ATOMIC MODELING OF PROTEIN DYNAMICS FROM CRYO-EM WITH CO-FOLDING PRIORS

Anonymous authors
Paper under double-blind review

ABSTRACT

Understanding how proteins and other biomacromolecular machines function requires characterizing not only individual structures, but their full conformational landscapes. New protein foundation models provide powerful priors over atomic structure, and recent work has leveraged experimental measurements to guide the generation of individual structures at inference time. However, recovering atomic conformational landscapes from experimental data remains an open problem. We introduce CryoCHEF, a method for all-atom modeling of protein **C**onformational **H**eterogeneous **E**nsembles using co-**F**olding priors. CryoCHEF proposes a new, general framework for mapping learned conformational representations from experimental cryo-EM data into the latent space of a pretrained co-folding model. Specifically, CryoCHEF learns a conformation-conditioned neural field over the pair representations of a pretrained co-folding model, inducing a continuous atomic ensemble conditioned on experimental data. Across synthetic benchmarks and experimental datasets, CryoCHEF recovers continuous structural ensembles with side-chain-level detail, captures known global and local conformational motions, and enables efficient generation of atomic trajectories across experimentally observed conformational landscapes.

1 INTRODUCTION

Proteins and other biomacromolecular machines function through conformational changes spanning local rearrangements to large-scale domain motions. Understanding these systems therefore requires characterizing not only individual structures, but the conformational landscapes they populate. Yet experimentally determining these landscapes at atomic resolution remains challenging.

Recent advances in protein structure prediction and generative modeling provide a powerful new route toward modeling structural ensembles. Protein co-folding models encode strong priors over atomic structure and can generate diverse conformations from sequence alone (Jumper et al., 2021; Abramson et al., 2024; Candido et al., 2026). However, the distributions learned by these models do not necessarily reflect the conformational landscapes populated by a protein under experimental conditions. Recent methods have incorporated experimental measurements from cryo-EM, X-ray crystallography, and NMR to guide structure generation at inference time, enabling the recovery of atomic structures consistent with individual experimental observations (Raghu et al., 2025; Fadini et al., 2026; Maddipatla et al., 2026; Li et al., 2026). However, these approaches primarily condition structure generation on individual experimental measurements, rather than learning an atomic representation of an experimentally observed conformational landscape.

Cryo-electron microscopy (cryo-EM) is particularly well suited for experimentally resolving conformational heterogeneity, as a single dataset contains images of many individual molecules spanning different conformational states. Recent approaches have combined individual cryo-EM density maps with protein structure models at inference time to provide atomic interpretations of this heterogeneity (Raghu et al., 2025; Li et al., 2026). However, applying inference-time guidance independently to individual maps does not readily scale to the continuous conformational ensembles recovered from heterogeneous cryo-EM datasets. In parallel, a growing number of cryo-EM heterogeneous reconstruction methods can recover latent representations of protein conformational variation from hundreds of thousands to millions of particle images (Zhong et al., 2021; Punjani & Fleet, 2021;

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.

Chen & Ludtke, 2021; Schwab et al., 2024; Li et al., 2024; Ducrocq et al., 2025), but lifting these experimentally learned conformational landscapes to atomic resolution remains an open problem.

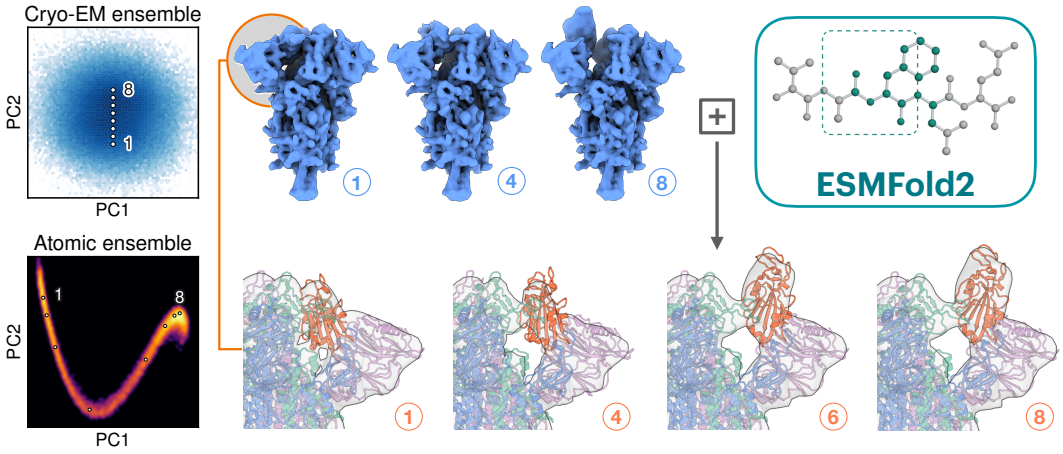

Figure 1: **CryoCHEF samples atomic models of the SARS-CoV-2 spike conditioned on an experimental cryo-EM dataset.** Leveraging ESMFold2 as a biophysical prior, CryoCHEF fits atomic ensembles to conformational landscapes reconstructed from cryo-EM. Top: PCA of cryoDRGN latent embeddings of 206k particles (density), shown alongside three decoded volumes. Points 1–8 are evenly spaced on a straight line along PC2, with the other components held at their mean. Bottom left: negative log probability density ($-\ln P$) over the first two PCs of Cα pairwise distances for 10,000 structures predicted from the latents of randomly chosen particles, with the eight traversal structures projected onto it (white). Bottom right: all-atom structures predicted by our model for each latent (chain-A receptor-binding domain in orange, other chains coloured by chain).

Here, we introduce **CryoCHEF**, a method for all-atom modeling of protein **C**onformational **H**eterogeneous **E**nsembles using co-**F**olding priors. CryoCHEF recovers atomic-resolution conformational landscapes from heterogeneous cryo-EM data using the structural priors of protein foundation models (Figure 1). Specifically, CryoCHEF learns a conformation-conditioned neural field over the pair representations of a pre-trained cofolding model to enable precise control over its output, while simultaneously inducing an experiment-conditioned continuous distribution over atomic structures. This amortizes atomic modeling across the conformational landscape rather than independently fitting or guiding structure prediction against individual density maps at inference time. Furthermore, by decoupling atomic modeling from heterogeneous reconstruction, CryoCHEF can operate on latent representations learned by different reconstruction methods without requiring a built-in structural model. In this work, we adapt ESMFold2 (Candido et al., 2026) to condition on cryoDRGN (Zhong et al., 2021) and DynaMight (Schwab et al., 2024) latent representations, enabling continuous generation of all-atom structures across experimentally observed conformational landscapes.

2 Methods

In cryo-EM, heterogeneous reconstruction methods learn a low-dimensional latent space in which each point z corresponds to one conformation, and a decoder that renders any z as a 3D density map. These volumes can encode large scale motion in the datasets but are generally low resolution and lack fine details. Protein co-folding models trained on the PDB produce detailed, physically plausible atomic structures, but often predict one static structure per sequence. CryoCHEF enables heterogeneity analysis at a atomic resolution by fine-tuning a protein foundation model to produce ensembles supported by experimental data.

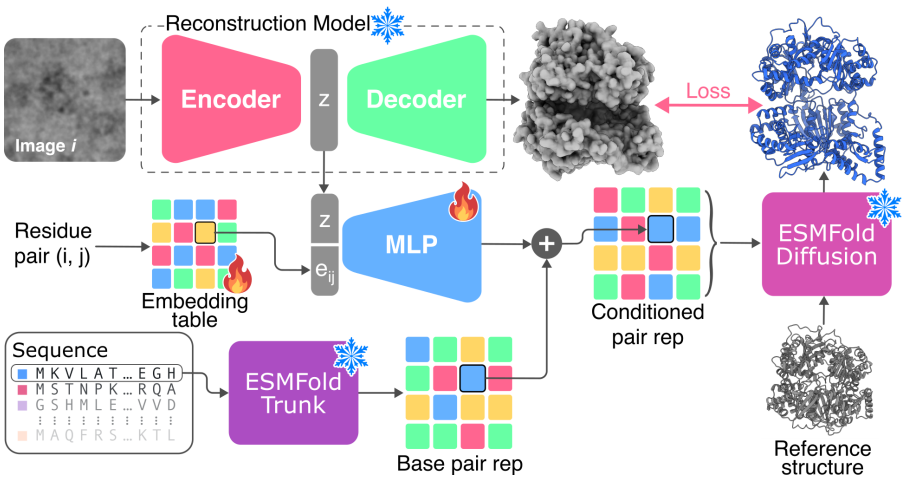

Figure 2: **CryoCHEF architecture**. We map the learned conformational manifold of a cryo-EM reconstruction method to the representation space of a pretrained co-folding model. For each residue pair (i, j), a learned MLP takes as input a learnable embedding e_{ij} and a conformation latent z, emitting a bias to the ESMFold2 pair state. This conditional neural field over pair states enables control of the structural conformations rendered by the ESMFold2 diffusion head.

