Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.26.000403. This version posted November 26, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

CoCoFlex: steering a pretrained structure generator with cryo-EM particle images to model continuous conformational variation

Junwen Liao[1], Mingxu Hu[2*], Chenglong Bao[3,4,5*]

[1]Qiuzhen College, Tsinghua University, 100084, Beijing, China.
[2]Institute of Bio-Architecture and Bio-Interactions, Shenzhen Medical Academy of Research and Translation, 518100, Shenzhen, China.
[3]Yau Mathematical Sciences Center, Tsinghua University, 100084, Beijing, China.
[4]Yanqi Lake Beijing Institute of Mathematical Sciences and Applications, 101408, Beijing, China.
[5]State Key Laboratory of Membrane Biology, School of Life Sciences, Tsinghua University, 100084, Beijing, China.

*Corresponding author(s). E-mail(s): humingxu@smart.org.cn; clbao@mail.tsinghua.edu.cn;

Abstract

Cryogenic electron microscopy (cryo-EM) images capture macromolecules in many conformations, and understanding how these molecules function requires atomic models of their continuous motions. Current heterogeneity analyses mostly describe such variation as density maps or latent coordinates, and individual particle images are too noisy to determine atomic detail on their own, making it appealing to incorporate pretrained structural priors learned from known protein structures. Here we present CoCoFlex, which steers a frozen pretrained protein structure generator with particle images: a latent variable for each particle and a shared modulator are optimized so that the generator supplies atomic detail while the images determine how structures vary across particles. In a retrospective comparison on CAHRA Challenge 2, a community benchmark of continuous conformational change, CoCoFlex reproduced the reference atomic ensemble more accurately than the challenge-winning submission. In four experimental datasets, the domain motions inferred by CoCoFlex were supported by density maps reconstructed from particle groups defined by their latent variables or from classifications initialized with these maps. CoCoFlex thus provides a

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.26.000403. This version posted November 26, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

route from cryo-EM particle images to atomic ensembles of continuous molecular motion.

Keywords: cryo-EM, conformational heterogeneity, CAHRA, diffusion models, structural priors

1 Introduction

Proteins and their assemblies often bind partners, catalyze reactions, transport substrates and transmit signals by changing conformation, so explaining how they work requires knowing how their structures vary[1]. In single-particle cryogenic electron microscopy (cryo-EM), each particle image is a projection of an individual molecule frozen in its own conformation, so a single dataset can sample structural variation that is often continuous rather than confined to a few discrete states[2]. Atomic models of this variation would show which residues, interfaces and domains move, allow direct comparison with simulations and predicted structures, and provide experimental training data for methods that predict conformational ensembles[3]. Such models are difficult to obtain, because atomic detail usually emerges only after many noisy particle images are aligned and averaged, whereas continuous variation leaves few particles in any one conformation[4]. This problem is now the focus of a community benchmark: Challenge 2 of the recent Community-Wide Assessment of Heterogeneous Reconstruction Algorithms (CAHRA)[5] asked participants to represent the continuous conformational variability in a simulated cryo-EM dataset with a set of atomic models.

Methods for analyzing cryo-EM heterogeneity have progressed from separating particles into discrete classes toward describing continuous variation in three-dimensional density. Multi-body refinement represents changes as motions of selected regions within a complex[6]; 3D variability analysis and RECOVAR infer low-dimensional density variation through a linear subspace model or through covariance estimation and kernel regression[7,8]; cryoDRGN and e2gmm learn neural representations of heterogeneous density[9,10]; and 3DFlex and DynaMight model motion by deforming a shared density map or a Gaussian representation of the consensus structure[4,11]. These methods reveal continuous changes in molecular shape, but their outputs are density maps or latent coordinates. Atomic models must then be built into many reconstructed states, each supported by only a fraction of the particles, and model building becomes increasingly difficult as map resolution decreases[3,12].

Structure prediction offers strong priors for this step. AlphaFold2 (AF2) and AlphaFold3 predict molecular structures from sequence and other molecular inputs[13,14], and ModelAngelo and DiffModeler combine density-map features with sequence or predicted structures to build atomic models[15,16]. Predicted structures can also be refined with experimental guidance, either through iterative AlphaFold-assisted rebuilding[17] or, as in ROCKET, by optimizing the input representation of a pretrained predictor against experimental observations[18]. For heterogeneous data,

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.26.000403. This version posted November 26, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

CryoBoltz and CryoSampler use ensembles of reconstructed maps to guide structure prediction or to fine-tune an ensemble model[3,19]. In these cryo-EM applications, however, the structural prior is introduced after density reconstruction, as noted previously[12], so the resulting models depend on the quality of the reconstructed maps. Because individual particle images are too noisy to determine atomic detail on their own, it is appealing instead to bring pretrained structural priors to the particle images themselves.

Several approaches already connect particle images to atomic structures through structural priors. HEMNMA and MDSPACE fit reference-based motions through normal-mode analysis and molecular-dynamics (MD) flexible fitting[20,21]; AtomVAE and Cryofold explored continuous distributions using residue-level transformations or coarse-grained representations[22,23]; CryoSTAR regularizes conformational changes with an input atomic model[24]; cryoSPHERE models the motion of segments derived from a reference structure[25]; Gaussian-mixture deformations can be transferred to an existing atomic model[26]; and CryoDyna combines learned multiscale deformations with atomic backmapping[27]. Other strategies refine an initialized set of conformations and their population weights against particle images with MD constraints[28], or build a series of models from heterogeneous reconstructions[29]. These methods largely describe variation as deformations of a reference model or of initialized conformations, so the accessible conformations depend on the starting structures and on the chosen deformation model.

More recent methods couple particle images to pretrained structure predictors. CryoARC infers particle latents with an image encoder and decodes them with a trainable structure module initialized from AF2 parameters and conditioned on sequence-derived representations[30]. Fold'EM infers atomic structures directly from particle images through inference-time guidance of a generative prior; in its continuous-heterogeneity experiments, particle latent coordinates from cryoDRGN, RECOVAR or B-RECOVAR define anchor sets, independently optimized Pairformer representations are interpolated, and the resulting structures undergo coordinate refinement and relaxation[12]. Our earlier CoCoFold fine-tunes an AlphaFold-based structure module against particle observations[31], whereas CoCoFold2 keeps a diffusion-based predictor frozen and optimizes a canonical structural representation against the images but describes a single structure[32]. These designs either update the pretrained network, which can erode the stereochemical knowledge it encodes, or derive particle heterogeneity from external latent coordinates. A method that keeps a pretrained structure generator intact while learning, jointly from all particle images, how its output varies across particles has been lacking.

Here we present CoCoFlex, which steers a frozen pretrained structure generator with particle images. Building on CoCoFold2 canonical refinement, CoCoFlex fixes a canonical structural representation of the target and assigns each particle a latent variable that is optimized directly, as in an auto-decoder[33,34], rather than inferred by an image encoder. A shared modulator combines each latent with the canonical representation to condition the frozen diffusion module of Protenix-mini, a compact model with reduced inference cost[35] from the Protenix family of open-source AlphaFold3 reproductions[36]; the conditioning blocks of the modulator follow those of CryoARC[30].

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.26.000403. This version posted November 26, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

Only the particle latents and the shared modulator are trained, by comparing projections of the generated structures with the observed particle images. The pretrained generator therefore supplies atomic detail, whereas the images determine how the structures vary across particles.

We evaluate CoCoFlex on simulated and experimental data. In a retrospective comparison on CAHRA Challenge 2[5], CoCoFlex reproduces the reference atomic ensemble more accurately than the challenge-winning submission, the cryoSPHERE-based ensemble. A second simulated benchmark derived from an MD trajectory of transforming growth factor-β-induced protein (TGFBIp) extends this assessment to a different pattern of structural variation, and on both benchmarks the relaxed CoCoFlex ensembles show favorable stereochemistry. In experimental datasets of merozoite surface protein 1 (MSP-1)[37], human epidermal growth factor receptor 2 (HER2) bound to trastuzumab and pertuzumab antigen-binding fragments (Fabs)[38], the transient receptor potential vanilloid 1 (TRPV1) channel[39] and the *Plasmodium falciparum* chloroquine resistance transporter (PfCRT) bound to a Fab[40], the domain motions inferred by CoCoFlex are supported by density maps reconstructed from particle groups defined by their latent variables or from classifications initialized with these maps. Together, these results show how particle images can steer a pretrained structure generator to model continuous conformational variation.

2 Results

2.1 CoCoFlex overview: steering a pretrained structure generator with particle images

CoCoFlex uses cryo-EM particle images to steer a frozen pretrained structure generator, producing an ensemble of atomic models that describes continuous conformational variation. The pretrained generator supplies atomic detail, whereas the particle images constrain how the structures vary across particles. CoCoFlex extends CoCoFold2[32] by combining a canonical structural representation of the target with a learned latent variable for each particle. We refer to quantities that are learned by CoCoFold2 and then fixed as canonical, and to CoCoFlex parameters that are trained jointly for all particles as shared.

CoCoFlex is trained in two phases (Fig. 1). In the first phase, CoCoFold2 canonical refinement[32] integrates sequence-derived structural information with particle images. The pretrained Protenix-mini trunk computes a pair representation from the target sequence, and a target-specific pair bias added to this representation is optimized against the particle images through the pretrained Protenix-mini diffusion module, together with the learned canonical parameters of a Gaussian projector[32]. This phase yields a canonical pair representation and a canonical Gaussian projector, which provide the common structural and imaging context for heterogeneous modeling.

In the second phase, CoCoFlex is trained as an auto-decoder[33,34]: each particle is assigned its own latent variable, which is optimized directly rather than inferred by an image encoder. A shared modulator combines the canonical pair representation with each particle's latent variable to generate a per-particle representation, which the pretrained Protenix-mini diffusion module, kept fixed, decodes into atomic coordinates for

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.26.000403. This version posted November 26, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

that particle. Each predicted structure is passed through the image-formation model, in which the canonical Gaussian projector renders it with the particle's upstream pose and the particle's contrast transfer function (CTF) is then applied, and the result is compared with the observed particle image using Fourier-ring correlation (FRC). The resulting image objective optimizes the particle latents and the shared modulator, while the canonical pair representation, canonical Gaussian projector and pretrained Protenix-mini modules remain fixed. After training, the learned particle latents generate an ensemble of atomic structures.

2.2 CoCoFlex outperforms competing methods and the winning submission on CAHRA Challenge 2

CAHRA Challenge 2[5] uses a simulated cryo-EM dataset of human presequence protease. Its 100 reference conformations were selected from an MD trajectory to span the protein's domain opening and closing motion, using an opening-angle collective variable defined by the relative positions of the D1 and D4 domains around the interdomain linker. The selected structures were used to generate simulated particle images, providing an atomic reference ensemble for evaluation. We compared CoCoFlex with CryoARC, CryoSTAR and the challenge-winning cryoSPHERE-based ensemble, which combined cryoSPHERE clustering with subsequent structural modeling. For the heavy-atom comparison, CoCoFlex and CryoARC structures underwent the same restrained relaxation; CryoSTAR outputs Cα models, which were first completed to all heavy atoms with PULCHRA[41] and then relaxed in the same way; and the cryoSPHERE-based ensemble was evaluated as submitted to the challenge. All four predicted ensembles largely spanned the reference opening-angle range (Fig. 2a): the reference structures spanned 64.0–94.8°, compared with 65.6–96.8° for CoCoFlex (Supplementary Video 1), 65.4–99.2° for CryoARC, 68.5–96.8° for CryoSTAR and 66.8–94.2° for the cryoSPHERE-based ensemble.

We quantified structural recovery across all heavy atoms using the three official CAHRA Challenge 2 ensemble metrics (Fig. 2b). Structural accuracy (SAC) averages the root-mean-square deviation (RMSD) between each predicted structure and its closest reference after alignment, describing the structural error of the predicted ensemble. Local flexibility comparison (LFC) measures the average absolute difference in per-atom root-mean-square fluctuations (RMSF) between the predicted and reference ensembles, describing agreement in local fluctuation amplitudes. Proportion of captured variance (PCV) measures how much of the reference ensemble's structural variance is captured by the directions of variation in the predicted ensemble. Lower SAC and LFC and higher PCV therefore indicate better agreement with the reference ensemble. CoCoFlex, CryoARC, CryoSTAR and the cryoSPHERE-based ensemble achieved SAC values of 2.003, 2.460, 2.640 and 2.380 Å; LFC values of 0.287, 0.378, 0.709 and 0.391 Å; and PCV values of 0.837, 0.805, 0.748 and 0.824, respectively. The violin plots show the distributions of per-structure RMSDs and per-atom RMSF differences underlying the SAC and LFC scores. Because SAC matches each prediction to its nearest reference and can remain low even if parts of the reference ensemble are unsampled, we complemented these metrics with coverage evaluated from the reference side, matching each reference to its closest predicted structure (Fig. 2b). At

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.26.000403. This version posted November 26, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

an RMSD threshold of 2.2 Å, CoCoFlex covered 90% of the 100 reference structures, compared with 2% for CryoARC, 0% for CryoSTAR and 41% for the cryoSPHERE-based ensemble; the mean reference-to-nearest-prediction RMSD was 2.006, 2.373, 2.752 and 2.234 Å, respectively.

We next assessed local stereochemical quality using Ramachandran outliers, side-chain rotamer outliers and clashscore, which also form part of the CAHRA evaluation (Fig. 2c). Across the 100 structures per ensemble, the reference ensemble had mean values of 0.68%, 2.37% and 1.16, respectively. The relaxed CoCoFlex ensemble achieved 0.13%, 1.54% and 4.72, compared with 35.75%, 15.27% and 95.41 for relaxed CryoARC, 2.13%, 6.74% and 26.37 for PULCHRA-completed and relaxed CryoSTAR, and 1.30%, 5.07% and 3.28 for the cryoSPHERE-based ensemble. Under the same restrained relaxation, CoCoFlex retained substantially better local geometry than CryoARC, whereas CryoSTAR's stereochemistry primarily reflects the PULCHRA completion and relaxation steps applied to its Cα outputs.

To examine how the overall LFC differences distribute along the 1,014-residue sequence, we compared per-residue heavy-atom RMSF profiles against the reference ensemble (Fig. 2d). CoCoFlex tracked the principal peaks and troughs of the reference fluctuation profile most closely across the sequence, whereas CryoARC and the cryoSPHERE-based ensemble underestimated fluctuations at several mobile peaks and CryoSTAR deviated more broadly, consistent with its higher LFC. Across the flexible Switch C region (model residues 483–527) [42], CoCoFlex closely reproduced the reference fluctuation peak, whereas the other methods underestimated local flexibility to varying degrees.

We also examined how the structural metrics changed during the first ten training epochs for different particle counts and modulator sizes (Extended Data Fig. 1). Across the four configurations, LFC decreased from about 2.7 Å after the first epoch to 0.32–0.44 Å after the tenth and PCV increased from 0.72–0.74 to 0.84, whereas SAC increased slightly, from 2.01 Å to 2.08–2.17 Å.

The original and relaxed structures, their paired geometric distributions and their changes in ensemble metrics are compared in Extended Data Fig. 2a,b. Relaxation reduced Ramachandran outliers and clashscores for all three relaxed ensembles but reduced rotamer outliers only for CoCoFlex, from 6.72% to 1.54%; those of CryoARC and PULCHRA-completed CryoSTAR increased from 9.34% to 15.27% and from 4.85% to 6.74%, respectively.

CryoARC, like CoCoFlex, was run with the upstream particle poses held fixed. In CryoARC runs with pose updates on CAHRA and TGFBIp, the particle orientations drifted by more than 100° from their initial values and the sampled structures lost nearly all secondary structure (Extended Data Fig. 3; Methods).

2.3 CoCoFlex achieves the lowest structural error and highest captured variance on a simulated TGFBIp benchmark

To examine recovery in a second structural landscape, we constructed a simulated benchmark from an MD trajectory of the human TGFBIp R555W mutant (PDB 7ASG). We selected 100 reference conformations and simulated cryo-EM particle images following the image-simulation procedure used for CAHRA Challenge 2 [5]. The

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.26.000403. This version posted November 26, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

100 reference conformations display changes in the relative positions of their terminal regions (Fig. 3a). Principal-component analysis (PCA) of the reference Cα coordinates showed that the first three principal components (PCs) accounted for 67.0%, 12.1% and 6.5% of the variance, compared with 86.1%, 2.8% and 1.7% for the CAHRA reference ensemble (Fig. 3a). Thus, a single component that follows the domain-opening motion dominates the CAHRA variation, whereas the TGFBIp ensemble combines a dominant terminal-segment motion along PC1 with substantial variation along PC2 and PC3 (18.6% of the variance, compared with 4.5% for CAHRA). This benchmark therefore tests whether methods recover structural variation beyond a single dominant motion.

As for CAHRA, CoCoFlex and CryoARC structures underwent the same restrained relaxation, and CryoSTAR Cα outputs were completed with PULCHRA and then relaxed. The terminal-segment distance used to select the references spanned 64.0–86.3 Å in the reference ensemble, compared with 71.3–81.1 Å for CoCoFlex (Supplementary Video 2), 71.7–80.0 Å for CryoARC and 72.3–79.2 Å for CryoSTAR (Fig. 3a). We compared the three methods (cryoSPHERE training became numerically unstable on this dataset; Methods) using 100 structures per method and a common panel of 4,567 heavy atoms (Fig. 3b). CoCoFlex, CryoARC and CryoSTAR had SAC values of 2.256, 2.568 and 2.631 Å; LFC values of 1.613, 1.487 and 1.895 Å; and PCV values of 0.840, 0.565 and 0.327, respectively. PCV separated the methods more widely than on CAHRA (0.748–0.837), consistent with reference variance that is distributed over several PCs rather than concentrated along one dominant direction. Coverage evaluated from the reference side gave mean reference-to-nearest-prediction RMSDs of 2.644, 3.113 and 3.502 Å, respectively (Fig. 3b).

