression objective from the physically rendered map to the experimental reference map. The BBDM variant trains the denoiser on time-indexed intermediate states across the bridge between the paired maps. Table 17 shows stronger spatial and Fourier-shell agreement for BBDM together with closer radial-spectrum agreement. These metrics distinguish spatial organization from the spherically averaged spectrum. The BBDM row uses the default one-transition stochastic sampler; the regression row uses its native one-evaluation prediction.

Table 17: Comparison of the default one-transition stochastic BBDM sampler with the same-backbone direct-regression ablation. The physically rendered map, Fourier-FiLM conditioning, backbone, and loss weights are shared.

Method	Statistic	Map–map CC	Masked FSC AUC	FSC@target	SSIM	Cosine similarity	log-RAPS MAE ↓	PCC
CryoNet.Density	Mean	**0.774**	**0.449**	**0.479**	**0.853**	**0.784**	**0.731**	**0.751**
	Median	**0.831**	**0.463**	**0.503**	**0.879**	**0.835**	**0.603**	**0.807**
Direct regression	Mean	0.743	0.334	0.370	0.849	0.755	0.758	0.673
	Median	0.792	0.339	0.393	0.871	0.800	0.662	0.718

J.3 SAMPLING TRAJECTORY

We compare two zero-transition controls with the default one-transition stochastic sampler. Direct evaluates the exact physically rendered map boundary state. Noisy Direct adds a boundary-scale Gaussian perturbation while retaining the clean physically rendered map as the spatial condition. The default sampler draws one posterior bridge-state transition and evaluates the shared denoiser twice. Direct and Noisy Direct are closely matched across most statistics. The one-transition sampler gives the best unrounded value in 13 of the 14 mean/median comparisons; Noisy Direct leads mean Cosine similarity.

Table 18: Sampler controls on the same 1,923 primary-test cases. Direct and Noisy Direct are zero-transition, 1-NFE controls; the default one-transition stochastic sampler uses 2 NFE. Bold marks the best unrounded value for each statistic.

Sampler	Transitions	NFE	Statistic	Map–map CC	Masked FSC AUC	FSC@target	SSIM	Cosine similarity	log-RAPS MAE ↓	PCC
Direct	0	1	Mean	0.766	0.412	0.470	0.851	0.785	0.765	0.744
			Median	0.822	0.415	0.495	0.879	0.834	0.680	0.800
Noisy Direct	0	1	Mean	0.766	0.412	0.469	0.851	**0.785**	0.765	0.744
			Median	0.822	0.415	0.496	0.879	0.834	0.680	0.799
One-transition stochastic	1	2	Mean	**0.774**	**0.449**	**0.479**	**0.853**	0.784	**0.731**	**0.751**
			Median	**0.831**	**0.463**	**0.503**	**0.879**	**0.835**	**0.603**	**0.807**

J.3.1 TRANSITION-COUNT ABLATION

We evaluate $J \in \{1, 2, 3, 4, 5, 10, 15, 20\}$ posterior bridge-state transitions on the complete test set. Every transition uses the full BBDM posterior, and a schedule with J transitions requires $J + 1$ evaluations of the shared denoiser. All schedules use the same fixed seed and 1,923 test cases. Relative time is measured on one NVIDIA H20 using 12 output-size representatives and five repeats per case, normalized to $J = 1$. One stochastic transition gives the strongest overall spatial and Fourier-shell fidelity. A small number of additional transitions can improve selected texture or radial-spectrum measures, while longer paths show lower agreement across the main spatial and spectral metrics. Measured cost increases monotonically with NFE.

Table 19: Full-test stochastic transition-count evaluation. Every row uses the same 1,923 cases; J posterior transitions require $J + 1$ denoiser evaluations. Relative time uses the single-H20 timing cohort described above. Bold marks the best unrounded value for each statistic.

J transitions	NFE	Relative time	Statistic	Map–map CC	Masked FSC AUC	FSC@target	SSIM	Cosine similarity	log-RAPS MAE ↓	PCC
1	2	1.00×	Mean	**0.774**	**0.449**	**0.479**	0.853	**0.784**	0.731	**0.751**
			Median	**0.831**	**0.463**	**0.503**	0.879	**0.835**	0.603	**0.807**
2	3	1.39×	Mean	0.771	0.445	0.469	0.854	0.781	0.709	0.749
			Median	0.828	0.458	0.492	0.880	0.832	0.582	0.804
3	4	1.81×	Mean	0.763	0.442	0.465	0.856	0.770	**0.684**	0.746
			Median	0.820	0.453	0.491	0.883	0.822	**0.564**	0.800
4	5	2.24×	Mean	0.758	0.438	0.462	**0.857**	0.764	0.684	0.744
			Median	0.814	0.450	0.487	0.884	0.817	0.568	0.797
5	6	2.67×	Mean	0.753	0.433	0.455	0.857	0.760	0.696	0.741
			Median	0.809	0.442	0.481	**0.885**	0.813	0.583	0.794
10	11	4.81×	Mean	0.739	0.420	0.440	0.856	0.745	0.723	0.732
			Median	0.793	0.425	0.464	0.883	0.795	0.610	0.784
15	16	6.95×	Mean	0.730	0.410	0.426	0.854	0.737	0.746	0.724
			Median	0.784	0.412	0.449	0.881	0.785	0.630	0.776
20	21	9.08×	Mean	0.722	0.403	0.419	0.853	0.729	0.773	0.718
			Median	0.776	0.403	0.442	0.879	0.778	0.657	0.770

We then use a deterministic mean-path intervention to test how evaluation placement affects the bridge. The two three-transition schedules have identical NFE and differ in whether they enter the

maximum-variance midpoint. This intervention tests bridge geometry; the table above reports the default stochastic posterior behavior.

