

CoCoFold2: scalable latent refinement of diffusion-based protein structure predictions from limited-particle cryo-EM data

Junwen Liao,¹ Mingxu Hu^{2,*} and Chenglong Bao^{3,4,5,*}

¹Qizhen College, Tsinghua University, 100084, Beijing, China

²Institute of Bio-Architecture and Bio-Interactions, Shenzhen Medical Academy of Research and Translation, 518100, Shenzhen, China

³Yau Mathematical Sciences Center, Tsinghua University, 100084, Beijing, China

⁴Yanqi Lake Beijing Institute of Mathematical Sciences and Applications, 101408, Beijing, China

⁵State Key Laboratory of Membrane Biology, School of Life Sciences, Tsinghua University, 100084, Beijing, China

*Corresponding authors. humingxu@smart.org.cn and clbao@mail.tsinghua.edu.cn

Abstract

Motivation: Limited-particle cryo-EM data may be insufficient for reliable de novo model building yet still provide structural constraints for correcting a corresponding sequence-based prediction. Extending this refinement setting to long protein sequences while still achieving meaningful improvements over the initial prediction remains challenging.

Results: We present CoCoFold2, a deterministic latent-refinement framework that learns a target-specific latent perturbation through a differentiable particle forward model while keeping the pretrained Protenix-v1 diffusion prior and a single diffusion realization fixed. To support long-sequence refinement, CoCoFold2 implements component-parallel refinement, as demonstrated on the four-chain, 1,683-residue merozoite surface protein 1 (MSP-1) assembly (PDB ID 6ZBH) and the 4,392-residue calcium-bound transient receptor potential melastatin 8 (TRPM8) tetramer (PDB ID 6O77). We evaluated the fixed-stochasticity latent-refinement procedure on an all-eligible post-cutoff benchmark comprising nine public protein-only EMPIAR datasets and one in-house dataset. Relative to unrefined Protenix-v1 predictions, fixed-stochasticity latent refinement yielded lower backbone root-mean-square deviation (RMSD) and higher completeness. It remained competitive with CryoAtom2 and CoCoFold, which served as map-based and prior-guided comparison methods, respectively. In three selected targets, the fixed-stochasticity procedure outperformed controls that resampled diffusion stochasticity or fine-tuned diffusion-module parameters.

Availability and Implementation: Source code is available at <https://github.com/jwliamath/CoCoFold2>.

Contact: humingxu@smart.org.cn; clbao@mail.tsinghua.edu.cn.

Keywords cryo-EM, protein structure prediction, structure refinement, AlphaFold, diffusion models

Introduction

Accurate sequence-based predictors now provide strong atomic structural hypotheses for proteins and biomolecular assemblies (Jumper et al., 2021; Abramson et al., 2024; Zhang et al., 2026). Their output is nevertheless a model-implied structure conditioned on the supplied molecular components, rather than a direct measurement of the state adopted by a specimen under a particular experimental condition. Even high-confidence predictions can differ from experimental structures in local conformation, domain orientation or assembly geometry (Terwilliger et al., 2024). We therefore treat a prediction as a structural prior that narrows the plausible model space while leaving target-specific state determination to experiment.

Single-particle cryo-EM estimates a three-dimensional density by combining many low-dose, low-signal-to-noise projection images. Limited particle counts, uneven viewing directions and unresolved conformational or compositional variation can produce anisotropic or poorly resolved maps that do not support unambiguous de novo atomic tracing (Stagg et al., 2014; Naydenova and Russo, 2017; Zhong et al., 2021; Lyumkis, 2019). The raw images nevertheless retain noisy, particle-specific projection evidence about the specimen. In this intermediate information regime, a particle subset may be insufficient for reliable de novo model building yet collectively informative enough to distinguish and correct nearby structural hypotheses. We study limited-particle cryo-EM in this refinement regime.

Existing approaches combine experimental observations with structural priors at different levels of the inference process.

Map-based model-building methods infer atomic models from reconstructed three-dimensional densities, often using sequence information, learned structural patterns or predicted structures; ModelAngelo, CryoAtom2 and DiffModeler extend the range of maps that can be interpreted (Jamali et al., 2024; Su et al., 2026; Wang et al., 2024). Iterative AlphaFold-assisted rebuilding likewise feeds map-corrected models back into prediction (Terwilliger et al., 2022). At the particle level, likelihood-based approaches use individual images to score candidate structures or reweight prior conformational ensembles, showing that raw images can evaluate structural hypotheses without treating a consensus map as the sole observation (Cossio and Hummer, 2013; Tang et al., 2023).

Prior-guided refinement methods move closer to the setting considered here. CoCoFold couples limited raw-particle supervision to an AlphaFold2 structure prior by fine-tuning its structure module (Liao et al., 2026), whereas ROCKET refines AlphaFold2/OpenFold predictions by optimizing conditioning features using cryo-EM, cryo-ET or crystallographic data (Fadini et al., 2026). Together, these methods show that experimental information can guide prediction from reconstructed maps or raw particles and can act on coordinates, model parameters or internal representations.

Recent work has begun to incorporate experimental information into diffusion-based protein structure predictors. CryoBoltz guides reverse diffusion with constraints from reconstructed cryo-EM densities, and experiment-guided AlphaFold3 injects likelihood gradients from cryo-EM maps, crystallographic data or NMR measurements during sampling to generate measurement-consistent ensembles (Raghu et al., 2025; Maddipatla et al., 2026b). CryoSampler trains an auxiliary conformational-ensemble model on top of a frozen Boltz-2 reference using ensembles of reconstructed cryo-EM maps, whereas EmbedOpt and inference-time optimization steer frozen AF3-style predictors by optimizing conditioning embeddings within or across stochastic diffusion trajectories (Shenoy et al., 2026; Li et al., 2026; Maddipatla et al., 2026a). These studies show that diffusion sampling and internal representations can be guided by experiment, but they largely address map-guided sampling, ensemble generation or related inference objectives rather than conditional refinement of one structure from limited raw particles.

This conditional refinement setting has two technical requirements. First, it must remain tractable for long multichain assemblies: particle-level optimization of a full assembly becomes increasingly demanding as sequence length and chain number grow, whereas independent component optimization can lose assembly context. A practical method therefore needs a component-parallel strategy that reduces the per-optimization cost while retaining sufficient information about the shared assembly. Second, refinement must translate weak particle evidence into meaningful improvements in backbone accuracy and model completeness. Resampling the diffusion realization between latent updates changes the decoded structure and its gradients even at the same latent perturbation, which can make experimental corrections harder to accumulate. Fixing one realization keeps successive updates directed at the same conditional objective, providing a basis for more effective refinement of the initial prediction. The key question is whether fixed-stochasticity latent refinement can deliver lower backbone RMSD and higher completeness under limited-particle conditions. Addressing both computational scalability and refinement effectiveness would extend the practical use of frozen diffusion priors to long assemblies.

Here we introduce CoCoFold2, a deterministic latent-refinement framework for this setting. CoCoFold2 caches the conditional representations produced by Protenix-v1, keeps all pretrained network parameters and one diffusion sampling realization fixed, and learns a separate target-specific latent perturbation, Δz , for each experimental dataset. It optimizes against individual particle images through a differentiable projection and contrast-transfer-function (CTF) forward model, retaining particle-specific pose and imaging information rather than using a single reconstructed density map as the refinement objective. We evaluate the method on nine public protein-only EMPIAR datasets and one in-house dataset, using unrefined Protenix-v1 predictions as starting models and CryoAtom2 and CoCoFold as comparison methods. Across this benchmark, fixed-stochasticity latent refinement reduced mean backbone RMSD and increased mean completeness relative to the unrefined predictions, while remaining competitive with both comparison methods. The modeled assemblies span 1,084–2,532 residues and two to six chains, demonstrating practical applicability to large multichain assemblies. In three selected targets with inaccurate initial predictions, fixed-stochasticity latent refinement also achieved lower backbone RMSD and higher completeness than controls that resampled diffusion stochasticity or fine-tuned the diffusion-module parameters. Finally, we assess approximate component-parallel refinement using the MSP-1 assembly as a shared-target control and the 4,392-residue TRPM8 tetramer as a larger-capacity example.