We also assessed local stereochemistry in these ensembles (Fig. 3c). The reference ensemble had mean Ramachandran-outlier percentage, rotamer-outlier percentage and clashscore of 1.28%, 2.35% and 1.01, respectively. The relaxed CoCoFlex ensemble achieved 0.07%, 2.25% and 4.46, compared with 32.03%, 13.11% and 62.56 for relaxed CryoARC and 1.91%, 6.26% and 21.74 for PULCHRA-completed and relaxed CryoSTAR. Paired comparisons of the original and relaxed structures, single-model controls and ensemble metrics are shown in Extended Data Fig. 2c,d. As for CAHRA, relaxation reduced rotamer outliers only for CoCoFlex, from 8.16% to 2.25%, whereas those of CryoARC and CryoSTAR increased from 2.12% to 13.11% and from 5.83% to 6.26%, respectively. These comparisons show improved CoCoFlex stereochemistry with small changes in SAC and LFC (from 2.246 to 2.256 Å and from 1.611 to 1.613 Å), whereas PCV decreased from 0.862 to 0.840.

Per-residue heavy-atom RMSF profiles showed pronounced reference fluctuations in both terminal segments (residues 1–100 and 501–594). All three predicted ensembles fluctuated less than the reference throughout the sequence, with the largest deficits in these segments (Fig. 3d), consistent with their narrower terminal-segment distance ranges.

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.26.000403. This version posted November 26, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

2.4 CoCoFlex particle groups match the six published MSP-1 conformations

We next examined MSP-1, a protein on the surface of the malaria parasite *P. falciparum*. MSP-1 comprises four noncovalently associated fragments (p83, p30, p38 and p42) generated by proteolytic processing. A previous cryo-EM study resolved six maps and corresponding atomic models of this processed complex: a main conformation and five alternative conformations[37]. These structures revealed substantial flexibility in a wing domain, whose position varies relative to the coiled-coil region. After alignment of the structures on the coiled-coil region, the wing centroids of the most widely separated pair, alternative conformations 4 and 1 (PDB 6ZBG and 6ZBE), differed by 8.3 Å (Fig. 4a). The six published models provide a structural reference for examining the conformational variation inferred from the experimental particle images.

We selected 2,000 particles from each of the six published source classes by pose-proportional sampling. These source classes describe the original particle sets, whereas the clusters below are learned from the combined particles. From the resulting 12,000 experimental particles, CoCoFlex learned particle latents that formed a continuous distribution in the first two PCs, which accounted for 49.7% and 21.9% of the latent variance, respectively (Fig. 4b). To examine this distribution, we grouped the particle latents into six clusters and selected a representative structure for each group. This grouping provided six CoCoFlex-derived particle sets for structural and density comparisons. To relate the learned groups to the published conformations, we compared the wing centroids of the six relaxed representatives with those of the six published structures after alignment on the coiled-coil region (Fig. 4c). The one-to-one assignment with the smallest total distance matched each group to a different published structure, with matched distances of 0.4–1.6 Å: Groups 1–6 corresponded to alternative conformations 4 and 5, the main conformation and alternative conformations 3, 2 and 1, respectively. The groups are numbered along latent PC1, and their matched structures follow the published wing positions along the axis from alternative conformation 4 to alternative conformation 1, linking the latent coordinate to the published conformational change. On this axis, where the main conformation and alternative conformation 3 nearly coincide, only the wing centroids of Groups 3 and 4 lie in reversed order.

To visualize this variation, we compared the relaxed representatives of Group 1 and Group 6. Following alignment of the coiled-coil region, their wing centroids differed by 8.4 Å. Density maps reconstructed from the two groups differed correspondingly in the wing region, and in each map the wing density correlated more strongly with the representative of its own group than with that of the other group (model–map correlation coefficient (CC) 0.780 versus 0.153 for the Group 1 map and 0.770 versus 0.100 for the Group 6 map; Fig. 4d).

We also ran CryoARC, CryoSTAR and cryoSPHERE on the same 12,000 particles and compared six relaxed representatives from each method with the six published structures across 9,584 common heavy atoms; CryoSTAR and cryoSPHERE outputs were first completed with PULCHRA (Fig. 4e). Because each predicted ensemble contains six representatives, PCV was calculated using up to five nonzero principal directions of the predicted structures. In this evaluation, the CoCoFlex ensemble

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.26.000403. This version posted November 26, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

achieved a SAC of 2.173 Å, an LFC of 0.179 Å and a PCV of 0.863. The corresponding values were 3.109 Å, 0.379 Å and 0.472 for CryoARC; 3.446 Å, 0.305 Å and 0.552 for CryoSTAR; and 6.972 Å, 1.220 Å and 0.337 for cryoSPHERE. The six published structures had no Ramachandran or rotamer outliers and a mean clashscore of 1.10; the relaxed CoCoFlex representatives had 0.72%, 5.03% and 13.17, compared with 39.28%, 16.21% and 81.42 for CryoARC, 5.50%, 10.19% and 39.23 for CryoSTAR and 3.36%, 6.74% and 30.65 for cryoSPHERE (Fig. 4f). Paired comparisons of the representatives before and after relaxation are shown in Extended Data Fig. 2e,f; relaxation again reduced rotamer outliers only for CoCoFlex, from 12.98% to 5.03%, whereas those of the other three methods increased by 0.33–1.52 percentage points.

The two MSP-1 endpoint reconstructions had global resolutions of 6.25 and 6.74 Å, estimated from corrected half-map Fourier shell correlation (FSC) curves (Extended Data Fig. 4a and Extended Data Table 1). To quantify wing motion relative to the coiled coil, we summarized regional translation and rotation as a rigid displacement field. The rigid-field RMS amplitudes were 9.23 Å for the models and 9.40 Å for the maps, with a model–map directional cosine of 0.997 (Extended Data Fig. 4b,c). Fields estimated separately from the two half-map series agreed with each other (cosine 0.999).

2.5 CoCoFlex models distinct Fab motions in the HER2 complex

The HER2–trastuzumab–pertuzumab complex contains two Fabs bound at distinct sites on the extracellular receptor. The original study reported continuous conformational heterogeneity involving the elbow regions between the variable and constant domains of both Fabs, together with HER2 domain IV[38]. This identifies interdomain flexibility as a feature of the complex and motivates distinguishing motion of a Fab relative to HER2 from bending within the Fab itself.

We examined these changes in trastuzumab (chains A and B) and pertuzumab (chains C and D), using HER2 (chain E) as the alignment reference. To separate the changes associated with different latent directions, we calculated PCs of the 30,000 learned particle latent variables and decoded 21 conformations along each direction between its 5th and 95th score percentiles, with the remaining latent-PC scores fixed at zero. The conformations at the 5th percentile, at the midpoint of this range and at the 95th percentile are labeled −, 0 and +, respectively.

The mean-latent complex, decoded from the mean of the particle latents, places the trastuzumab and pertuzumab Fabs in the context of the receptor (Fig. 5a). After alignment on HER2, latent PC1 primarily showed a swing of the trastuzumab Fab relative to the receptor (Fig. 5b,c and Supplementary Video 3). The endpoint root-mean-square (RMS) Cα displacement of this Fab was 10.94 Å. This measure summarizes how far its Cα atoms move between the − and + conformations after receptor alignment. PC3 also moved the trastuzumab Fab, with an endpoint RMS displacement of 6.94 Å (Fig. 5b), and in addition bent it internally: after alignment on the antigen-binding variable domains, the distal constant domains changed position relative to them, exposing motion around the Fab elbow (Fig. 5d). Their relative motion is consistent with the Fab elbow flexibility described in the original study[38]. PC4 moved

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.26.000403. This version posted November 26, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

both Fabs, the pertuzumab Fab most, with an endpoint Cα RMS displacement of 5.59 Å relative to HER2, compared with 4.16 Å for the trastuzumab Fab (Fig. 5b,e). Each direction thus also moved the other Fab, but by less (Fig. 5b), so different latent directions emphasize different regional motions decoded by the same CoCoFlex model.

To examine the corresponding image evidence, we first reconstructed five groups of 6,000 training particles ranked along each of PC1, PC3 and PC4. These maps provided references for five-class focused classification of the full 708,607-particle collection. We locally refined selected classes around the trastuzumab (AB) Fab for PC1 and PC3 and the pertuzumab (CD) Fab for PC4. The displayed maps show differences in the Fab density envelopes across the selected classes (Fig. 5c–e). In PC1 and PC3, each selected map correlated most strongly with its paired −, 0 or + model in the fixed three-state comparison (Fig. 5c,d). PC4 showed weaker state discrimination: the map paired with the 0 model had a model–map CC of 0.833–0.834 with all three models (Fig. 5e). Because the three classes per axis were selected and ordered by their regional CC with the axis models (Methods), this pattern is not independent of the models. The unselected PC-group maps, however, also correlated most strongly with their paired axis models for PC1 and PC3 (diagonal CC 0.688–0.715 and 0.674–0.682, respectively), and their displacement fields agreed closely in direction with the model fields (cosines 0.94–0.99; Extended Data Fig. 4c).

Within the fixed evaluation regions, the local-resolution medians of the class maps were 4.98–6.42 Å for the AB Fab (PC1 and PC3) and 3.46–3.77 Å for the CD Fab (PC4), compared with 8.44–8.75 Å and 6.72–6.88 Å for the PC-group maps (Extended Data Tables 1 and 2). The class-map rigid-field RMS amplitudes were 14.86, 9.11 and 2.58 Å for PC1, PC3 and PC4, respectively, compared with 10.95, 6.78 and 5.57 Å for the model fields, with model–map directional cosines of 0.808, 0.879 and 0.626 (Extended Data Fig. 4b,c). The class maps thus moved broadly in the directions of the PC1 and PC3 models, with larger amplitudes, whereas the PC4 maps agreed less in direction and moved less than the models.

The remaining HER2 direction, latent PC2, showed a non-monotonic response in both Fabs (Extended Data Fig. 5a). After alignment on HER2, the Cα RMS displacement of the pertuzumab constant domains from the − model increased to 4.46 Å at the midpoint (0) and decreased to 0.75 Å at the + end. The trastuzumab Fab followed a similar course, reaching 3.93 Å at the midpoint and 2.41 Å at the + end. Along this direction, the pertuzumab constant domains therefore return almost to their − position by the + end, whereas the trastuzumab Fab returns only partly.

2.6 CoCoFlex models peripheral motions around membrane-protein cores

We next examined peripheral flexibility in TRPV1 and PfCRT–Fab particles prepared in nanodiscs[39,40]. The TRPV1 tetramer, bound here to the double-knot toxin DkTx, has four cytoplasmic ankyrin-repeat domains (ARDs) arranged around its pore, whereas the PfCRT complex has a Fab extending from the membrane-protein core (Fig. 6a,b). These architectures allow peripheral components to move relative to the membrane-associated core.

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.26.000403. This version posted November 26, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

The cytoplasmic domains of TRPV1 are a documented source of continuous flexibility in this dataset. In an analysis of the same EMPIAR-10059 data, 3DFlex identified coordinated bending of opposing peripheral domains and twisting of all four domains around the pore axis[4]. We therefore examined whether the learned atomic conformations resolved motion of these domains relative to the pore core.

Along latent PC2, the four TRPV1 ARDs underwent a shared twisting motion around the pore, accompanied by inward radial displacement from the − to the + conformation (Fig. 6c,d and Supplementary Video 4). After pore-core alignment, the endpoint Cα RMS displacements ranged from 5.74 to 6.57 Å across the four ARDs (Fig. 6c). PC2 also moved the DkTx F chain, by 4.43 Å between the − and + models, a motion that the ARD-focused classification below did not test. The pore-axis twisting resembles the geometric mode previously reported by 3DFlex[4].

In the original PfCRT study, a bound Fab facilitated cryo-EM structure determination of the relatively small membrane protein[40]. The complex provides a complementary setting in which to distinguish motion of the antigen-binding variable domains near the transporter from that of the distal constant domains. PfCRT latent PC2 primarily showed a swing of the distal Fab constant domains, while the proximal variable domains changed less (Fig. 6e and Supplementary Video 5). With the membrane-protein core as the alignment reference, the endpoint Cα RMS displacements were 5.25 Å for the distal region and 0.79 Å for the proximal region (Fig. 6c).

To examine image support for these regional motions, we first divided the training particles into five approximately equal groups ranked by latent PC2. For TRPV1, these group reconstructions supplied references for five-class focused classification of all 218,786 filtered particles using a mask covering the four ARDs, followed by local refinement of selected classes with the same regional target. The displayed class maps show changes in the ARD A envelope (Fig. 6d). PfCRT retains the original PC2 grouping: its 16,905 particles form five groups of 3,381, whose maps show density-envelope changes in the distal Fab (Fig. 6e). Regional model–map CC compares each displayed map series with the fixed −, 0 and + models (Fig. 6d,e); the corresponding displacement fields are summarized in Extended Data Fig. 4b,c. The TRPV1 − and + maps each favored their paired endpoint model over the opposite endpoint (CC 0.635 versus 0.488, and 0.570 versus 0.444, respectively). As for HER2, these classes were selected by their CC with the axis models, whereas the unselected PC-group maps also correlated most strongly with their paired axis models (diagonal CC 0.608–0.629) and had displacement fields closely aligned with the model fields (cosines 0.98–0.99; Extended Data Fig. 4c). In PfCRT, each endpoint map correlated more strongly with its paired model than with the opposite endpoint model.

The TRPV1 ARD A local-resolution medians were 5.92–6.29 Å (Extended Data Table 2). Across the four ARDs, model–map directional cosines ranged from 0.620 to 0.724; map rigid-field RMS amplitudes were 3.28–4.91 Å, compared with 5.70–6.54 Å for the models. Across HER2 and TRPV1, the class-map fields generally showed larger amplitudes but lower model–map directional cosines than the original PC-group fields (Extended Data Fig. 4b,c). The PfCRT groups had corrected global FSC resolutions of 3.57–3.67 Å (Extended Data Fig. 4a and Extended Data Table 1). Their distal-Fab

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.26.000403. This version posted November 26, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

map displacement field had an amplitude of 4.93 Å and a directional cosine of 0.990 with the model field. The PfCRT distal-Fab density thus followed the model direction closely, whereas the TRPV1 class maps agreed only partly in direction and moved less than the models.

The other TRPV1 directions also moved the ARDs and the DkTx F chain (Extended Data Fig. 5b). PC1 displaced residues A351–375 by 1.79 Å between the − and + models, but its largest changes were in the four ARDs, which deviated from the − model by 3.42–4.05 Å at the midpoint. PC4 changed all four ARDs non-monotonically, with maxima of 2.31–2.62 Å before the midpoint; ARD A deviated by up to 2.49 Å and was 1.66 Å from the − model at the + end. Along both directions, the DkTx F chain also deviated by more than 3 Å at the midpoint (3.39 and 3.07 Å). The other PfCRT directions likewise showed regional variation (Extended Data Fig. 5c,d). Along PC1, the A31–55 region moved away from the − model by up to 2.37 Å at the midpoint and was 1.11 Å from it at the + end. PC3 and PC4 displaced the A-chain N terminus (A1–49, 2.10 Å) and Fab B121–145 (2.37 Å), respectively; along PC3, the overlapping segment A31–55 moved further (2.66 Å). In all three PfCRT directions, the distal Fab moved more than these segments (2.82–3.05 Å), whereas the proximal Fab changed by less than 1 Å.

DkTx stabilizes the open channel through interactions with the channel and surrounding lipids[39]. Along latent PC3, the decoded DkTx F chain moved relative to the pore and changed internally, with an endpoint Cα RMS displacement of 4.15 Å after pore-core alignment; the ARDs and the DkTx E chain moved less (at most 2.96 and 1.14 Å; Extended Data Fig. 5e). Reconstructions of the first and fifth PC3 particle groups had global resolutions of 4.20–4.30 Å, estimated from corrected half-map FSC curves (Extended Data Fig. 4a). In these maps, the rigid-field RMS amplitude of the F-chain neighborhood was only 0.31 Å, compared with 3.83 Å for the models (Extended Data Fig. 4b). A previous analysis of the same dataset reported variable DkTx density and possible compositional heterogeneity[43], which may contribute to the weak image support for this toxin motion.

3 Discussion

CoCoFlex treats a pretrained structure generator as a fixed prior and learns only how particle images condition it. The images therefore determine how structures vary across particles, whereas atomic detail comes from what the generator learned from known structures. This division distinguishes CoCoFlex from approaches that deform a reference model[24,25], apply structural priors to reconstructed maps[3,19], take particle heterogeneity from external latent coordinates[12] or update pretrained network parameters against the images[30,31].

