

## APPENDIX

### A LOG-LIKELIHOOD FUNCTIONS

We define here the log-likelihood functions, as mentioned in Algorithm 1.

**Structure Completion.** For structure completion, we use the log-likelihood function defined in (10):

$$f(\mathbf{y}, \mathbf{z}) = -\left\| \begin{pmatrix} \mathbf{I}_m & 0 \\ 0 & 0 \end{pmatrix} \mathbf{V}^T \mathbf{z} - \mathbf{S}^+ \mathbf{U}^T \mathbf{y} \right\|_2^2. \quad (1)$$

**Model Refinement.** In the model refinement task, we combine three sources of information with the following three log-likelihood functions:

$$\begin{aligned} f_m(\bar{\mathbf{x}}, \mathbf{z}) &= -\left\| \begin{pmatrix} \mathbf{I}_m & 0 \\ 0 & 0 \end{pmatrix} \mathbf{V}^T \mathbf{z} - \mathbf{S}^+ \mathbf{U}^T \bar{\mathbf{x}} \right\|_2^2 && \text{(incomplete model)} \\ f_s(\mathbf{s}, \mathbf{z}) &= \log p_\theta(\mathbf{s} | \mathbf{R}\mathbf{z}, t = 0) && \text{(sequence)} \\ f_d(\mathbf{y}, \mathbf{z}) &= -\|\mathbf{y} - \mathcal{B}(\Gamma(\mathbf{R}\mathbf{z}, \mathbf{s}, \chi))\|_2^2, \text{ where } \chi \sim p_\theta(\chi | \mathbf{R}\mathbf{z}, \mathbf{s}, t = 0) && \text{(cryo-EM density)} \end{aligned} \quad (2)$$

$\Gamma$  is an operator turning an all-atom model into a density map with a user-defined resolution, following the `pdb2mrc` implementation of EMAN Ludtke et al. (1999). Given  $\mathbf{X} = (\mathbf{x}, \mathbf{s}, \chi) = [X_1, \dots, X_A]^T \in (\mathbb{R}^3)^A$ , the Cartesian coordinates of  $A$  heavy atoms,

$$\Gamma(\mathbf{X}) : x \in \mathbb{R}^3 \mapsto \sum_{i=1}^A Z_i \exp\left(-\frac{\|X_i - x\|_2^2}{2s^2}\right) \in \mathbb{R}, \quad (3)$$

where  $Z_i$  is the atomic number of atom  $i$  and  $s = \text{resolution}/(\sqrt{2\pi})$ . The norm  $\|\mathbf{y} - \Gamma(\mathbf{X})\|_2^2$  is computed directly in Fourier space using Parseval’s theorem, up to a time-dependent resolution  $r_t$ .

**Pairwise Distances to Structure.** We use the following log-likelihood function:

$$f(\mathbf{y}, \mathbf{z}) = -\|\mathbf{y} - \|\mathbf{D}\mathbf{R}\mathbf{z}\|_2\|_2^2. \quad (4)$$

### B OPTIMIZATION PARAMETERS

**Structure Completion.** All experiments are ran for 1,000 epochs, with a learning rate  $\lambda$  of 0.3 and a momentum  $\rho$  of 0.9. The time schedule is linear:

$$t = 1 - \text{epoch}/1000. \quad (5)$$

**Model Refinement.** All experiments are ran for 4,000 epochs. The learning rates are

$$\lambda_m = 0.1, \quad \lambda_s = \begin{cases} 0 & \text{if epoch} < 3,000 \\ 1 \times 10^{-5} & \text{otherwise} \end{cases}, \quad \lambda_d = 0.01, \quad (6)$$

and all the momenta are set to 0.9.  $r_t$  is set to 5 Å for the first 3,000 epochs and linearly decreases to 1.5 Å during the last 1,000 epochs. The side chain angles ( $\chi$ ) are sampled every 100 epochs. The time schedule is

$$t = 1 - \sqrt{\text{epoch}/4,000}. \quad (7)$$

**Pairwise Distances to Structure.** All experiments are ran for 1,000 epochs, with a learning rate  $\lambda = 200/n$ , where  $n$  is the number of known pairwise distances, and a momentum  $\rho$  of 0.99. The time schedule is

$$t = 1 - \sqrt{\text{epoch}/1,000}. \quad (8)$$

## C COMPUTE TIME

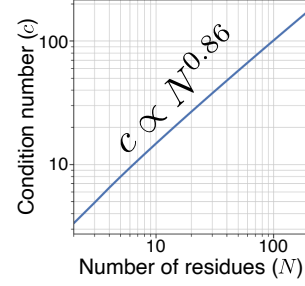
In each experiment, we run 8 replicas in parallel on a single NVIDIA A100 GPU. The typical compute time depends on the task, the number of residues in the protein, and the diffusion model. With Chroma, for which the compute time scales sub-quadratically with the number of residues (Ingraham et al., 2023), the typical compute time for protein with 100 residues is 2 minutes for the structure completion task, 10 minutes for the “distances to structure” task and 30 minutes for the model refinement task.