2.1 Heterogeneous cryo-EM reconstruction

Given a dataset of 2D cryo-EM projection images $\{I_i\}_{i=1}^N$, modern heterogeneous reconstruction algorithms learn a continuous latent variable model over 3D electron scattering potential volumes, $\mathcal{V}_\theta(z) : \mathbb{R}^{d_z} \to \mathbb{R}^{D^3}$, where z represents a protein conformation. This function may be parameterized as a neural field (Zhong et al., 2021), deformation field of a canonical volume (Punjani & Fleet, 2023; Schwab et al., 2024), or a combination of eigenbases (Punjani & Fleet, 2021; Gilles & Singer, 2025), amongst others. Typically, the posterior $p_\theta(z|I_i)$ is parameterized as a second neural network and is learned jointly with $\mathcal{V}_\theta$ by variational inference. Once trained, the aggregate posterior $p_\theta(z)$ is a low-dimensional representation of the conformational manifold, while the decoded volumes capture the 3D protein states. Our atomic modeling method is designed to work downstream of any latent variable reconstruction model, assuming access only to a representative set of latents and their corresponding decoded volumes, $\{(z_i, V_i)\}_{i=1}^M$.

2.2 Steering ESMFold2

ESMFold2 predicts structure in two stages (section A.1). A folding trunk turns the sequence into a pair representation $\mathbf{c}$, with one feature vector for every pair of residues. A diffusion head then generates all-atom coordinates by iterative denoising, conditioned on $\mathbf{c}$. This can be approximately described as

$$\mathbf{c} = \text{FoldingTrunk}(\mathbf{s}), \qquad \mathbf{x} = \text{DiffusionModule}(\mathbf{c}; \epsilon), \tag{1}$$

where $\mathbf{s}$ is input features into ESMFold2, ϵ denotes the sampling noise, and $\mathbf{x}$ is the output coordinates of ESMFold2.

We observe empirically that folding models are strongly conditioned by the pair representation and diffusion model is responsible for projecting the representation into coordinate space. By learning a model over only the pair representation, we enable control of a heterogeneous ensemble while the diffusion module provides a strong geometric prior. As shown in Figure 2, CryoCHEF learns a latent-dependent bias on the trunk's pair representation,

$$\mathbf{c}_z^{(ij)} = \mathbf{c}_{\text{base}}^{(ij)} + \hat{b}_\phi(\bar{z}, e_{ij}),$$

Here $\mathbf{c}_{\text{base}}$ is computed once from the sequence, $\bar{z}$ is a normalized latent embedding from the reconstruction model, e_{ij} is an embedding of residue pair (i, j), and $\hat{b}_\phi$ is a neural field evaluated

independently for every residue pair (section A.2). The only parameters learned by CryoCHEF are the $L \times L$ embedding table e and $\hat{b}_\phi$, which is parameterized by an MLP.

2.3 Reward Finetuning

ESMFold2 predicts a structure in atomic coordinates, which must be made consistent with a reference 3D volume generated by the heterogeneity analysis. We therefore train b_ϕ via reward fine-tuning. A differentiable forward model renders a structure into a density map, and the reward is its Fourier-space correlation with the target map V_i, restricted to frequencies below a cutoff (eq. (2)). A bond-geometry term keeps covalent structure physical.

Training this loss through backpropagation through a full reverse-diffusion rollout is intractable. Following PIDM (Bastek et al., 2025), we instead roll out the frozen sampler once and pick a few noise levels along the trajectory. At each, we evaluate the reward on the denoiser's one-step prediction of the clean structure, and backpropagate only through that single denoiser call into b_ϕ (section A.3).

A structure is rendered on the target map's grid and in its frame as the band-limited Fourier transform $\tilde{\Gamma}(\mathbf{x})$ of point scatterers at its heavy atoms, weighted by atomic number. The map reward is the normalized cross-correlation over Fourier components within a shell of radius $1/K$, with the DC component removed:

$$\mathcal{R}_{\text{map}}(\mathbf{x}; V, K) = \frac{\operatorname{Re} \sum_{0<|k|\le 1/K} \tilde{V}_k \, \tilde{\Gamma}_k(\mathbf{x})^*}{\sqrt{\sum_{0<|k|\le 1/K} |\tilde{V}_k|^2} \, \sqrt{\sum_{0<|k|\le 1/K} |\tilde{\Gamma}_k(\mathbf{x})|^2}}. \tag{2}$$

The resolution cutoff K follows a resolution march: one rung per epoch, linear in spatial frequency from 20 Å to a final resolution of 4 Å (protease, TRPV1), 10 Å (spike, integrin), or 15 Å (HER2). The total reward is

$$\mathcal{R}(\mathbf{x}) = \mathcal{R}_{\text{map}}(\mathbf{x}) + \alpha_{\text{bond}} \, \mathcal{R}_{\text{bond}}(\mathbf{x}) + \alpha_{\text{lddt}} \, \mathcal{R}_{\text{lddt}}(\mathbf{x}),$$

where $\mathcal{R}_{\text{bond}}$ is the negative AlphaFold2 bond-length violation loss (Jumper et al., 2021) ($\alpha_{\text{bond}} = 0.5$) and $\mathcal{R}_{\text{lddt}}$ is an optional negative smooth Cα-lDDT loss to the reference structure with distance thresholds $\{2.0, 4.0\}$ (Abramson et al., 2024) ($\alpha_{\text{lddt}} = 0.5$ for spike, 0 otherwise).

To train stably, we implement:

1. **Partial rollouts from a reference structure.** Rollouts start from a moderately noised reference structure rather than pure noise, as in SDEdit (Meng et al., 2022). This shortens the rollout and fixes the docking interfaces in multi-chain complexes.
2. **Fixed diffusion noise.** The noise is held fixed across examples, so reward differences between maps come from the latent alone.
3. **Resolution marching.** The frequency cutoff tightens over training, so the model first matches the coarse shape of each map and then finer detail.

The diffusion head generates structures in an arbitrary rotation. Before scoring, each predicted structure must be aligned to the target volume. We first superimpose the structure onto an output from a previous training step and refine the result with a short gradient descent optimization to maximize the map reward at the current resolution. The intermediate predictions that receive the rewards are superimposed into this pose (section A.4).

2.4 Inference

At inference, we can obtain stronger results by modulating the strength of the pair representation bias in a timestep-dependent manner. During the reverse rollouts, we can modulate the conditioned pair representation each timestep as

$$\mathbf{c}_z^{(t)} = \mathbf{c}_{\text{base}} + \beta^{(t)} \cdot b_\phi(z) \tag{3}$$

Akin to classifier-free guidance (Ho & Salimans, 2022), this allows us to tradeoff steering fidelity with geometric quality. We find that higher $\beta^{(t)}$ at earlier timesteps encourages strong conformational steering, while lower $\beta^{(t)}$ at later timesteps allows the model to sample local geometry from the pretrained prior instead of overfitting to the map (section A.5).

3 RESULTS

3.1 EXPERIMENTAL SETUP

We evaluate our method on two protein systems with synthetically generated cryo-EM images – presequence protease (Liang et al., 2022) and TRPV1 (Gao et al., 2016)), and three systems with real experimental cryo-EM images – SARS-CoV-2 spike (Walls et al., 2020), HER2-Fabs (Ruedas et al., 2024), and Integrin-Fabs (Campbell et al., 2020). We train CryoCHEF on density maps reconstructed from CryoDRGN (Zhong et al., 2021) and DynaMight (Schwab et al., 2024).

Baselines. For the synthetic datasets, we baseline three structure-guided cryo-EM reconstruction models – CryoSPHERE (Ducrocq et al., 2025), CryoSTAR (Li et al., 2024), and e2GMM (Chen & Ludtke, 2021), in addition to ESMFold2's native ensemble. We also train CryoCHEF on an "Oracle" map ensemble, where high-resolution maps are simulated directly from the ground truth structures, and the latents are the first 8 principal components of the structures' pairwise Cα distance matrix. This experiment decouples the accuracy of the upstream reconstruction model from the accuracy of our method.

Metrics. For the synthetic datasets, we predict 1,000 structures corresponding to k-means centers of each method's learned latent space, then compute structural and distributional accuracy against the ground truth structures used to generate the image dataset. We compute Cα-RMSD precision, Cα-RMSD recall, and "matched" Cα-RMSD, where a sample is scored against the ground truth structure which generated its image. This assesses both structural accuracy and the heterogeneous reconstruction method's ability to map a noisy image to the correct conformation. Metrics are computed after Kabsch alignment on Cα atoms. For distributional accuracy, we first compute the PCA of the Cα pairwise distances of the structures. We score the cosine similarity of the reference distribution PC1 against the sampled distribution PC1, and the accuracy of the PC1 variance, $\min(r, 1/r)$ where $r = \sigma^2_{\text{sampled}}/\sigma^2_{\text{ref}}$. We additionally score the Wasserstein distance between the first 8 principal components of the two distributions, with the PC transform fit on the reference distribution. Finally, we compute clashscore to measure geometric quality.

For the real datasets, we predict 100 structures corresponding to k-means centers of cryoDRGN's latent space, and compute Fourier shell correlation (FSC) curves against the corresponding decoded maps, and backprojections of the corresponding particle subsets.