Across the simulated benchmarks and MSP-1, the clearest difference between CoCoFlex and CryoARC, which also builds on a pretrained predictor, lay in local stereochemistry, and we attribute it to keeping the generator fixed. A pretrained generator implicitly encodes the backbone and side-chain geometry of the structures on which it was trained. When only its conditioning is optimized, image gradients move the outputs among structures that the generator can produce rather than changing

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.26.000403. This version posted November 26, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

the generator itself; consistent with this, CoCoFold2 refinement of the canonical representation slightly improved the stereochemistry of the initial Protenix-mini prediction (Extended Data Fig. 2a,c). Heterogeneous training then introduced clashes, raising the mean clashscore of the unrelaxed CoCoFlex ensembles to 41.0 and 44.7 on CAHRA and TGFBIp, compared with 18.4 and 16.7 for the refined single models, but these values remained far below those of CryoARC (921 and 746). CryoARC instead trains a structure module initialized from AF2 parameters, and before relaxation its structures had far more Ramachandran outliers and clashes than those of CoCoFlex. Restrained relaxation then widened this difference. It reduced all three stereochemical measures only for CoCoFlex and, on the simulated benchmarks, brought its Ramachandran and rotamer outliers below those of the reference ensembles while changing SAC and PCV only slightly, indicating that the generated backbones already lay close to physically realizable conformations. For the other methods, whose unrelaxed structures had higher clashscores than those of CoCoFlex, relaxation increased rotamer outliers, and the CryoARC structures retained many Ramachandran outliers. Because CoCoFlex and CryoARC also differ in their pretrained models, decoders and objectives, and CryoARC uses an image encoder, this attribution remains an interpretation rather than the result of a controlled comparison. Stereochemical quality reports physical plausibility, whereas agreement with the underlying conformations is assessed by the ensemble metrics and particle-image evidence.

On the simulated benchmarks, CoCoFlex recovered more of the reference variation than the other methods, most clearly in the reference-side coverage on CAHRA. Optimizing each particle latent directly, as in an auto-decoder[33,34], rather than predicting it with an image encoder may contribute, because particle latents are then not limited by the amortization error of an encoder; we did not test this separately. The recovered distributions were nevertheless narrower than the reference where variation was concentrated in flexible terminal segments: on TGFBIp, all three methods spanned only part of the reference terminal-segment distance range and underestimated terminal fluctuations, and CryoARC had a lower LFC than CoCoFlex. Each CoCoFlex structure is a deterministic function of a four-dimensional latent, decoded with diffusion noise shared by all particles, so local fluctuations that are not coupled to the learned latent directions cannot be represented; a prior learned from known structures may also pull the extremes of a distribution toward more typical conformations. The benchmarks also have a limited scope. Their reference ensembles were selected from MD trajectories along chosen coordinates, and most predicted ensembles were represented by structures selected by farthest-point sampling (FPS) or K-means clustering, so the metrics assess structural coverage rather than conformational populations. In addition, the CAHRA comparison is retrospective, whereas the challenge submissions were prepared without access to the reference ensemble. Training endpoints and representative selection were fixed without reference to the benchmark ensembles, and the baselines were run by us with their published settings, so their results may not represent their best achievable performance.

In the experimental datasets, the evidence supports variation at the level of domains and subunits. Density maps from latent-defined particle groups, or from classifications started from them, changed broadly in the directions of the corresponding

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.26.000403. This version posted November 26, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

model motions, most closely for the MSP-1 wing and the PfCRT Fab, whereas the HER2 and TRPV1 class maps agreed less in direction (cosines 0.62–0.88) and differed in amplitude, and HER2 PC4 discriminated weakly among the fixed model states. The particle groups were drawn from the particles used to learn the latents, and the half-map comparisons test reproducibility within the same grouping; classification of the full HER2 and TRPV1 collections extended the comparison to particles outside the training sets but started from CoCoFlex-derived references, and the displayed classes were selected and ordered by their agreement with the axis models. Because heterogeneous training used image information only to 4 Å resolution, the atomic detail of individual particles, such as side-chain conformations, is determined mainly by the generator, and our model–map comparisons, made primarily at 6 Å and through regional rigid-body displacement fields, test domain positions rather than atomic detail. Latent principal directions are likewise a linear summary of a nonlinear mapping from latents to structures: individual directions often moved several regions together, and some produced non-monotonic changes, so they need not correspond to physical modes. The DkTx F chain marks the limit of this support: its decoded motion had little counterpart in the density (rigid-field amplitude 0.31 Å versus 3.83 Å for the models).

Several assumptions of the current implementation define where CoCoFlex applies. Particle poses and CTF parameters are taken from upstream refinement and held fixed. On the simulated benchmarks, poses were supplied with the challenge data or refined from the simulation poses, so robustness to pose errors coupled to large peripheral motions remains to be established, and the CryoARC diagnostics suggest that joint pose optimization is not straightforward (Extended Data Fig. 3). All particles share one molecular topology, and unmodeled components, such as bound small molecules, lipids and the nanodisc, contribute to the images without being represented; occupancy changes, as suggested for DkTx[43], therefore cannot be captured. Training used subsets of the larger particle collections and is computationally demanding: ten epochs of the large modulator on 60,000 CAHRA particles took about 36 h on eight GPUs (Extended Data Fig. 1). The per-particle pair representation also grows quadratically with the number of tokens, and assigning latents to new particles requires further optimization. Joint pose refinement, latent variables for composition, per-particle diffusion noise to express local fluctuations and more efficient conditioning would extend the approach to more systems. CoCoFlex ensembles could also provide experimentally grounded training data for methods that predict conformational ensembles[3], although such use should account for the prior that these ensembles share with the pretrained generator.

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.26.000403. This version posted November 26, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

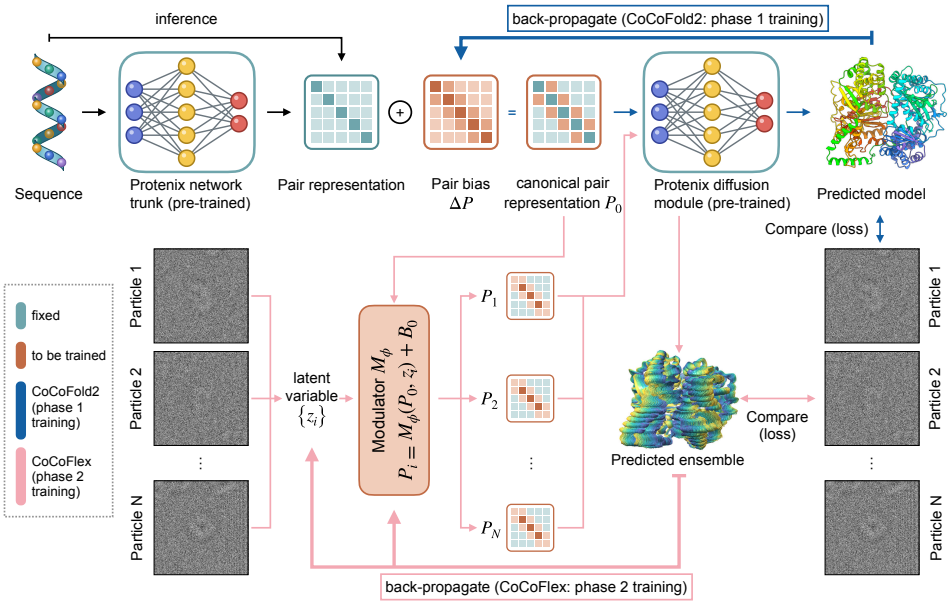


Fig. 1 Overview of CoCoFlex. Teal elements are pretrained and remain fixed, orange elements are optimized, and blue and pink arrows indicate the first and second training phases, respectively. In the first phase (CoCoFold2 canonical refinement), the pretrained Protenix-mini trunk computes the pair representation from the target sequence in a single inference pass. A target-specific pair bias ΔP is added to obtain the canonical pair representation P_0, which conditions the pretrained Protenix-mini diffusion module. The predicted model is passed through the image-formation model, comprising a Gaussian projector with the particle poses and CTFs, and compared with the particle images; the image loss is back-propagated to update ΔP and the Gaussian-projector parameters, after which P_0 and the projector are fixed. In the second phase (CoCoFlex heterogeneous training), each particle i has its own learnable latent variable z_i. The modulator M_ϕ combines P_0 and z_i into a per-particle pair representation $P_i = M_\phi(P_0, z_i) + B_0$, where the fixed offset B_0 makes a zero latent reproduce P_0 before training. The same diffusion module decodes $P_1, \ldots, P_N$ into a predicted ensemble, and each structure is projected with its particle pose and CTF and compared with the corresponding image by FRC. The FRC loss is back-propagated to update M_ϕ and the latents $\{z_i\}$, while the trunk, diffusion module and P_0 remain fixed. The predicted model and predicted ensemble show human presequence protease from CAHRA Challenge 2 for illustration.

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.26.000403. This version posted November 26, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

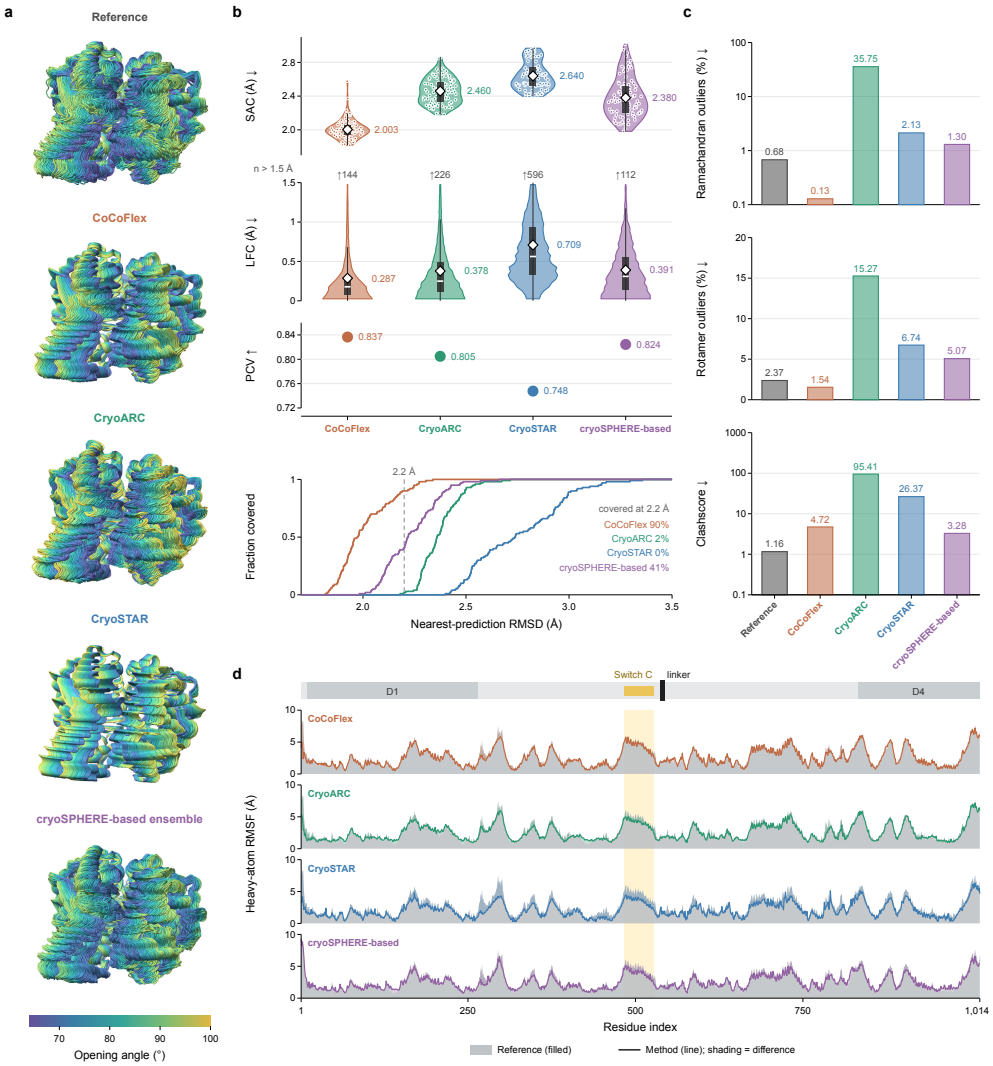


Fig. 2 Structural recovery and stereochemical quality in CAHRA Challenge 2. Each ensemble contains 100 structures. CoCoFlex and CryoARC outputs are shown after restrained relaxation; CryoSTAR Cα outputs were completed to all heavy atoms with PULCHRA and then relaxed in the same way; reference and cryoSPHERE-based ensembles retain their supplied coordinates (Methods). **a**, Reference, CoCoFlex, CryoARC, CryoSTAR and cryoSPHERE-based ensembles superposed on all Cα atoms and colored by opening angle (64–100°). **b**, Official CAHRA heavy-atom metrics (SAC, LFC and PCV; arrows indicate the favorable direction) and reference coverage. Violins show per-structure nearest-reference RMSDs (SAC, overlaid with the 100 individual structures) and per-atom absolute RMSF differences (LFC); narrow inner boxes mark the interquartile range, white lines mark medians, whiskers extend to 1.5 times the interquartile range, and open diamonds with labels give the means (SAC and LFC). The LFC axis displays 0–1.5 Å, with counts above ($n > 1.5$ Å) giving the number of atoms exceeding 1.5 Å. Bottom, cumulative fraction of reference structures covered versus nearest-prediction RMSD; the dashed line and percentages indicate coverage at 2.2 Å. **c**, Local stereochemical quality (Ramachandran outliers, rotamer outliers and clashscore; ensemble means of 100 structures; ↓ indicates lower is better). Ramachandran outliers and clashscore use logarithmic axes. **d**, Residue-wise heavy-atom RMSF, shown with one track per method: gray shading gives the reference profile, colored lines give the method profiles and light fill between them highlights the difference. The top bar spans residues 1–1,014 and marks D1 (residues 10–265), the linker vertex (539–541, black tick) and D4 (832–1,014) used to define the opening angle; the yellow segment and vertical band mark Switch C (model residues 483–527). Original-output geometry and paired before–after ensemble metrics are shown in Extended Data Fig. 2a,b.

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.26.000403. This version posted November 26, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

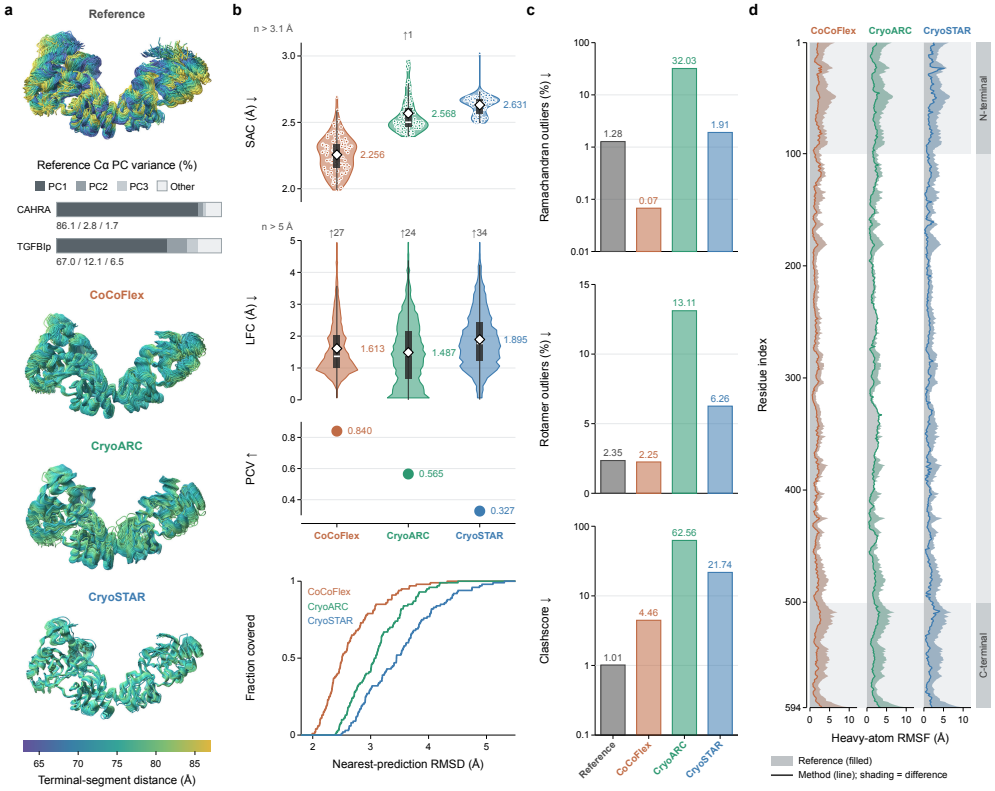


Fig. 3 Structural recovery and stereochemical quality on an MD-derived TGFBIp benchmark. Each ensemble contains 100 structures. CoCoFlex and CryoARC outputs are shown after restrained relaxation; CryoSTAR Cα outputs were completed to all heavy atoms with PULCHRA and then relaxed in the same way; the reference ensemble retains its MD coordinates (Methods). **a**, Reference, CoCoFlex, CryoARC and CryoSTAR ensembles superposed on all Cα atoms and colored by terminal-segment centroid distance (Cα centroids of residues 1–100 and 501–594; 63–87 Å). Bars below the reference ensemble give the fraction of reference Cα coordinate variance explained by the first three PCs for the CAHRA and TGFBIp reference ensembles. **b**, Heavy-atom SAC, LFC and PCV on 4,567 common heavy atoms (arrows indicate the favorable direction) and reference coverage. Violins show per-structure nearest-reference RMSDs (SAC, overlaid with the individual structures) and per-atom absolute RMSF differences (LFC); narrow inner boxes mark the interquartile range, white lines mark medians, whiskers extend to 1.5 times the interquartile range, and open diamonds with labels give the means (SAC and LFC). The SAC and LFC axes display 1.9–3.1 Å and 0–5 Å, respectively; counts above the violins give the number of structures (SAC) or atoms (LFC) beyond these ranges. Bottom, cumulative fraction of reference structures covered versus nearest-prediction RMSD. **c**, Local stereochemical quality (Ramachandran outliers, rotamer outliers and clashscore; ensemble means of 100 structures; ↓ indicates lower is better). Ramachandran outliers and clashscore use logarithmic axes. **d**, Residue-wise heavy-atom RMSF, shown with one track per method: gray shading gives the reference profile, colored lines give the method profiles and light fill between them highlights the difference. Residues run from top to bottom; the right-hand bar and light bands mark the N- and C-terminal segments (residues 1–100 and 501–594) that define the terminal-segment distance. Original-output geometry and paired before–after ensemble metrics are shown in Extended Data Fig. 2c,d.