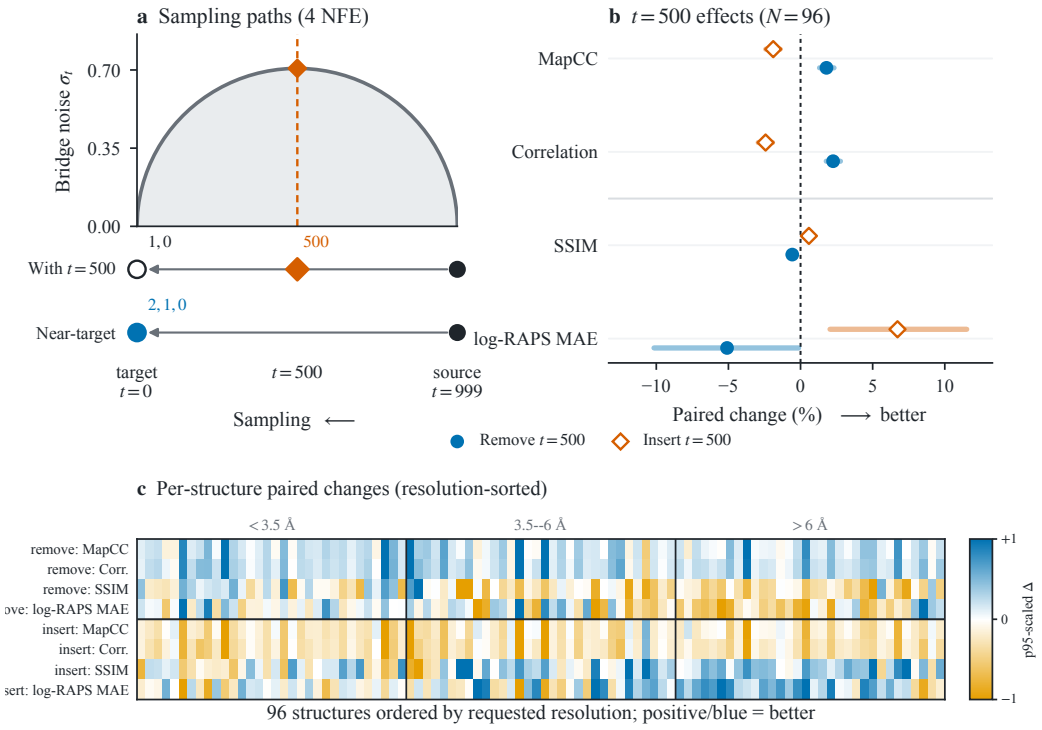

Figure 15: Bridge-region coverage under a deterministic mean-path schedule intervention. (a) The midpoint and near-target three-transition schedules use the same 4 NFE but differ in whether they query the maximum-variance bridge midpoint. (b) Paired mean changes on 96 resolution-balanced test structures; error bars are 95% paired-bootstrap confidence intervals, and positive values indicate better agreement for every metric. (c) Casewise changes sorted by requested resolution. For each metric, the two intervention rows share the same symmetric absolute-change scale; blue denotes better agreement and amber denotes worse agreement.

J.3.2 Complete Single-Structure Transition Trajectory

We evaluate every integer $J = 1, \ldots, 200$ on EMD-22341/PDB-*7ji0*. The default $J = 1$ gives the best value on all seven metrics for this case. Every setting restarts from the same physically rendered map, uses the full stochastic BBDM posterior and the same seed, and requires $J + 1$ NFE. Figure 16 shows that spatial and spectral agreement generally decreases as the path lengthens.

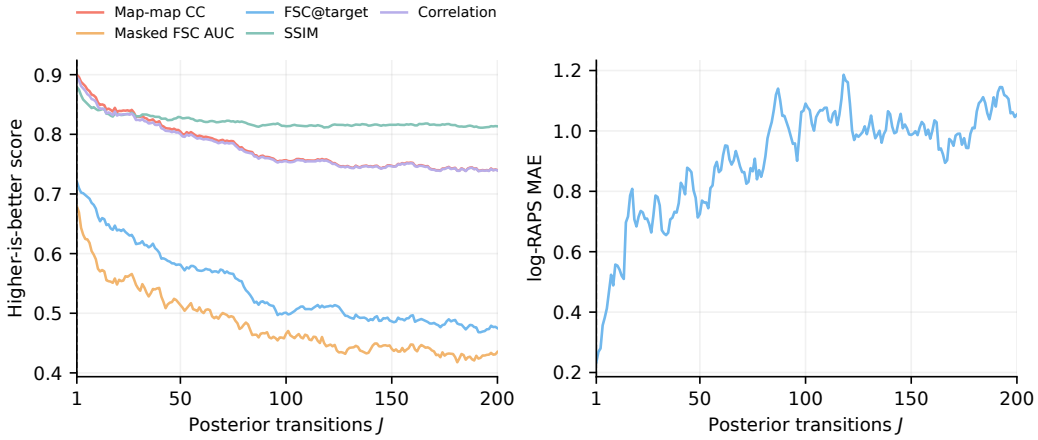

Figure 16: Integer-by-integer stochastic bridge-transition trajectories for EMD-22341/PDB-*7ji0*, $J = 1, \ldots, 200$. The dashed line marks the default $J = 1$.