3.2 SYNTHETIC DATASETS

Protease. We generate a synthetic dataset of 100k images from 2,000 structures obtained by MD simulation (Feathers et al., 2026). CryoCHEF captures the dominant opening and closing motion of the protein when fit on both CryoDRGN and DynaMight maps (Figure 3a). CryoCHEF trained on CryoDRGN volumes achieves the strongest structural accuracy metrics, with a precision of nearly 1Å, while CryoCHEF trained on DynaMight volumes achieves the strongest distributional accuracy (Table 1), capturing 99.7% of the variance along the first PC. CryoSPHERE and CryoSTAR also capture the dominant motion. While we observed the e2GMM volumes to capture the motion, the decoded structures do not show flexibility. The base ensemble from ESMFold2 shows flexibility in the right direction (96.3% PC1 cosine similarity) but not the full range of the motion (38.8% PC1 variance captured). In Supplemental Table 3, we show that an ensemble of 10k structures predicted from random latents improves distributional metrics compared to the 1,000 structure training ensemble.

TRPV1. We generate a synthetic dataset of 100k images from 10k structures obtained by MD simulation (Astore et al., 2025). In Figure 3b, we show that CryoCHEF fits the inward and outward movement of the cytosolic domains in the first principal component of the ensemble, while describing the domain lengthening in its second. CryoCHEF achieves the best structural accuracy metrics, though the dominant motion is only partly captured, e.g. 68.1% PC1 cosine similarity and 69.1% PC1 variance accuracy when trained on CryoDRGN volumes with $\beta = 1.0$ at inference time. Fitting to the Oracle reconstructions does not noticably improve metrics, indicating that the co-folding prior is likely restricting the learned conformational manifold. e2GMM only shows mild flexibility. CryoSTAR learns the correct motions, but local inaccuracies lead to lower RMSD. CryoSPHERE loss diverged under two settings of the number of segments. An inference-time conditioning strength

(section 2.4) linearly decreasing from $\beta^{t_{\text{start}}} = 1.0$ to $\beta^{t_{\text{end}}} = 0.1$ slightly improves structural and distributional metrics, while dramatically reducing clashscore.

Table 1: **Structural and distributional accuracy on synthetic datasets.** Values are mean ± std across samples. Arrows indicate whether higher (↑) or lower (↓) is better. Clashscore cannot be computed for CryoSPHERE and CryoSTAR as they produce Cα traces. Best value per column (excluding Oracle) is bolded. $\beta = [1.0, 0.1]$ is a linearly decreasing inference-time conditioning strength from $\beta^{t_{\text{start}}} = 1.0$ to $\beta^{t_{\text{end}}} = 0.1$.

Dataset	*Method*	*Precision Cα-RMSD* (↓)	*Recall Cα-RMSD* (↓)	*Matched Cα-RMSD* (↓)	*Clashscore* (↓)	*PC1 cos. sim.* (↑)	*PC1 var. cap.* (↑)	*All-PC $\mathcal{W}2$* (↓)
CAHRA	CryoCHEF + Oracle, $\beta = 1.0$	1.062 ± 0.098	1.131 ± 0.111	1.194 ± 0.153	16.664 ± 0.958	0.987	0.889	0.247
	CryoCHEF + Oracle, $\beta = [1.0, 0.1]$	1.516 ± 0.111	1.622 ± 0.137	1.726 ± 0.126	7.327 ± 0.536	0.975	0.808	0.380
	CryoCHEF + CryoDRGN, $\beta = 1.0$	$\mathbf{1.024 \pm 0.094}$	$\mathbf{1.200 \pm 0.187}$	$\mathbf{1.271 \pm 0.238}$	17.650 ± 0.696	**0.989**	0.823	0.537
	CryoCHEF + CryoDRGN, $\beta = [1.0, 0.1]$	1.471 ± 0.126	1.647 ± 0.196	1.709 ± 0.226	$\mathbf{7.666 \pm 0.297}$	0.974	0.767	0.642
	CryoCHEF + DynaMight, $\beta = 1.0$	1.100 ± 0.120	1.280 ± 0.201	1.454 ± 0.412	18.625 ± 1.378	0.983	0.932	**0.459**
	CryoCHEF + DynaMight, $\beta = [1.0, 0.1]$	1.516 ± 0.097	1.703 ± 0.173	1.851 ± 0.367	7.824 ± 0.444	0.973	**0.997**	0.543
	e2GMM	1.700 ± 0.030	4.020 ± 1.929	4.383 ± 1.960	86.301 ± 2.079	0.438	0.009	2.106
	CryoSPHERE	1.790 ± 0.148	1.939 ± 0.189	2.076 ± 0.247	–	0.970	0.703	0.605
	CryoSTAR	1.591 ± 0.128	1.657 ± 0.167	3.091 ± 1.266	–	0.977	0.684	0.743
	ESMFold2	2.113 ± 0.209	2.031 ± 0.213	–	22.048 ± 1.175	0.963	0.388	2.046
TRPV1	CryoCHEF + Oracle, $\beta = 1.0$	3.542 ± 0.212	4.179 ± 0.351	5.855 ± 1.285	71.618 ± 3.740	0.720	0.738	0.790
	CryoCHEF + Oracle, $\beta = [1.0, 0.1]$	3.302 ± 0.212	4.087 ± 0.391	5.591 ± 1.317	13.539 ± 1.290	0.755	0.626	0.800
	CryoCHEF + CryoDRGN, $\beta = 1.0$	3.623 ± 0.188	4.490 ± 0.384	5.738 ± 1.284	154.746 ± 7.113	0.681	0.691	**0.863**
	CryoCHEF + CryoDRGN, $\beta = [1.0, 0.1]$	3.272 ± 0.206	$\mathbf{4.316 \pm 0.431}$	$\mathbf{5.385 \pm 1.346}$	21.055 ± 1.285	**0.699**	0.519	0.868
	CryoCHEF + DynaMight, $\beta = 1.0$	3.586 ± 0.068	4.706 ± 0.394	5.754 ± 1.148	196.705 ± 3.894	0.601	0.597	0.951
	CryoCHEF + DynaMight, $\beta = [1.0, 0.1]$	$\mathbf{3.221 \pm 0.064}$	4.464 ± 0.442	5.439 ± 1.140	34.926 ± 1.410	0.601	0.450	0.955
	e2GMM	3.828 ± 0.029	5.416 ± 0.499	5.546 ± 0.524	219.684 ± 1.620	0.016	0.015	1.030
	CryoSPHERE	–	–	–	–	–	–	–
	CryoSTAR	5.697 ± 0.192	6.386 ± 0.350	7.356 ± 0.730	–	0.667	**0.992**	1.136
	ESMFold2	5.116 ± 0.302	6.504 ± 0.530	–	$\mathbf{13.498 \pm 2.589}$	0.077	0.125	1.914

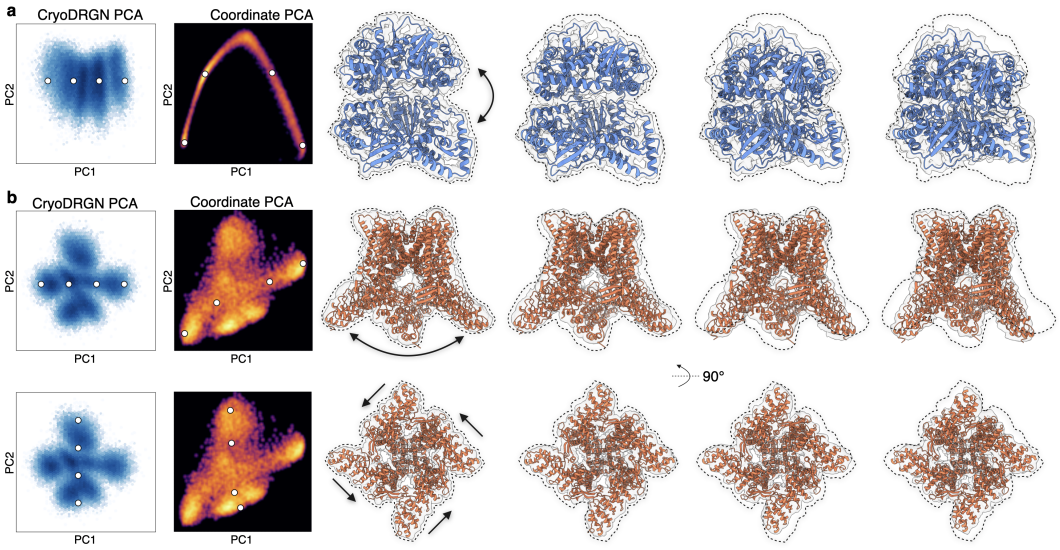

Figure 3: **CryoCHEF learns conformational landscapes from synthetic cryo-EM datasets.** (a) Results on synthetic dataset of human presequence protease (Liang et al., 2022). (b) Results on synthetic dataset of TRPV1 Gao et al. (2016). (Left) CryoDRGN latents of each dataset projected onto first two PCs. White dots are equispaced latents traversed along the principal axes. (Middle) PCA projection of 10k sampled structures from CryoCHEF, with predicted traversal structures projected as white dots. (Right) Traversal structures overlaid with corresponding cryoDRGN maps.

3.3 Real Datasets

HER2-Fabs. We first perform heterogeneous reconstruction of the full 708,609 particle dataset (Ruedas et al., 2024) using cryoDRGN, then train CryoCHEF on 1,000 latents and decoded volumes sampled by k-means clustering. Running inference on latents sampled along the first cryoDRGN

principal component produces a structural trajectory corresponding to rotation of Trastuzumab Fab relative to Pertuzumab Fab (Figure 4a). Likewise, the traversal of the second cryoDRGN PC in structural space shows a hinging of HER2 domain IV with the two Fabs being brought in proximity or apart (Figure 4b). In addition to these global motions, local flexibility in the Trastuzumab Fab is observed. From 10k structures predicted from random latents, we fit PCA on just the Fab coordinates. The first PC correlates to the distance between the constant and variable region centroids of the Fab, ranging from 28Å to 35Å (Figure 4c). Map-model FSCs go to around 8Å against both cryoDRGN and backprojected maps (Figure 4d; Table 4).