Materials and methods

Overview and optimized variables

CoCoFold2 performs target-specific refinement of a sequence-derived structure prior using limited-particle cryo-EM observations. These observations may be insufficient for reliable de novo model building but remain informative enough to correct a strong sequence-derived prior (Fig. 1a). For each target, Protenix-v1 is run once to obtain a predicted structure and cached decoder-conditioning representations. During refinement, the pretrained network remains frozen, while a latent perturbation and two particle-renderer parameters are optimized by backpropagating an observation-space loss through the differentiable renderer and diffusion decoder (Fig. 1b). The output is a single refined atomic model for each experimental dataset.

Let D_θ denote the Protenix-v1 diffusion decoder with pretrained parameters θ , let z denote the active cached pairwise conditioning tensor supplied to that decoder, and let ξ_0 denote the fixed sampling state defined below. CoCoFold2 generates

$$x(\Delta z) = D_\theta(z + \Delta z; \xi_0), \quad x(\Delta z) \in \mathbb{R}^{N_{atom} \times 3}, \quad (1)$$

where N_{atom} is the number of generated atoms and the target-specific latent perturbation Δz has the same shape as z and is initialized to zero. The perturbation is added element-wise to the active tensor immediately before the diffusion decoder, and the alternative conditioning tensor is left unchanged. The Protenix-v1 parameters, cached representations, input features and diffusion schedule are excluded from the optimizer. The trainable quantities in the particle implementation are Δz , the per-atom rendering amplitudes w , and the two-dimensional Gaussian widths σ .

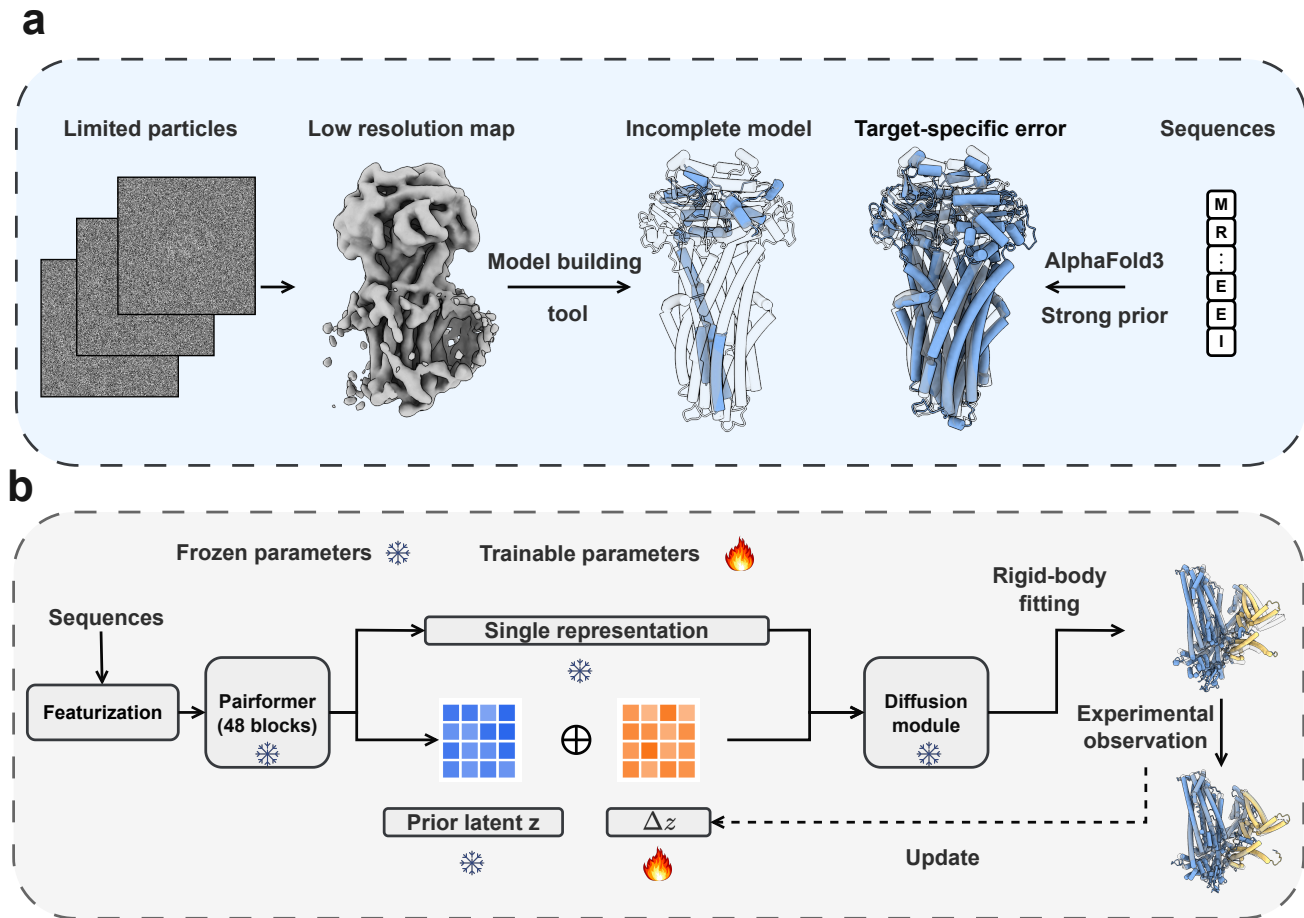


Figure 1 CoCoFold2 refines frozen structure-prediction priors by deterministic latent optimization. **a**, Weak cryo-EM observations create an underdetermined modeling problem. Sparse particle images yield low-resolution density, making model building difficult. Sequence-based predictors such as AlphaFold-style models provide strong structural priors, but these priors can retain errors in domain placement, local geometry or assembly. **b**, Pipeline of CoCoFold2. The input sequence is passed through a frozen Pairformer and diffusion module to produce a prior latent representation and structure. CoCoFold2 optimizes a trainable latent update, Δz , while all network parameters remain fixed. The decoded structure is rigidly fitted to the experimental observation, and the experimental objective back-propagates gradients to Δz .

Fixed stochasticity

A central design choice is to keep the diffusion sampling state fixed during latent refinement. The output of a diffusion decoder depends not only on $z + \Delta z$, but also on stochastic variables such as the initial noise and the random draws used during denoising. If these variables are resampled between optimization steps, the same value of Δz can produce different structures, so successive gradients correspond to different conditional objectives (Supplementary Note S1). CoCoFold2 therefore fixes the sampling state at ξ_0 , making $\Delta z \mapsto D_\theta(z + \Delta z; \xi_0)$ reproducible within the specified implementation.

Before each sampler call, NumPy and PyTorch CPU/CUDA generators are reset to seed 42, so the same initial Gaussian noise and random-number stream are reused. This branch also disables random reorientation. The cached schedule, one diffusion sample and five denoising steps are unchanged during optimization and checkpoint reconstruction; with $\gamma_0 = \gamma_{\min} = 0$, the per-step stochastic perturbation has zero amplitude. Particle shuffling changes the observations in an update but not the mapping in Eq. 1. Fixed

stochasticity therefore defines one conditional objective, not an expectation over diffusion realizations.

Particle forward model and optimization

Coordinate-frame initialization.