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.26.000403. This version posted November 26, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

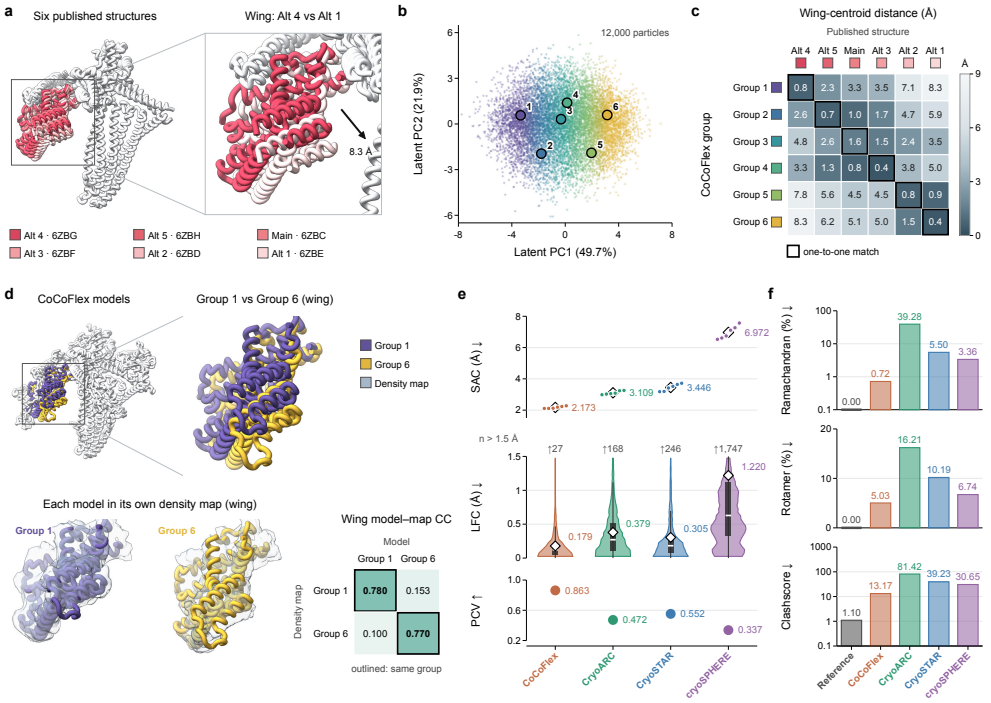


Fig. 4 Structural variation across MSP-1 particles and reconstructed densities. **a**, Six published MSP-1 structures aligned on the coiled-coil region; the wings (chain A residues 1–385) are colored from coral to blush in order of their position along the axis between the most widely separated pair, alternative conformations 4 (6ZBG) and 1 (6ZBE), which the close-up compares. The arrow indicates the direction of wing-centroid displacement and is enlarged for visibility; 8.3 Å is the three-dimensional displacement. **b**, PC projection of 12,000 particle latents, colored by each particle's PC1 score, and the six CoCoFlex cluster centers; Groups 1–6 refer to the learned latent clusters. **c**, Wing-centroid distances between the six relaxed CoCoFlex group representatives and the six published structures after coiled-coil alignment; outlined cells mark the one-to-one assignment with the smallest total distance. **d**, Relaxed Group 1 and Group 6 representatives (full complex and wing close-up), each wing shown in the density reconstructed from its own particle group, and wing model–map Pearson CC at 6 Å (rows, maps; columns, models; outlined cells pair each map with its own group). **e**, Heavy-atom SAC, LFC and PCV for six representatives per method against the six published structures on 9,584 common heavy atoms (arrows indicate the favorable direction). SAC points show the six representatives; LFC violins show per-atom absolute RMSF differences, with counts above giving the number of atoms beyond the displayed 0–1.5 Å range; open diamonds with labels give the means. **f**, Ramachandran outliers, rotamer outliers and clashscore (means of six structures; ↓ indicates lower is better). Ramachandran outliers and clashscore use logarithmic axes, and zero values are labeled at the baseline. CoCoFlex and CryoARC representatives are shown after restrained relaxation; CryoSTAR and cryoSPHERE outputs were completed with PULCHRA and then relaxed (Methods).

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.26.000403. This version posted November 26, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

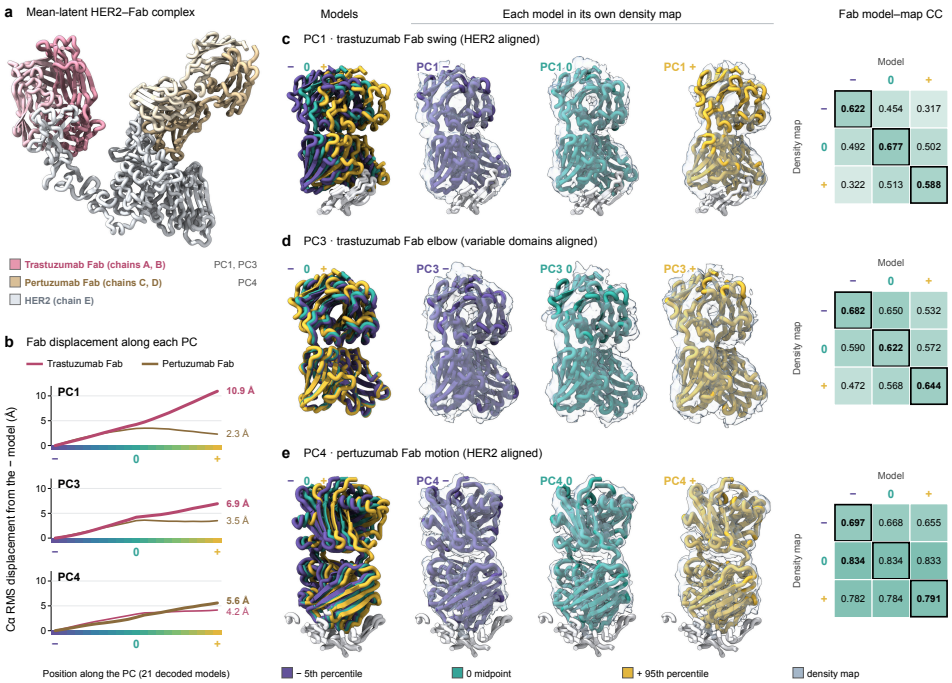


Fig. 5 Distinct Fab motions around HER2. **a**, Mean-latent complex of HER2 (chain E) with the trastuzumab Fab (chains A and B) and the pertuzumab Fab (chains C and D); the latent PCs that mainly move each Fab are indicated. **b**, Cα RMS displacement of each Fab from the − model across the 21 conformations decoded along PC1, PC3 and PC4, after alignment on HER2. Thick lines mark the Fab shown in the corresponding row (c–e), and values give the endpoint displacements. **c**–**e**, PC1 (c), PC3 (d) and PC4 (e). From left: the superposed −, 0 and + models; each model in the locally refined density map of its paired class; and regional Fab model–map Pearson CC at 6 Å (rows, maps; columns, models). Outlined cells mark the displayed pairings, and bold type marks actual row maxima. Models are aligned on HER2 in c and e, with nearby HER2 residues in white, and on the trastuzumab variable domains in d, exposing constant-domain motion around the elbow. −, 0 and + denote the 5th percentile, the midpoint and the 95th percentile of the PC score, with other latent-PC scores zero, and are colored purple, teal and gold. Each map receives the alignment of its paired model and is shown within 3 Å of the displayed Fab atoms, with surface fragments smaller than 6 Å hidden.

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.26.000403. This version posted November 26, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

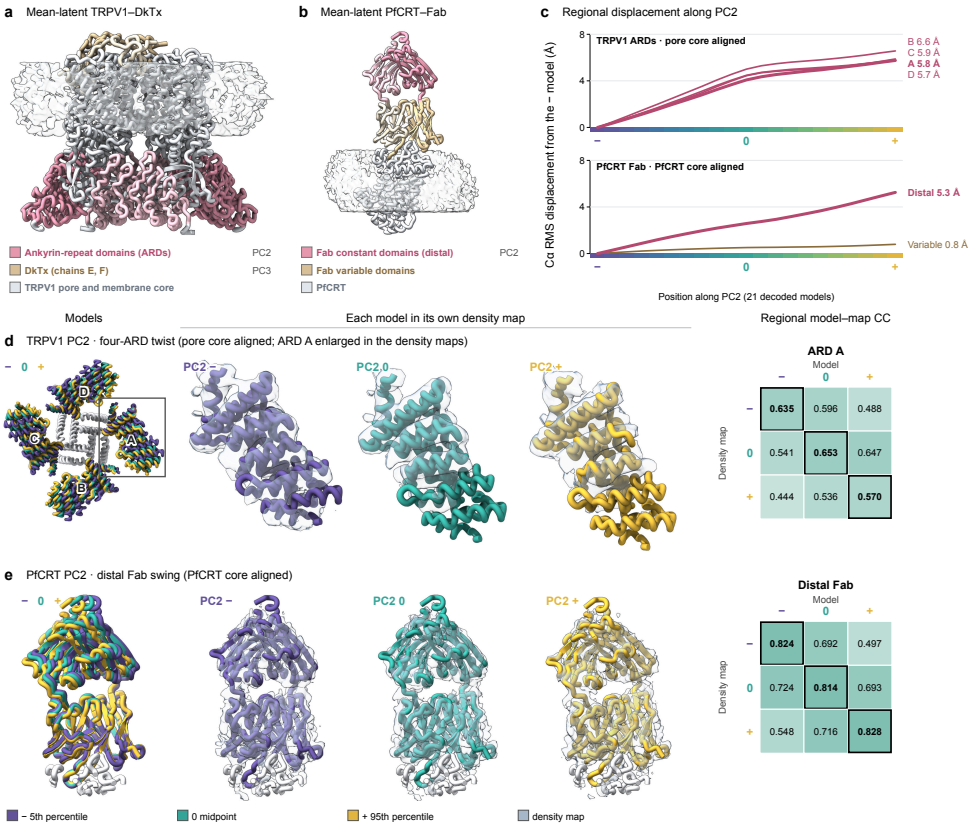


Fig. 6 Peripheral structural variation around membrane-protein cores. a,b, Mean-latent TRPV1–DkTx (a) and PfCRT–Fab (b) models with nanodisc context from the middle PC2-group map; the latent PCs discussed for each component are indicated. In a, the ARDs are rose and the DkTx chains (E and F) sand; in b, the Fab constant (distal) domains are rose and the variable domains sand. **c**, Cα RMS displacement from the − model across the 21 conformations decoded along PC2, for the four TRPV1 ARDs after pore-core alignment and for the distal (constant) and proximal (variable) regions of the PfCRT Fab after alignment on the PfCRT core. Thick lines mark the regions of the CC matrices in d and e, and values give the endpoint displacements. **d,e**, TRPV1 PC2 (d) and PfCRT PC2 (e). From left: the superposed −, 0 and + models; each model in its own density map; and regional model–map Pearson CC at 6 Å for ARD A (d) and the distal Fab (e) (rows, maps; columns, models). Outlined cells mark the displayed pairings, and bold type marks actual row maxima. TRPV1 is viewed from the cytoplasmic side; the outlined region of the first tile, ARD A, is enlarged in the density maps, which are locally refined maps of focused classes from all filtered particles, using a mask covering the four ARDs. PfCRT maps are reconstructions of the first, third and fifth of five PC2 particle groups (3,381 particles each). −, 0 and + denote the 5th percentile, the midpoint and the 95th percentile of the PC2 score, with other latent-PC scores zero, and are colored purple, teal and gold. Each map receives the alignment of its paired model and is shown within 3 Å of the displayed atoms, with surface fragments smaller than 6 Å hidden.

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.26.000403. This version posted November 26, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

Methods

CoCoFlex model

CoCoFlex extends the canonical-refinement framework of CoCoFold2 [32] to the atomic conformations of individual particles. Its inputs are a target sequence or a saved prediction cache, the particle images, and their orientations, translations and CTF parameters, which are estimated upstream. All particles share one molecular topology and one set of atom identities. Canonical refinement yields the canonical pair representation P_0, on which heterogeneous modeling builds. Heterogeneous modeling is formulated as an auto-decoder [33]: it jointly learns a table of particle latents and a shared modulator, without an image encoder. After training, each particle is decoded from its learned latent and the saved structural context.

Latent-conditioned structure generation

For T structural tokens, canonical pair representation $P_0 \in \mathbb{R}^{T\times T\times c}$ and particle latent $z_i \in \mathbb{R}^4$, the modulator M_ϕ conditions a frozen Protenix-mini diffusion module [35]:

$$B_0 = P_0 - M_{\phi_0}(P_0, 0), \tag{1}$$

$$P_i = M_\phi(P_0, z_i) + B_0, \tag{2}$$

$$X_i^{\mathrm{raw}} = \mathcal{D}_{\theta_0}(P_i; \mathcal{C}_0, \xi_0). \tag{3}$$

Here c is the number of pair channels, ϕ_0 denotes the initial modulator parameters, $\mathcal{C}_0$ comprises the fixed topology and structural conditioning, and ξ_0 is the fixed diffusion noise. The offset B_0 makes the initial modulator reproduce P_0 for a zero latent; it is computed once and stays fixed during optimization and sampling.

The modulator uses pair-conditioning blocks based on CryoARC [30]. A linear projection maps each latent vector to four tokens. Each block applies latent cross-attention along both pair axes, followed by a pair transition, with residual updates and a final pair layer normalization. The small and large modulators have one and four blocks, respectively; both use four attention heads, with hidden widths per head of 64 and 128. The Protenix-mini atom encoder, central transformer and atom decoder have depths of 1, 8 and 1, respectively, and sampling uses five denoising steps.

Heterogeneous training optimizes only the modulator and the particle latents. Gradients propagate through the fixed diffusion module and projector, while P_0, the other conditioning inputs, the Gaussian parameters, the map placement, the particle poses and the CTFs remain fixed. All particles share the same diffusion noise during training and sampling.

Fixed-frame Gaussian projection

Before canonical refinement, the initial structure is placed in the frame of the reference density map. Every decoded conformation X_i^{raw} is transformed into this frame with the same placement, giving X_i. For particle orientation Ω_i and in-plane shift $\boldsymbol{\delta}_i$, an atom-centered Gaussian projector gives $J_i = \mathcal{G}_{\gamma_0}(X_i; \Omega_i, \boldsymbol{\delta}_i)$, whose parameters γ_0,

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.26.000403. This version posted November 26, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

the Gaussian amplitudes and detector-plane widths, come from canonical refinement and stay fixed. The predicted Fourier image is

$$\widehat{F}_i(k) = C_i(k)\,\mathcal{F}[J_i](k), \tag{4}$$

where C_i is the CTF of particle i, calculated from the supplied voltage, defocus, astigmatism, spherical aberration, amplitude contrast, phase shift and pixel size. The CTF is applied once to the predicted projection, which is then compared with the observed image in the contrast convention of the input data. Structural alignments used for evaluation and visualization are separate from this fixed imaging frame.

Fourier-ring objective and latent regularization

For an even $D \times D$ image with pixel size h and resolution cutoff d_{cut}, both in Å,

$$r_{\mathrm{cut}} = \left\lfloor \frac{D}{2} \min\left(1, \frac{2h}{d_{\mathrm{cut}}}\right) \right\rfloor, \qquad \mathcal{R}_r = \{k : r \le \|k\|_2 < r+1\}. \tag{5}$$

Frequencies k are integer pixel-frequency coordinates on the centered Fourier grid. The zero-frequency ring is excluded, so rings $1, \ldots, r_{\mathrm{cut}} - 1$ contribute. For the observed transform $F_i = \mathcal{F}[I_i]$ of image I_i,

$$\ell_i = -\frac{1}{r_{\mathrm{cut}}} \sum_{r=1}^{r_{\mathrm{cut}}-1} \frac{\mathrm{Re}\left[\sum_{k\in\mathcal{R}_r} \widehat{F}_i(k)\overline{F_i(k)}\right]}{\sqrt{\left(\sum_{k\in\mathcal{R}_r} |\widehat{F}_i(k)|^2\right)\left(\sum_{k\in\mathcal{R}_r} |F_i(k)|^2\right)} + 10^{-8}}. \tag{6}$$

The ring sum is normalized by r_{cut}. For a global particle batch $\mathcal{B}$, the objective is

$$\mathcal{L} = \frac{1}{|\mathcal{B}|} \sum_{i\in\mathcal{B}} \ell_i + \lambda \frac{\|\bar{z}\|_2^2}{4}, \qquad \bar{z} = \frac{1}{N}\sum_{i=1}^{N} z_i, \qquad \lambda = 0.05, \tag{7}$$

where N is the number of particles, so the mean penalty uses all particle latents rather than only those in the batch. Latents are initialized from $\mathcal{N}(0, 0.35^2)$, clipped to $[-0.95, 0.95]$ at initialization and then optimized without constraints.

Initialization and training

Dataset-specific preparation is summarized in Extended Data Table 3.

Initial structures and map placement

Automated map placement uses a multistart search over proper rigid-body transformations, starting from the principal axes of the structure and of the map and including 24 proper axis permutations and the identity. The rotation and translation are optimized against the atomic-number-weighted mean density at heavy-atom positions. A search on a map smoothed with a Gaussian of 5 Å standard deviation is followed by

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.26.000403. This version posted November 26, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

refinement of up to six candidates at 2 Å, and the highest-scoring placement initializes canonical refinement.