## D CORRELATED DIFFUSION

We use the “ $R_g$ -confined globular polymer” correlation matrix, as defined in (Ingraham et al., 2023):

$$\mathbf{R} = a \begin{bmatrix} \nu b^0 & & & & & \\ \nu b^1 & b^0 & & & & \\ \nu b^2 & b^1 & b^0 & & & \\ \vdots & & & \ddots & \ddots & \\ \nu b^{N-2} & & & & b^1 & b^0 \\ \nu b^{N-1} & & \dots & & b^2 & b^1 & b^0 \end{bmatrix}, \quad (9)$$

where  $a$  is a free parameter,  $\nu = \frac{1}{\sqrt{1-b^2}}$  and  $b$  is chosen such that the radius of gyration  $R_g$  scales with the number of residues as  $R_g \sim rN^\alpha$  ( $r \approx 2.0 \text{ \AA}$ ,  $\alpha \approx 0.4$  (Tanner, 2016; Hong & Lei, 2009)). The condition number of  $\mathbf{R}$  is shown in Figure S1 as a function of the number of residues.



**Figure S1:** Condition number of  $\mathbf{R}$  vs. number of residues  $N$ .

## E OTHER DIFFUSION MODELS

ADP-3D can be extended to leverage diffusion models that do not operate on Cartesian coordinates. If the diffusion model operates in a Riemannian manifold (like  $SE(3)$ ), the gradient step on the log-likelihood needs to be generalized using the exponential map ( $\hat{\mathbf{z}}_0 = \exp_{\mathbf{z}_0}(\sum_i \mathbf{v}_i)$ ).

For diffusion models that do not use a correlated diffusion process, the correlation matrix  $\mathbf{R}$  and the preconditioning strategy described in Section 4.2 can be ignored.

## F TARGET STRUCTURES

Information on the proteins used as target structures are given in Table S1.

## G ADDITIONAL RESULTS

### G.1 STRUCTURE COMPLETION

We provide additional results in Table S2. The Evidence Lower Bound (ELBO) is computed with Chroma and is a lower bound of the learned log-prior. Note that, as the subsampling factor increases (i.e., as the problem becomes less constrained), the gap between our approach (MAP estimation) and the baseline (low-temperature posterior sampling) decreases. For highly undersampled structures, the RMSD with the target structure is high but the generated structure remains plausible under the learned prior distribution.

### G.2 EXPERIMENTS WITH RFDIFFUSION

We provide additional results using RFDiffusion (Watson et al., 2023) as a prior for the “structure completion” and “distances to structure” tasks (Table S3). RFDiffusion describes backbone structures as collections of frames in  $\mathbb{R}^3 \times SO(3)$ . We show that using a prior consistently leads to more accurate structures than just performing gradient descent on the neg-log-likelihood.

**Table S1:** Information on target structures. We indicate the number of modeled residues (residues included in the atomic model provided in the PDB file) and the number of deposited residues (number of residues in the sequence). We selected all single-chain proteins from the CASP15 database. TBP = to be published.

PDB	Modeled	Deposited	Resolution	Release date	∈ CASP15	Reference
7r5b	127	127	1.77 Å	2023-02-08	No	Warstat et al. (2023)
7qum	130	130	1.50 Å	2023-03-01	No	Davison et al. (2022)
7pzt	160	160	1.84 Å	2022-11-02	Yes	TBP
8em5	92	103	1.95 Å	2023-09-27	Yes	Cuthbert et al. (2024)
8ok3	121	125	1.50 Å	2023-10-25	Yes	Caulton et al. (2024)
7roa	117	136	1.82 Å	2022-10-12	Yes	Cruz et al. (2022)
8ork	460	460	1.64 Å	2023-12-06	Yes	De Rose et al. (2023)
8c6z	573	630	1.85 Å	2023-05-31	Yes	Dierick et al. (2024)
8dys	409	603	1.80 Å	2022-08-17	Yes	TBP
7uww	632	635	1.61 Å	2023-06-14	Yes	Photenhauer et al. (2024)
8h2n	709	821	4.41 Å	2023-10-11	Yes	Chaban et al. (2024)
7zcx	1040	1424	3.10 Å	2023-06-14	Yes	Gambelli et al. (2024)
8sxa	1153	1325	3.30 Å	2024-01-17	Yes	Pourmal et al. (2024)
7xkw	499	561	3.10 Å	2022-04-20	No	Ye et al. (2024)
7wsm	464	520	3.25 Å	2022-01-30	No	Yuan et al. (2022)
7ug0	418	418	2.55 Å	2022-03-23	No	Huang et al. (2023)