The initial Protenix-v1 prediction was rigidly placed in the reconstruction and particle-pose frame in ChimeraX. A Kabsch fit between the deterministic decoder coordinates and this fitted copy, with matched atom ordering, defines R_0 and t_0 :

$$\tilde{x}(\Delta z) = x(\Delta z)R^T + 1t^T. \quad (2)$$

By default, $(R, t) = (R_0, t_0)$ remains fixed. A Kabsch transform replaces it only when $\text{tr}(R_{\text{current}}R^T) < 2.5$, providing a discrete flip safeguard outside the optimizer. Upstream particle rotations, translations and CTF parameters remain fixed.

Differentiable renderer and particle loss.

For particle i , the upstream pose gives projected atom centers $c_{ij} \in \mathbb{R}^2$, which are rendered as

$$G_i(u) = \sum_{j=1}^{N_{atom}} w_j \exp \left[-\frac{1}{2} \sum_{d \in \{x,y\}} \left(\frac{u_d - c_{ij,d}}{\sigma_{j,d}} \right)^2 \right], \quad (3)$$

where the amplitudes w_j are initialized to atomic number, widths $\sigma_{j,d}$ to $3/(\pi\sqrt{2})$, and Gaussians are truncated at five standard deviations. After pixel-size conversion, grid scaling and in-plane translation, the dataset-specific sign is applied and the Fourier transform is multiplied by the particle CTF.

Let $\hat{P}_i$ denote the CTF-modulated predicted image, $\hat{Y}_i$ the observed image, and B_q the integer Fourier ring of radius q . The active implementation computes

$$\text{FRC}_i(q) = \frac{\text{Re} \left[\sum_{k \in B_q} \hat{P}_i(k) \overline{\hat{Y}_i(k)} \right]}{\left(\sum_{k \in B_q} |\hat{P}_i(k)|^2 \right)^{1/2} \left(\sum_{k \in B_q} |\hat{Y}_i(k)|^2 \right)^{1/2} + 10^{-8}}. \quad (4)$$

For an outer batch of size B , microbatch Fourier ring correlation (FRC) terms are accumulated before one optimizer step:

$$\mathcal{L}_{\text{FRC}} = -\frac{1}{BR_{\text{max}}} \sum_{i=1}^B \sum_{q=1}^{R_{\text{max}}-1} \text{FRC}_i(q). \quad (5)$$

Here, $R_{\text{max}} = \lfloor (H/2)(2a/r_{\text{cut}}) \rfloor$, where H , a and r_{cut} denote the box size, pixel size and requested real-space cutoff, respectively. We set $r_{\text{cut}} = 2a$, which gives the Nyquist radius.

Renderer parameters are constrained by the soft penalties

$$\mathcal{R}_{\sigma} = \text{mean}_{j,d} [\text{ReLU}(\sigma_{j,d} - 0.8) + \text{ReLU}(0.1 - \sigma_{j,d})], \quad (6)$$

$$\mathcal{R}_w = \text{mean}_j [\text{ReLU}(w_j - 20) + \text{ReLU}(1 - w_j)]. \quad (7)$$

These terms constrain the renderer, not protein coordinates, and the particle objective is $\mathcal{L}_{\text{particle}} = \mathcal{L}_{\text{FRC}} + \mathcal{R}_{\sigma} + \mathcal{R}_w$.

Optimization and checkpointing.

AdamW learning rates for Δz , w and σ are 10^{-2} , 10^{-2} and 5×10^{-3} ; other arguments use PyTorch defaults. Network parameters have `requires_grad=False` and are absent from the optimizer, while the decoder computational graph is retained to update Δz . Gradients are accumulated over the microbatches of one shuffled outer batch before an optimizer step. Ten epochs are run, and a complete state checkpoint and coordinate file are written after every epoch. All reported CoCoFold2, Resampled, Fine-tuning and component-parallel endpoints were taken from the final (10th) epoch, without reference-based checkpoint selection.

Datasets, comparison methods and evaluation

We screened post-30 September 2021 EMPIAR entries for protein-only targets with an atomic reference, particle images and pose/CTF metadata, excluding ambiguous particle-map or PDB-map links, severe heterogeneity, extended assemblies, missing inputs and cases outside the prespecified computational scope. Nine public targets and the prespecified in-house human Cav1.2 L-type calcium-channel complex (PDB ID 8WE6) form the benchmark (Supplementary Tables S1 and S2). The public targets were the

Danio rerio tRNA ligase complex (PDB ID 9I8V), human hormone-sensitive lipase (HSL) dimer (PDB ID 8ZVQ), trypanosomal aquaglyceroporin 2 (AQP2) tetramer (PDB ID 8OFZ), Salmonella ArnC glycosyltransferase (PDB ID 8VXH), assimilatory sulfite reductase minimal dimer (PDB ID 9C91), RhlR-PqsE quorum-sensing complex (PDB ID 8DQ0), human neurokinin 1 receptor (NK1R)-miniGs/q70 complex (PDB ID 7RMI), the MSP-1 assembly, and the Patched1 (PTCH1)-TI23 nanobody complex (PDB ID 7K65). For the AQP2 tetramer, the deposited coordinate model contains one 312-residue protomer, whereas the particle stack and reconstructed map represent a homotetramer; the refinement assembly therefore contains four protomer copies and 1,248 modeled residues.

The primary comparisons were unrefined Protenix-v1 using the same sequence, CryoAtom2 using the sequence and a map reconstructed from the same particle subset, and CoCoFold using the same particles and metadata. ROCKET was not run in the ten-target benchmark because its audited workflow did not cover these multichain and sequence-length inputs; it appears only in labeled exploratory cases. Map-derived, cryo-ET-like and X-ray operators are described in Supplementary Note S2. EmbedOpt was not included because its public cryo-EM evaluation is chain-specific and its released workflow uses a deposited structure for reward-time Kabsch alignment, whereas CoCoFold2 includes multichain targets and uses only the Protenix-v1 prediction for alignment.

ModelAngelo evaluation utilities align and residue-match each output to the deposited reference with a 3 Å maximum matching distance. We report backbone and Cα RMSD, recall, precision and completeness (Supplementary Table S3, Fig. 2).

Fixed-stochasticity and optimization-variable ablations

The ablation benchmark comprises the MSP-1 assembly, the RhlR-PqsE complex and the HSL dimer, selected before the ablation analysis because their unrefined Protenix-v1 models had substantial correction headroom in initial backbone RMSD and completeness (Supplementary Table S4). Four endpoints are reported for each target. *Protenix* is the unrefined deterministic prediction. *CoCoFold2* optimizes Δz , w and σ under the fixed sampling state described above. *Resampled* denotes the resampled-stochasticity ablation, not a separate prediction method: the deterministic branch is disabled, so the initial diffusion noise is not reset between sampler calls and random reorientation is applied during denoising. *Fine-tuning* replaces the latent update with optimization of diffusion-module parameters. The same target particle inputs and retrospective evaluator are used across conditions.

Approximate component-parallel multi-GPU refinement

For assemblies for which repeated full-complex forward and backward passes are impractical on a single GPU, complete-chain groups are assigned one per GPU (Fig. 3a). Each GPU owns a frozen decoder, local conditioning cache, Δz_k and renderer parameters. In *Independent-component*, sequences are partitioned before separate cache generation, omitting cross-component Pairformer and diffusion conditioning. In *Contextual-component*, the full assembly is passed once through the Pairformer; component single rows and exact diagonal blocks of full-complex z are extracted and atom-local features are rebuilt. This retains context encoded by the Pairformer, although component-local decoding still omits cross-component

diffusion attention and joint coordinate generation. Neither path is equivalent to full-complex refinement.

All components are placed in one experimental coordinate frame. For each particle, GPU k renders an unshifted local projection P_{ik} using a shared projected origin. An autograd-aware distributed sum constructs

$$P_i = \sum_{k=1}^K P_{ik}. \quad (8)$$

The common translation, sign, CTF and FRC loss are applied once after this autograd-aware sum. The replicated loss is divided by GPU count and renderer penalties use global atom/width counts, so one particle-space objective returns gradients to every Δz_k without updating decoder weights.