For CAHRA and TGFBIp, the initial predicted structure was placed by this automated search. For MSP-1, a Protenix-V1 prediction placed automatically in the map guided Protenix-mini initialization, which was followed by canonical refinement. For HER2, the initial Protenix-mini prediction contained overlapping chains; rigid-body searches against the reconstructed map, followed by adjustment of these chains, yielded a corrected assembly that served as the Protenix-mini guidance target. For TRPV1, a Protenix-V1 assembly with a more suitable placement of the two DkTx chains guided Protenix-mini initialization. For PfCRT, the initial predicted assembly was misassembled; its components were aligned separately to the reconstructed map, and the resulting coordinates served as the target for coordinate-guided initialization (Extended Data Table 3).

Coordinate-guided initialization

When an alternative prediction or a corrected assembly is better placed in the consensus map, its coordinates guide Protenix-mini initialization through the representation-optimization framework of CoCoFold2[32]. The target is placed in the map frame and matched to the Protenix-mini prediction by atom identity. An additive bias to the cached Protenix-mini pair representation is then optimized against the target coordinates, with gradients propagated through the diffusion generator, whose pretrained weights stay frozen. The loss is a Huber loss with a 1 Å threshold on the matched Cα coordinates, averaged within each chain and then equally across chains. The guided pair representation initializes image-based canonical refinement, whose output provides P_0 for heterogeneous training.

Canonical refinement

Canonical refinement optimizes a target-specific pair bias and the Gaussian-projector parameters against a subset of the particle images, with the pretrained generator weights kept fixed. The pair bias (ΔP in Fig. 1) corresponds to the pair-representation correction of CoCoFold2[32]. The image-loss cutoff is twice the pixel size, and the learning rates are 0.01 for the pair bias and Gaussian amplitudes and 0.005 for the Gaussian widths. For CAHRA, the 17,000-particle subset for canonical refinement was drawn by pose-proportional sampling, that is, without replacement (seed 42) so as to preserve the half-set proportions and the joint occupancy of Rot, cos(Tilt) and Psi in $12 \times 6 \times 6$ bins, using largest-remainder allocation.

Heterogeneous training

Heterogeneous training uses Adam with learning rates of 0.001 for the modulator and 0.05 for the particle latents and default values for the other parameters, a global batch size of 32 and a particle mini-batch size of one. The combined gradient norm is clipped at 1 before each optimizer step. During the first 200 updates, only the modulator is optimized, with zero latent inputs, and the particle latents keep their random initial values; all later updates optimize the modulator and the particle latents jointly. These

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.26.000403. This version posted November 26, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

warmup updates are included in the reported update and epoch counts. Training resumed from a checkpoint restores the saved optimizer state and structural context.

All main runs used eight GPUs and the large modulator, except for TRPV1, which used the small modulator. For CAHRA and TGFBIp, the image-loss cutoff was 6 Å for ten epochs, changed to 4 Å during epoch 11 and stayed at 4 Å thereafter; training ended at epoch 14 (26,250 updates) for CAHRA and at epoch 15 for TGFBIp. HER2, PfCRT and TRPV1 were trained for six epochs at 6 Å, with the transition during epoch 7, and training ended at epoch 10. For TGFBIp and TRPV1, the cutoff was interpolated linearly in resolution during the transition. MSP-1 was trained for 200 warmup and 800 joint updates at 6 Å, a 200-update transition interpolated in reciprocal resolution and 600 updates at 4 Å, ending after 4.80 epochs including warmup.

Datasets

Particle availability and training-set sizes are listed in Extended Data Table 4, and canonical-refinement subsets and training endpoints in Extended Data Table 3.

CAHRA Challenge 2

For CAHRA Challenge 2[5], we used EMPIAR-14006 with the pose and CTF metadata supplied by the competition; the challenge provides 100 reference conformations of a 1,014-residue protein. Input images were 256 pixels at 1 Å per pixel, and predicted projections were multiplied by -1 to match the image contrast.

MD-derived TGFBIp system

For TGFBIp R555W (PDB 7ASG), we selected 100 frames from the 1,001 frames spanning 0–100 ns of ATLAS replicate 1[44]. The selection variable was the distance between the Cα centroids of MD residues 1–100 and 501–594. One hundred equally spaced target values across the range of this distance were assigned one-to-one to distinct frames by minimizing the squared differences between target and frame values. The selected frames keep their MD coordinates and have equal weight in the reference ensemble.

Following the CAHRA workflow[5], we used Parakeet 0.6.5[45] to generate 380 micrographs of $4,000 \times 4,000$ pixels at 1 Å per pixel, each containing three molecules of every conformation in ice 500 Å thick. The simulation used 300 kV, 45 electrons per Å^2, spherical and chromatic aberration coefficients of 2.7 mm and a single exposure fraction. Defocus values alternated between a normal distribution with mean $-20{,}000$ Å and standard deviation 5,000 Å and a ten-component normal mixture with means of $-10{,}000$ to $-19{,}000$ Å and standard deviation 500 Å. Of the 114,000 particles, 100,053 had complete 256-pixel extraction boxes. Particles were extracted in RELION with a 96-pixel background radius, background normalization and contrast inversion; incomplete boxes were discarded, and overlaps between projected particles were kept.

Consensus processing started from a backprojection of all particles with their CTFs at the true poses of the simulation, followed by low-pass filtering and C1 Blush refinement[46] with a 180 Å particle diameter and an initial 10 Å cutoff. The initial angular

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.26.000403. This version posted November 26, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

sampling and search width were 1.875° and 3.75°, the translation range and step were 2 and 1 pixels, the soft edge was five pixels and the seed was 42. CoCoFlex training used these *truth-initialized consensus-refined poses*.

Experimental MSP-1

For MSP-1, we sampled 2,000 particles pose-proportionally (seed 42) from each of the six EMPIAR-10437 source lists, which comprise the main class and five alternative classes. Images were 288 pixels at 1.073 Å per pixel. The correspondence between each particle image and its pose, CTF and shift was preserved, and orientations were converted once between conventions by $\Omega \mapsto \Omega\,\mathrm{diag}(1, 1, -1)$. The poses from the published classes were used unchanged, and the source labels were not part of the training input. The six published structures[37] served as references.

Experimental HER2, TRPV1 and PfCRT

For HER2–trastuzumab–pertuzumab, we used EMPIAR-11665, associated with PDB 8PWH[38], with 280-pixel images at 1.16 Å per pixel. Two particle images with extreme intensities were excluded from the 708,609 inputs before the remaining 708,607 particles were reconstructed with their upstream poses.

For TRPV1, we used EMPIAR-10059 from the nanodisc study associated with model 5IRX[39], with 192-pixel images at 1.2156 Å per pixel. These particles contain rat TRPV1 bound to DkTx and resiniferatoxin (RTX) in lipid nanodiscs; our models include the four TRPV1 subunits and both DkTx chains but not RTX. One particle image with extreme intensity was excluded from the 218,787 inputs, leaving a filtered collection of 218,786 particles.

For PfCRT–Fab, we used the particles of EMPIAR-10330 from the PfCRT structure study associated with PDB 6UKJ[40], with 300-pixel input images at 1.035 Å per pixel.

Baselines and comparison scope

Baselines were run from their published implementations and standard configurations, with the dataset-specific settings below. Each baseline kept its own representation, initialization, image loss and geometry objectives.

CryoARC

CryoARC 0.1.0[30] uses separately obtained AF2 predictions and embeddings, an image encoder, a four-dimensional latent model and a structure module. Its implementation supports both pose optimization and fixed supplied poses. In preliminary CAHRA and TGFBIp runs with pose updates, the image loss and the rotation from the initial poses increased, the Kullback–Leibler (KL) divergence decreased and the sampled structures lost nearly all secondary structure (Extended Data Fig. 3). The reported CAHRA, TGFBIp and MSP-1 comparisons used the fixed-pose option of CryoARC, restarted from the initial AF2 parameters with the same prepared particle inputs and initial placement, and the checkpoints saved after 15 completed epochs.

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.26.000403. This version posted November 26, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

Following the authors' settings, all three fixed-pose runs used a five-layer convolutional encoder of width 32, four-dimensional latents, the pair stack, a four-layer modulator of width 32 and a trainable four-block structure module, with the angle parameters frozen. The inputs were real-space images, and the image loss was computed in the Fourier domain. The chi-angle and violation weights were each 0.01, and the KL weight was 0.25. The encoder and decoder learning rate was 8×10^{-5}, with a ten-step linear warmup followed by a constant rate; the runs used eight GPUs, a batch of one per GPU and seed 42.

The pose-update diagnostics used a model learning rate of 8×10^{-5} and a pose learning rate and weight decay of 10^{-4} each, without a pretraining epoch. The recorded per-epoch values are CryoARC's mean $1 - \text{corr}$ image loss, the unweighted KL divergence of the posterior from a standard normal and the mean orientation change from the initial poses. Each structural example is the lower-median structure by Cα radius of gyration among 100 samples, aligned to the AF2 model and shown with a common view and scale.

CryoSTAR and cryoSPHERE

For MSP-1, CryoSTAR 0.1.0[24] and cryoSPHERE 0.6.0[25] used the same 12,000 particles and the supplied starting structure in their prepared map-centered frame, with Cα representations and eight-dimensional variational-autoencoder latents. For all three datasets (60,000 particles each for CAHRA and TGFBIp), CryoSTAR used reference initialization, a 4 Å image-loss bandwidth, a batch size of 32, a learning rate of 10^{-4} and the checkpoint after 24,000 updates. cryoSPHERE used 20 learned segments, a batch size of 32, a 4 Å bandwidth, main and segmentation learning rates of 3×10^{-5} and 3×10^{-4}, respectively, and the checkpoint after all 300 epochs. cryoSPHERE training on TGFBIp became numerically unstable, so no ensemble from that run entered the comparison.

The cryoSPHERE-based ensemble submitted to CAHRA Challenge 2 combines cryoSPHERE clustering, targeted MD, side-chain relaxation and density-guided modeling of side chains and disordered regions.

Evaluation of structural ensembles

Selecting representative conformations

Representative sets were selected from the complete particle-latent table of each method: the learned latents for CoCoFlex and the saved encoder representations for encoder-based baselines. FPS uses Euclidean distances in the full, unwhitened latent space without PCA; it starts from the particle farthest from the latent mean and then repeatedly adds the particle whose distance to its nearest selected particle is largest. Ties are broken by row order, which makes FPS deterministic. K-means clustering uses Lloyd's algorithm in the same space, with K-means++ initialization, seed 42, ten initializations, a tolerance of 10^{-4} and at most 300 iterations. Each cluster is represented by the particle whose latent vector is closest to the cluster centroid.

For the two MD-based simulated benchmarks, we chose between 100 FPS representatives (FPS100) and 100 K-means representatives (K-means100) by the structural

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.26.000403. This version posted November 26, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

variation of the decoded structures. For each candidate set, all 4,950 pairwise Cα RMSDs were calculated after proper whole-model alignment on a fixed set of common atom identities, and the set with the larger median was used. This criterion favors a broader range of conformations without using the reference coordinates. For CAHRA, FPS100 was selected for CoCoFlex, CryoARC and CryoSTAR; for TGFBIp, K-means100 was selected for CoCoFlex and FPS100 for both baselines. The cryoSPHERE-based ensemble consists of its 100 submitted structures.

For the CAHRA configuration comparison (Extended Data Fig. 1), the 30,000-particle runs used a pose-proportional subset of the main 60,000-particle set. Each checkpoint was scored with FPS100, and the curves report cumulative heterogeneous-training time on eight GPUs, excluding prediction, canonical refinement, decoding, evaluation and queueing.

For MSP-1, we used six K-means representatives from the epoch-4.80 model to evaluate the recovery of the six published class-derived conformations; this fit used at most 100 iterations and otherwise the settings above. The representatives have equal weight in the coordinate metrics, and the cluster fractions are recorded separately. The learned clusters and the published particle classes are distinct groupings.

Stereochemical validation and restrained relaxation

Ramachandran outliers, rotamer outliers and clashscore were evaluated with Phenix 2.1-6048[47] and MolProbity[48], using identical procedures for the CAHRA and TGFBIp reference ensembles, the unrelaxed and relaxed CoCoFlex and CryoARC ensembles, the PULCHRA-completed and relaxed CryoSTAR ensembles and the CAHRA cryoSPHERE-based ensemble. Validation used the complete available heavy-atom models, including OXT. Outlier percentages are calculated over evaluable residues, and clashscore counts serious overlaps per 1,000 atoms. Reduce added hydrogens for these diagnostics without Asn/Gln/His flips, and Probe evaluated the contacts; hydrogens were excluded from coordinate scoring. The single-model controls were the initial Protenix-mini prediction, its CoCoFold2-refined model after one epoch and the unrelaxed full AF2 prediction supplied to CryoARC before modulator initialization. For MSP-1, the same procedures were applied to the six published structures and to the six representatives of each method before and after relaxation.

CryoSTAR Cα models were first completed to all heavy atoms with PULCHRA 3.06[41], which preserves the Cα positions. Each CoCoFlex, CryoARC and completed CryoSTAR conformation was then minimized independently in Phenix with harmonic restraints to its starting coordinates on all available heavy atoms (standard deviation 0.5 Å, without top-out). Standard covalent and nonbonded geometry terms were combined with Ramachandran restraints and rotamer correction allowing favored and allowed states. Minimization comprised two macrocycles of at most 100 L-BFGS iterations each and used no maps, reference conformations or generating labels. The starting coordinates underwent the standard Phenix model interpretation, including normalization of symmetric side-chain atom names. No structure was discarded. The main simulated-benchmark figures show CoCoFlex, CryoARC and CryoSTAR after restrained relaxation, whereas the reference ensembles and the CAHRA cryoSPHERE-based ensemble are shown with their supplied coordinates. Unrelaxed and relaxed

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.26.000403. This version posted November 26, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

structures are compared in Extended Data Fig. 2. For MSP-1, cryoSPHERE Cα models were completed with PULCHRA in the same way, relaxation was applied to the assembled four-chain models, and Fig. 4c–f use the relaxed representatives.

Coordinate identities and atom panels

CAHRA ensembles were scored with CAHRA-metrics 0.0.1, with no histidine-tag exclusion, a maximum PCV rank of 100 and the option `selection=protein`. CoCoFlex, CryoSTAR (after PULCHRA completion) and the cryoSPHERE-based ensemble were scored on the canonical 8,079 standard-protein heavy atoms; CryoARC was scored on 8,078 common atoms, excluding OXT, through an adapter to the core metric functions. CoCoFlex coordinates were reordered by atom identity and written in PDB format for scoring. The references were first aligned to one another, and each prediction was then fitted to its nearest aligned reference.

For MSP-1, chains were matched uniquely by their complete sequences, and atoms were identified by canonical chain, `label_seq_id` and atom name. The common atom panels were fixed from the canonical structure and all six references before any prediction was read, and every prediction had to contain all panel atoms. The 1,154-Cα and 9,584-heavy-atom panels were aligned and scored independently with whole-assembly fits and equal weights for atoms and representatives. MSP-1 was scored with an independent implementation, in which the references were fitted to the first reference and the predictions to their nearest aligned reference.

Ensemble metrics

Let $\mathcal{E}_{\mathrm{pred}}$ and $\mathcal{E}_{\mathrm{ref}}$ denote the predicted and reference (ground-truth) ensembles, with n_{pred} and n_{ref} structures, respectively, and let R_{ab} be the whole-assembly RMSD after proper rigid alignment (rotation determinant $+1$) of prediction a to reference b. Then

$$\mathrm{SAC} = \frac{1}{n_{\mathrm{pred}}} \sum_{a} \min_{b} R_{ab}, \qquad R_{\mathrm{cov}} = \frac{1}{n_{\mathrm{ref}}} \sum_{b} \min_{a} R_{ab}, \tag{8}$$

where R_{cov} is the mean RMSD from each reference to its nearest prediction. The coverage at threshold τ, $C(\tau) = n_{\mathrm{ref}}^{-1} \left|\{b : \min_a R_{ab} \le \tau\}\right|$, is the fraction of references whose nearest prediction lies within τ; Figs. 2b and 3b plot $C(\tau)$.

For an aligned ensemble Y with n_Y structures and n_{atom} common atoms,

$$f_j(Y) = \sqrt{\frac{1}{n_Y} \sum_{a} \left\| Y_{aj} - \frac{1}{n_Y} \sum_{a'} Y_{a'j} \right\|_2^2}, \qquad \mathrm{LFC} = \frac{1}{n_{\mathrm{atom}}} \sum_{j} \left| f_j(\mathcal{E}_{\mathrm{pred}}) - f_j(\mathcal{E}_{\mathrm{ref}}) \right|. \tag{9}$$

For PCV, the aligned coordinate ensembles are flattened and mean-centered separately before PCA. With Q_{pred} and Q_{ref} containing the orthonormal predicted and reference coordinate directions and Σ_{ref} the reference singular values,

$$\mathrm{PCV}(\mathcal{E}_{\mathrm{pred}} \leftarrow \mathcal{E}_{\mathrm{ref}}) = \frac{\| Q_{\mathrm{pred}}^{\mathsf{T}} Q_{\mathrm{ref}} \Sigma_{\mathrm{ref}} \|_F^2}{\| \Sigma_{\mathrm{ref}} \|_F^2}, \tag{10}$$

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.26.000403. This version posted November 26, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

so the reference variance supplies the denominator. For MSP-1, up to five nonzero predicted directions are compared with all valid reference directions; singular values are thresholded at $\max(10^{-8}\,\text{Å}, 10^{-7}\sigma_{\max})$, where $\sigma_{\max}$ is the largest singular value, and the rank is bounded by the number of structures minus one, so six centered references have rank at most five. PCV is zero when the predicted variance is zero and undefined when the reference variance is zero. This coordinate PCA is calculated from generated structures and is distinct from the four-dimensional latent representation.

