CryoNet.Refine: Fully automated cryo-EM model refinement through physics-informed deep learning

Fuyao Huang[1,2,5], Kui Xu[3,5], Kunting Mu[1,5], Xiaozhu Yu[2,5], Ruixue Wan[4], and Qiangfeng Cliff Zhang[1,*]

[1]State Key Laboratory of Membrane Biology, Beijing Frontier Research Center for Biological Structures, Tsinghua-Peking Joint Center for Life Sciences, Ministry of Education Key Laboratory of Bioinformatics, Center for Synthetic and Systems Biology, Center for Life Sciences and Artificial intelligence, School of Life Sciences, Tsinghua University; Beijing, 100084, China.

[2]State Key Laboratory of Membrane Biology, Membrane Structure and Artificial Intelligence Biology Branch, Hangzhou, 311308, China.

[3]State Key Laboratory of Membrane Biology, Beijing Tsinghua Institute for Frontier Interdisciplinary Innovation, Beijing, 102202, China.

[4]State Key Laboratory of Gene Expression, Westlake Laboratory of Life Sciences and Biomedicine, Zhejiang Key Laboratory of Structural Biology, School of Life Sciences, Westlake University, 18 Shilongshan Road, Hangzhou 310024, Zhejiang Province, China.

[5]These authors contributed equally to this work.

*Correspondence: qczhang@tsinghua.edu.cn (Q.C.Z.)

ABSTRACT

Translating cryo-electron microscopy (cryo-EM) density maps into accurate atomic models requires refinement that maximizes map–model agreement and stereochemical validity, yet current workflows often remain laborious and expert-dependent. Here we introduce CryoNet.Refine, a physics-informed deep learning method that couples an AlphaFold 3-derived coordinate update module with a fully differentiable composite loss for fully automated atomic model refinement. We assess performance using a composite quality score (CS-score) integrating density-fit and geometry metrics. Across benchmarks spanning 225 AlphaFold 3–predicted starting models and 675 PDB-deposited structures of two deposition eras, CryoNet.Refine consistently outperforms *Phenix.real_space_refine*. On unrefined predicted models, CryoNet.Refine raises the mean CS-score from 0.64 to 0.91—versus 0.72 for *Phenix.real_space_refine*. On PDB-deposited structures—including post-2018 entries already optimized by *Phenix.real_space_refine* and expert manual correction—CryoNet.Refine achieves higher CS-scores in 90.9% of cases. Strikingly, it lowers the mean MolProbity score from 1.84 to 1.36, whereas *Phenix.real_space_refine* increases it to 2.02. Half-map cross-validation confirms that these gains reflect genuine structural signal rather than overfitting. CryoNet.Refine is available as open-source software and a web server to support reproducible, high-throughput cryo-EM refinement.

INTRODUCTION

Cryo-electron microscopy (cryo-EM) now routinely delivers three-dimensional density maps of macromolecular structures at near-atomic resolution[1-3]. However, translating a three-dimensional reconstructed density map into a stereochemically accurate atomic model remains a major bottleneck[4-9]. A critical step in this pipeline is model refinement. Fundamentally, refinement aims to adjust the atomic coordinates of an initially built model to maximize its fit to the experimental map while strictly adhering to stereochemical and geometric constraints[10-12].

Existing refinement tools combine map-model fitting with stereochemical restraints and empirical force fields. For instance, real-space refinement engines (*e.g.*, *Phenix.real_space_refine*[10]) optimize model coordinates under geometric restraints, and are often followed by manual inspection and correction in interactive tools (*e.g.*, Coot[13], ISOLDE[14]). More comprehensive energy-function approaches (*e.g.*, Rosetta-based refinement[15]) can further improve models but often require case-specific weighting and careful tuning. While these methods are highly effective, performance can depend on resolution, starting model quality, and the selection of refinement weights; fully automated refinement remains challenging.