The MSP-1 assembly is split A versus B-D and the 4,392-residue TRPM8 tetramer is split A-B versus C-D (Supplementary Table S5). Contextual-component still requires one full $N_{token} \times N_{token}$ Pairformer pass; it distributes only repeated component-local decoding, rendering and backpropagation. Cache generation and refinement must therefore be profiled separately.

Results

Comparison under scarce-particle conditions

We first evaluated CoCoFold2 on an all-eligible post-cutoff benchmark. We screened public EMPIAR entries released after 30 September 2021 for protein-only targets with an atomic reference, particle images and pose/CTF metadata; we excluded ambiguous particle-map or PDB-map links, severe heterogeneity, extended assemblies, missing inputs and cases outside the prespecified computational scope. The benchmark comprised nine public datasets and one prespecified in-house dataset of the human Cav1.2 L-type calcium-channel complex (PDB ID 8WE6). The selected particle subsets contained 1.2k–2.6k particles, the corresponding maps had resolutions of 4.05–9.70 Å, and the modeled assemblies contained 1,084–2,532 residues (Supplementary Tables S1 and S2).

Protenix-v1 served as the prediction-only frozen-prior baseline. CryoAtom2 was run with the sequence and a map reconstructed from the same particle subset, whereas CoCoFold2 and CoCoFold used the selected raw particles and associated metadata. ROCKET was not included in this ten-target benchmark because its audited workflow did not cover the relevant multichain and sequence-length inputs; it was used only in the labeled exploratory comparisons. Thus, the CryoAtom2 comparison uses a reconstructed map and the CoCoFold2 comparison uses raw particles, as specified by each method's input modality.

We report backbone and C α RMSD, recall, precision and completeness using the ModelAngelo evaluation utilities after aligning and residue-matching each output to the reference with a maximum matching distance of 3 Å. RMSD measures the geometric deviation of matched backbone or C α atoms from the deposited reference; recall is the fraction of reference residues or structural elements correctly recovered, precision is the fraction of predicted residues or structural elements correctly placed, and completeness is the fraction of the reference sequence or structure successfully modeled. All endpoints were taken from the final optimization epoch without reference-based checkpoint selection.

Across the ten real-particle benchmark cases, CoCoFold2 shifted the performance distribution toward lower RMSD values and higher

completeness (Fig. 2a, Supplementary Table S3). Compared with Protenix, CoCoFold2 reduced the mean backbone RMSD from 1.45 Å to 1.07 Å and the mean C α RMSD from 1.34 Å to 1.03 Å. Mean completeness increased from 0.68 to 0.89, while recall and precision were maintained or improved in most cases. Relative to CryoAtom2, whose input was a map reconstructed from the same limited-particle subset, CoCoFold2 achieved lower backbone RMSD in 6 of 10 cases and higher completeness in 9 of 10 cases. This is a contextual comparison of different observation representations rather than a same-input head-to-head comparison. Compared with CoCoFold across the 10 cases, CoCoFold2 achieved lower backbone RMSD in 7 of 10 cases, lower C α RMSD in 8 of 10 cases, and higher completeness in 9 of 10 cases. Mean backbone RMSD and mean completeness were 1.41 Å and 0.83 for CoCoFold. These results support CoCoFold2 as a particle-guided refinement strategy that improves frozen diffusion-prior predictions while remaining competitive with prior-guided model fine-tuning and improving upon it in several cases.

Case-level paired comparisons showed improvements across multiple benchmark cases (Fig. 2b). The largest gains were observed for initially incomplete or mispositioned predictions, including the MSP-1 assembly, the RhlR-PqsE complex, the Cav1.2 complex and the HSL dimer. In these cases, the Protenix prior was close enough to the correct structural basin for correction, but still required constraints from the corresponding particle images to improve domain placement or model completeness. By contrast, cases with already accurate Protenix predictions or stronger background-dominated particle signals showed smaller improvements, suggesting that CoCoFold2 is most effective when the frozen prior is approximately correct but under-constrained by sequence information alone.

The aggregate benchmark shows that CoCoFold2 improves model accuracy and completeness across limited-particle cryo-EM datasets, but summary metrics alone do not reveal which structural errors are corrected. We therefore inspected representative cases by overlaying the predicted models with the corresponding reference structures. Across the examples shown in Fig. 2c, CryoAtom2 frequently produced incomplete traces under weak experimental signal, whereas the sequence-based prior from Protenix provided a more complete initial model but retained errors in local placement, domain orientation or assembly geometry. CoCoFold2 corrected a subset of these errors while preserving the overall fold of the prediction prior.

For the MSP-1 assembly, the Protenix prior captured the global architecture but showed local deviations in peripheral and interfacial regions. After CoCoFold2 refinement, the corresponding regions were closer to the reference structure (Fig. 2c,d). For the RhlR-PqsE complex, the comparison highlights a different type of correction. The Protenix model captured the overall architecture of the complex but exhibited errors in the relative placement of structural elements. CoCoFold2 improved the agreement of these regions with the reference structure, suggesting that latent refinement can adjust domain or subunit-level geometry when the limited-particle observation contains sufficient information to constrain the update. In contrast, the model-building result remained less complete, consistent with the difficulty of tracing a full atomic model directly from sparse experimental signal. The HSL dimer example further illustrates that CoCoFold2 is most useful when the prediction prior is already close enough to the correct fold for experimental refinement to be informative. In this case, CoCoFold2 preserved the overall topology of the prior while improving the agreement of several

regions with the reference. These examples support the interpretation that CoCoFold2 primarily corrects deviations between a strong sequence-derived prior and the experimentally observed structure rather than generating structures de novo from weak cryo-EM data.

Component-parallel refinement of large assemblies

Having established overall refinement gains on the real-particle benchmark, we next tested whether component-parallel refinement could extend the workflow to long multichain assemblies.

Refining large protein assemblies imposes a substantially greater practical memory burden than producing a single static prediction. Each CoCoFold2 iteration repeatedly evaluates the frozen diffusion decoder, transforms the generated coordinates, renders particle projections using the upstream poses and CTF parameters, and backpropagates the resulting particle-space loss to the latent variables. As the number of tokens and atoms increases, retaining the activations required for repeated forward and backward passes becomes increasingly demanding even when the Pairformer representations are cached. We therefore implemented an approximate component-parallel refinement strategy that distributes the repeated decoder and rendering computations across multiple GPUs (Fig. 3a). In the Independent-component strategy, the input sequences are first divided into components, and each component is processed separately by the frozen Pairformer to produce component-only cached representations. In the Contextual-component strategy, the complete assembly is processed once by the frozen Pairformer, after which contextualized single representations and diagonal pair blocks are extracted for the same components. Each GPU then optimizes a component-specific perturbation, Δz_k , through a frozen decoder. The resulting component projections are summed in a common coordinate frame, and one global particle-space loss propagates gradients to all Δz_k . We used the MSP-1 assembly, for which a full-complex CoCoFold2 result was available, to assess the quality retained by the two approximations, and used the 4,392-residue TRPM8 tetramer to examine their practical capacity beyond the current full-complex coordinate-generation path.

For the component-parallel MSP-1 experiment, we used approximately 3,000 CryoSieve-selected particles and divided the assembly into chain A and chains B–D (Fig. 3b,c). Full-complex refinement reduced the backbone RMSD from 1.41 to 0.61 Å and increased completeness from 0.54 to 0.97. Independent-component refinement reduced the backbone RMSD from 1.59 to 0.99 Å and increased completeness from 0.44 to 0.81, whereas Contextual-component refinement reduced the backbone RMSD from 1.27 to 0.66 Å and increased completeness from 0.85 to 0.96. Thus, the contextual result approached the full-complex refinement and remained more accurate and complete than the independently conditioned result. The isolated-chain comparison in Fig. 3c helps explain this difference. Under Independent-component initialization, chain B deviated markedly from the reference after full-assembly cross-component context was omitted; the global particle-space objective corrected much of this error but left visible residual deviations. By contrast, the contextual initialization retained a more reference-like chain-B conformation because its cached representation had already been conditioned on the complete assembly. These observations are consistent with the view that full-assembly prior context improves component initialization and subsequent refinement in this case.