For the paired comparisons of unrelaxed and relaxed structures, we scored 1,014 Cα and 8,078 heavy atoms for CAHRA and 594 Cα and 4,567 heavy atoms for TGFBIp, excluding OXT, and 1,154 Cα and 9,584 heavy atoms for MSP-1; PULCHRA-completed CryoSTAR models of CAHRA were scored on the canonical 8,079 heavy atoms, including OXT. Each panel was aligned and scored against the same references. Simulated benchmarks used the public PCV core (rank limit 100, seed 5). The PCV rank limit was five for MSP-1. Unrelaxed and relaxed scores were calculated on identical conformations and atom panels. In the structure-level distributions, each observation is one sampled conformation of the evaluated ensemble.

Reference-based structural analyses

Each CAHRA opening angle is the angle between the heavy-atom mass-weighted centers of D1 (residues 10–265) and D4 (residues 832–1,014), with the corresponding center of the linker (residues 539–541) as the vertex. Switch C comprises natural human presequence protease residues 506–550[42], which correspond to model residues 483–527. Residue-wise RMSF is the mean heavy-atom RMSF of each residue. For the reference variance fractions in Fig. 3a, the Cα coordinates of each reference ensemble, 1,014 atoms for CAHRA and 594 for TGFBIp, were aligned to its first structure and mean-centered before PCA.

Latent-space and regional analyses

Latent PCs and axis traversals

For HER2, TRPV1 and PfCRT, PCA was applied to the complete, mean-centered four-dimensional latent table without scaling or whitening. The score of particle i along unit principal direction v_p is

$$s_{ip} = (z_i - \bar{z})^{\mathsf{T}} v_p. \tag{11}$$

Each principal direction has the same sign for structure sampling and particle grouping. Each axis traversal decodes the 21 latent vectors $\bar{z} + t\,v_p$, with t equally spaced between the 5th and 95th percentiles of the observed scores and the scores along the other PCs held at zero; we refer to the decoded structures as axis models. All four latent PCs were inspected, and the main figures emphasize the directions that illustrate distinct regional motions in the decoded structures.

For MSP-1, PCA was applied to the complete mean-centered latent table, with the sign of PC1 following the centroid order. Clusters are ordered by the projections of their centroids onto the latent PCs, whose signs are set deterministically.

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.26.000403. This version posted November 26, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

Regional structural displacement

For HER2, the displayed axes are PC1, PC3 and PC4 (Fig. 5) and PC2 (Extended Data Fig. 5a), with whole-complex and Fab views aligned on HER2 chain E. The pertuzumab constant domains comprise C108–214 and D120–222. The internal-Fab view instead aligns the trastuzumab antigen-binding variable domains (A1–107 and B1–120) to the mean-latent model.

For TRPV1, the displayed axes are PC2 (Fig. 6) and PC1, PC3 and PC4 (Extended Data Fig. 5b,e). The ARDs comprise model residues 5–254 (natural residues 110–359). The pore alignment uses model residues 470–495 and 537–557 of chains A–D (natural residues 575–600 and 665–685), and the F-chain displacement is calculated over its 75 Cα atoms after the same alignment.

For PfCRT, the displayed directions are PC2 (Fig. 6) and PC1, PC3 and PC4 (Extended Data Fig. 5c,d). The membrane-core anchor comprises residues A50–116 and A126–408, the proximal Fab B1–129 and C1–106, and the distal Fab B130–241 and C107–215, each restricted to the Cα atoms available in all models.

For HER2, TRPV1 and PfCRT, each model is aligned by a single proper rigid transformation on the receptor or membrane-core Cα anchor defined above. In this common frame, without refitting the moving region U, the direct displacement of axis model m from the low-score model is $[n_U^{-1}\sum_j \|\mathbf{x}_j^{(m)} - \mathbf{x}_j^{(\mathrm{low})}\|^2]^{1/2}$, where $\mathbf{x}_j^{(m)}$ are the positions of the n_U Cα atoms of U in model m. Endpoint values compare the high-score and low-score models; for non-monotonic traversals, the reported intermediate model is the one with the largest displacement from the low-score model. Figs. 5b and 6c report this displacement for all 21 axis models of each traversal, over the Cα atoms of each Fab (HER2), each ARD (TRPV1) or the proximal and distal Fab regions (PfCRT). Extended Data Fig. 5 reports it in the same way for the pertuzumab constant domains (HER2 PC2), A351–375, ARD A and the F chain (TRPV1) and A31–55, A1–49 and B121–145 (PfCRT), together with the two Fabs (HER2), the four ARDs and the DkTx chains (TRPV1) or the proximal and distal Fab regions and the other named segments (PfCRT). The separately labeled internal-Fab view aligns the antigen-binding variable domains instead.

For MSP-1, the six published models were aligned to the main conformation (6ZBC) over 330 common Cα atoms in coiled-coil helices H30–H37, and wing centroids were calculated from the 209 common Cα atoms of the wing-domain helices. For the group-to-structure correspondence (Fig. 4c), the wing centroids of the six relaxed group representatives were compared with those of the six published models after the same coiled-coil alignment, and groups and structures were paired one-to-one by minimizing the total centroid distance.

Density-map analyses

PC-group reconstructions

For each PC, the training particles were ranked by s_{ip} and divided into five non-overlapping PC groups of approximately equal size, with ties ordered by their row in the training table. Each PC group was reconstructed with CryoSPARC Homogeneous Reconstruction Only[49] using C1 symmetry and fixed poses. The exported particles

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.26.000403. This version posted November 26, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

kept their identities, orientations, translations, CTFs and half-set assignments. Comparisons of PC-group maps used the first, third and fifth groups, which cover the 0–20%, 40–60% and 80–100% score windows. Training-pool sizes are listed in Extended Data Table 4, and the input counts of the particle groups and displayed maps in Extended Data Tables 1 and 2. Whereas each axis model holds the scores along the other PCs at zero, each PC-group map averages the remaining latent variation within its score window.

Full-collection classification and regional refinement

For HER2 and TRPV1, the PC-group maps provided the initial references for five-class CryoSPARC 3D Classification[49] of the full filtered particle collections, which used the upstream poses and CTF parameters and performed no alignment. The displayed HER2 maps used a focus on the AB Fab for PC1 and PC3 and on the CD Fab for PC4, and the displayed TRPV1 PC2 maps used a regional focus covering all four ARDs. Three representative classes per axis were selected and ordered by their regional model–map CC with the 21 axis models, so the CC patterns of the selected classes are not independent of the models. Each selected class map was paired with one of the fixed −, 0 and + axis models and labeled accordingly (Figs. 5 and 6 and Extended Data Table 2); the deposited maps are labeled Low, Medium and High. The classes and their full-collection particle assignments differ from the PC groups and are recorded in the source data.

Full-size class maps were reconstructed with the class assignments and poses from classification. For each selected class, Split Volumes paired the class map with its particle subset, and these pairs provided the initial volume and particles for C1 Local Refinement. Within each axis, a common static mask covered the target region: the AB Fab for HER2 PC1 and PC3, the CD Fab for HER2 PC4 and all four ARDs for TRPV1 PC2. Local Refinement adjusted the particle orientations and shifts around the mask center. Local-resolution fields were estimated from the corresponding half-map pairs with the same regional mask, an FSC threshold of 0.5 and an annealing factor of zero. For PfCRT PC2 and TRPV1 PC3, the fixed-pose PC-group maps were used instead.

Half-map assessment

Corrected FSC resolutions were taken from the CryoSPARC half-map FSC curves of each reconstruction or local-refinement job, corrected by noise substitution; for locally refined maps, these curves used the corresponding refinement masks. The regional half-map FSC summarizes half-map agreement over a fixed evaluation region of interest (ROI) by its uncorrected 0.143 crossing; core registration places the same soft evaluation region on the native grid of each half map. The ROI local-resolution median summarizes the spatial distribution of voxel-wise resolution estimates obtained at a local FSC of 0.5. Local resolution was estimated with the reconstruction masks for PC-group maps and with the regional refinement masks for class maps, and both were summarized over the same fixed evaluation ROI within each axis. Medians use valid positive estimates, and their spatial coverage is recorded in the source data. Both

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.26.000403. This version posted November 26, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

measures are reported in Extended Data Tables 1 and 2. For the Fourier coefficients $\hat{\rho}_1$ and $\hat{\rho}_2$ of the two half maps in frequency shell S_q,

$$\mathrm{FSC}(q) = \frac{\mathrm{Re}\left[\sum_{\mathbf{k}\in S_q} \hat{\rho}_1(\mathbf{k})\hat{\rho}_2(\mathbf{k})^*\right]}{\sqrt{\sum_{\mathbf{k}\in S_q} |\hat{\rho}_1(\mathbf{k})|^2 \sum_{\mathbf{k}\in S_q} |\hat{\rho}_2(\mathbf{k})|^2}}. \tag{12}$$

Cross-comparison of models and density maps

Model–map CC and the displayed motion measurements use the full map reconstructed from both half-sets, whereas FSC uses the two half maps. Model densities were calculated with Phenix 2.1-6048 and cctbx from electron scattering factors, with unit occupancy and zero atomic displacement parameters after removal of hydrogens. Within each region, every fixed model was compared with every displayed map in the common reference frame, without fitting individual models to individual maps. Each MSP-1 map received the coiled-coil alignment transformation of its representative. The CC matrices in the main figures use the same maps as the corresponding density panels.

Regional masks cover the union of the relevant heavy-atom positions, with a weight of one within 4 Å and a cosine taper to zero at 8 Å; each comparison uses one mask for all of its models and maps.

Real-space CC is the Pearson correlation over voxels with a mask weight of at least 0.5, after the model and experimental densities are low-pass filtered with a common Fourier cutoff of 6 Å. The HER2, TRPV1 and PfCRT comparisons use the fixed −, 0 and + axis models, and the MSP-1 comparisons the Group 1 and Group 6 representatives.

Regional model and density motion

Regional motion is measured relative to the HER2 receptor, the TRPV1 pore core, the PfCRT membrane core or the MSP-1 coiled coil. Map displacement fields are estimated after the maps are smoothed with a Gaussian of 8 Å full width at half maximum (FWHM), a smoothing distinct from the Fourier cutoffs used for CC.

For each region and anchor, rigid map-to-map transformations are estimated independently by maximizing a weighted local density correlation with a common reference map: the middle PC-group map for all displayed class maps, or the Group 1 map for MSP-1. Correlations are evaluated on a 3 Å grid within 8 Å of the reference Cα positions of the region, with unit weight up to 4 Å and a cosine taper to 8 Å. Powell optimization starts from the identity and is bounded to ±15° per rotation-vector component and ±8 Å per component of the centroid translation. The displayed fields come from the map series of each axis; the PC-group maps and their half maps were also fitted independently to assess reconstruction repeatability. Model transformations are proper Kabsch fits of the corresponding regional and anchor Cα coordinates.

For reference positions $\mathbf{x}_j$ ($j = 1, \ldots, n_U$) of region U, let $\mathcal{T}_{U,m}$ and $\mathcal{T}_{A,m}$ map the reference state to state m for the region and for its anchor A. The anchor-relative

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.26.000403. This version posted November 26, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

positions and the endpoint displacement field are

$$\mathbf{y}_{m,j} = \mathcal{T}_{A,m}^{-1}(\mathcal{T}_{U,m}(\mathbf{x}_j)), \qquad \mathbf{d}_j = \mathbf{y}_{\mathrm{high},j} - \mathbf{y}_{\mathrm{low},j}, \tag{13}$$

and the rigid-field amplitude and directional agreement are

$$\mathrm{RMS} = \sqrt{\frac{1}{n_U}\sum_j \|\mathbf{d}_j\|^2}, \qquad \cos\theta = \frac{\sum_j \mathbf{d}_j^{\mathrm{model}} \cdot \mathbf{d}_j^{\mathrm{map}}}{\sqrt{\sum_j \|\mathbf{d}_j^{\mathrm{model}}\|^2 \sum_j \|\mathbf{d}_j^{\mathrm{map}}\|^2}}. \tag{14}$$

The same cosine compares fields estimated separately from the two half-map series. Rigid-field amplitudes describe the transformations at common reference positions and differ from the direct coordinate displacements defined above. Map motion metrics are calculated from the maps shown in the corresponding density panels or, for TRPV1 PC3, from the first and fifth PC-group maps.

Visualization

Structures and densities were visualized with UCSF ChimeraX[50], which was also used to render the structures in Supplementary Videos 1–5. For HER2, TRPV1 and PfCRT, the overview panels show the mean-latent model. Displayed maps were background-subtracted and normalized using a fixed receptor or membrane-core region. HER2 contours equal the median of the solvent background plus one core-region standard deviation of the native map; TRPV1 class maps and PfCRT group maps are contoured at 1.2 core-region standard deviations, with the same rule for all compared states.

Each class map received one rigid core-registration transformation into the frame of the middle PC-group map, followed by a scene transformation. In Figs. 5 and 6, the scene transformation of each map is the display alignment of its paired model, and the axis models keep their common anchor alignment.

Map surfaces are shown without additional low-pass filtering or surface smoothing. In Figs. 5 and 6, they are limited to 3 Å around the displayed Fab or ARD atoms, with surface fragments smaller than 6 Å hidden. In the TRPV1 and PfCRT overviews (Fig. 6a,b), the middle PC2-group map provides nanodisc context: it is restricted to a disc around the membrane-core centroid (radius 87 Å for TRPV1 and 66 Å for PfCRT; half-thickness 20 Å), weighted from 0 at 6 Å to 1 at 11 Å from any model atom, and shown at 0.8 core-region standard deviations with surface fragments smaller than 10 Å hidden. Views within each row share orientation, and the TRPV1 density views enlarge ARD A. In Extended Data Fig. 5, each locator shares the orientation of the enlarged views of its panel.

For MSP-1, the displayed wing densities are limited to 3 Å around the wing atoms, with surface fragments smaller than 6 Å hidden. The endpoint group representatives are aligned on the same coiled coil, and their transformations are applied to the paired maps, which share a contour determined from the means and standard deviations of all six maps.

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.26.000403. This version posted November 26, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

Supplementary information

Supplementary Videos 1–5 are provided as separate MP4 files.

Supplementary Video 1

Domain opening in the CAHRA reference and relaxed CoCoFlex ensembles. The video steps through 21 reference structures (light gray) with opening angles of 65.6–94.8°, from the most closed to the most open and back, each superposed on a relaxed CoCoFlex structure whose opening angle differs by at most 0.6°. CoCoFlex structures are colored by opening angle (64–100°), and the panel at left gives the opening angles of the displayed pair.

Supplementary Video 2

Structural variation of the relaxed CoCoFlex ensemble on the TGFBIp benchmark. Each panel steps through 21 relaxed CoCoFlex structures in order of their scores on the first (left) or second (right) PC of the reference ensemble, from the lowest to the highest score and back. Insets plot the reference (gray) and relaxed CoCoFlex (brown) structures on the first two reference PCs and mark the displayed structure. Displayed structures are colored by terminal-segment centroid distance (Cα centroids of residues 1–100 and 501–594; 63–87 Å), whose value is given at lower right.

Supplementary Video 3

Fab motions decoded along latent PCs of the HER2 complex. The video first shows the complex, with the trastuzumab Fab rose, the pertuzumab Fab sand and HER2 light gray. It then steps through the 21 conformations decoded along latent PC1 and PC3 (trastuzumab Fab) and PC4 (pertuzumab Fab), from the 5th to the 95th percentile of each PC score and back, with the Fab colored from purple through teal to gold. Models are aligned on HER2 for PC1 and PC4, with nearby HER2 residues white, and on the trastuzumab variable domains for PC3. Insets locate the displayed conformation among the particle latents.

Supplementary Video 4

ARD motion decoded along TRPV1 latent PC2. The video steps through the 21 conformations decoded along latent PC2, from the 5th to the 95th percentile of the PC2 score and back, viewed along the pore axis and then, after the view rotates, from the side with only chains A and C shown. The ARDs are colored from purple through teal to gold, and other regions are white. Insets locate the displayed conformation among the particle latents.

Supplementary Video 5

Fab motion decoded along PfCRT latent PC2. The video first shows the PfCRT–Fab complex, with the Fab constant domains rose, the Fab variable domains sand and PfCRT light gray. It then steps through the 21 conformations decoded along

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.26.000403. This version posted November 26, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

latent PC2, from the 5th to the 95th percentile of the PC2 score and back, after alignment on the PfCRT core. The distal Fab (constant domains) is colored from purple through teal to gold, and the proximal Fab and PfCRT are white. Insets locate the displayed conformation among the particle latents.

Declarations

Funding

This work was supported by the National Key Research and Development Program of China (grant no. 2025YFA1309602 to M.H.), the National Natural Science Foundation of China (grant nos. 12625119 and W2612010 to C.B.) and the Shenzhen Medical Academy of Research and Translation (to M.H.).

Acknowledgments

We acknowledge the support in data analysis provided by the Computing Labware for Electron-microscopy Visualization and Experimental Research (CLEVER) at the Shenzhen Medical Academy of Research and Translation.

Competing interests

The authors declare no competing interests.

Data availability

The particle datasets used in this study are available through EMPIAR: CAHRA (EMPIAR-14006), MSP-1 (EMPIAR-10437), TRPV1 (EMPIAR-10059), PfCRT (EMPIAR-10330) and HER2 (EMPIAR-11665). The TGFBIp MD trajectory is available through ATLAS[44]. The published structures used in this study are available from the PDB under accession codes 6ZBC, 6ZBD, 6ZBE, 6ZBF, 6ZBG and 6ZBH (MSP-1), 7ASG (TGFBIp), 8PWH (HER2), 5IRX (TRPV1) and 6UKJ (PfCRT). The generated atomic ensembles, particle-group reconstruction maps, locally refined maps and numerical source data are available from Zenodo at https://doi.org/10.5281/zenodo.24097400.