Integrin-Fabs. We run *ab initio* cryoDRGN (Levy et al., 2025) on an unposed image stack of 84,266 particles (Campbell et al., 2020) to obtain alignments, then run standard cryoDRGN to resolve heterogeneity. We then discard a population of broken particles, then run a fresh *ab initio* cryoDRGN reconstruction on 51,367 particles. We train CryoCHEF on 100 latents and decoded volumes sampled by k-means clustering. Running inference on latents sampled along the first cryoDRGN principal component produces a structural trajectory corresponding to hinging of the calf region relative to the thigh, and rotation of the headpiece relative to the leg (Figure 5a). From 10k structures predicted from random latents, we also visualize the distribution of the flexible αV knee angle between the calf and the thigh, which ranges from 100° to 112° (Figure 5b). For this dataset, we observe that the map-model FSC curve against the cryoDRGN decoded maps are much higher than against the backprojections, which are reconstructed from smaller particle counts.

SARS-CoV-2 spike. We run cryoDRGN on a 206,260 particle subset (Walls et al., 2020) from the original analysis of this dataset (Zhong et al., 2021), then train CryoCHEF on 1,000 latents and decoded volumes sampled by k-means clustering. Running inference on latents sampled along the first cryoDRGN principal component produces a structural trajectory corresponding to the opening and closing motion of one receptor binding domain (RBD) (Figure 1). While the cryoDRGN latent space PCA is a diffuse circle, PCA fit to 10k predicted structures reveals a manifold where the two endpoints of the motion, RBD up and RBD down, are most densely populated as is expected for this dataset.

3.4 Ablations

We report training ablations on the synthetic benchmarks in Supplemental Table 5. Fixing the training resolution at the Nyquist limit, instead of marching from low to high resolution, has a mild mixed effect on structural and distributional accuracy, as does ablating the bond loss. Both lead to worse clashscore on TRPV1. Removing the pose refinement strategy, instead always aligning samples to the reference, leads to a considerable degradation of structural quality in both datasets.

4 Related Work

Continuous heterogeneity in cryo-EM. Methods such as cryoDRGN, 3D variability analysis (3DVA), e2GMM, DynaMight, CryoSTAR, and CryoSPHERE recover continuous conformational heterogeneity from cryo-EM particle images (Zhong et al., 2021; Punjani & Fleet, 2021; Chen & Ludtke, 2021; Schwab et al., 2024; Li et al., 2024; Ducrocq et al., 2025). These methods differ in how they represent molecular motion. CryoDRGN and 3DVA represent heterogeneity primarily in density space, whereas e2GMM and DynaMight use deformable Gaussian representations and can incorporate atomic models to constrain or refine the reconstruction (Chen & Ludtke, 2021; Schwab et al., 2024). CryoSTAR and CryoSPHERE more directly incorporate atomic or coarse-grained structural representations and geometric constraints into heterogeneous reconstruction (Li et al., 2024; Ducrocq et al., 2025). In these approaches, atomic or pseudo-atomic modeling is part of the heterogeneous reconstruction itself. CryoCHEF instead decouples these problems: it operates downstream of an existing reconstruction and maps its learned conformational space into the representation of a pretrained folding model. This allows the same atomic interpretation framework to be applied across reconstruction methods, as demonstrated here with cryoDRGN and DynaMight.

Experiment-guided protein structure prediction. A complementary line of work combines pretrained protein structure models with experimental data. ROCKET optimizes AlphaFold2 representations using differentiable likelihoods from cryo-EM, cryo-ET, and X-ray data (Fadini et al., 2026). CryoBoltz and EmbedOpt guide Boltz-1 diffusion with cryo-EM density to sample structures

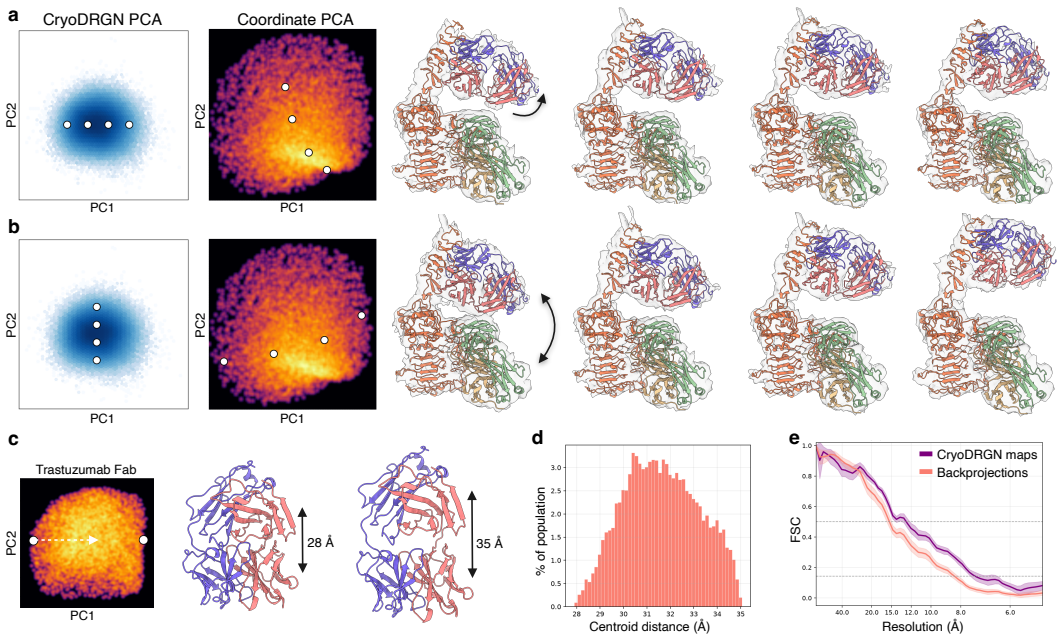

Figure 4: **CryoCHEF captures global and local flexibility in an experimental dataset of HER2 bound to two Fabs.** (a) Left: CryoDRGN latent embeddings of 708,609 particles projected onto first two PCs. White dots are 4 equispaced points along the first PC. Middle: Pairwise Cα distances of 10k sampled structures from CryoCHEF projected onto first two PCs, with traversal structures projected as white dots. Right: Traversal structures overlaid with corresponding cryoDRGN maps. (b) Same as (a) with structures sampled along second PC. (c) PC projection of only the trastuzumab chains in the 10k structures. First PC corresponds to inter-centroid distance between constant and variable region of the Fab. (d) Distribution of inter-centroid distances across the 10k sampled structures. (e) Map-model FSC curves between 100 sampled structures against corresponding cryoDRGN maps and backprojections.

consistent with a target map (Raghu et al., 2025; Li et al., 2026). Another approach formulates structure determination as posterior sampling with AlphaFold3, using NMR, X-ray, and cryo-EM measurements to generate measurement-consistent structural ensembles (Maddipatla et al., 2026). CryoCHEF extends this direction to continuous cryo-EM heterogeneity. Rather than guiding structure prediction independently against individual maps, CryoCHEF learns a mapping from the conformational latent space of a heterogeneous reconstruction to the folding-model representation, yielding a continuous all-atom ensemble.

5 Discussion

This work introduces a method for fast, amortized learning of atomic structural ensembles from heterogeneous cryo-EM data. By learning a continuous correspondence between experimentally recovered conformational variation and the structural prior of a pretrained co-folding model, CryoCHEF generates all-atom structures across an entire conformational landscape without independently fitting or guiding a structure model for each state.

On synthetic benchmarks with known ground truth, CryoCHEF achieves strong structural and distributional accuracy. On experimental datasets, CryoCHEF resolves motions spanning multiple spatial scales, including domain rotations and hinging in the HER2–antibody complex and integrin, local antibody flexibility in the HER2 complex, and receptor-binding-domain opening and closing in the SARS-CoV-2 spike protein.