The TRPM8 tetramer is a four-chain assembly containing 4,392 protein residues. We refined the AB and CD components on separate GPUs using 11,020 CryoSieve-selected particles (Fig. 3d,e). Contextual-component refinement reduced the backbone RMSD from 2.20 to 0.98 Å and increased completeness from 0.27 to 0.86. Independent-component refinement reduced the backbone RMSD from 2.23 to 0.91 Å and increased completeness from 0.20 to 0.87. Both strategies therefore produced substantial corrections, while the independent result was slightly better in the final whole-structure metrics in this single run. In the Protenix-v1 implementation used here, the full-assembly Pairformer calculation produced the 4,392-token pair representation, but the subsequent full-complex coordinate-generation path failed because of an indexing overflow. The extracted contextual component caches matched their corresponding source slices exactly in our numerical audit and contained no non-finite values, allowing the full-assembly prior to be used for component-local refinement despite the absence of a valid full-complex coordinate output. The TRPM8 experiment therefore demonstrates the feasibility of the component-parallel workflow at this larger size.

Fixed-stochasticity latent refinement improves structural correction

We then isolated the contribution of the fixed diffusion realization and the optimized variables in selected stress-test cases.

We performed a focused ablation on the MSP-1 assembly, the RhlR–PqsE complex and the HSL dimer, three cases from the ten-target benchmark whose unrefined Protenix-v1 models combined appreciable backbone error with incomplete model completeness (Supplementary Table S4). We examined two design choices. With the target-specific latent perturbation, Δz , as the optimized structural variable, we compared a fixed diffusion realization with resampling of ξ during optimization. With ξ fixed, we compared latent refinement with a control that set $\Delta z = 0$ and instead updated the Protenix-v1 diffusion-module parameters.

Holding the latent-update parameterization fixed, the fixed-stochasticity condition yielded lower backbone RMSD and higher completeness than resampling for all three selected targets (Fig. 4a; Supplementary Table S4). Across the three cases, the descriptive mean backbone RMSD decreased from 1.53 Å with resampling to 0.97 Å with a fixed realization, while mean completeness increased from 0.74 to 0.95. In the representative global and local comparisons, the fixed-realization result placed the highlighted region closer to the reference than the resampled control (Fig. 4b,c).

With stochasticity fixed, latent refinement also yielded lower backbone RMSD and higher completeness than diffusion-module fine-tuning in all three cases (Fig. 4a). The descriptive mean backbone RMSD was 0.97 Å for CoCoFold2 and 1.51 Å for fine-tuning, and the corresponding mean completeness values were 0.95 and 0.67. The representative overlays show that the fine-tuned structure retained a larger local displacement in the highlighted region, whereas latent refinement placed this region closer to the reference (Fig. 4b,c). Thus, within these selected cases and the tested optimization budget, adapting a persistent latent perturbation was more effective than updating the diffusion-module parameters.

For the single-structure refinement objective considered here, fixing ξ establishes a consistent mapping from each latent update to atomic coordinates, allowing successive updates to be evaluated under the

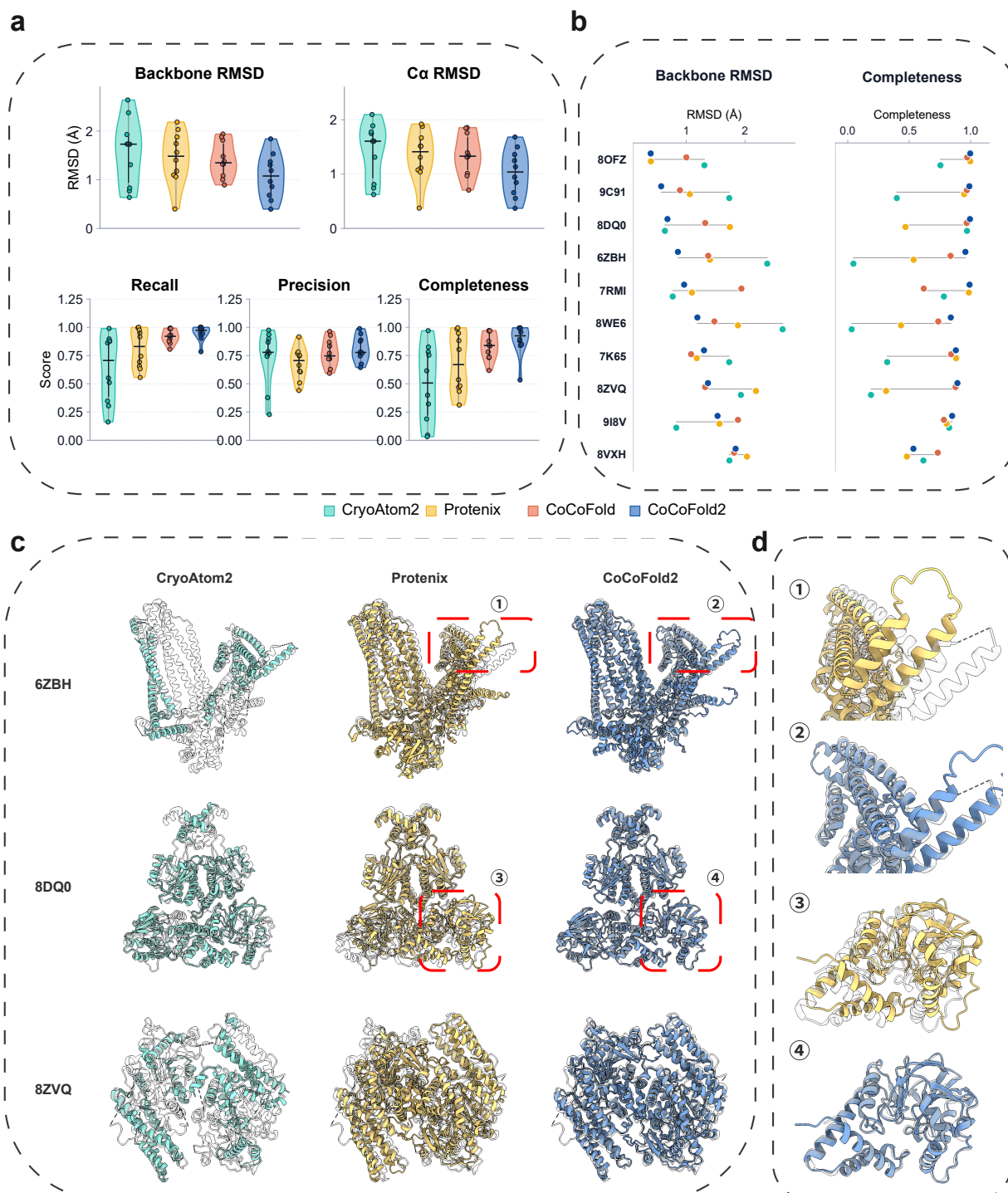


Figure 2 Performance on real limited-particle cryo-EM datasets. **a**, Distribution of structure-quality metrics across ten real limited-particle datasets for CryoAtom2, Protenix, CoCoFold and CoCoFold2. **b**, Case-level paired comparison of backbone RMSD and completeness. Each row corresponds to one benchmark case, and points indicate the performance of the compared methods. Connecting lines indicate the range of method performance within each case and highlight cases in which CoCoFold2 improves geometric accuracy or model completeness under limited-particle conditions. **c**, Whole-structure comparisons on three limited-particle cryo-EM cases. Predicted models from CryoAtom2, Protenix and CoCoFold2 are overlaid on the reference structures, shown in gray. Method predictions are colored by method: CryoAtom2 in cyan, Protenix in yellow and CoCoFold2 in blue. Red brackets mark regions with pronounced differences between the prediction prior and the CoCoFold2-refined model. **d**, Local views of representative corrected regions. Enlarged views of the numbered regions in c show examples in which the Protenix prior contains local displacement, incomplete placement or incorrect relative geometry, whereas CoCoFold2 better follows the reference structure after experimental latent refinement. The gray model denotes the reference structure.