While artificial intelligence has profoundly accelerated upstream steps of cryo-EM structure determination—such as automated particle picking[16,17], map reconstruction[18] and enhancement[19,20], and *de novo* model building[4,5,21-24]—its potential for directly refining atomic coordinates has remained largely untapped. This gap stems primarily from the challenges of purely data-driven neural networks to enforce complex stereochemical rules. To overcome this limitation, computational strategies must move beyond pure pattern recognition and integrate physical laws into the learning process. Recent advances in physics-informed deep learning—popularized by physics-informed neural networks (PINNs)[25-27]—however, demonstrate exactly this: by embedding physical equations directly into neural network loss functions, models are forced to generate outputs that strictly adhere to real-world physical constraints. In structural biology, modern structure prediction systems have already begun to incorporate such physical and geometric reasoning to produce stereochemically consistent models (*e.g.*, AlphaFold2[28], RoseTTAFold[29]). More recently, these architectures have evolved into highly capable structure generators, leveraging learned structural priors to construct accurate 3D coordinates[30-34], establishing the necessary foundation for automated, AI-driven refinement.

Here we introduce CryoNet.Refine, a physics-informed deep learning approach for cryo-EM atomic model refinement, coupling (i) a coordinate update module—repurposed from the AlphaFold 3[35] diffusion module architecture—to predict atomic coordinate updates based on diffusion-trained structural priors, with (ii) a fully differentiable density generator that synthesizes density maps from

atomic coordinates, enabling direct optimization of map–model agreement, and (iii) a composite differentiable loss that enforces both experimental density fit as well as stereochemistry penalties that constrain bonds, angles, torsions, and steric clashes. Through extensive benchmarks on a dataset of AlphaFold 3-predicted starting models (225 structures), as well as datasets totaling 675 PDB-deposited structures (pre-2018: 325; post-2018: 350), we demonstrate that CryoNet.Refine consistently outperforms traditional methods in both map-model correlation and geometric quality. By successfully merging experimental data restraints with neural network-based optimization, CryoNet.Refine establishes a highly effective system for cryo-EM structure model refinement.

RESULTS

CryoNet.Refine architecture and loss function

CryoNet.Refine performs cryo-EM atomic model refinement using a physics-informed neural network that predicts coordinate updates, evaluated by a composite differentiable objective that measures both map–model agreement and stereochemical validity (Fig. 1A). The framework takes as input an experimental cryo-EM density map, an initial atomic model (derived from *de novo* model building or structure prediction), and the corresponding amino acid sequence. A coordinate encoder extracts pairwise geometric features from the input model, while a sequence embedder integrates residue identities with their corresponding local reference conformer geometries into an initial sequence representation. These representations are then fed into a coordinate update module that predicts refined coordinates. Rather than learning these complex spatial and chemical representations from scratch, CryoNet.Refine leverages the deep structural priors embedded in Boltz-2[36] (an open-source reimplementation of the AlphaFold 3 architecture[35]) to initialize these core components. Crucially, we repurpose this architecture from a conventional structure prediction pipeline into a refinement engine that predicts structural updates to be evaluated against the density map, generating gradients that directly optimize the network's weights.

To enable this gradient-based evaluation—specifically, to translate discrete atomic coordinates into a format directly comparable to the experimental cryo-EM volume—we introduce a fully differentiable density generator. Following the *molmap* algorithm in ChimeraX[37,38], the generator constructs a synthetic density map by placing Gaussian kernels at each atom position, with the kernel width scaled according to the map resolution. The density contributions from all atoms are accumulated via summation at each grid point, yielding a smooth, continuous 3D volume. We implement this module using the *autograd* function[39], which explicitly defines the gradient flow from each output voxel back to the input atomic coordinates. This ensures that every step—from Gaussian kernel placement to grid sampling and intensity accumulation—is fully differentiable. Finally, the

synthetic map is then compared with the experimental map via a density loss:

$$\mathcal{L}_{den} = 1 - CC$$

where CC denotes the cosine similarity computed over the overlapping region.