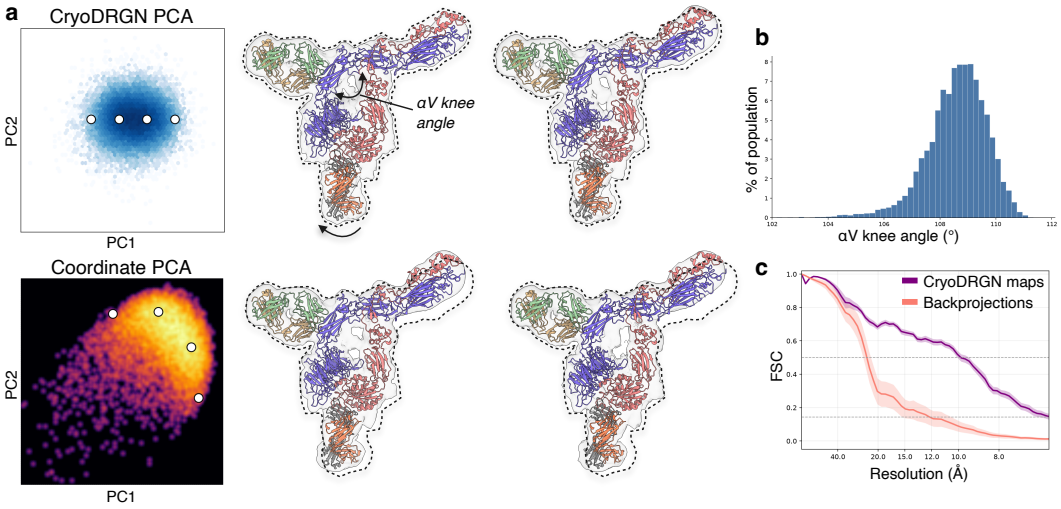

Figure 5: **CryoCHEF recovers motion trajectories in an experimental dataset of integrin.** (a) Top left: CryoDRGN latents of 51,367 particles projected onto first two PCs. White dots indicate 4 equispaced points traversed along the first PC. Bottom left: PC projection of pairwise Cα distances of 10k structures sampled from CryoCHEF, with traversal structure projections overlaid. Right: Predicted traversal structures overlaid with corresponding cryoDRGN maps. (b) Distibution of the αV knee angle between the leg and calf domains across the 10k structures. (c) Map-model FSC curves between 100 k-means sampled structures against cryoDRGN maps and backprojections.

A key feature of CryoCHEF is the separation of heterogeneous reconstruction from atomic modeling, establishing a general framework for ensemble atomic modeling. Reconstruction methods supply the conformational representation and density maps, while co-folding models supply a prior over atomic structure. This modularity allows advances in either component to be incorporated independently: Advances in heterogeneous reconstruction could resolve a broader range of motions and provide more informative maps, while stronger co-folding priors could improve atomic geometry and coverage of conformational states. Together, these advances could extend CryoCHEF to larger and more complex molecular systems and increasingly diverse modes of conformational heterogeneity.

The current approach remains limited by its reliance on an accurate initial structure prediction, choices of training and inference hyperparameters, and computational requirements that can become prohibitive for larger molecular systems. Important directions are therefore to improve the accuracy and computational efficiency of the foundation models underlying this approach and to automate hyperparameter selection. We envision a future with a fully automated route from cryo-EM measurements to experimentally supported atomic ensembles of protein motion, enabling large-scale analysis of protein dynamics.

AI Use Statement

We used generative AI tools throughout the research process. All AI-assisted work was extensively reviewed by a human author. We take responsibility for the final content of this work, including all text, claims, and artifacts produced with the aid of generative AI.

Reproducibility Statement

Details needed to reproduce experiments are described in sections 2 (Methods), 3 (Results), and A (Supplementary Material). Code will be released upon publication.

REFERENCES

Josh Abramson, Jonas Adler, Jack Dunger, Richard Evans, Tim Green, Alexander Pritzel, Olaf Ronneberger, Lindsay Willmore, Andrew J. Ballard, Joshua Bambrick, Sebastian W. Bodenstein, David A. Evans, Chia-Chun Hung, Michael O'Neill, David Reiman, Kathryn Tunyasuvunakool, Zachary Wu, Akvilė Žemgulytė, Eirini Arvaniti, Charles Beattie, Ottavia Bertolli, Alex Bridgland, Alexey Cherepanov, Miles Congreve, Alexander I. Cowen-Rivers, Andrew Cowie, Michael Figurnov, Fabian B. Fuchs, Hannah Gladman, Rishub Jain, Yousuf A. Khan, Caroline M. R. Low, Kuba Perlin, Anna Potapenko, Pascal Savy, Sukhdeep Singh, Adrian Stecula, Ashok Thillaisundaram, Catherine Tong, Sergei Yakneen, Ellen D. Zhong, Michal Zielinski, Augustin Žídek, Victor Bapst, Pushmeet Kohli, Max Jaderberg, Demis Hassabis, and John M. Jumper. Accurate structure prediction of biomolecular interactions with alphafold 3. *Nature*, 630(8016):493–500, 06 2024. ISSN 1476-4687. doi: 10.1038/s41586-024-07487-w. URL https://doi.org/10.1038/s41586-024-07487-w.

Miro A. Astore, David Silva-Sánchez, Robert Blackwell, Pilar Cossio, and Sonya M. Hanson. Comparing cryo-em methods and molecular dynamics simulation to investigate heterogeneity in ligand-bound trpv1. *Nature Communications*, 17(1):1067, 2025. ISSN 2041-1723. doi: 10.1038/s41467-025-67821-2. URL https://doi.org/10.1038/s41467-025-67821-2.

Jan-Hendrik Bastek, WaiChing Sun, and Dennis M. Kochmann. Physics-informed diffusion models. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=tpYeermigp.

Melody G. Campbell, Anthony Cormier, Saburo Ito, Robert I. Seed, Andrew J. Bondesson, Jianlong Lou, James D. Marks, Jody L. Baron, Yifan Cheng, and Stephen L. Nishimura. Cryo-EM Reveals Integrin-Mediated TGF-β Activation without Release from Latent TGF-β. *Cell*, 180(3):490–501.e16, 2020. doi: 10.1016/j.cell.2019.12.030. URL http://dx.doi.org/10.1016/j.cell.2019.12.030.

Salvatore Candido, Thomas Hayes, Alexander Derry, Roshan Rao, Zeming Lin, Robert Verkuil, Bryan Z. Wu, Jin Sub Lee, Elise S. Bruguera, Jehan A. Keval, Mykhailo Kopylov, John E. Pak, Wesley Wu, Neil Thomas, Samson Mataraso, Alvin Hsu, Ashton C. Trotman-Grant, Kilian Fatras, Allan dos Santos Costa, Rohil Badkundri, Halil Akın, Deniz Oktay, Jonathan Deaton, Elizabeth Montabana, Hrishita Sitwala, Yue Yu, Marius Wiggert, Dylan Alexander Carlin, Anthony W. Goering, Tomasz Blazejewski, McCullen Sandora, Michael Hla, Tina Z. Jia, Leon H. Kloker, Nicholas J. Sofroniew, Masatoshi Uehara, Jassi Pannu, Sharrol Bachas, Daniel S. Liu, Tom Sercu, and Alexander Rives. Language modeling materializes a world model of protein biology. *bioRxiv*, 2026. doi: 10.64898/2026.06.03.729735. URL https://www.biorxiv.org/content/early/2026/06/04/2026.06.03.729735.

Muyuan Chen and Steven J. Ludtke. Deep learning-based mixed-dimensional gaussian mixture model for characterizing variability in cryo-em. *Nature Methods*, 18(8):930–936, 08 2021. ISSN 1548-7105. doi: 10.1038/s41592-021-01220-5. URL https://doi.org/10.1038/s41592-021-01220-5.

Gabriel Ducrocq, Lukas Grunewald, Sebastian Westenhoff, and Fredrik Lindsten. cryoSPHERE: Single-particle HEterogeneous REconstruction from cryo EM. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=n8O0trhost.

Alisia Fadini, Minhuan Li, Airlie J. McCoy, Suresh Banjara, Hiroki Okumura, Eve Napier, Pietro Fontana, Amir R. Khan, Luca Jovine, Thomas C. Terwilliger, Randy J. Read, Doeke R. Hekstra, and Mohammed AlQuraishi. Alphafold as a prior: experimental structure determination conditioned on a pretrained neural network. *Nature Methods*, 23(4):785–795, 04 2026. ISSN 1548-7105. doi: 10.1038/s41592-026-03047-4. URL https://doi.org/10.1038/s41592-026-03047-4.

J. Ryan Feathers, Robert Heeter, Geoffrey Woollard, Sonya M Hanson, Pilar Cossio, Joel Greer, Tom Burnley, and Ellen Zhong. The cahra challenge: A community-wide assessment of cryo-em heterogeneous reconstruction algorithms. *bioRxiv*, 2026. doi: 10.64898/2026.09.15.751515. URL https://www.biorxiv.org/content/early/2026/09/24/2026.09.15.751515.

Yuan Gao, Erhu Cao, David Julius, and Yifan Cheng. Trpv1 structures in nanodiscs reveal mechanisms of ligand and lipid action. *Nature*, 534(7607):347–351, Jun 2016. ISSN 1476-4687. doi: 10.1038/nature17964. URL https://doi.org/10.1038/nature17964.

Marc Aurèle Gilles and Amit Singer. Cryo-EM heterogeneity analysis using regularized covariance estimation and kernel regression. *Proceedings of the National Academy of Sciences*, 122(9), 2025. doi: 10.1073/pnas.2419140122. URL http://dx.doi.org/10.1073/pnas.2419140122.

Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance, 2022. URL https://arxiv.org/abs/2207.12598.

John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A. A. Kohl, Andrew J. Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Jonas Adler, Trevor Back, Stig Petersen, David Reiman, Ellen Clancy, Michal Zielinski, Martin Steinegger, Michalina Pacholska, Tamas Berghammer, Sebastian Bodenstein, David Silver, Oriol Vinyals, Andrew W. Senior, Koray Kavukcuoglu, Pushmeet Kohli, and Demis Hassabis. Highly accurate protein structure prediction with alphafold. *Nature*, 596(7873):583–589, 08 2021. ISSN 1476-4687. doi: 10.1038/s41586-021-03819-2. URL https://doi.org/10.1038/s41586-021-03819-2.