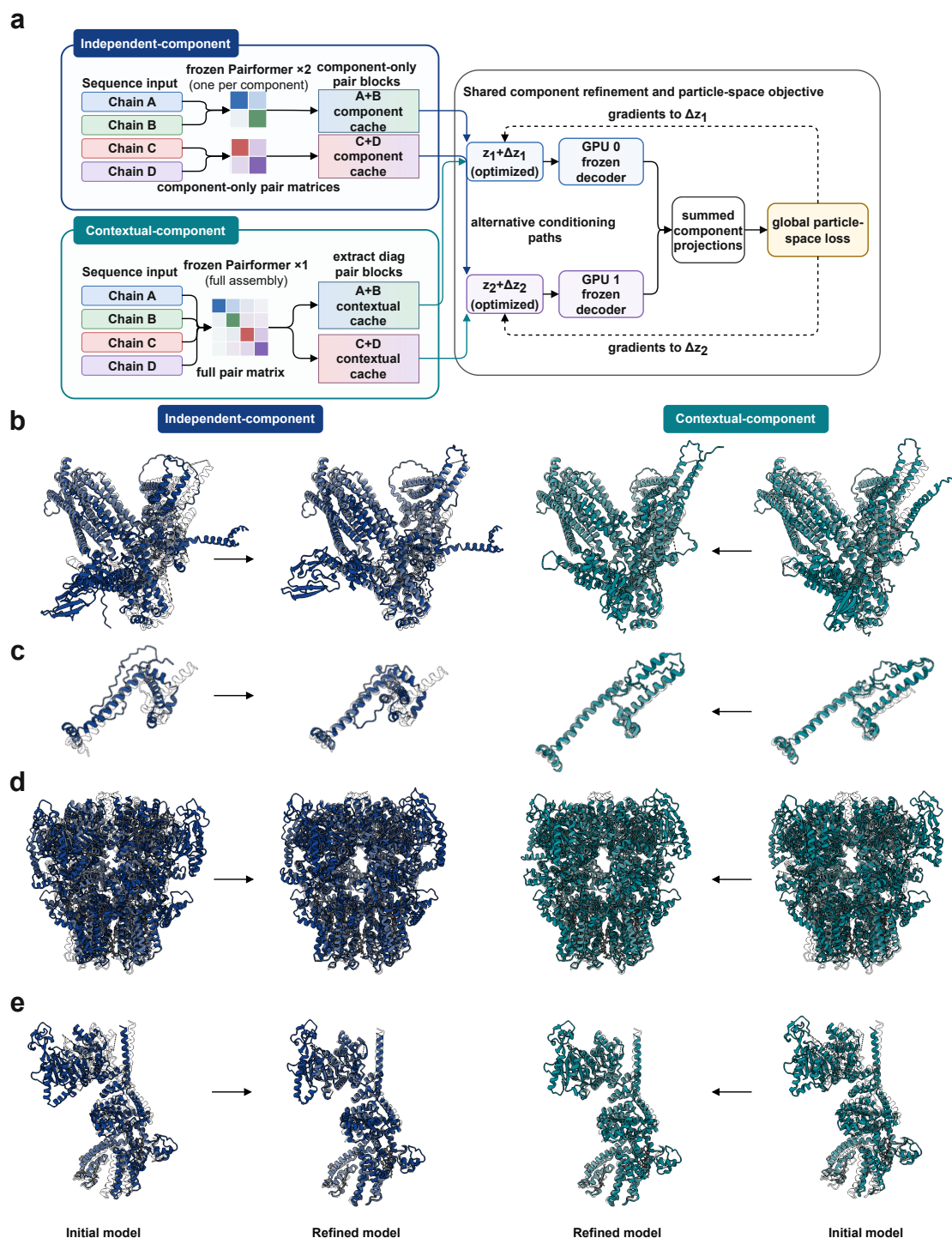


Figure 3 Component-parallel CoCoFold2 refinement for large multichain assemblies. **a**, Schematic comparison of the Independent-component and Contextual-component conditioning strategies. In the Independent-component strategy, the frozen Pairformer is applied separately to each component, omitting cross-component prior context. In the Contextual-component strategy, the complete assembly is first processed by the frozen Pairformer, after which diagonal contextualized component representations, z_1 and z_2 , are extracted. In both strategies, component-specific latent perturbations, Δz_1 and Δz_2 , are optimized through frozen diffusion decoders and a shared global particle-space objective. **b**, Whole-assembly comparison of the initial and refined MSP-1 structures using the Independent-component and Contextual-component strategies. **c**, Corresponding isolated chain B views of the MSP-1 assembly. **d**, Whole-assembly comparison for the larger TRPM8 tetramer. **e**, Corresponding isolated-chain A views of the TRPM8 tetramer. The experimental reference is shown in gray; the Independent-component and Contextual-component results are shown in navy and teal, respectively.