**Table S2:** Additional structure completion results. We compare our method to Chroma (Ingraham et al., 2023) with a SubstructureConditioner. We report the RMSD in Å (best of 8) and the associated Evidence Lower Bound (ELBO) provided by Chroma as a proxy for realism.

PDB (# res.)	Metric	Method	Subsampling factor						
			2	4	8	16	32	64	128
7r5b (127)	RMSD (↓)	Chroma*	0.41	1.34	3.73	9.34	15.95	16.27	–
		<b>ADP-3D</b>	<b>0.10</b>	<b>0.47</b>	<b>2.21</b>	<b>4.47</b>	<b>9.45</b>	<b>11.54</b>	–
	ELBO (↑)	Chroma*	<b>7.79</b>	6.72	7.27	<b>7.56</b>	7.40	<b>9.10</b>	–
		<b>ADP-3D</b>	7.38	<b>7.87</b>	<b>7.48</b>	6.65	<b>7.64</b>	7.98	–
7qum (130)	RMSD (↓)	Chroma*	0.28	0.81	2.14	3.46	7.70	11.49	14.45
		<b>ADP-3D</b>	<b>0.21</b>	<b>0.27</b>	<b>1.13</b>	<b>2.68</b>	<b>5.60</b>	<b>7.76</b>	<b>13.01</b>
	ELBO (↑)	Chroma*	<b>5.87</b>	6.08	7.16	7.13	<b>8.43</b>	8.89	<b>9.27</b>
		<b>ADP-3D</b>	5.64	<b>6.47</b>	<b>7.87</b>	<b>7.56</b>	7.99	<b>8.76</b>	8.98
7pzt (160)	RMSD (↓)	Chroma*	0.50	1.61	4.82	10.36	16.15	16.81	16.93
		<b>ADP-3D</b>	<b>0.23</b>	<b>0.67</b>	<b>2.64</b>	<b>9.39</b>	<b>14.26</b>	<b>15.47</b>	<b>14.07</b>
	ELBO (↑)	Chroma*	<b>5.99</b>	5.55	5.39	7.47	<b>8.52</b>	<b>8.74</b>	<b>9.04</b>
		<b>ADP-3D</b>	5.79	<b>6.44</b>	<b>6.92</b>	<b>7.79</b>	8.29	7.81	8.64

### G.3 GRADIENT DESCENT FOR LINEAR CONSTRAINTS

In Figure S2, we compare different gradient-based techniques for minimizing over  $\mathbf{z}$  the following objective function:

$$\mathcal{L}(\mathbf{z}) = \|\mathbf{M}\mathbf{x}^* - \mathbf{M}\mathbf{R}\mathbf{z}\|_2^2. \quad (10)$$

$\mathbf{x}^*$  is the backbone structure of the ATAD2 protein (PDB: 7qum).  $\mathbf{M}$  is a masking matrix with a subsampling factor of 2. The preconditioning strategy is described in Section 4.2. Gradient descent with preconditioning and momentum leads to the fastest convergence.

### G.4 COMPARISON TO DPS

We adapted the Diffusion Posterior Sampling (DPS) algorithm (Chung et al., 2022) to tackle the structure completion problem. The central idea of DPS is to combine a reverse diffusion step with

**Table S3:** Results for the “structure completion” and “distances to structure” tasks on additional targets with two different priors (Chroma and RFdiffusion). We report the lowest RMSD over 8 runs and the number of outputs with less than 1.5 Å RMSD between parenthesis. We compare our method to gradient descent on the negative log-likelihood (“No prior”).