W. Kabsch. A solution for the best rotation to relate two sets of vectors. *Acta Crystallographica Section A*, 32(5):922–923, 1976. doi: 10.1107/s0567739476001873. URL http://dx.doi.org/10.1107/S0567739476001873.

Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the design space of diffusion-based generative models, 2022. URL https://arxiv.org/abs/2206.00364.

Axel Levy, Rishwanth Raghu, J. Ryan Feathers, Michal Grzadkowski, Frédéric Poitevin, Jake D. Johnston, Francesca Vallese, Oliver Biggs Clarke, Gordon Wetzstein, and Ellen D. Zhong. CryoDRGN-AI: neural ab initio reconstruction of challenging cryo-EM and cryo-ET datasets. *Nature Methods*, 22(7):1486–1494, 2025. doi: 10.1038/s41592-025-02720-4. URL http://dx.doi.org/10.1038/s41592-025-02720-4.

Minhuan Li, Jiequn Han, Pilar Cossio, and Luhuan Wu. Robust inference-time steering of protein diffusion models via embedding optimization, 2026. URL https://arxiv.org/abs/2602.05285.

Yilai Li, Yi Zhou, Jing Yuan, Fei Ye, and Quanquan Gu. Cryostar: leveraging structural priors and constraints for cryo-em heterogeneous reconstruction. *Nature Methods*, 21(12):2318–2326, 12 2024. ISSN 1548-7105. doi: 10.1038/s41592-024-02486-1. URL https://doi.org/10.1038/s41592-024-02486-1.

Wenguang G. Liang, Juwina Wijaya, Hui Wei, Alex J. Noble, Jordan M. Mancl, Swansea Mo, David Lee, John V. Lin King, Man Pan, Chang Liu, Carla M. Koehler, Minglei Zhao, Clinton S. Potter, Bridget Carragher, Sheng Li, and Wei-Jen Tang. Structural basis for the mechanisms of human presequence protease conformational switch and substrate recognition. *Nature Communications*, 13(1), 2022. doi: 10.1038/s41467-022-29322-4. URL http://dx.doi.org/10.1038/s41467-022-29322-4.

Advaith Maddipatla, Nadav Sellam Bojan, Meital Bojan, Volodymyr Masalitin, Sanketh Vedula, Paul Schanda, Ailie Marx, and Alex M. Bronstein. Experiment-guided AlphaFold3 resolves measurement-consistent protein ensembles. *Nature Biotechnology*, 2026. doi: 10.1038/s41587-026-03166-5. URL http://dx.doi.org/10.1038/s41587-026-03166-5.

Chenlin Meng, Yutong He, Yang Song, Jiaming Song, Jiajun Wu, Jun-Yan Zhu, and Stefano Ermon. Sdedit: Guided image synthesis and editing with stochastic differential equations, 2022. URL https://arxiv.org/abs/2108.01073.

Michael J. O'Brien, David Silva-Sánchez, Geoffrey Woollard, Kwanghwi Je, Sonya M. Hanson, Daniel J. Needleman, Pilar Cossio, Erik Henning Thiede, and Miro A. Astore. CryoJAX: a cryo-electron microscopy image-simulation library in JAX. *Acta Crystallographica Section D*, 82(3), Mar 2026. doi: 10.1107/S2059798326000550. URL https://doi.org/10.1107/S2059798326000550.

Hayden Prairie, Zachary Novack, Taylor Berg-Kirkpatrick, and Daniel Y. Fu. Parcae: Scaling laws for stable looped language models, 2026. URL https://arxiv.org/abs/2604.12946.

Ali Punjani and David J. Fleet. 3D variability analysis: Resolving continuous flexibility and discrete heterogeneity from single particle cryo-EM. *Journal of Structural Biology*, 213(2): 107702, 2021. doi: 10.1016/j.jsb.2021.107702. URL http://dx.doi.org/10.1016/j.jsb.2021.107702.

Ali Punjani and David J. Fleet. 3DFlex: determining structure and motion of flexible proteins from cryo-EM. *Nature Methods*, 20(6):860–870, 2023. doi: 10.1038/s41592-023-01853-8. URL http://dx.doi.org/10.1038/s41592-023-01853-8.

Rishwanth Raghu, Axel Levy, Gordon Wetzstein, and Ellen D. Zhong. Multiscale guidance of protein structure prediction with heterogeneous cryo-em data. In *NeurIPS*, 2025.

Rémi Ruedas, Rémi Vuillemot, Thibault Tubiana, Jean-Marie Winter, Laura Pieri, Ana-Andreea Arteni, Camille Samson, Slavica Jonic, Magali Mathieu, and Stéphane Bressanelli. Structure and conformational variability of the HER2-trastuzumab-pertuzumab complex. *Journal of Structural Biology*, 216(2):108095, 2024. doi: 10.1016/j.jsb.2024.108095. URL http://dx.doi.org/10.1016/j.jsb.2024.108095.

Johannes Schwab, Dari Kimanius, Alister Burt, Tom Dendooven, and Sjors H. W. Scheres. Dynamight: estimating molecular motions with improved reconstruction from cryo-em images. *Nature Methods*, 21(10):1855–1862, 10 2024. ISSN 1548-7105. doi: 10.1038/s41592-024-02377-5. URL https://doi.org/10.1038/s41592-024-02377-5.

Alexandra C. Walls, Young-Jun Park, M. Alejandra Tortorici, Abigail Wall, Andrew T. McGuire, and David Veesler. Structure, function, and antigenicity of the sars-cov-2 spike glycoprotein. *Cell*, 181(2):281–292.e6, 2020. ISSN 0092-8674. doi: https://doi.org/10.1016/j.cell.2020.02.058. URL https://www.sciencedirect.com/science/article/pii/S0092867420302622.

Ellen D Zhong, Tristan Bepler, Bonnie Berger, and Joseph H Davis. Cryodrgn: reconstruction of heterogeneous cryo-em structures using neural networks. *Nature Methods*, 18(2):176–185, 02 2021. doi: 10.1038/s41592-020-01049-4. URL https://doi.org/10.1038/s41592-020-01049-4.

A SUPPLEMENTARY MATERIAL

A.1 PROTEIN STRUCTURE PREDICTION WITH ESMFOLD2

Given a sequence $\mathbf{s} \in \{0, \dots, 19\}^L$, ESMFold2 embeds it with the protein language model ESMC (Candido et al., 2026) and combines these embeddings with additional input features, through an AlphaFold3-style input embedder, to initialize the pair representation $\mathbf{c} \in \mathbb{R}^{L \times L \times d_c}$. A folding trunk of triangle-multiplication and SwiGLU feed-forward blocks refines $\mathbf{c}$ and is applied recurrently following the Parcae looping scheme (Prairie et al., 2026). For computational efficiency, we relied on a Triton kernel implementation that fused operations inside triangle multiplication and the transition Feed Forward Network. The kernels also supported a chunking variant that allowed us to scale the sequence protein length up to 4000 residues. The model we use has no separate trunk single representation, so the diffusion head is conditioned only on per-residue input features and on $\mathbf{c}$.

The diffusion head generates coordinates $\mathbf{x} \in \mathbb{R}^{N \times 3}$ for all N heavy atoms in the EDM framework (Karras et al., 2022), with the denoiser preconditioned as

$$D_\theta(\mathbf{x}; \sigma, \mathbf{c}) = c_{\text{skip}}(\sigma)\, \mathbf{x} + c_{\text{out}}(\sigma)\, F_\theta\big(c_{\text{in}}(\sigma)\, \mathbf{x},\, c_{\text{noise}}(\sigma),\, \mathbf{c}\big),$$

where F_θ is a transformer. We refer to (Candido et al., 2026; Abramson et al., 2024) for a definition of the c functions. Sampling starts from $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \sigma_{\text{max}}^2 \mathbf{I})$ and integrates the direction $\mathbf{d} = (\mathbf{x} - D_\theta(\mathbf{x}; \sigma, \mathbf{c}))/\sigma$ over a decreasing schedule $\sigma_{\text{max}} = \sigma_0 > \cdots > \sigma_K \approx 0$, using the AlphaFold3 stochastic sampler (Abramson et al., 2024) with $K = 20$ steps.

A.2 LATENT-CONDITIONED PAIR BIAS

Predicting a full $L \times L \times d_c$ bias directly from z is impractical, so $\hat{b}_\phi$ is a neural field over residue pairs:

$$\mathbf{c}_z^{(ij)} = \mathbf{c}_{\text{base}}^{(ij)} + \hat{b}_\phi\big(\bar{z}, e_{ij}\big),$$

where $\bar{z}$ is the latent normalized by a single global mean and standard deviation, which preserves the relative scale of the latent dimensions, and $e_{ij} \in \mathbb{R}^{32}$ is a learned embedding of the residue pair. $\hat{b}_\phi$ is an MLP with 3 hidden layers of width 256 and ReLU activations. Its output layer is zero-initialized, so training starts from the unmodified ESMFold2 prediction, and the bias is symmetrized, $\hat{b}^{(ij)} \leftarrow \frac{1}{2}(\hat{b}^{(ij)} + \hat{b}^{(ji)})$. The bias is applied only at noise levels $\sigma \in [1, 32]$ Å; at higher noise levels the base representation is used. The embedding table and the MLP are the only trainable parameters.