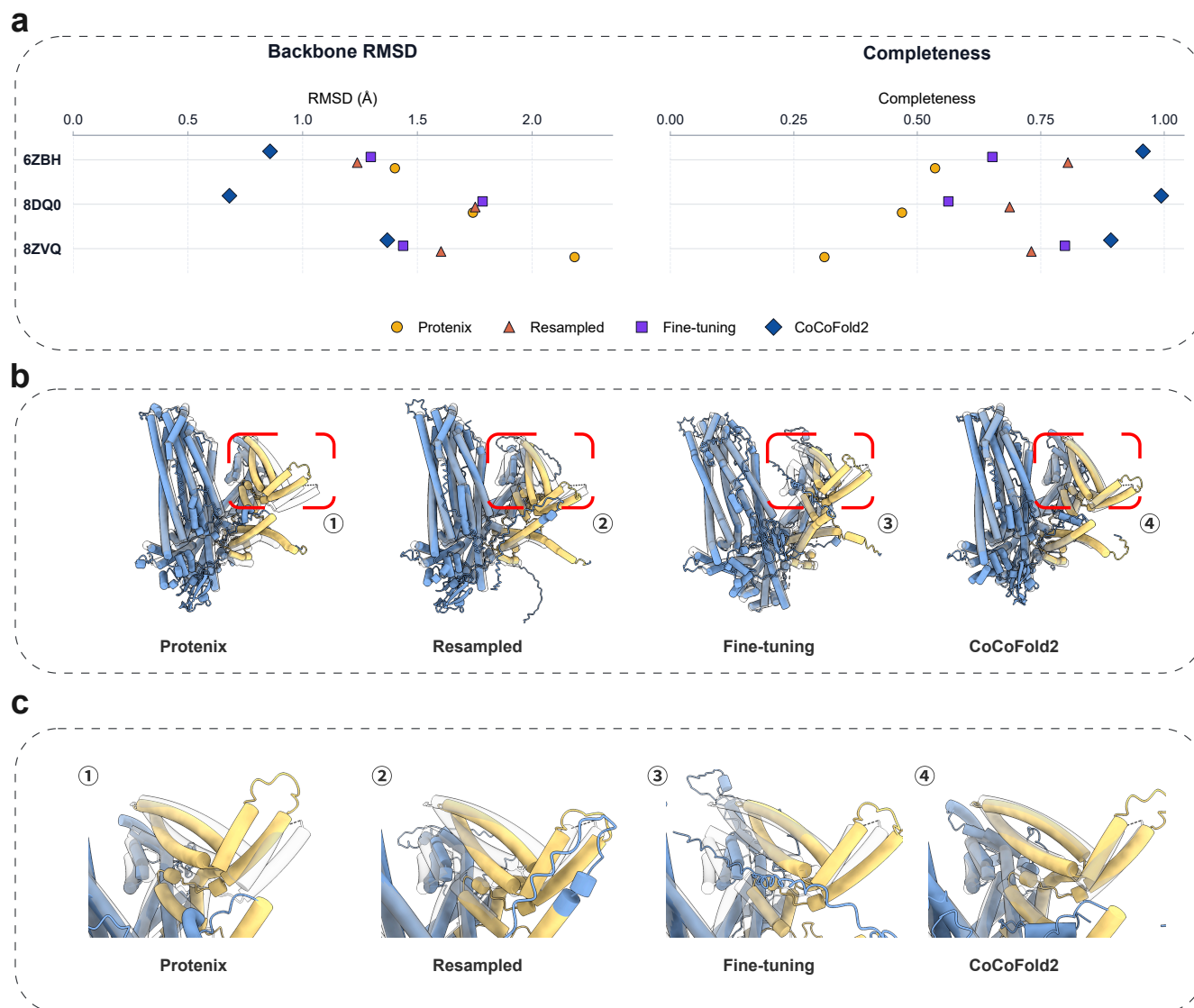


Figure 4 Ablation of sampling stochasticity and optimization variables. a, Backbone RMSD and completeness for the MSP-1 assembly, RhlR-PqsE complex and HSL dimer. Three benchmark targets selected for appreciable error and incomplete completeness in their initial Protenix-v1 predictions. *Protenix* denotes the unrefined prediction. *Resampled* optimizes the target-specific latent perturbation while resampling the diffusion stochasticity during optimization. *Fine-tuning* keeps the diffusion realization fixed, sets $\Delta z = 0$ and updates the selected diffusion-module parameters. *CoCoFold2* optimizes Δz and the renderer parameters under one fixed diffusion realization. Each point represents the reported endpoint for one target and condition; lower backbone RMSD and higher completeness are better. **b, Whole-structure overlays for the MSP-1 assembly under the four conditions.** The deposited reference structure is shown in gray. Blue and yellow identify corresponding structural regions. Red brackets and numbered labels identify the regions enlarged in c. **c, Local views of the numbered regions in b.** These views show the different structural corrections produced by the resampled-stochasticity, fine-tuning and fixed-stochasticity latent-refinement conditions. Numerical endpoints are reported in Supplementary Table S4.

same diffusion realization. The three stress-test cases support this fixed-stochasticity design under the tested protocol, but they do not imply that fixed stochasticity is generally preferable for expectation-based objectives.

Discussion

CoCoFold2 uses particle-level observations to correct an informative Protenix-v1 prediction prior without updating pretrained network

weights. Across ten real limited-particle datasets, it improved backbone agreement and completeness relative to the unrefined prior and compared favorably with map-based model building and particle-guided fine-tuning under their respective input modalities. These results support a practical regime in which a particle subset that is insufficient for reliable de novo tracing still provides enough information to refine a nearby structural hypothesis. CoCoFold2 is therefore most relevant when an informative prior is available but remains underconstrained by sequence information alone.

For this single-structure objective, fixing the diffusion realization makes the coordinate response to Δz reproducible, while latent optimization confines experimental adaptation to a persistent target-specific variable. In the three selected targets with substantial correction headroom, this design gave better RMSD and completeness endpoints than resampled stochasticity or network fine-tuning. The conditional variance analysis offers a possible explanation, not a convergence result for the AdamW procedure used here. These findings concern conditional refinement; posterior or ensemble objectives may instead require diversity across diffusion realizations.

Component-parallel refinement was evaluated on the MSP-1 assembly, for which a full-complex result was available, and on the 4,392-residue TRPM8 tetramer. For the MSP-1 assembly, retaining full-assembly Pairformer context before component extraction produced a result closer to full-complex refinement than independent conditioning, indicating a quality cost when intercomponent prior information is removed. The TRPM8 result establishes feasibility at the larger tested size. Contextual refinement still requires one full-complex Pairformer pass, and neither component strategy retains cross-component diffusion attention.

Several limitations remain. First, CoCoFold2 depends on the quality of the initial prediction prior: if the prior is far from the correct structural basin, or if the experimental signal is dominated by noise, background, pose error or CTF error, latent refinement may fail or produce only limited improvement. Second, the present formulation refines a single structure and does not explicitly model conformational ensembles or continuous heterogeneity. Future work should therefore focus on heterogeneity-aware refinement, in which multiple latent corrections could represent alternative conformational states rather than forcing all signal into a single structure. Supplementary latent-scaling and cross-quality analyses characterize case-specific corrections, whereas the homologous-target reuse and map-derived, cryo-ET-like and X-ray experiments are exploratory in this stage (Supplementary Notes S2–S4).

Conflicts of interest

The authors declare that they have no competing interests.

Funding

This work was supported by the National Key Research and Development Program of China (grant no. 2025YFA1309602) and the National Natural Science Foundation of China (grant nos. 12625119, W2612010).

Data availability

The real limited-particle datasets were downloaded from EMPIAR and comprise the *Danio rerio* tRNA ligase complex (PDB ID 9I8V), human hormone-sensitive lipase (HSL) dimer (PDB ID 8ZVQ), trypanosomal aquaglyceroporin 2 (AQP2) tetramer (PDB ID 8OFZ), *Salmonella* ArnC glycosyltransferase (PDB ID 8VXH), assimilatory sulfite reductase minimal dimer (PDB ID 9C91), RhlR–PqsE quorum-sensing complex (PDB ID 8DQ0), human neurokinin 1 receptor (NK1R)–miniGs/q70 complex (PDB ID 7RMI), merozoite surface protein 1 (MSP-1) assembly (PDB ID 6ZBH) and Patched1

(PTCH1)–TI23 nanobody complex (PDB ID 7K65). The human Cav1.2 L-type calcium-channel complex (PDB ID 8WE6) was collected in-house. The map-derived datasets were generated from EMDB maps. The cryo-ET-like and X-ray scope tests used GroEL (PDB ID 8P4P) and the hepatocyte growth factor activator inhibitor 1 (HAI-1) extracellular domain (PDB ID 5H7V), respectively. The homologous-target diagnostic used tuna apo P-glycoprotein (PDB ID 9NFJ) and human apo P-glycoprotein (PDB ID 8GMG). The accession codes and corresponding information for all datasets are listed in the Supplementary Tables. Code and tutorials are available at <https://github.com/jwliaomath/CoCoFold2>.

Author contributions statement

Conceptualization, J.L., M.H., and C.B.; methodology, J.L., M.H., and C.B.; investigation, J.L., M.H., and C.B.; writing—original draft, J.L.; writing—review & editing, J.L., M.H., and C.B.; funding acquisition, M.H. and C.B.; resources, M.H. and C.B.; supervision, M.H. and C.B.