PDB	Prior	Fraction of given coordinates			Fraction of given distances	
		1/2	1/8	1/32	1/8	1/16
8em5	No prior	6.46 (0)	10.23 (0)	15.79 (0)	11.46 (0)	14.60 (0)
	RFdiffusion	1.20 (8)	3.50 (0)	16.45 (0)	1.65 (0)	4.75 (0)
	Chroma	<b>0.09</b> (8)	<b>1.57</b> (0)	<b>14.26</b> (0)	<b>1.29</b> (3)	<b>1.86</b> (0)
8ok3	No prior	6.17 (0)	8.98 (0)	19.67 (0)	13.11 (0)	14.66 (0)
	RFdiffusion	1.16 (8)	3.73 (0)	18.48 (0)	3.87 (0)	9.62 (0)
	Chroma	<b>0.10</b> (6)	<b>1.58</b> (0)	<b>11.43</b> (0)	<b>1.10</b> (5)	<b>1.19</b> (3)
7roa	No prior	6.69 (0)	8.59 (0)	15.16 (0)	13.75 (0)	14.59 (0)
	RFdiffusion	0.62 (8)	4.01 (0)	15.04 (0)	2.40 (0)	3.16 (0)
	Chroma	<b>0.13</b> (8)	<b>2.89</b> (0)	<b>11.57</b> (0)	<b>1.08</b> (5)	<b>1.44</b> (4)
8ork	No prior	6.39 (0)	9.70 (0)	<b>15.12</b> (0)	23.68 (0)	22.41 (0)
	RFdiffusion	0.98 (8)	3.47 (0)	16.67 (0)	<b>0.87</b> (2)	2.29 (0)
	Chroma	<b>0.17</b> (6)	<b>1.76</b> (0)	15.94 (0)	1.30 (4)	<b>1.34</b> (3)
8c6z	No prior	6.32 (0)	10.20 (0)	17.00 (0)	23.11 (0)	24.13 (0)
	RFdiffusion	1.16 (8)	4.13 (0)	17.78 (0)	3.69 (0)	4.12 (0)
	Chroma	<b>0.16</b> (8)	<b>2.40</b> (0)	<b>15.22</b> (0)	<b>1.42</b> (4)	<b>1.46</b> (2)
8dys	No prior	6.08 (0)	9.37 (0)	<b>14.74</b> (0)	18.72 (0)	19.25 (0)
	RFdiffusion	1.23 (8)	4.13 (0)	17.78 (0)	<b>1.61</b> (0)	2.98 (0)
	Chroma	<b>0.22</b> (3)	<b>2.42</b> (0)	18.74 (0)	1.65 (0)	<b>1.63</b> (0)
7uww	No prior	6.41 (0)	9.55 (0)	18.14 (0)	25.31 (0)	26.78 (0)
	RFdiffusion	1.30 (8)	4.55 (0)	21.24 (0)	1.90 (0)	3.13 (0)
	Chroma	<b>0.15</b> (8)	<b>2.67</b> (0)	<b>16.12</b> (0)	<b>1.48</b> (1)	<b>1.47</b> (1)
8h2n	No prior	6.35 (0)	9.37 (0)	16.57 (0)	26.25 (0)	26.24 (0)
	RFdiffusion	1.04 (8)	4.44 (0)	20.30 (0)	4.43 (0)	5.10 (0)
	Chroma	<b>0.17</b> (8)	<b>2.22</b> (0)	<b>15.59</b> (0)	<b>1.49</b> (1)	<b>1.52</b> (0)
7zcx	No prior	6.35 (0)	9.43 (0)	18.14 (0)	41.77 (0)	40.22 (0)
	RFdiffusion	1.31 (8)	4.36 (0)	22.95 (0)	6.09 (0)	5.81 (0)
	Chroma	<b>0.21</b> (4)	<b>2.36</b> (0)	<b>17.33</b> (0)	<b>1.99</b> (0)	<b>1.74</b> (0)
8sxa	No prior	6.12 (0)	9.19 (0)	14.32 (0)	34.06 (0)	34.43 (0)
	RFdiffusion	0.95 (8)	4.06 (0)	15.70 (0)	1.87 (0)	2.64 (0)
	Chroma	<b>0.19</b> (8)	<b>2.02</b> (0)	<b>10.78</b> (0)	<b>1.65</b> (0)	<b>1.86</b> (0)

a gradient-ascent step on the log-likelihood. Instead of computing the gradient from the iterate  $\mathbf{x}_t$  (w.r.t.  $\mathbf{x}_t$ ), DPS suggests to denoise the current estimate ( $\hat{\mathbf{x}}_0 = \hat{\mathbf{x}}_\theta(\mathbf{x}_t)$ ) and compute the gradient from the iterate  $\hat{\mathbf{x}}_0$  (w.r.t.  $\mathbf{x}_t$ ). The magnitude of the gradient step is controlled by a parameter named  $\zeta'$ .