A.3 PIDM TRAINING

Each rollout starts from the reference structure $\mathbf{x}_{\text{ref}}$ noised, as in SDEdit (Meng et al., 2022), to the highest schedule noise level at or below 32 Å ($\sigma_{\text{start}} = 23.9$ Å):

$$\mathbf{x}_{\sigma_{\text{start}}} = \mathbf{x}_{\text{ref}} + \sigma_{\text{start}}\, \boldsymbol{\varepsilon}, \qquad \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}).$$

The frozen sampler integrates it down to $\sigma \approx 0$ under $\mathbf{c}_z$, with $\boldsymbol{\varepsilon}$ and all sampler noise fixed across examples. From the rollout we draw, without replacement, $m = 4$ states $\mathbf{x}_{\sigma_j}$ inside the band $[1, 32]$ Å. The gradient is

$$\nabla_{\phi,e} \sum_{j=1}^{m} \mathcal{R}\Big(\mathcal{A}\big(D_\theta(\mathbf{x}_{\sigma_j}; \sigma_j, \mathbf{c}_z)\big)\Big),$$

where $\mathcal{A}$ is the alignment of section A.4. Gradients flow through one denoiser evaluation per sampled state and not through the rest of the rollout. We optimize with Adam at a constant learning rate of 10^{-3}, gradient clipping at 1.0, batch size 8, for 1000 steps, without EMA.

A.4 ALIGNMENT

For example i, the terminal prediction $\mathbf{x}_0^{(i)}$ is superposed by Cα-weighted Kabsch alignment (Kabsch, 1976) onto a cached pose $\mathbf{x}_{0,\text{cache}}^{(i)}$, then refined by 10 Adam steps on $\mathcal{R}_{\text{map}}$ at the current cutoff K over a rotation vector (learning rate 0.01) and a translation (learning rate 0.1), keeping the best pose seen. This gives $\mathbf{x}_{0,\text{aligned}}^{(i)}$. The cache is initialized to $\mathbf{x}_{\text{ref}}$ and overwritten with $\mathbf{x}_{0,\text{aligned}}^{(i)}$ every step. The m reward-carrying predictions are Kabsch-superposed onto $\mathbf{x}_{0,\text{aligned}}^{(i)}$ before scoring.

A.5 INFERENCE SCHEDULE

Inference uses the same 20-step sampler, partial-noising start, and fixed noise as training. For all datasets except Integrin-Fabs, the bias strength β is interpolated linearly over the in-band steps, from $\beta = 1$ at the first to $\beta = 0.1$ at the last (table 2).

Table 2: Bias strength at each in-band inference step.

σ (Å)	23.9	12.7	6.3	2.9	1.2
β	1.0	0.78	0.55	0.33	0.1

For Integrin-Fabs, we find that a schedule of $\beta = \{1.0, 1.0, 1.0, 0.0, 0.0\}$ provides the best tradeoff of steering and geometry.

A.6 ALGORITHMS

We provide pseudocode summarizing the CryoCHEF training and inference procedures described in Section 2. Algorithm 1 details the training loop, including the per-residue-pair rasterization of the conditional bias $\hat{b}_\phi$, the partially-noised reverse diffusion rollout, the Kabsch alignment and map-based refinement procedure used to orient predicted structures relative to the reconstruction frame, and the reward-based gradient update applied to the bias network ϕ and residue-pair embedding table e. Algorithm 2 details inference given a trained model and a target conformational latent z, including the timestep-dependent modulation of the pair representation bias via $\beta^{(t)}$, which we found necessary to balance conformational steering against sample geometric quality (Section 2).

Algorithm 1 CryoCHEF Training

Require: Sequence $\mathbf{s}$; frozen cryo-EM encoder ζ; frozen folding trunk; frozen diffusion head D_θ; dataset $\{(\gamma_i, V_i)\}_{i=1}^{N}$; reference structure $\mathbf{x}_{\text{ref}}$; training band $[t_{\text{start}}, t_{\text{end}}]$; residue-pair embedding table e; bias network b_ϕ

Ensure: Trained parameters ϕ, e

$\mathbf{c}_{\text{base}} \leftarrow \textsc{Trunk}(\mathbf{s})$ ▷ frozen pair representation
$\varepsilon \sim \mathcal{N}(0, \sigma_{t_{\text{start}}}\mathbf{I})$ ▷ fixed noise realization, reused for all rollouts
for $i = 1, \ldots, N$ **do**
  $z_i \leftarrow \zeta(\gamma_i)$
  $\mathbf{x}_{0,\text{cache}}^{(i)} \leftarrow \mathbf{x}_{\text{ref}}$
end for
while not converged **do**
  Sample minibatch $\mathcal{B}$
  $\mathcal{R}_{\text{total}} \leftarrow 0$
  for $i \in \mathcal{B}$ **do**
    for $(j, k) \in [L] \times [L]$ **do**
      $\mathbf{c}_z^{(i,jk)} \leftarrow \mathbf{c}_{\text{base}}^{(jk)} + \hat{b}_\phi(z_i, e_{jk})$
    end for
    $\mathbf{x}_{t_{\text{start}}} \leftarrow \mathbf{x}_{\text{ref}} + \varepsilon$
    Sample m timesteps $t(1), \ldots, t(m) \in [t_{\text{start}}, t_{\text{end}}]$
    Simulate reverse rollout $(\mathbf{x}_{t_{\text{start}}}, \ldots, \mathbf{x}_0)$ via $D_\theta(\cdot\,; \sigma_t, \mathbf{c}_z^{(i)})$
    $\mathbf{x}_{0,\text{aligned}}^{(i)} \leftarrow \textsc{Kabsch}(\mathbf{x}_0, \mathbf{x}_{0,\text{cache}}^{(i)})$
    $\mathbf{x}_{0,\text{aligned}}^{(i)} \leftarrow \textsc{RefineAlign}(\mathbf{x}_{0,\text{aligned}}^{(i)}; \nabla\mathcal{R}_{\text{map}}, K \leftarrow K_{\text{current}})$
    for $m' = 1, \ldots, m$ **do**
      $\mathbf{x}_{t(m'),\text{aligned}}^{(i)} \leftarrow \textsc{Kabsch}(\mathbf{x}_{t(m')}, \mathbf{x}_{0,\text{aligned}}^{(i)})$
      $\mathcal{R}_{\text{total}} \leftarrow \mathcal{R}_{\text{total}} + \mathcal{R}_{\text{map}}(\mathbf{x}_{t(m'),\text{aligned}}^{(i)}) + \alpha_{\text{bond}}\mathcal{R}_{\text{bond}}(\cdot) + \alpha_{\text{lddt}}\mathcal{R}_{\text{lddt}}(\cdot)$
    end for
    $\mathbf{x}_{0,\text{cache}}^{(i)} \leftarrow \mathbf{x}_{0,\text{aligned}}^{(i)}$
  end for
  $\phi, e \leftarrow \phi, e - \eta\,\nabla_{\phi,e}\,\mathcal{R}_{\text{total}}$ ▷ gradient accumulated over minibatch
end while

Algorithm 2 CryoCHEF Inference

Require: Sequence $\mathbf{s}$; conformational latent z; frozen $\mathbf{c}_{\text{base}}$; trained b_ϕ, e; guidance schedule $\beta^{(t)}$; fixed noise realization ε; reference structure $\mathbf{x}_{\text{ref}}$; training band start t_{start}; noise schedule $\sigma_{t_{\text{start}}} > \cdots > \sigma_0 \approx 0$

Ensure: Predicted all-atom structure $\mathbf{x}_0$ for conformation z

for $(j, k) \in [L] \times [L]$ **do**
  $b_z^{(jk)} \leftarrow \hat{b}_\phi(z, e_{jk})$
end for
$\mathbf{x}_{t_{\text{start}}} \leftarrow \mathbf{x}_{\text{ref}} + \varepsilon$ ▷ same partial noising as training, Alg. 1
for $t = t_{\text{start}}, \ldots, 1$ **do**
  $\mathbf{c}_z^{(t)} \leftarrow \mathbf{c}_{\text{base}} + \beta^{(t)} \cdot b_z$
  $\mathbf{d} \leftarrow \left(\mathbf{x}_t - D_\theta(\mathbf{x}_t; \sigma_t, \mathbf{c}_z^{(t)})\right)/\sigma_t$
  $\mathbf{x}_{t-1} \leftarrow \textsc{Integrate}(\mathbf{x}_t, \mathbf{d}, \sigma_t \to \sigma_{t-1})$
end for
return $\mathbf{x}_0$

A.7 Additional Experimental Details

We describe the generation of the synthetic datasets, and the settings for the density reconstruction models, CryoCHEF, and baselines.

A.7.1 PROTEASE

An all-atom MD simulation with 2,000 frames was obtained from Feathers et al. (2026). We used CryoJAX (O'Brien et al., 2026) to generate 50 cryo-EM images per structure, for a total of 100k images. Images were simulated at a box size of 192 at 1Å/pix, with poses uniformly drawn from $\mathrm{SO}(3)$, defocus uniformly drawn from $[0.5, 0.25]\mu m$ with no astigmatism, in-plane translations uniformly drawn from $[-15, 15]$ Å independently in the x and y directions, and noise added at SNR 0.01. Images were then downsampled to a box size of 96 (2.0Å/pix).