Code availability

CoCoFlex is currently available as a module in CoCo (https://cocoproject.org.cn). Its source code and documentation are being prepared and will be made available at https://github.com/jwliaomath/cocoflex.

Contributions

C.B., M.H. and J.L. conceived the study. J.L. developed the methodology, implemented the software, performed the computational experiments and analyzed the data. M.H. and J.L. prepared the figures. C.B. and M.H. supervised the project, provided resources and acquired funding. J.L. wrote the first draft of the manuscript, and all authors reviewed and edited it.

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.26.000403. This version posted November 26, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

References

[1] Henzler-Wildman, K. & Kern, D. Dynamic personalities of proteins. *Nature* **450**, 964–972 (2007).

[2] Toader, B., Sigworth, F. J. & Lederman, R. R. Methods for cryo-EM single particle reconstruction of macromolecules having continuous heterogeneity. *Journal of Molecular Biology* **435**, 168020 (2023).

[3] Shenoy, J. *et al.* 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. arXiv preprint arXiv:2605.09832.

[4] Punjani, A. & Fleet, D. J. 3DFlex: determining structure and motion of flexible proteins from cryo-EM. *Nature Methods* **20**, 860–870 (2023). URL https://www.nature.com/articles/s41592-023-01853-8.

[5] Feathers, J. R. *et al.* The CAHRA Challenge: a community-wide assessment of cryo-EM heterogeneous reconstruction algorithms (2026). URL https://www.biorxiv.org/content/10.64898/2026.09.15.751515v3. bioRxiv preprint, version 3, 10 October 2026.

[6] Nakane, T., Kimanius, D., Lindahl, E. & Scheres, S. H. W. Characterisation of molecular motions in cryo-EM single-particle data by multi-body refinement in RELION. *eLife* **7**, e36861 (2018). URL https://elifesciences.org/articles/36861.

[7] Punjani, A. & Fleet, D. J. 3D variability analysis: Resolving continuous flexibility and discrete heterogeneity from single particle cryo-EM. *Journal of Structural Biology* **213**, 107702 (2021). URL https://linkinghub.elsevier.com/retrieve/pii/S1047847721000071.

[8] Gilles, M. A. & Singer, A. Cryo-EM heterogeneity analysis using regularized covariance estimation and kernel regression. *Proceedings of the National Academy of Sciences* **122**, e2419140122 (2025). URL https://pnas.org/doi/10.1073/pnas.2419140122.

[9] Zhong, E. D., Bepler, T., Berger, B. & Davis, J. H. CryoDRGN: reconstruction of heterogeneous cryo-EM structures using neural networks. *Nature Methods* **18**, 176–185 (2021). URL https://www.nature.com/articles/s41592-020-01049-4.

[10] Chen, M. & Ludtke, S. J. Deep learning-based mixed-dimensional Gaussian mixture model for characterizing variability in cryo-EM. *Nature Methods* **18**, 930–936 (2021). URL https://www.nature.com/articles/s41592-021-01220-5.

[11] Schwab, J., Kimanius, D., Burt, A., Dendooven, T. & Scheres, S. H. W. DynaMight: estimating molecular motions with improved reconstruction from cryo-EM images. *Nature Methods* **21**, 1855–1862 (2024). URL https://www.

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.26.000403. This version posted November 26, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

nature.com/articles/s41592-024-02377-5.

[12] Maddipatla, A. *et al.* Fold'EM: direct atomic structure inference from cryo-EM particles (2026). URL https://arxiv.org/abs/2610.01358v2. arXiv preprint arXiv:2610.01358.

[13] Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021). URL https://www.nature.com/articles/s41586-021-03819-2.

[14] Abramson, J. *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500 (2024). URL https://www.nature.com/articles/s41586-024-07487-w.

[15] Jamali, K. *et al.* Automated model building and protein identification in cryo-EM maps. *Nature* **628**, 450–457 (2024). URL https://www.nature.com/articles/s41586-024-07215-4.

[16] Wang, X., Zhu, H., Terashi, G., Taluja, M. & Kihara, D. DiffModeler: large macromolecular structure modeling for cryo-EM maps using a diffusion model. *Nature Methods* **21**, 2307–2317 (2024). URL https://www.nature.com/articles/s41592-024-02479-0.

[17] Terwilliger, T. C. *et al.* Improved AlphaFold modeling with implicit experimental information. *Nature Methods* **19**, 1376–1382 (2022). URL https://www.nature.com/articles/s41592-022-01645-6.

[18] Fadini, A. *et al.* AlphaFold as a prior: experimental structure determination conditioned on a pretrained neural network. *Nature Methods* **23**, 785–795 (2026). URL https://www.nature.com/articles/s41592-026-03047-4.

[19] Raghu, R., Levy, A., Wetzstein, G. & Zhong, E. D. Multiscale guidance of protein structure prediction with heterogeneous cryo-EM data (2025). URL http://www.proceedings.com/085713-0344.html. Advances in Neural Information Processing Systems 38 (NeurIPS 2025), pp. 11329–11357.

[20] Jin, Q. *et al.* Iterative elastic 3D-to-2D alignment method using normal modes for studying structural dynamics of large macromolecular complexes. *Structure* **22**, 496–506 (2014). URL https://linkinghub.elsevier.com/retrieve/pii/S0969212614000136.

[21] Vuillemot, R. *et al.* MDSPACE: extracting continuous conformational landscapes from cryo-EM single particle datasets using 3D-to-2D flexible fitting based on molecular dynamics simulation. *Journal of Molecular Biology* **435**, 167951 (2023). URL https://linkinghub.elsevier.com/retrieve/pii/S0022283623000074.

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.26.000403. This version posted November 26, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

[22] Rosenbaum, D. *et al.* Inferring a continuous distribution of atom coordinates from cryo-EM images using VAEs (2021). URL https://arxiv.org/abs/2106.14108. arXiv preprint arXiv:2106.14108.

[23] Zhong, E. D., Lerer, A., Davis, J. H. & Berger, B. Exploring generative atomic models in cryo-EM reconstruction (2021). URL https://arxiv.org/abs/2107.01331. arXiv preprint arXiv:2107.01331.

[24] Li, Y., Zhou, Y., Yuan, J., Ye, F. & Gu, Q. CryoSTAR: leveraging structural priors and constraints for cryo-EM heterogeneous reconstruction. *Nature Methods* **21**, 2318–2326 (2024). URL https://www.nature.com/articles/s41592-024-02486-1.

[25] Ducrocq, G. C. J., Grunewald, L., Westenhoff, S. & Lindsten, F. cryoSPHERE: Single-Particle HEterogeneous REconstruction from cryo EM (2025). URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/2ac2eac5098dba08208807b65c5851cc-Abstract-Conference.html. International Conference on Learning Representations (ICLR), 2025.

[26] Chen, M., Toader, B. & Lederman, R. Integrating molecular models into CryoEM heterogeneity analysis using scalable high-resolution deep Gaussian mixture models. *Journal of Molecular Biology* **435**, 168014 (2023). URL https://linkinghub.elsevier.com/retrieve/pii/S0022283623000700.

[27] Zhang, C. *et al.* CryoDyna: Multiscale end-to-end modeling of cryo-EM macromolecule dynamics with physics-aware neural network (2025). URL https://arxiv.org/abs/2510.16510. arXiv preprint arXiv:2510.16510.

[28] Silva-Sánchez, D. *et al.* Cryo-electron microscopy structural ensemble optimization using individual particles. *Journal of Chemical Theory and Computation* **22**, 4922–4938 (2026). URL https://pubs.acs.org/doi/10.1021/acs.jctc.6c00053.

[29] Chen, M. Building molecular model series from heterogeneous CryoEM structures using Gaussian mixture models and deep neural networks. *Communications Biology* **8**, 798 (2025). URL https://www.nature.com/articles/s42003-025-08202-9.

[30] Vuillemot, R. & Grudinin, S. CryoARC: Atomic-resolution conformational landscapes of protein assemblies from cryo-EM single particles with evolutionary priors (2026). URL https://www.biorxiv.org/content/10.64898/2026.05.25.727696v1. bioRxiv preprint, version 1, 26 May 2026.

[31] Liao, J. *et al.* Fine-tuning AlphaFold with limited cryo-EM observations. *Communications Chemistry* **9**, 95 (2026). URL https://www.nature.com/articles/s42004-026-01899-7.

[32] Liao, J., Hu, M. & Bao, C. CoCoFold2: scalable latent refinement of diffusion-based protein structure predictions from limited-particle cryo-EM data (2026).

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.26.000403. This version posted November 26, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

URL https://langtaosha.org.cn/lts/preprint/view/338. LangTaoSha Preprint Server.

[33] Park, J. J., Florence, P., Straub, J., Newcombe, R. & Lovegrove, S. DeepSDF: learning continuous signed distance functions for shape representation (2019). URL https://doi.org/10.1109/CVPR.2019.00025. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 165–174.

[34] Levy, A. *et al.* CryoDRGN-AI: neural ab initio reconstruction of challenging cryo-EM and cryo-ET datasets. *Nature Methods* **22**, 1486–1494 (2025).

[35] Gong, C. *et al.* Protenix-Mini: efficient structure predictor via compact architecture, few-step diffusion and switchable pLM (2025). URL https://arxiv.org/abs/2507.11839v1. arXiv preprint arXiv:2507.11839.

[36] ByteDance AML AI4Science Team *et al.* Protenix – advancing structure prediction through a comprehensive AlphaFold3 reproduction (2025). URL https://www.biorxiv.org/content/10.1101/2025.01.08.631967v1. bioRxiv preprint.

[37] Dijkman, P. M. *et al.* Structure of the merozoite surface protein 1 from *Plasmodium falciparum*. *Science Advances* **7**, eabg0465 (2021). URL https://www.science.org/doi/10.1126/sciadv.abg0465.

[38] Ruedas, R. *et al.* Structure and conformational variability of the HER2-trastuzumab-pertuzumab complex. *Journal of Structural Biology* **216**, 108095 (2024). URL https://linkinghub.elsevier.com/retrieve/pii/S1047847724000352.

[39] Gao, Y., Cao, E., Julius, D. & Cheng, Y. TRPV1 structures in nanodiscs reveal mechanisms of ligand and lipid action. *Nature* **534**, 347–351 (2016). URL https://www.nature.com/articles/nature17964.

[40] Kim, J. *et al.* Structure and drug resistance of the *Plasmodium falciparum* transporter PfCRT. *Nature* **576**, 315–320 (2019). URL https://www.nature.com/articles/s41586-019-1795-x.

[41] Rotkiewicz, P. & Skolnick, J. Fast procedure for reconstruction of full-atom protein models from reduced representations. *Journal of Computational Chemistry* **29**, 1460–1465 (2008).

[42] Liang, W. G. *et al.* Structural basis for the mechanisms of human presequence protease conformational switch and substrate recognition. *Nature Communications* **13**, 1833 (2022).

[43] Astore, M. A., Silva-Sánchez, D., Blackwell, R., Cossio, P. & Hanson, S. M. Comparing cryo-EM methods and molecular dynamics simulation to investigate heterogeneity in ligand-bound TRPV1. *Nature Communications* **17**, 1067 (2026). URL https://www.nature.com/articles/s41467-025-67821-2.

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.26.000403. This version posted November 26, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

[44] Vander Meersche, Y., Cretin, G., Gheeraert, A., Gelly, J.-C. & Galochkina, T. ATLAS: protein flexibility description from atomistic molecular dynamics simulations. *Nucleic Acids Research* **52**, D384–D392 (2024). URL https://academic.oup.com/nar/article/52/D1/D384/7438909.

[45] Parkhurst, J. M. *et al.* Parakeet: a digital twin software pipeline to assess the impact of experimental parameters on tomographic reconstructions for cryo-electron tomography. *Open Biology* **11**, 210160 (2021). URL https://royalsocietypublishing.org/doi/10.1098/rsob.210160.

[46] Kimanius, D. *et al.* Data-driven regularization lowers the size barrier of cryo-EM structure determination. *Nature Methods* **21**, 1216–1221 (2024). URL https://www.nature.com/articles/s41592-024-02304-8.

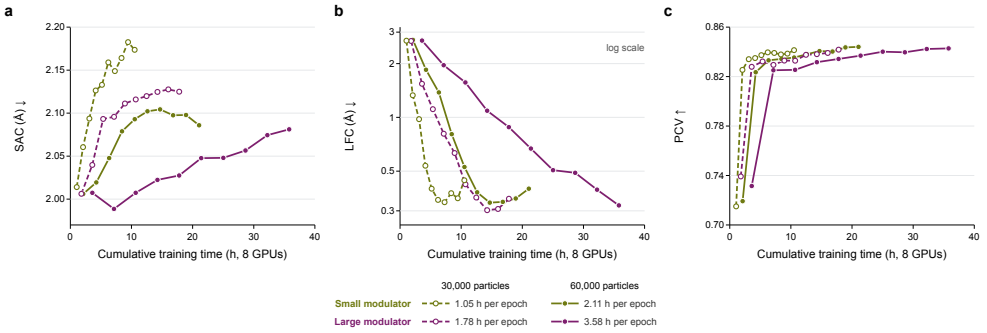
[47] Liebschner, D. *et al.* Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallographica Section D: Structural Biology* **75**, 861–877 (2019). URL https://journals.iucr.org/d/issues/2019/10/00/di5033/.

[48] Williams, C. J. *et al.* MolProbity: More and better reference data for improved all-atom structure validation. *Protein Science* **27**, 293–315 (2018). URL https://onlinelibrary.wiley.com/doi/10.1002/pro.3330.

[49] Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nature Methods* **14**, 290–296 (2017). URL https://www.nature.com/articles/nmeth.4169.

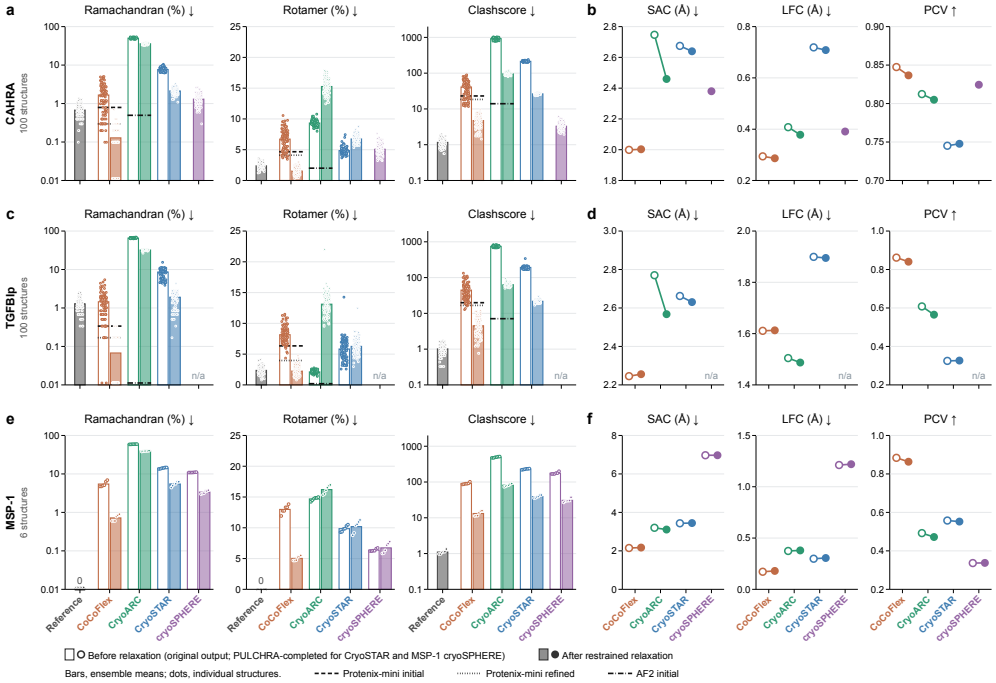
[50] Pettersen, E. F. *et al.* UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Science* **30**, 70–82 (2021). URL https://doi.org/10.1002/pro.3943.

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.26.000403. This version posted November 26, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0



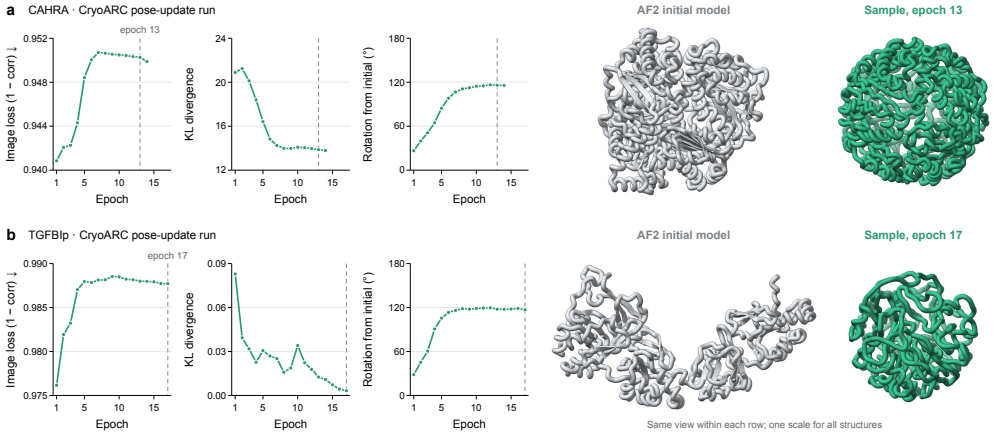
Extended Data Fig. 1 Training cost and structural metrics for four CoCoFlex configurations on CAHRA. a–c, SAC (a), LFC (b) and PCV (c) of the checkpoints saved after each of the first ten epochs of heterogeneous training with the 6 Å cutoff, plotted against cumulative heterogeneous-training time on eight GPUs (Methods). Colors distinguish the small and large modulators (one and four pair-conditioning blocks, respectively); dashed lines with open circles show training on 30,000 particles and solid lines with filled circles training on 60,000 particles. Each checkpoint was evaluated using 100 FPS-selected structures. The legend gives the mean training time per epoch for each configuration. LFC is shown on a logarithmic axis; arrows indicate the favorable direction.