Table 20: Single-structure transition-count values. Relative time is the median of 5 measurements per setting on one NVIDIA H20, normalized to $J = 1$.

J	NFE	Relative time	Map–map CC	Masked FSC AUC	FSC@target	SSIM	Cosine similarity	log-RAPS MAE ↓	PCC
1	2	1.00×	**0.900**	**0.679**	**0.721**	**0.879**	**0.895**	**0.234**	**0.921**
2	3	1.47×	0.896	0.669	0.710	0.875	0.891	0.269	0.918
3	4	1.94×	0.888	0.643	0.702	0.866	0.881	0.280	0.909
4	5	2.41×	0.884	0.632	0.701	0.858	0.877	0.356	0.905
5	6	2.88×	0.881	0.624	0.698	0.855	0.873	0.382	0.902
10	11	5.23×	0.862	0.590	0.675	0.843	0.855	0.558	0.882
15	16	7.58×	0.841	0.555	0.648	0.834	0.835	0.697	0.861
20	21	9.92×	0.837	0.548	0.639	0.834	0.831	0.684	0.857
50	51	24.02×	0.805	0.514	0.581	0.829	0.801	0.725	0.824
100	101	47.40×	0.757	0.466	0.501	0.814	0.755	1.091	0.776
150	151	70.84×	0.747	0.440	0.486	0.816	0.746	0.987	0.766
200	201	94.43×	0.740	0.435	0.475	0.813	0.739	1.054	0.759

Figure 17 reports measured time for all 200 settings. On one NVIDIA H20, $J = 200$ takes $94.43\times$ the $J = 1$ time, closely tracking the increase in NFE.

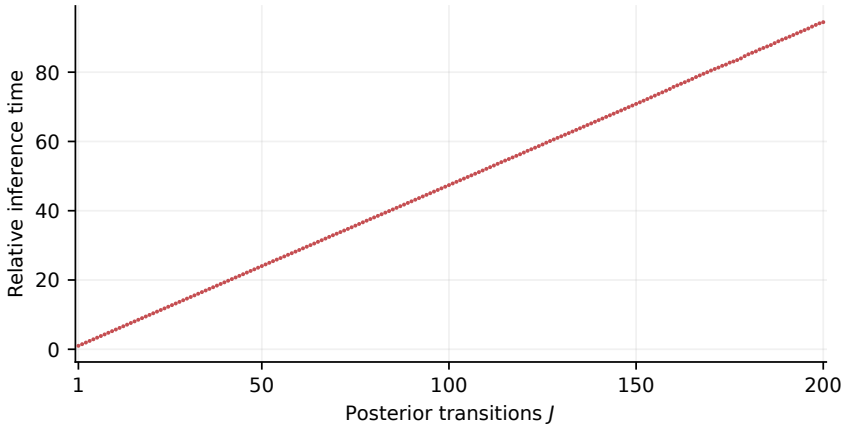

Figure 17: Measured relative inference time for EMD-22341/PDB-*7ji0* at every integer $J = 1, \ldots, 200$ on one NVIDIA H20; each point is the median of 5 synchronized measurements normalized to $J = 1$.

J.3.3 Why One Transition Is the Default Stochastic Schedule

Transition count and network evaluations. A posterior bridge-state transition moves from one scheduled state to the next. A schedule with J transitions contains $J + 1$ evaluations of the shared BBDM denoiser. The default therefore uses $J = 1$, the schedule $[999, 0]$, and two network evaluations (2 NFE). At each scheduled state, the EMA denoiser $f_{\bar{\theta}}$ predicts $Z_t - X_0$, so $\widehat{X}_{0,t} = Z_t - f_{\bar{\theta}}(Z_t, t, Y_r, r)$ is already a complete experimental reference map estimate. The first estimate parameterizes a stochastic posterior draw at a bridge state near the experimental reference map boundary, and the second evaluation corrects that sampled state. This short path retains the generative posterior while using the same resolution-conditioned BBDM-FiLM mapping on both sides of the transition.

Longer schedules enter the high-variance bridge interior. For the linear bridge used in our experiments, $\delta_{\max} = 1$, so a forward state has standard deviation $\sigma_t = \sqrt{\delta_t} = \sqrt{2(m_t - m_t^2)}$. The variance peaks at $m_t = 0.5$ and decreases toward the physically rendered map and experimental reference map boundaries. Table 21 shows the geometry of the first three stochastic schedules. One and two transitions remain near the low-variance boundaries; three transitions first add the midpoint $t = 500$, where σ_t is about 15.8 times its boundary value.

Table 21: Bridge regions traversed by the first three stochastic schedules. Scores are full-test means over 1,923 structures. Bold marks the best value among the displayed schedules; the complete transition-count results appear in Table 19.