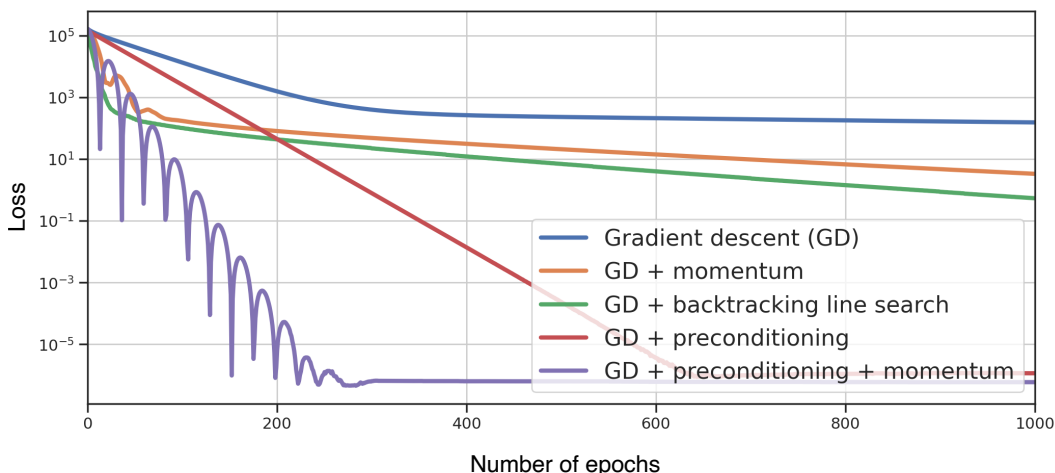
We compare ADP-3D to the adaptation of DPS for the structure completion task on PDB:8ok3 with a subsampling factor of 4. We compare the distribution of final RMSDs over 64 replicas and the total runtime. We perform a sweep over  $\zeta'$  for DPS. Both methods are ran over 1,000 steps. As shown in Figure S3, ADP-3D leads to more accurate reconstructions. Moreover, since ADP-3D does not need to compute gradients through the denoiser, it is approximately six times faster than DPS, per iteration.

## G.5 ABLATION STUDY FOR MODEL REFINEMENT

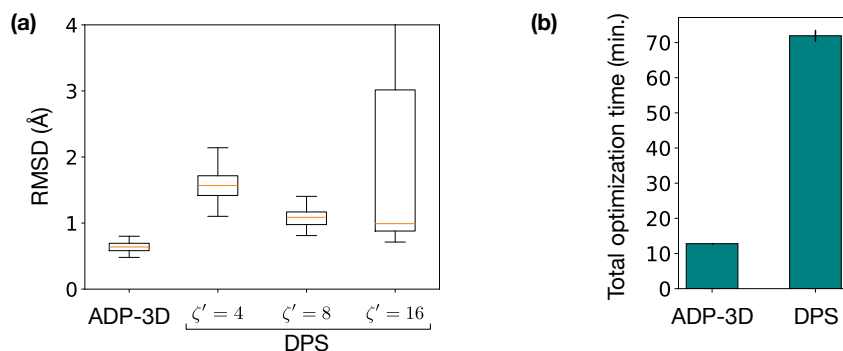
In Figure S4, we analyze the importance of the different input measurements for atomic model refinement. Removing the partial atomic model leads to the largest drop in accuracy. The cryo-EM density map is the second most important measurement, followed by the generative prior and the sequence.

## G.6 INFLUENCE OF RESOLUTION ON MODEL REFINEMENT

We use ChimeraX (Meng et al., 2023; Tang et al., 2007) to simulate the 3D density maps of the TecA bacterial toxin (PDB:7pzt), at different resolutions. We run ModelAngelo (Jamali et al., 2024) on the simulated maps, using the known sequence and the default parameters. We show our results in Figure S5. For both input resolutions (2.0 Å and 4.0 Å), our method improves on ModelAngelo’s



**Figure S2: Gradient Descent for Linear Constraint.** Convergence speed of different optimization techniques on a linear inverse problem (structure completion with a subsampling factor of 2), without the diffusion prior. Using preconditioning with momentum leads to the fastest convergence. The “loss” corresponds to the sum of squared distances between unmasked atom coordinates.



**Figure S3: Comparison to DPS.** We compare ADP-3D to DPS (Chung et al., 2022) for the structure completion task on PDB: 8ok3, with a subsampling factor of 4. (a) Distribution of final RMSDs over 64 replicas. For DPS, we perform a sweep over the parameter  $\zeta'$ , controlling the magnitude of the gradient step. (b) Comparison of runtime for the same experiment.