To ensure stereochemical realism, we design a set of differentiable geometry penalties regarding: bond angles and lengths RMSD ($\mathcal{L}_{\text{angle}}$, $\mathcal{L}_{\text{bond}}$), steric clash penalties ($\mathcal{L}_{\text{non\_bond}}$, $\mathcal{L}_{\text{clash}}$), C_β deviations ($\mathcal{L}_{C\beta}$), Ramachandran backbone outliers ($\mathcal{L}_{\text{rama}}$), Z-scores ($\mathcal{L}_{\text{ramaz}}$), and side-chain rotamer outliers ($\mathcal{L}_{\text{rot}}$). Together, these terms form a composite objective:

$$\mathcal{L} = \gamma_{\text{den}}\mathcal{L}_{\text{den}} + \gamma_{\text{angle}}\mathcal{L}_{\text{angle}} + \gamma_{\text{bond}}\mathcal{L}_{\text{bond}} + \gamma_{\text{non\_bonded}}\mathcal{L}_{\text{non\_bonded}} + \gamma_{\text{clash}}\mathcal{L}_{\text{clash}} + \gamma_{C_\beta}\mathcal{L}_{C_\beta}$$

$$+\gamma_{\text{rama}}\mathcal{L}_{\text{rama}} + \gamma_{\text{ramaz}}\mathcal{L}_{\text{ramaz}} + \gamma_{\text{rot}}\mathcal{L}_{\text{rot}}$$

By minimizing this composite loss, gradients are backpropagated through the density generator and into the coordinate update module, teaching it to generate coordinate updates that simultaneously optimize map–model agreement and stereochemistry.

Training and inference of CryoNet.Refine

As mentioned above, we initialized CryoNet.Refine with weights from the Boltz-2 model. We then trained it on a large corpus of 5,061 pairs of atomic models and density maps (*i.e.*, model-map pairs), each consisting of an experimental map from EMDB[40] and a corresponding atomic model from the AlphaFold Database[41] (AFDB). Pairs passing a $CC_{mask} \geq 0.2$ threshold were retained, ensuring minimum map-model correspondence while preserving diversity in resolution, molecular size, and conformation. The use of AlphaFold Database models—rather than manually refined PDB deposits[42]—was deliberate: these models contain various types of structural inaccuracies (such as locally misaligned backbones or unoptimized side-chain conformations) typical of unrefined computational predictions, providing training examples that resemble the inputs encountered at inference time.

During training, the sequence embedder, coordinate encoder, and pairformer module were kept frozen, and only the coordinate update module was updated. Each training sample was cropped to a maximum of 1,000 residues and processed through 10 recycle iterations (Fig. 1A, green dashed line). At each iteration, the coordinate update module generated a structural update, which was immediately evaluated by the composite objective to compute the loss and backpropagate the gradient signal. Training was performed for 100 epochs with a constant learning rate of 1.8×10^{-4} and all loss weights fixed throughout training.

At inference, CryoNet.Refine performs test-time, in-depth optimization on an input model-map pair, mirroring the training loop but deeply adapting the module through up to 300 recycle iterations—or until convergence. This extension from training (10 recycle iterations) to inference (up

to 300 recycle iterations) reflects the design: because each recycle step is conditioned only on the current coordinates, the training of the coordinate update module generalizes to longer trajectories to maximize adaptation to the input model. Finally, we observed that map-model correlation coefficients typically plateau within 50–100 recycle iterations (Fig. S1C), at which point early stopping terminates refinement.