J	NFE	Scheduled timesteps	$\max_k \sigma_{t_k}$	Map–map CC ↑	log-RAPS MAE ↓
1	2	$[999, 0]$	0.0447	**0.774**	0.731
2	3	$[999, 1, 0]$	0.0632	0.771	0.709
3	4	$[999, 500, 1, 0]$	0.7071	0.763	**0.684**

The full stochastic sweep separates two effects. The one-transition stochastic sampler gives the strongest spatial and Fourier-shell fidelity, including Map–map CC, Masked FSC AUC, FSC@target, Cosine similarity, and PCC. A few added transitions can improve SSIM and the radially averaged spectrum, while longer paths gradually reduce both spatial and spectral agreement. The early trade-off indicates that radial amplitude and local texture can move closer to the experimental reference map even as phase-sensitive spatial registration weakens. Every added transition also applies the same denoiser to another state constructed from its preceding prediction.

Coupled-noise error recurrence. We make the recursive effect explicit by coupling a learned trajectory Z_t and an oracle trajectory Z_t^*. The two trajectories share the corresponding Gaussian

draw at each posterior transition, while successive transitions use independent draws. Define $\Delta_t = Z_t - Z_t^*$ and write the learned experimental reference map estimate at the current state as

$$\widehat{X}_{0,t} = \mathrm{clip}_{[0,1]}[Z_t - f_{\bar{\theta}}(Z_t, t, Y_r, r)] = X_0 - E_t, \tag{59}$$

where E_t is the finite experimental reference map estimation error after the implemented clipping. With $b_{t\to s} = \sqrt{\widetilde{\delta}_{t\to s}}$, the full stochastic posterior update is

$$Z_s = (1 - m_s)\widehat{X}_{0,t} + m_s Y_r + a_{t\to s}\Big[Z_t - (1 - m_t)\widehat{X}_{0,t} - m_t Y_r\Big] + b_{t\to s}\epsilon_s. \tag{60}$$

Subtracting the oracle update under the shared draw ϵ_s gives

$$\Delta_s = a_{t\to s}\Delta_t + [a_{t\to s}(1 - m_t) - (1 - m_s)]\, E_t. \tag{61}$$

With full BBDM posterior stochasticity and $\delta_t = 2m_t(1 - m_t)$, Equation 29 gives $a_{t\to s} = m_s/m_t$. The recurrence then simplifies to

$$\boxed{\Delta_s = \frac{m_s}{m_t}\Delta_t - \left(1 - \frac{m_s}{m_t}\right) E_t}. \tag{62}$$

Within a coupled pair, the shared random increment cancels from the learned-oracle difference, but it remains in each conditional transition. Independent Gaussian draws therefore produce different transition states. Table 22 reports the resulting nonzero pairwise differences between generated density maps. Equation 62 tracks a separate quantity: finite error shifts the conditional mean of the next state, which changes the input received by the next denoiser evaluation.

Training constructs each random bridge state directly from the paired (X_0, Y_r) and a Gaussian field. During sampling, later states depend on earlier experimental reference map estimates. Starting from the same physically rendered map gives $\Delta_{999} = 0$, after which every transition combines the accumulated state shift with the current E_t. The Gaussian term supplies sample variability; the experimental reference map estimation error remains a state-dependent output of the nonlinear denoiser and per-evaluation clipping. Longer schedules repeat this feedback. In our schedule, they also query the high-variance bridge interior, where a larger stochastic perturbation and an inherited mean shift jointly determine the next input. The recursive construction of inference states from preceding estimates is consistent with error-propagation analyses of iterative diffusion sampling (Ning et al., 2023; 2024; Li & van der Schaar, 2024).

Oracle estimation-error check. We replace the learned EMA denoiser with the exact predictor $f^*(Z_t) = Z_t - X_0$ and test $J \in \{1, 2, 3, 4, 5, 10, 15, 20\}$, corresponding to NFE $\{2, 3, 4, 5, 6, 11, 16, 21\}$, under both deterministic and fully stochastic updates with per-evaluation clipping. Every setting recovers the experimental reference map, with maximum absolute error at most 7.45×10^{-9}. Exact prediction sets $E_t = 0$, and Equation 62 preserves $\Delta_t = 0$ along the coupled trajectories, so the terminal evaluation recovers X_0 for every tested path. An analytic-update unit test compares the implemented posterior transition with Equation 29 under the deterministic and full-stochastic limits and agrees to numerical precision. The recurrence attributes the learned–oracle shift to finite E_t and recursively constructed inputs. The original BBDM (Li et al., 2023) likewise reports that gains saturate as sampling becomes longer; related diffusion analyses connect iterative degradation to train-inference input mismatch (Ning et al., 2023; 2024; Li & van der Schaar, 2024), while consistency models (Song et al., 2023) obtain short paths through an explicitly trained path-consistency objective. For the present model, one stochastic bridge-state transition connects two low-variance states adjacent to the bridge boundaries, introduces posterior sample variation once, and limits recursive state feedback to a single update.

J.4 Conditional Sample Diversity

We draw five samples with independent posterior-transition noise for one test structure from each resolution interval.