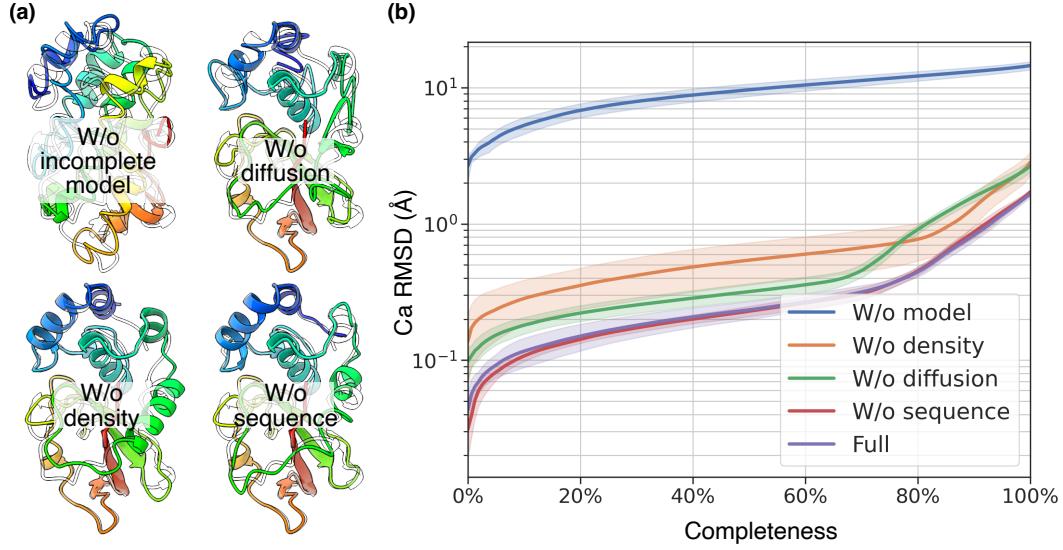
accuracy for a fixed completeness level (and reaches a higher completeness for the same accuracy). As the resolution of the map improves (gets smaller), the completeness of the output of ModelAngelo increases and the output of ADP-3D improves.

## G.7 VARIABILITY OF THE OUTPUT ON THE MODEL REFINEMENT TASK

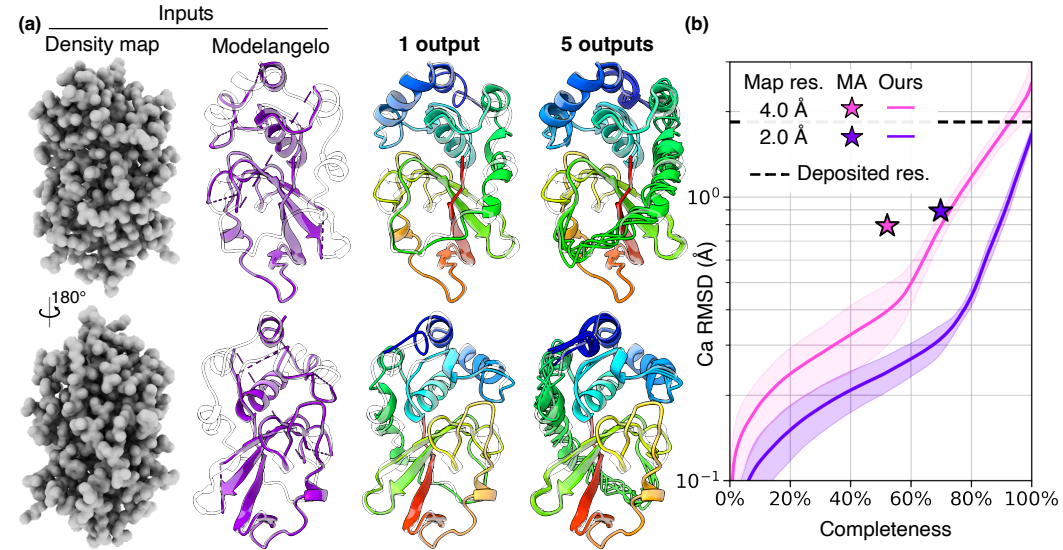
In Figure S6, we show the spread of final RMSDs obtained over 40 replicas of the same experiment (refinement of 7pzt with 2 Å density map). Some of the runs fail to reach sub-2Å RMSD, but we show that these outliers can be removed by measuring how well the structures fit into the input density map (using the log-likelihood function). This outlier removal process can therefore operate without access to ground truth information, making it straightforward to apply on new molecules.

## H ADDITIONAL VIDEO

We provide one additional video showing the predicted structure throughout the diffusion process, for the three tasks explored in this paper.