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.26.000403. This version posted November 26, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0



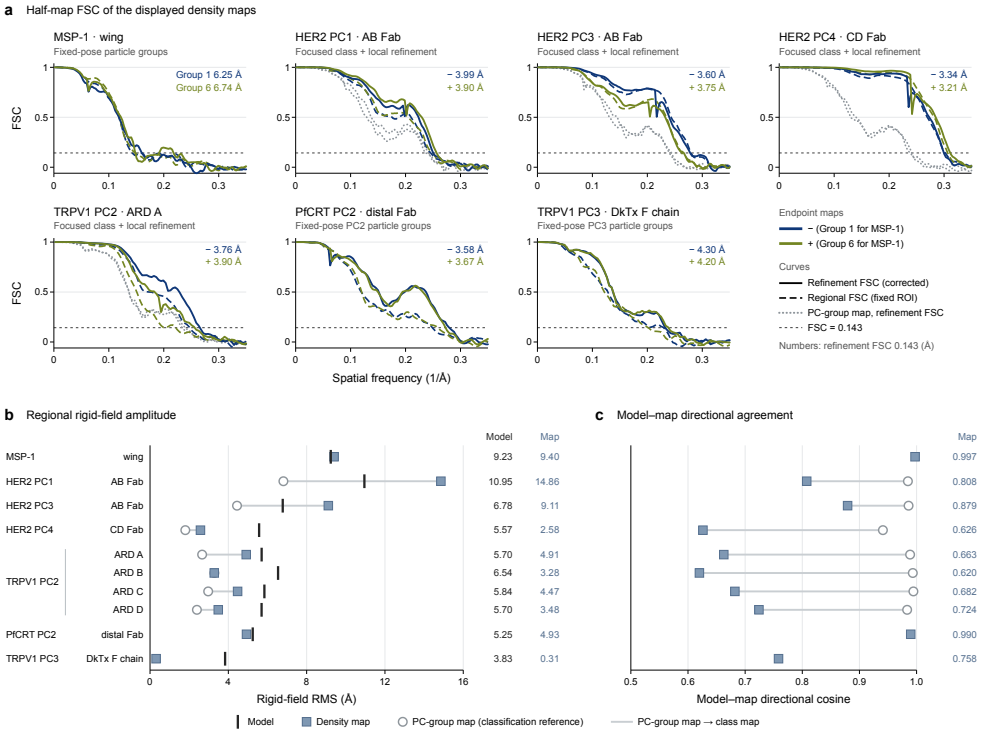
Extended Data Fig. 2 Effects of restrained relaxation on stereochemistry and ensemble recovery. a,b, CAHRA. **c,d**, TGFBIp. **e,f**, MSP-1. Open and filled symbols show structures before and after restrained relaxation, respectively. Before relaxation, CoCoFlex and CryoARC structures are the original outputs; the Cα outputs of CryoSTAR and, for MSP-1, cryoSPHERE are shown after completion with PULCHRA. Reference structures and the CAHRA cryoSPHERE-based ensemble retain their supplied coordinates and are shown once, as filled symbols. CAHRA and TGFBIp ensembles contain 100 structures; for MSP-1, six representatives per method are compared with the six published structures. n/a, no cryoSPHERE ensemble entered the TGFBIp comparison (Methods). **a,c,e**, Ramachandran outliers, rotamer outliers and clashscore (↓ indicates lower is better). Bars give ensemble means and dots individual structures. Ramachandran outliers and clashscore use logarithmic axes, on which zero values are drawn at the bottom; zero means are labeled. In a and c, black lines mark the single-model controls in the group of the method that starts from them: the Protenix-mini initial (dashed) and Protenix-mini refined (dotted) models for CoCoFlex and the AF2 initial model (dash-dotted) for CryoARC. **b,d,f**, Heavy-atom SAC, LFC and PCV (arrows indicate the favorable direction); lines connect the two states of each method. The heavy-atom panels contain 8,078 common atoms excluding OXT for CAHRA (8,079 for CryoSTAR and the cryoSPHERE-based ensemble), 4,567 for TGFBIp and 9,584 for MSP-1. Values after relaxation match those in Figs. 2b,c, 3b,c and 4e,f.

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.26.000403. This version posted November 26, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0



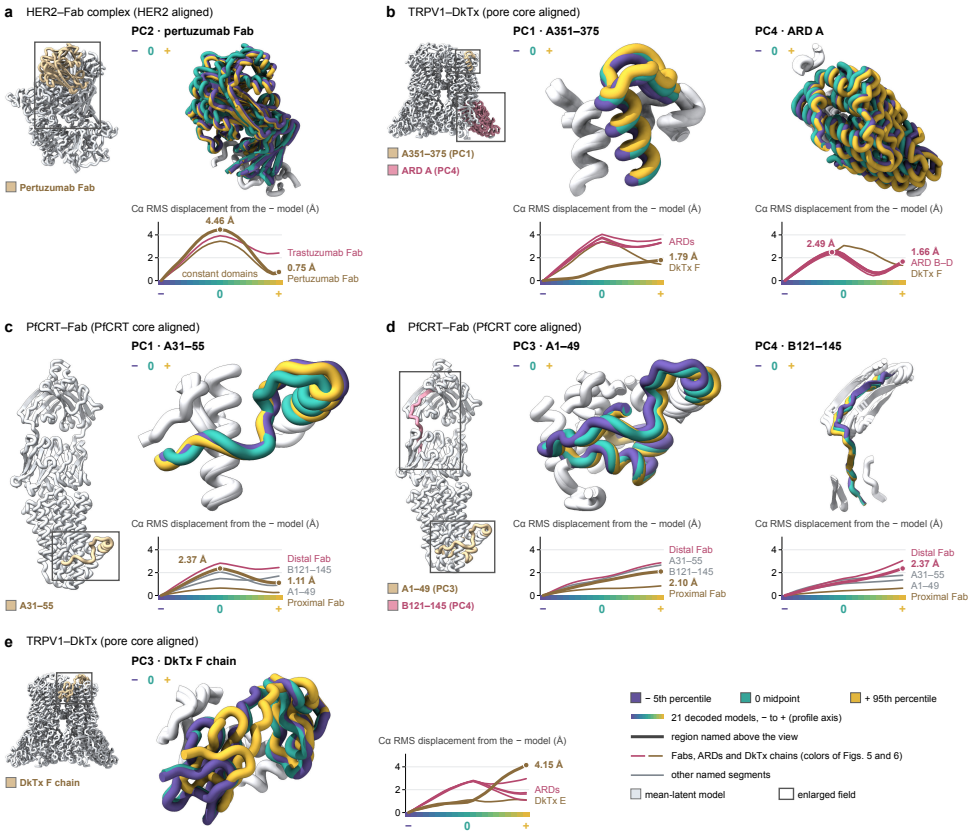
Extended Data Fig. 3 CryoARC pose-update diagnostics on two simulated datasets. a,b, Epoch-aggregated image loss, KL divergence and mean particle-pose rotation relative to the initial orientations, recorded in CryoARC pose-update runs on CAHRA (a) and TGFBIp (b) (Methods). Image loss is the native mean $1-\mathrm{corr}$ objective; ↓ indicates that lower values are better. Dashed lines mark the checkpoint of each structure example (epoch 13 for CAHRA and epoch 17 for TGFBIp). At right, the AF2 initial model (gray) and a pose-update sample (green) are shown after whole-structure Cα alignment of the sample to the AF2 model, in the views of Fig. 2a (CAHRA) and Fig. 3a (TGFBIp) and at one common scale. Each sample is the lower median by Cα radius of gyration among 100 structures saved at that checkpoint.

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.26.000403. This version posted November 26, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0



Extended Data Fig. 4 Quality and regional motion of the density maps. a, Half-map FSC curves of the endpoint density maps. HER2 PC1, PC3 and PC4 and TRPV1 PC2 maps are locally refined maps of focused classes from all filtered particles; MSP-1 maps are fixed-pose reconstructions of Groups 1 and 6, and PfCRT PC2 and TRPV1 PC3 maps are fixed-pose reconstructions of the first and fifth of five PC particle groups. Plot titles name the fixed evaluation ROI. Navy and olive denote the − and + endpoint maps, respectively (Groups 1 and 6 for MSP-1). Solid curves show CryoSPARC noise-substitution-corrected refinement FSC under each reconstruction or local-refinement mask, with their 0.143 resolutions at upper right; dashed curves show uncorrected FSC within the soft evaluation ROI. Gray dotted curves show the refinement FSC of the original PC-group maps that served as classification references; the thin dark dashed line marks FSC = 0.143. **b**, Endpoint regional rigid-field RMS displacement of the fixed models (ticks), density maps (squares) and original PC-group maps (open circles), relative to the MSP-1 coiled coil, HER2 receptor, TRPV1 pore core or PfCRT membrane core (Methods). Gray lines connect each PC-group map to the class map that replaced it, and numbers at right give the model and density-map values. **c**, Directional cosines between model and map displacement fields, with the same symbols and the density-map values at right. Panels b and c include all four TRPV1 ARDs; motion measurements use an 8 Å Gaussian FWHM.

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.26.000403. This version posted November 26, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0



Extended Data Fig. 5 Additional structural variation along latent principal directions. **a**, HER2 latent PC2: the pertuzumab Fab (chains C and D) after alignment on HER2. **b**, TRPV1 latent PC1 (A351–375) and PC4 (ARD A) after pore-core alignment. **c,d**, PfCRT latent PC1 (A31–55; c), PC3 (A1–49; d) and PC4 (Fab B121–145; d) after alignment on the PfCRT core. **e**, TRPV1 latent PC3: the DkTx F chain after pore-core alignment. In each panel, the mean-latent complex at left locates the regions, colored as in the key below it; outlines mark the fields of the enlarged views, which share its orientation. Each enlarged view superposes the −, 0 and + models, with the region colored and nearby residues in white. Below each view, the Cα RMS displacement from the − model is plotted across the 21 conformations decoded along that PC. The thick line shows the region named above the view (in a, the pertuzumab constant domains C108–214 and D120–222), with values at its maximum, where this precedes the + end, and at the + end. Thin lines show other regions of the same complex: the Fabs, ARDs and DkTx chains in the colors of Figs. 5 and 6, and the other named segments in gray. −, 0 and + denote the 5th percentile, the midpoint and the 95th percentile of the PC score, with other latent-PC scores zero, and are colored purple, teal and gold.

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.26.000403. This version posted November 26, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

Dataset	Axis	Map	Particles	Corrected FSC 0.143 (Å)	Regional half-map FSC 0.143 (Å)	ROI local-resolution median (Å) (local FSC 0.5)
MSP-1	Clusters	Group 1	2,252	6.25	7.01	8.26
MSP-1	Clusters	Group 6	2,394	6.74	6.76	8.10
HER2	PC1	−	6,000	4.23	6.83	8.75
HER2	PC1	0	6,000	4.17	6.03	8.53
HER2	PC1	+	6,000	4.13	6.04	8.47
HER2	PC3	−	6,000	4.10	6.60	8.44
HER2	PC3	0	6,000	4.15	7.35	8.60
HER2	PC3	+	6,000	4.16	6.70	8.69
HER2	PC4	−	6,000	4.20	4.11	6.79
HER2	PC4	0	6,000	4.18	4.11	6.88
HER2	PC4	+	6,000	4.15	4.09	6.72
TRPV1	PC2	−	4,376	4.18	6.39	7.60
TRPV1	PC2	0	4,376	4.17	6.57	7.70
TRPV1	PC2	+	4,375	4.28	6.19	7.40
PfCRT	PC2	−	3,381	3.58	4.02	7.52
PfCRT	PC2	0	3,381	3.57	3.89	7.32
PfCRT	PC2	+	3,381	3.67	3.87	7.44

Extended Data Table 1 Quality of fixed-pose particle-group reconstructions. Each map was reconstructed from one group of training particles with their upstream poses fixed. MSP-1 Groups 1 and 6 are the first and last of the six learned latent clusters, which are numbered along latent PC1 (Fig. 4b,d). For HER2, TRPV1 and PfCRT, training particles were ranked by their score on the indicated latent PC and divided into five approximately equal groups; −, 0 and + denote the first, third and fifth groups (0–20%, 40–60% and 80–100% score windows), which are compared with the −, 0 and + models. The MSP-1 and PfCRT maps are shown in Figs. 4d and 6e; for HER2 and TRPV1, the five group reconstructions of each axis provided initial references for the focused classification in Extended Data Table 2. Corrected FSC is the CryoSPARC noise-substitution-corrected half-map FSC 0.143 resolution under the reconstruction mask. Regional half-map FSC is the uncorrected 0.143 crossing over a fixed evaluation ROI: the MSP-1 wing, the HER2 AB Fab (PC1 and PC3) or CD Fab (PC4), TRPV1 ARD A or the PfCRT distal Fab. The ROI local-resolution median summarizes the voxel-wise local-resolution estimates (local FSC 0.5) with valid positive values within the same ROI. Particle counts are the reconstruction inputs; half-set counts and the spatial coverage of local-resolution estimates are given in the source data. Groups within an axis are disjoint, whereas groups from different axes can share particles.

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.26.000403. This version posted November 26, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

Dataset	Axis	Map	Particles	Corrected FSC 0.143 (Å)	Regional half-map FSC 0.143 (Å)	ROI local-resolution median (Å) (local FSC 0.5)
HER2	PC1	−	129,148	3.99	4.07	6.42
HER2	PC1	0	146,727	3.64	3.63	5.39
HER2	PC1	+	141,945	3.90	4.01	6.25
HER2	PC3	−	150,933	3.60	3.58	4.98
HER2	PC3	0	137,908	3.76	3.69	6.09
HER2	PC3	+	134,006	3.75	3.81	6.33
HER2	PC4	−	138,643	3.34	3.31	3.77
HER2	PC4	0	149,220	3.28	3.28	3.58
HER2	PC4	+	154,102	3.21	3.26	3.46
TRPV1	PC2	−	49,918	3.76	4.04	5.92
TRPV1	PC2	0	46,289	3.64	3.99	6.02
TRPV1	PC2	+	38,646	3.90	4.94	6.29

Extended Data Table 2 Quality of locally refined HER2 and TRPV1 maps. Each map is a class selected from five-class focused classification of the full filtered particle collection (708,607 HER2 or 218,786 TRPV1 particles) and locally refined with a mask covering the AB Fab (HER2 PC1 and PC3), the CD Fab (HER2 PC4) or the four ARDs (TRPV1 PC2). −, 0 and + denote the maps paired with the −, 0 and + models in Figs. 5c–e and 6d. Corrected FSC is the CryoSPARC noise-substitution-corrected half-map FSC 0.143 resolution under the local-refinement mask. Regional half-map FSC and ROI local-resolution medians are summarized as in Extended Data Table 1, over the AB Fab, CD Fab or ARD A evaluation ROI. Particle counts are the refinement inputs. Classes within an axis are disjoint, whereas classes from different axes can share particles.

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.26.000403. This version posted November 26, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

		Canonical refinement		
Dataset	Initial preparation	Particles	Epochs	Training epochs
CAHRA	Automated map placement	17,000	1	14
TGFBIp	Automated map placement	17,000	1	15
MSP-1	Map-placed Protenix-V1 prediction guides Protenix-mini initialization	12,000	2	4.80
HER2	Map-fitted assembly with corrected chain overlaps guides Protenix-mini initialization	17,000	1	10
TRPV1	Protenix-V1 assembly guides Protenix-mini initialization	10,000	1	10
PfCRT	Component-wise map-aligned prediction guides Protenix-mini initialization	16,905	2	10

Extended Data Table 3 Initialization and training for each dataset. For MSP-1, HER2, TRPV1 and PfCRT, the listed coordinate target guides Protenix-mini initialization of the structure representation (Methods). Canonical-refinement subsets for HER2 and TRPV1 are nested within their training sets (Extended Data Table 4). Training epochs give the heterogeneous-training endpoint of the displayed results; resolution schedules are described in Methods.

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.11.26.000403. This version posted November 26, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

Dataset	Available particles	Training particles	Used (%)
CAHRA	1,753,470	60,000	3.42
TGFBIp	100,053	60,000	59.97
MSP-1	690,086	12,000	1.74
HER2	708,607	30,000	4.23
TRPV1	218,786	21,879	10.00
PfCRT	16,905	16,905	100.00

Extended Data Table 4 Particle availability and use in heterogeneous training. Available particles are the filtered preparation pools; training particles and the percentage used refer to heterogeneous training. For TGFBIp, 114,000 particles were placed and 100,053 had complete extraction boxes; the latter is the denominator. MSP-1 combines six source lists, from the main class (141,838 particles) and alternative classes 1–5 (112,939, 118,878, 124,774, 116,037 and 75,620 particles); each supplies 2,000 pose-proportionally sampled particles, and these source classes differ from the learned latent clusters. The PfCRT denominator is the final refined stack. TRPV1 contains 218,786 particles after removal of one extreme-intensity particle image from 218,787 inputs; HER2 contains 708,607 after removal of two such images from 708,609 inputs.