Table 22: Five-seed conditional sampling with the default one-transition stochastic sampler. Molecular support and background are defined by $M > 0.1$ and $M < 0.05$, respectively. Pairwise statistics average the ten seed pairs; full-volume target PCC is reported across seeds.

Structure	r (Å)	Full-volume target PCC mean±SD	Support variation energy (%)	Support/background variance ratio	Pairwise support PCC	Pairwise support MAE
EMD-22246/PDB-*6xlk*	3.30	0.828191 ± 0.000170	78.851	26.272	0.993931	0.006370
EMD-32562/PDB-*7wk8*	3.61	0.828770 ± 0.000263	86.200	31.650	0.994711	0.007226
EMD-8480/PDB-*5u0s*	7.80	0.793743 ± 0.000156	92.469	42.848	0.994373	0.009549

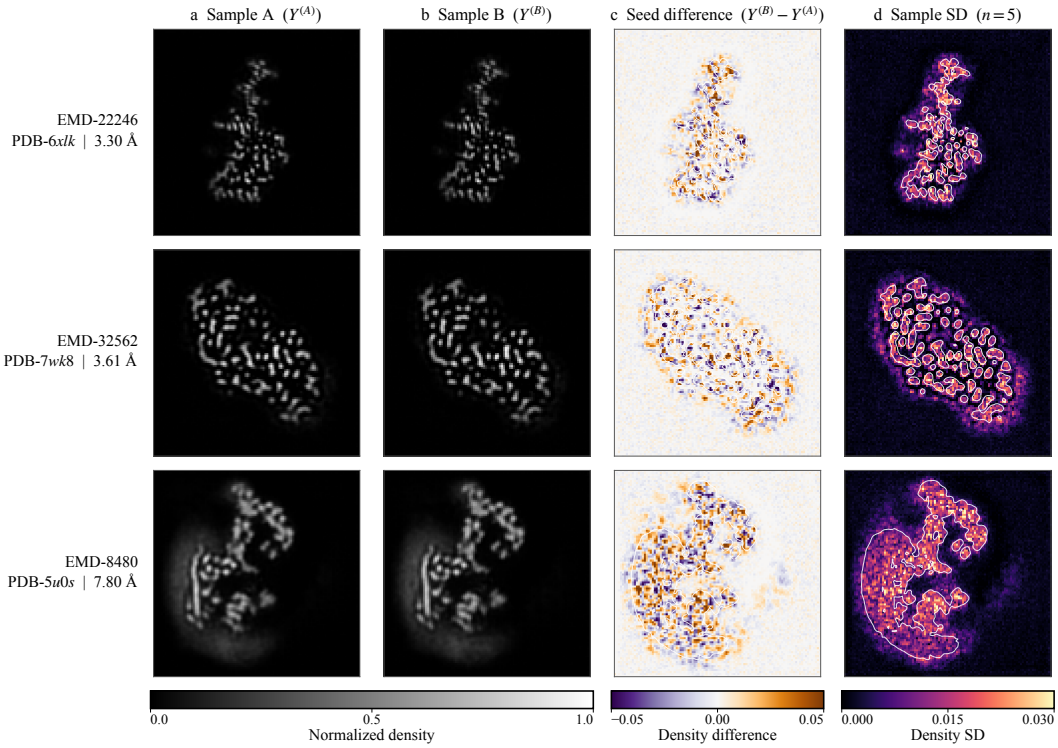

Figure 18: Conditional density variation from independent posterior-noise draws. Each row shows one structure from Table 22. Samples A and B are the retained pair with the largest full-volume MAE among five draws and share a fixed density scale. The signed difference and voxelwise standard deviation use shared scales across all rows; the white contour marks the five-sample mean density at 0.1.

The variation follows molecular features and becomes most pronounced in the 7.80 Å case. Across all three structures, most between-sample variation energy lies in molecular support, where the variance density is substantially higher than in the background. The high pairwise support PCC and nonzero support MAE show that independent draws preserve the global fold while changing local density details.

J.5 Spatial and Spectral Residual Analysis

Definition and sign. We analyze the paired correction $C = X_0 - Y_r$, a registered physical-to-experimental difference field analogous to map-model difference analyses (Joseph et al., 2020). Physically rendered map boundary supervision uses $Y_r - X_0 = -C$ as the BBDM residual relative to the experimental reference map. The following diagnostics report C, whose sign directly denotes density added to or removed from the physically rendered map.

Diagnostic definitions. We form the molecular mask from $Y_r > 10^{-5} \max(Y_r)$, apply $5 \times 5 \times 5$ maximum and uniform filters, normalize the resulting soft mask, and threshold it at 0.1. After centering C within this mask, lag-one correlation is the mean of the Pearson correlations

between valid adjacent voxel pairs along the three Cartesian axes; molecular-region energy is $\sum_{\mathbf{x}\in M} C(\mathbf{x})^2 / \sum_{\mathbf{x}} C(\mathbf{x})^2$. The local-operator diagnostics are the within-mask correlations of $|C|$ with $|\nabla Y_r|$ and of C with $\nabla^2 Y_r$; excess kurtosis is computed from the centered within-mask values. For the frequency diagnostics, both maps are multiplied by the soft mask and transformed with a 3D real FFT. We fit a real isotropic transfer in radial shells of width equal to the smallest reciprocal box length from cross-power divided by source power, clip it to $[-3, 3]$, and report the resulting reduction in Fourier residual energy; the spectral centroid is radial frequency weighted by residual power. The phase control uses fixed per-case random seeds to generate 16 surrogates that preserve $|\mathcal{F}(C)|$ and the DC component while randomizing phase. Its periodic lag-one statistic averages circular-shift correlations over the three axes, and each observed diagnostic is paired with the within-case mean over the 16 surrogates before cohort aggregation.