**Figure S4: Ablation Study.** For atomic model refinement, we combine three sources of conditioning information (incomplete model, density map and sequence) with the data-driven prior of the diffusion model. Here we highlight the importance of each conditioning information, and that of the generative prior. (a) Qualitative reconstructions with the target structure in transparency. (b) RMSD of alpha carbons vs. completeness. We use the same structure as in Fig. 4 (PDB: 7pzt), and a cryo-EM map at 2.0 Å resolution.

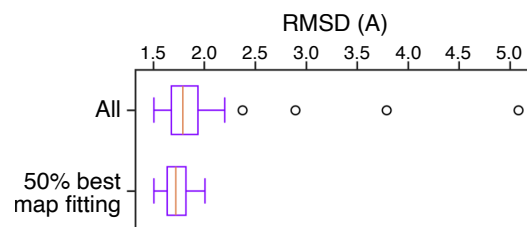


**Figure S5: Atomic Model Refinement.** Results on the TecA bacterial toxin (PDB: 7pzt, 160 residues). (a) Qualitative results. From left to right: the input density map at 2.0 Å resolution, the incomplete model given by ModelAngelo and our refined models (1 output and 5 outputs), overlaid on the target structure in transparency. (b) RMSD of alpha carbons vs. completeness (number of predicted residues / total number of residues) with ModelAngelo (MA) and our method. We run 5 experiments and report the mean of the lowest RMSD on  $\alpha$ -carbons over 8 replicas ( $\pm 1$  std). The spread of RMSD is further described in the supplements. The experimental (deposited) resolution is indicated with a dashed line.

## REFERENCES

Simon G Caulton, Carey Lambert, Jess Tyson, Paul Radford, Asmaa Al-Bayati, Samuel Greenwood, Emma J Banks, Callum Clark, Rob Till, Elisabete Pires, et al. Bdellovibrio bacteriovorus uses chimeric fibre proteins to recognize and invade a broad range of bacterial hosts. *Nature Microbiology*, 9(1):214–227, 2024.





**Figure S6:** Spread of final C $\alpha$ -RMSD for the model refinement task. By filtering out the structures that least fit the input density map, we can remove outliers.

Anastasiia Chaban, Leonid Minakhin, Ekaterina Goldobina, Brain Bae, Yue Hao, Sergei Borukhov, Leena Putzeys, Maarten Boon, Florian Kabinger, Rob Lavigne, et al. Tail-tape-fused virion and non-virion rna polymerases of a thermophilic virus with an extremely long tail. *Nature communications*, 15(1):317, 2024.

Hyungjin Chung, Jeongsol Kim, Michael T Mccann, Marc L Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. *arXiv preprint arXiv:2209.14687*, 2022.

Melissa R Cruz, Shane Cristy, Shantanu Guha, Giuseppe Buda De Cesare, Elena Evdokimova, Hiram Sanchez, Dominika Borek, Pedro Miramón, Junko Yano, Paul L Fidel Jr, et al. Structural and functional analysis of entv reveals a 12 amino acid fragment protective against fungal infections. *Nature communications*, 13(1):6047, 2022.

Bonnie J Cuthbert, Jessica Mendoza, Rodger de Miranda, Kadamba Papavinasasundaram, Christopher M Sasseti, and Celia W Goulding. The structure of mycobacterium thermoresistibile mmgs5 reveals a conserved disulfide bond across mycobacteria. *Metallomics*, 16(3):mfae011, 2024.

Gemma Davison, Mathew P Martin, Shannon Turberville, Selma Dormen, Richard Heath, Amy B Heptinstall, Marie Lawson, Duncan C Miller, Yi Min Ng, James N Sanderson, et al. Mapping ligand interactions of bromodomains brd4 and atad2 with fraglites and peplites-halogenated probes of druglike and peptide-like molecular interactions. *Journal of Medicinal Chemistry*, 65(22):15416–15432, 2022.

Simone A De Rose, Michail N Isupov, Harley L Worthy, Christina Stracke, Nicholas J Harmer, Bettina Siebers, Jennifer A Littlechild, HotSolute consortium, Bettina Siebers, Christopher Bräsen, et al. Structural characterization of a novel cyclic 2, 3-diphosphoglycerate synthetase involved in extremolyte production in the archaeon methanothermus fervidus. *Frontiers in Microbiology*, 14:1267570, 2023.

Evelien Dierick, Chana Callens, Yehudi Bloch, Savvas N Savvides, Sarah Hark, Stefan Pelzer, Richard Ducatelle, Filip Van Immerseel, and Evy Goossens. Clostridium perfringens chitinases, key enzymes during early stages of necrotic enteritis in broiler chickens. *PLoS pathogens*, 20(9):e1012560, 2024.