All images are backprojected to obtain the dataset's consensus map. ESMFold2 with 20 diffusion steps is used to predict a sample, which is docked into the consensus map and used as the reference structure.

CryoDRGN is trained on the images with latent dimension 8. K-means volumes are obtained through the cryoDRGN *analyze* pipeline. DynaMight is trained at latent dimension 8 and 2000 Gaussians. Inverse deformations are trained, after which volumes are obtained by the deformable backprojection pipeline. CryoCHEF is trained on these two sets of volumes.

CryoSTAR is trained with latent dimension 8 and *dist_keep_ratio*$= 0.8$, which demonstrated better flexibility than the default *dist_keep_ratio*$= 0.99$. Structures were obtained by k-means clustering the latent space and decoding with the stage 1 model. For e2GMM, *e2gmm_refine_new –trainmodel* is run for 20 iterations, with 2000 Gaussians and 4.0Å max resolution. *e2gmm_refine_new –heter* is run for 20 iterations, with latent dimension 8 and 8Å max resolution. *e2gmm_model_fit* is run between the initial structure and the dataset backprojection. Structures are generated from k-means clustered latents with *e2gmm_eval*. CryoSPHERE is trained with latent dimension 8 and 6 segments, as training diverged with 20 segments. Structures are generated from k-means clustered latents.

A.7.2 TRPV1

An all-atom MD simulation with 10,000 frames was obtained from Astore et al. (2025). We used CryoJAX (O'Brien et al., 2026) to generate 10 cryo-EM images per structure, for a total of 100k images. Images were simulated at a box size of 256 at 1Å/pix, with poses uniformly drawn from $\mathrm{SO}(3)$, defocus uniformly drawn from $[0.5, 0.25]\mu m$ with no astigmatism, in-plane translations uniformly drawn from $[-20, 20]$ Å independently in the x and y directions, and noise added at SNR 0.01. Images were then downsampled to a box size of 128 (2.0Å/pix).

All images are backprojected to obtain the dataset's consensus map. ESMFold2 with 20 diffusion steps is used to predict a sample, which is docked into the consensus map and used as the reference structure.

CryoDRGN is trained on the images with latent dimension 8. K-means volumes are obtained through the cryoDRGN *analyze* pipeline. DynaMight is trained at latent dimension 8 and 2000 Gaussians. Inverse deformations are trained, after which volumes are obtained by the deformable backprojection pipeline. CryoCHEF is trained on these two sets of volumes.

CryoSTAR is trained with latent dimension 8, *dist_keep_ratio*$= 0.8$, and *inter_chain_cutoff*$= 12.0$. Structures were obtained by k-means clustering the latent space and decoding with the stage 1 model. For e2GMM, *e2gmm_refine_new –trainmodel* is run for 20 iterations, with 5000 Gaussians and 6.0Å max resolution. *e2gmm_refine_new –heter* is run for 20 iterations, with latent dimension 8 and 8Å max resolution. *e2gmm_model_fit* is run between the initial structure and the dataset backprojection. Structures are generated from k-means clustered latents with *e2gmm_eval*. CryoSPHERE training diverged with both 6 and 20 segments.

A.7.3 HER2-FABS

Posed particle stacks are downloaded from EMPIAR-11665 (Ruedas et al., 2024). All 708,609 particles are downsampled to box size 128 (2.5375Å/pix) and processed with cryoDRGN for 25 epochs with latent dimension 8.

The PDB deposition associated with the dataset (PDB:8PWH) is used as the reference structure.

CryoCHEF is trained on 1000 k-means latents and volumes are obtained through the cryoDRGN *analyze* pipeline.

A.7.4 INTEGRIN-FABS

Unposed particle stacks are downloaded from EMPIAR-10345 (Campbell et al., 2020). All 84,266 particles are downsampled to box size 128 (3.152Å/pix) and processed with cryoDRGN *ab initio* for 30 epochs and latent dimension 8. The estimated poses are then input to a standard cryoDRGN training run for 25 epochs and latent dimension 8. Some 20 k-means volumes produced by the cryoDRGN *analyze* pipeline are observed to have broken density, so the corresponding particles are filtered out. A fresh *ab initio* cryoDRGN training run is performed on the remaining 51,367 particles for 30 epochs with latent dimension 8. The 100 k-means latents and volumes are obtained through the cryoDRGN *analyze* pipeline.

The filtered images are backprojected to obtain a consensus map. As the PDB deposition associated with the dataset (PDB:6UJB) is only partially modeled, it is provided as a template to ESMFold (20 diffusion steps) to produce a complete structure, which is docked into the consensus map. The cryo-EM reconstructions contain strong density for a second Fab which is not part of the deposited sequence. Therefore, we duplicate the existing Fab from the ESMFold structure and dock it into the consensus map. This structure of integrin with two Fabs is used as the reference.

CryoCHEF is trained on 100 k-means latents and volumes are obtained through the cryoDRGN *analyze* pipeline.

A.7.5 SARS-COV-2 SPIKE

A filtered stack of 206,260 particles (box size 128, 3.28125Å/pix) is obtained from the original cryoDRGN analysis Walls et al. (2020); Zhong et al. (2021). A new cryoDRGN training run is performed for 20 epochs and latent dimension 8.

The filtered images are backprojected to obtain a consensus map. ESMFold2 with MSA input and 50 diffusion steps is used to predict a structure, which is docked into the consensus map and used as the reference.

CryoCHEF is trained on 1,000 k-means latents and volumes are obtained through the cryoDRGN *analyze* pipeline.

A.8 Additional Results

Table 3: **Generalization to unseen latents at inference time.** (↑) or lower (↓) is better. Values are mean ± std across replicates.

Dataset	*Method*	*Precision CA-RMSD* ↓	*Recall CA-RMSD* ↓	*Matched CA-RMSD* ↓	*Clash-score* ↓	*PC1 cos sim* ↑	*PC1 var diff* ↑	*All-PC W2* ↓
CAHRA	CryoCHEF + CryoDRGN ($\beta = 1.0$), 1k train	1.024 ± 0.094	1.200 ± 0.187	1.271 ± 0.238	17.650 ± 0.696	0.989	0.823	0.537
	CryoCHEF + CryoDRGN ($\beta = 1.0$), 10k infer	1.020 ± 0.090	1.190 ± 0.179	1.281 ± 0.239	17.700 ± 0.674	0.989	0.983	0.263
TRPV1	CryoCHEF + CryoDRGN ($\beta = [1.0, 0.1]$), 1k train	3.272 ± 0.206	4.316 ± 0.431	5.385 ± 1.346	21.055 ± 1.285	0.699	0.519	0.868
	CryoCHEF + CryoDRGN ($\beta = [1.0, 0.1]$), 10k infer	3.278 ± 0.201	4.266 ± 0.428	5.449 ± 1.339	21.151 ± 1.394	0.695	0.422	0.857

Table 4: **Evaluation on real datasets.** AUC-FSC is area under the FSC curve, computed between 100 structures and decoded cryoDRGN maps or backprojected maps. Arrows indicate whether higher (↑) or lower (↓) is better. Values are mean ± std across structures.

Dataset	*AUC-FSC CryoDRGN* (↑)	*AUC-FSC Backprojection* (↑)	*Clashscore* (↓)
SARS-CoV-2 spike	0.334 ± 0.004	0.230 ± 0.010	45.441 ± 0.723
HER2-Fabs	0.218 ± 0.007	0.186 ± 0.008	41.232 ± 6.176
Integrin-Fabs	0.290 ± 0.008	0.159 ± 0.017	80.743 ± 2.654

Table 5: **Ablation study.** Arrows indicate whether higher (↑) or lower (↓) is better. Values are mean ± std across replicates.

Dataset	*Ablation*	*Precision CA-RMSD* ↓	*Recall CA-RMSD* ↓	*Matched CA-RMSD* ↓	*Clash-score* ↓	*PC1 cos sim* ↑	*PC1 var diff* ↑	*All-PC W2* ↓
CAHRA	CryoCHEF (full)	1.024 ± 0.094	1.200 ± 0.187	1.271 ± 0.238	17.650 ± 0.696	0.989	0.823	0.537
	- resolution marcher	1.028 ± 0.092	1.198 ± 0.190	1.288 ± 0.214	16.403 ± 0.788	0.988	0.924	0.491
	- bond loss	1.031 ± 0.097	1.196 ± 0.178	1.267 ± 0.225	17.654 ± 1.495	0.990	0.841	0.554
	- pose refinement	1.357 ± 0.090	1.541 ± 0.185	1.715 ± 0.284	24.682 ± 1.303	0.961	0.896	0.737
TRPV1	CryoCHEF (full)	3.623 ± 0.188	4.490 ± 0.384	5.738 ± 1.284	154.746 ± 7.113	0.681	0.691	0.863
	- resolution marcher	3.774 ± 0.164	4.663 ± 0.388	5.846 ± 1.240	189.633 ± 4.537	0.652	0.714	0.932
	- bond loss	3.717 ± 0.151	4.608 ± 0.393	5.857 ± 1.253	191.008 ± 8.630	0.595	0.819	0.882
	- pose refinement	5.418 ± 0.192	6.189 ± 0.356	7.101 ± 1.065	223.841 ± 7.725	0.590	0.633	1.304