Full-test spatial diagnostics. Table 23 summarizes all 1,923 test structures. The joint pattern is spatially correlated, concentrated around the molecule, aligned with source gradients and curvature, and heavy tailed. A radial-transfer fit measures the part organized by spatial frequency.

Table 23: Median correction-field diagnostics over 1,923 test structures.

Diagnostic	Median	Interpretation
Lag-one spatial correlation	0.635	Positive local dependence
Lag-one correlation after within-mask permutation	0.00004	Same amplitudes, shuffled locations
Energy in molecular region	0.845	Medians: volume .159; enrichment 4.69×
$\|C\|$-$\|\nabla Y_r\|$ correlation	0.490	Boundary and local-detail alignment
C-$\nabla^2 Y_r$ correlation	0.585	Blur/sharpen-like signed structure
Excess kurtosis	7.10	Heavy-tailed amplitude distribution
Isotropic radial-transfer energy explained	0.531	Radial frequency component

Resolution changes the characteristic scale of the correction. Median lag-one correlation is 0.575, 0.707, and 0.897 in the $r < 3.5$, $3.5 \le r \le 6$, and $r > 6$ Å strata, respectively; the corresponding spectral centroids are 0.235, 0.168, and 0.073 Å^{-1}. Coarser requested resolution therefore produces smoother, longer-range corrections, consistent with resolution-dependent, spatially varying frequency transfer (Urzhumtsev et al., 2022).

Spectrum-preserving phase control. We test molecular registration directly with a spectrum-preserving phase control. For a fixed resolution-balanced cohort of 96 structures, we generate 16 surrogates per case by randomizing Fourier phase while retaining the complete anisotropic 3D Fourier amplitude and DC component. The construction preserves the power spectrum and circular autocorrelation while randomizing spatial registration with Y_r. The worst relative Fourier-amplitude error across all surrogates is 7.1×10^{-16}.

Table 24: Observed corrections versus spectrum-preserving phase-randomized surrogates.

Diagnostic	Observed	Surrogate	Paired median difference	Observed > surrogate
$\|C\|$-$\|\nabla Y_r\|$ correlation	**0.4611**	-0.0002	0.4625	97.9%
Molecular-region energy fraction	**0.8101**	0.1570	0.6222	99.0%
C-$\nabla^2 Y_r$ correlation	**0.5512**	-0.0001	0.5511	100.0%
Periodic lag-one correlation	**0.7921**	**0.7921**	1.1×10^{-16}	57.3%

The first three paired differences have two-sided Wilcoxon $p \le 2.1 \times 10^{-17}$; periodic lag-one correlation is preserved to numerical precision. Under this control, preserved autocorrelation together with lost molecular alignment shows that registration to the physically rendered map depends on Fourier phase. Figure 19 summarizes this result: C has a colored spectrum, and its phase aligns the field with molecular boundaries and local curvature. Panel (c) orders the observed-minus-surrogate differences by requested resolution for the three source-alignment diagnostics; Table 24 reports the spectrum-preserving lag-one check.