Lavinia Gambelli, Mathew McLaren, Rebecca Conners, Kelly Sanders, Matthew C Gaines, Lewis Clark, Vicki AM Gold, Daniel Kattinig, Mateusz Sikora, Cyril Hanus, et al. Structure of the two-component s-layer of the archaeon sulfolobus acidocaldarius. *Elife*, 13:e84617, 2024.

Liu Hong and Jinzhi Lei. Scaling law for the radius of gyration of proteins and its dependence on hydrophobicity. *Journal of Polymer Science Part B: Polymer Physics*, 47(2):207–214, 2009.

Yun Huang, Krishna D Reddy, Clay Bracken, Biao Qiu, Wenhui Zhan, David Eliezer, and Olga Boudker. Environmentally ultrasensitive fluorine probe to resolve protein conformational ensembles by 19f nmr and cryo-em. *Journal of the American Chemical Society*, 145(15):8583–8592, 2023.

John B Ingraham, Max Baranov, Zak Costello, Karl W Barber, Wujie Wang, Ahmed Ismail, Vincent Frappier, Dana M Lord, Christopher Ng-Thow-Hing, Erik R Van Vlack, et al. Illuminating protein space with a programmable generative model. *Nature*, 623(7989):1070–1078, 2023.

- Kiarash Jamali, Lukas Käll, Rui Zhang, Alan Brown, Dari Kimanius, and Sjors HW Scheres. Automated model building and protein identification in cryo-em maps. *Nature*, pp. 1–2, 2024.
- Steven J Ludtke, Philip R Baldwin, and Wah Chiu. Eman: semiautomated software for high-resolution single-particle reconstructions. *Journal of structural biology*, 128(1):82–97, 1999.
- Elaine C Meng, Thomas D Goddard, Eric F Pettersen, Greg S Couch, Zach J Pearson, John H Morris, and Thomas E Ferrin. Ucsf chimeraX: Tools for structure building and analysis. *Protein Science*, 32(11):e4792, 2023.
- Amanda L Photenhauer, Rosendo C Villafuerte-Vega, Filipe M Cerqueira, Krista M Armbruster, Filip Mareček, Tiantian Chen, Zdzisław Wawrzak, Jesse B Hopkins, Craig W Vander Kooi, Štefan Janeček, et al. The ruminococcus bromii amylosome protein sas6 binds single and double helical alpha-glucan structures in starch. *Nature structural & molecular biology*, 31(2):255–265, 2024.
- Sergei Pourmal, Evan Green, Ruchika Bajaj, Ilan E Chemmama, Giselle M Knudsen, Meghna Gupta, Andrej Sali, Yifan Cheng, Charles S Craik, Deanna L Kroetz, et al. Structural basis of prostaglandin efflux by mrp4. *Nature structural & molecular biology*, 31(4):621–632, 2024.
- 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.
- John J Tanner. Empirical power laws for the radii of gyration of protein oligomers. *Acta Crystallographica Section D: Structural Biology*, 72(10):1119–1129, 2016.
- Robin Warstat, Mehrosh Pervaiz, Pierre Regenass, Marius Amann, Karin Schmidtkunz, Oliver Einsle, Manfred Jung, Bernhard Breit, Martin Hügler, and Stefan Günther. A novel pan-selective bromodomain inhibitor for epigenetic drug design. *European Journal of Medicinal Chemistry*, 249:115139, 2023.
- Joseph L Watson, David Juergens, Nathaniel R Bennett, Brian L Trippe, Jason Yim, Helen E Eisenach, Woody Ahern, Andrew J Borst, Robert J Ragotte, Lukas F Milles, et al. De novo design of protein structure and function with rfdiffusion. *Nature*, 620(7976):1089–1100, 2023.
- Dan Ye, Yi-Zhen Shao, Wen-Rui Li, Zhen-Jia Cui, Ting Gong, Jin-Ling Yang, Hai-Qiang Wang, Jun-Gui Dai, Ke-Ping Feng, Ming Ma, et al. Characterization and engineering of two highly paralogous sesquiterpene synthases reveal a regioselective reprotonation switch. *Angewandte Chemie International Edition*, 63(13):e202315674, 2024.
- Yafei Yuan, Fang Kong, Hanwen Xu, Angqi Zhu, Nieng Yan, and Chuangye Yan. Cryo-em structure of human glucose transporter glut4. *Nature Communications*, 13(1):2671, 2022.