References

- J. Abramson, J. Adler, J. Dunger, R. Evans, T. Green, A. Pritzel, O. Ronneberger, L. Willmore, A. J. Ballard, J. Bambrick, S. W. Bodenstein, D. A. Evans, C.-C. Hung, M. O'Neill, D. Reiman, K. Tunyasuvunakool, Z. Wu, A. Žemgulytė, E. Arvaniti, C. Beattie, O. Bertolli, A. Bridgland, A. Cherepanov, M. Congreve, A. I. Cowen-Rivers, A. Cowie, M. Figurnov, F. B. Fuchs, H. Gladman, R. Jain, Y. A. Khan, C. M. R. Low, K. Perlin, A. Potapenko, P. Savy, S. Singh, A. Stecula, A. Thillaisundaram, C. Tong, S. Yakneen, E. D. Zhong, M. Zielinski, A. Židek, V. Bapst, P. Kohli, M. Jaderberg, D. Hassabis, and J. M. Jumper. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, 630 (8016):493–500, June 2024. ISSN 0028-0836, 1476-4687. doi: 10.1038/s41586-024-07487-w. URL <https://www.nature.com/articles/s41586-024-07487-w>.
- P. Cossio and G. Hummer. Bayesian analysis of individual electron microscopy images: towards structures of dynamic and heterogeneous biomolecular assemblies. *Journal of Structural Biology*, 184(3):427–437, 2013. doi: 10.1016/j.jsb.2013.10.006. URL <https://doi.org/10.1016/j.jsb.2013.10.006>.
- A. Fadini, M. Li, A. J. McCoy, S. Banjara, H. Okumura, E. Napier, P. Fontana, A. R. Khan, L. Jovine, T. C. Terwilliger, R. J. Read, D. R. Hekstra, and M. AlQuraishi. AlphaFold as a Prior: Experimental Structure Determination Conditioned on a Pretrained Neural Network. *Nature Methods*, 23(4):785–795, 2026. doi: 10.1038/s41592-026-03047-4. URL <https://www.nature.com/articles/s41592-026-03047-4>.
- K. Jamali, L. Käll, R. Zhang, A. Brown, D. Kimanius, and S. H. W. Scheres. Automated model building and protein identification in cryo-EM maps. *Nature*, 628(8007):450–457, Apr. 2024. ISSN 0028-0836, 1476-4687. doi: 10.1038/s41586-024-07215-4. URL <https://www.nature.com/articles/s41586-024-07215-4>.
- J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Židek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman,

- E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstern, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, and D. Hassabis. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873):583–589, Aug. 2021. ISSN 0028-0836, 1476-4687. doi: 10.1038/s41586-021-03819-2. URL <https://www.nature.com/articles/s41586-021-03819-2>.
- M. Li, J. Han, P. Cossio, and L. Wu. Robust inference-time steering of protein diffusion models via embedding optimization, 2026. URL <https://arxiv.org/abs/2602.05285>.
- J. Liao, D. Zheng, H. Zhang, L. Zhang, M. Hu, and C. Bao. Fine-tuning AlphaFold with limited cryo-EM observations. *Communications Chemistry*, 9(1):95, Jan. 2026. ISSN 2399-3669. doi: 10.1038/s42004-026-01899-7. URL <https://www.nature.com/articles/s42004-026-01899-7>.
- D. Lyumkis. Challenges and opportunities in cryo-EM single-particle analysis. *Journal of Biological Chemistry*, 294(13):5181–5197, Mar. 2019. ISSN 00219258. doi: 10.1074/jbc.REV118.005602. URL <https://linkinghub.elsevier.com/retrieve/pii/S0021925820355666>.
- A. Maddipatla, A. Rzyayev, M. Pegoraro, M. Pacesa, P. Schanda, A. Marx, S. Vedula, and A. M. Bronstein. Inference-time optimization for experiment-grounded protein ensemble generation, 2026a. URL <https://arxiv.org/abs/2602.24007>.
- A. Maddipatla, N. Sellam Bojan, M. Bojan, V. Masalitin, S. Vedula, P. Schanda, A. Marx, and A. M. Bronstein. Experiment-guided AlphaFold3 resolves measurement-consistent protein ensembles. *Nature Biotechnology*, 2026b. doi: 10.1038/s41587-026-03166-5. URL <https://www.nature.com/articles/s41587-026-03166-5>.
- K. Naydenova and C. J. Russo. Measuring the effects of particle orientation to improve the efficiency of electron cryomicroscopy. *Nature Communications*, 8(1):629, 2017. doi: 10.1038/s41467-017-00782-3. URL <https://www.nature.com/articles/s41467-017-00782-3>.
- R. Raghu, A. Levy, G. Wetzstein, and E. D. Zhong. Multiscale guidance of protein structure prediction with heterogeneous cryo-EM data. In *Advances in Neural Information Processing Systems*, volume 38, 2025. doi: 10.52202/085713-0344. URL https://papers.nips.cc/paper_files/paper/2025/hash/0ee4999d856dc51b6e97b2853e69e9ea-Abstract-Conference.html.
- J. Shenoy, M. Astore, A. Levy, F. Poitevin, S. M. Hanson, and G. Wetzstein. Modeling atomic conformational ensembles of proteins via test-time supervision of Boltz-2 on cryo-EM density maps, 2026. URL <https://arxiv.org/abs/2605.09832>.
- S. M. Stagg, A. J. Noble, M. Spilman, and M. S. Chapman. ResLog plots as an empirical metric of the quality of cryo-EM reconstructions. *Journal of Structural Biology*, 185(3):418–426, 2014. doi: 10.1016/j.jsb.2013.12.010. URL <https://doi.org/10.1016/j.jsb.2013.12.010>.
- B. Su, L. Zhu, W. Li, W. Wang, J. Liang, K. Zhao, Z. Peng, and J. Yang. CryoAtom2: inverse folding-inspired deep learning enables accurate model building of protein-nucleic acid complexes in cryo-EM, 2026. URL <https://langtaosha.org.cn/lts/en/preprint/view/185>.
- W. S. Tang, D. Silva-Sánchez, J. Giraldo-Barreto, B. Carpenter, S. M. Hanson, A. H. Barnett, E. H. Thiede, and P. Cossio. Ensemble reweighting using cryo-EM particle images. *The Journal of Physical Chemistry B*, 127(24):5410–5421, 2023. doi: 10.1021/acs.jpcc.3c01087. URL <https://doi.org/10.1021/acs.jpcc.3c01087>.
- T. C. Terwilliger, B. K. Poon, P. V. Afonine, C. J. Schlicksup, T. I. Croll, C. Millán, J. S. Richardson, R. J. Read, and P. D. Adams. Improved AlphaFold modeling with implicit experimental information. *Nature Methods*, 19(11):1376–1382, 2022. doi: 10.1038/s41592-022-01645-6. URL <https://www.nature.com/articles/s41592-022-01645-6>.
- T. C. Terwilliger, D. Liebschner, T. I. Croll, C. J. Williams, A. J. McCoy, B. K. Poon, P. V. Afonine, R. D. Oeffner, J. S. Richardson, R. J. Read, and P. D. Adams. AlphaFold predictions are valuable hypotheses and accelerate but do not replace experimental structure determination. *Nature Methods*, 21(1):110–116, 2024. doi: 10.1038/s41592-023-02087-4. URL <https://www.nature.com/articles/s41592-023-02087-4>.
- X. Wang, H. Zhu, G. Terashi, M. Taluja, and D. Kihara. DiffModeler: large macromolecular structure modeling for cryo-EM maps using a diffusion model. *Nature Methods*, 21:2307–2317, 2024. doi: 10.1038/s41592-024-02479-0. URL <https://www.nature.com/articles/s41592-024-02479-0>.
- Y. Zhang, C. Gong, H. Zhang, W. Ma, Z. Liu, X. Chen, J. Guan, L. Wang, Y. Yang, Y. Xia, and W. Xiao. Protenix-v1: Toward high-accuracy open-source biomolecular structure prediction. *bioRxiv*, 2026. doi: 10.64898/2026.02.05.703733. URL <https://www.biorxiv.org/content/early/2026/02/22/2026.02.05.703733.1>.
- E. D. Zhong, T. Bepler, B. Berger, and J. H. Davis. CryoDRGN: reconstruction of heterogeneous cryo-EM structures using neural networks. *Nature Methods*, 18(2):176–185, Feb. 2021. ISSN 1548-7091, 1548-7105. doi: 10.1038/s41592-020-01049-4. URL <https://www.nature.com/articles/s41592-020-01049-4>.