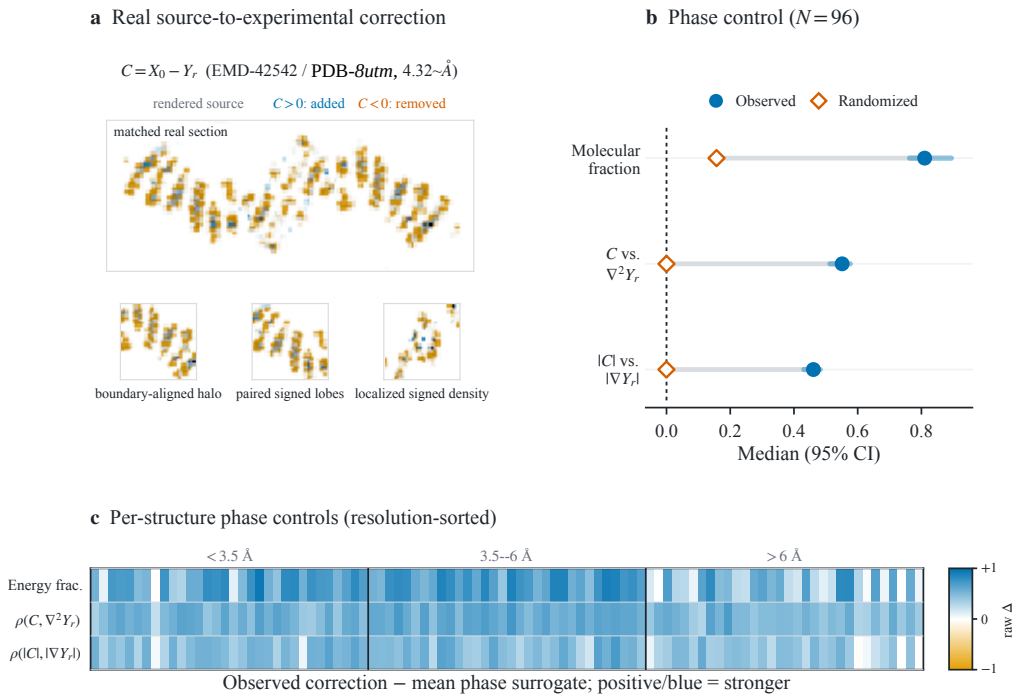

Figure 19: Real-case morphology and phase-randomized controls for the physical-to-experimental correction. (a) A section from EMD-42542/PDB-*8utm* overlays the signed correction C on the physically rendered map Y_r; the three crops show observed boundary-aligned, paired-sign, and localized morphologies. These signatures are consistent with the first-order organization in Equation 64; attribution to individual physical terms remains unresolved. (b) Cohort medians and paired-bootstrap intervals show that phase randomization removes molecular-support, gradient, and Laplacian alignment. (c) Raw per-structure observed-minus-surrogate differences, sorted by requested resolution, show the same pattern across the cohort for the three source-alignment diagnostics.

Why a structured correction appears visually fine-grained. Subtracting two maps with the same molecular envelope cancels much of their shared low-frequency content and emphasizes small differences in scale, position, and frequency response. For example, a small change in Gaussian blur gives

$$G_{\sigma^2+\Delta\sigma^2} * Y_r - G_{\sigma^2} * Y_r \approx \frac{\Delta\sigma^2}{2}\nabla^2(G_{\sigma^2} * Y_r), \qquad Y_r(\mathbf{x}-\Delta\mathbf{r}) - Y_r(\mathbf{x}) \approx -\Delta\mathbf{r}\cdot\nabla Y_r(\mathbf{x}). \quad (63)$$

The first term produces signed halos or ringing, and the second produces adjacent positive and negative lobes. Missing occupancy, ligands, solvent, or alternative conformations produce localized, often one-sided density. Removal of the shared coarse signal makes each of these structured effects appear visually fine grained.

A composite effective-physics model. Let O_{model} denote the object represented by the deposited atomic coordinates, ΔO unmodeled or differently occupied structure, H_G the global Gaussian source operator, and H_{eff} the aggregate experimental reconstruction and post-processing operator. A useful first-order organization is

$$C = X_0 - Y_r \approx (H_{\text{eff}} - H_G)O_{\text{model}} + H_{\text{eff}}\Delta O + n_{\text{recon}} + C_{\text{prep}}. \quad (64)$$

Here $H_{\text{eff}} - H_G$ collects effective-transfer differences associated with electron-scattering representation, local or anisotropic resolution, sharpening, filtering, masking, and reconstruction regularization (Bharadwaj & Jakobi, 2022; Urzhumtsev et al., 2022). The ΔO term covers occupancy and

atomic-displacement effects, unresolved flexibility, conformational or compositional heterogeneity, hydration, solvent, lipids, and unmodeled ligands. We use n_{recon} as a bookkeeping term for realization-dependent reconstruction variation from motion, dose and radiation damage, alignment uncertainty, detector response, and finite-particle estimation. Finally, C_{prep} records normalization and clipping of the paired deposited map and source.

At the particle-image level, the microscope CTF enters through filtering that varies with defocus, astigmatism, particle position, and higher-order aberrations; modern refinement estimates several of these quantities per particle (Zivanov et al., 2018). Particle weighting, regularization, and post-processing then combine them into the effective transfer of the deposited 3D map. The per-case isotropic transfer fit captures about half of the correction energy, showing that radial transfer is a substantial component. Molecularly localized and phase-aligned structure accounts for additional variation summarized jointly by Equation 64. The measured C forms a colored, source-registered composite correction. Spatially independent white noise would have lag-one correlation near zero and molecular-region energy near the region's volume fraction; the observed medians show strong departures from both signatures.

Independent half maps provide a route to a finer statistical decomposition. For half maps X_1 and X_2, define

$$X_{\mathrm{avg}} = \frac{X_1 + X_2}{2}, \qquad N_{\mathrm{half}} = \frac{X_1 - X_2}{2}, \qquad C_{\mathrm{shared}} = X_{\mathrm{avg}} - Y_r. \tag{65}$$

Shell-wise comparison of the FiLM-conditioned BBDM correction with C_{shared} and N_{half} can use the established signal/noise role of independent half maps to estimate repeatable experimental-domain structure and realization-dependent half-map variation (Read et al., 2023). The present paired-map diagnostics characterize their aggregate through source alignment, spatial correlation, and frequency transfer; half-map data would estimate the two contributions more directly.

K Database Construction and Access

K.1 PDB Selection and Generation Protocol

We queried the RCSB PDB current-holdings endpoint (RCSB Protein Data Bank, 2026a) on September 17, 2026 and obtained 260,089 released entries. We retained entries containing at least one finite, non-hydrogen atom with Cartesian coordinates in the first available atomic model; for alternate conformations, the primary or first available conformer was used. This criterion retained 260,044 atomic entries and removed 45 IHM sphere/bead entries without an `atom_site` Cartesian-coordinate table. We apply the frozen model independently to every retained entry at the 14 requested resolutions listed in Table 25, for 3,640,616 outputs in total. Each resolution is generated from the same deposited coordinates, so the collection separates controlled bandwidth variation from structural variation. The archive export retains the PDB identifier, requested resolution, voxel spacing, grid origin and dimensions, and preprocessing status required to trace every density back to its coordinate input.

K.2 Dataset Statistics and Metadata

Table 25: Scale and resolution distribution of the generated density map collection built from the RCSB PDB snapshot dated September 17, 2026.

Collection summary	Value
RCSB holdings entries	260,089
Eligible atomic entries	260,044
Excluded non-atomic IHM entries	45
Density maps	3,640,616
Total uncompressed storage	29.542 TB

Resolution (Å)	Maps	Share (%)
1.5	260,044	7.14
2.0	260,044	7.14
2.5	260,044	7.14
3.0	260,044	7.14
3.5	260,044	7.14
4.0	260,044	7.14
4.5	260,044	7.14
5.0	260,044	7.14
5.5	260,044	7.14
6.0	260,044	7.14
7.0	260,044	7.14
8.0	260,044	7.14
9.0	260,044	7.14
10.0	260,044	7.14
Total	3,640,616	100.00

K.3 Web Interface and Downloads

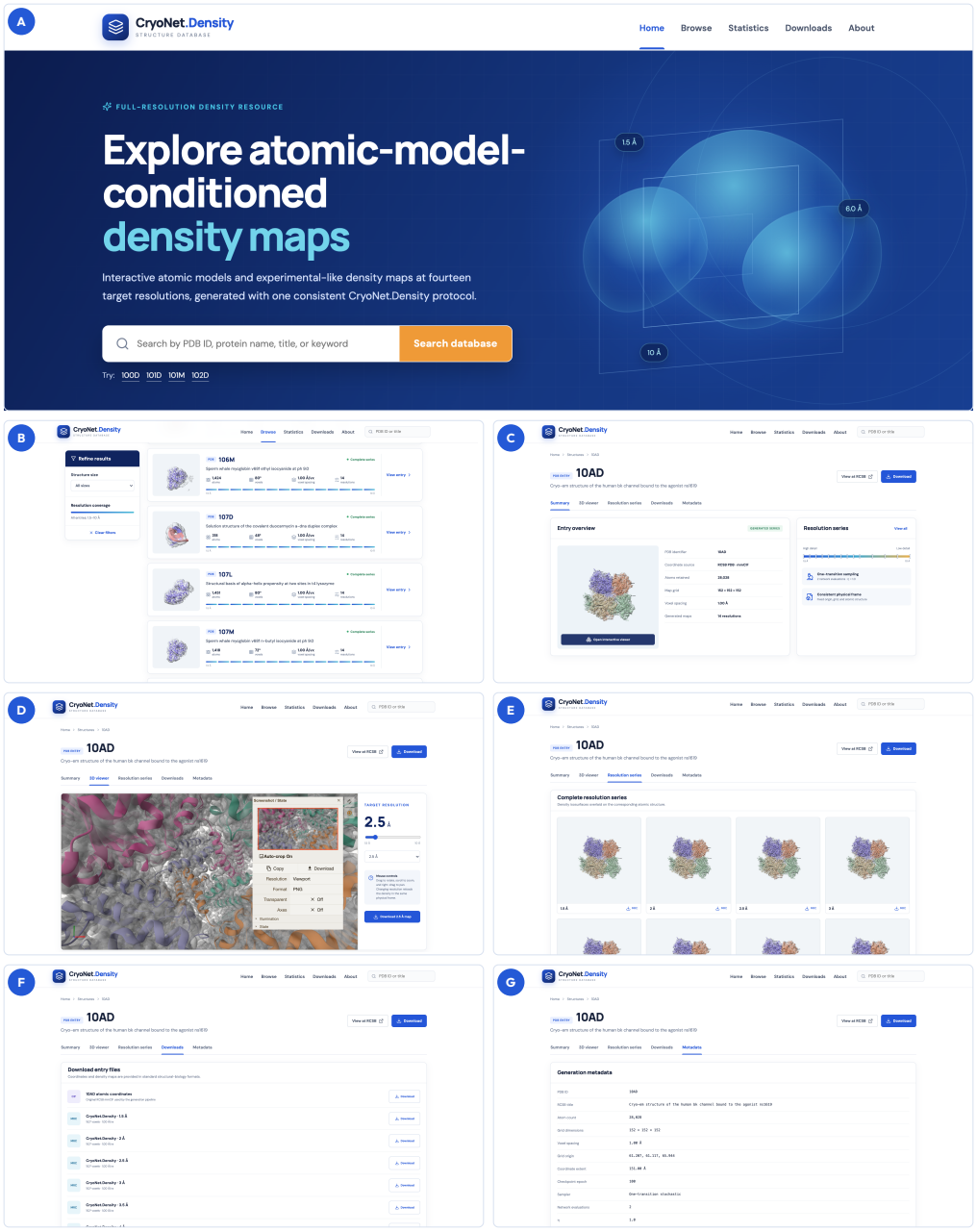

Figure 20: CryoNet.Density web interface. (A) Home page; (B) searchable collection browser; (C–G) summary, interactive 3D viewer, resolution series, downloads, and generation metadata for PDB-*10ad*.