Evaluation framework

We assessed refinement quality along two complementary axes: map-model correlation and stereochemical validity. For map-model agreement, we adopted the Q-score[43] as the primary density-fit metric to evaluate how well the density around each atom matches an expected Gaussian shape, with higher scores indicating better atomic resolvability. We note that, among the 20,568 deposited EMDB models (resolution $\leqslant$ 10 Å), 88.9% exceed the resolution-dependent Q_low_95% threshold (Fig. S2B), establishing a statistical reference for map–model consistency. We also computed six map–model correlation coefficients (CC_{mask}, CC_{box}, CC_{volume}, CC_{peaks}, CC_{mc}, and CC_{sc}) to quantify the overall and local volume overlap between the experimental and model-simulated maps.

For stereochemical validity, we evaluated twelve standard geometry metrics—covering Ramachandran statistics, rotamer preferences, clash score, C_β deviations, and bond/angle/dihedral RMSDs—as well as the MolProbity score[44] and the EMRinger score[45]. The MolProbity score is a composite metric that integrates clash, Ramachandran, and rotamer statistics into a single geometry summary, where a lower score reflects better overall physical realism. Meanwhile, the EMRinger score tests whether side-chain dihedral angles are corroborated by the experimental density. A higher EMRinger score (typically above 1.0) indicates that the modeled side-chains are in valid rotamer states and are accurate relative to the experimental data, providing a residue-level bridge between stereochemical and map-fit assessment.

To integrate both axes into a single measure, we introduce the Cryo-EM-Structure Score (CS-score), defined as the fraction of 16 indicators that satisfy their corresponding acceptance criteria. These comprise seven density-based metrics (the Q-score and the six correlation coefficients[46]) and nine geometry-based metrics (Ramachandran statistics, rotamer preferences, C_β deviations, bond/angle RMSDs, and clash score). To establish a baseline for interpretation, we calculated the CS-score for 20,568 deposited map–model pairs from the EMDB. The mean CS-score across this large-scale reference set is 0.769 (resolution $\leqslant$ 10 Å), and notably, this value remains largely stable across resolutions (Fig. S2A), indicating that this composite metric captures intrinsic model quality independent of map resolvability. The CS-score thus provides a single, interpretable measure of whether an atomic model simultaneously fits the experimental density and maintains expected stereochemistry.

CryoNet.Refine refines AlphaFold 3-predicted starting models

To benchmark CryoNet.Refine, we assembled a dataset of 225 cryo-EM targets (with structure models) spanning resolutions from 1.8 to 8.0 Å, encompassing both protein and protein-DNA/RNA structures (Fig. S2C, D). Initial models were generated using AlphaFold 3 and aligned to the deposited PDB models using the *matchmaker* tool in UCSF ChimeraX, and subsequently refined with either *Phenix.real_space_refine* (default settings) or CryoNet.Refine. Overall, both tools effectively improved the unrefined starting models: CryoNet.Refine raised the mean CS-score from 0.64 (AlphaFold 3 starting models) to 0.91, representing a 42.2% improvement, compared with only 12.5% for *Phenix.real_space_refine* (mean CS-score of 0.72, Fig. 2A). Specifically, CryoNet.Refine outperformed *Phenix.real_space_refine* in the vast majority of cases (215 out of 225, or 95.6%).

CryoNet.Refine substantially improved map-model agreement across all evaluated metrics, including both the Q-score and all six map–model correlation metrics. Specifically, it raised the Q-score pass rate (the fraction of refined models exceeding the Q_low_95% threshold) from 27.5% (initial models) to 92.4%, substantially higher than *Phenix.real_space_refine* (64.9%) (Fig. 2B), with the mean Q-score rising from 0.21 to 0.40 (vs. 0.33 for *Phenix.real_space_refine*) (Fig. 2C). CryoNet.Refine also improved all six map–model correlation metrics, achieving higher means and narrower interquartile ranges (Fig. 2C); for example, it increased the mean CC_{mask} from 0.35 to 0.61 (vs. 0.50 for *Phenix.real_space_refine*) (Fig. 2C).

Critically, these gains in density fit were accompanied by improvements in stereochemical quality across nine of twelve geometry metrics (Fig. 2D, S2E). For example, the mean MolProbity score decreased from 1.71 to 1.42 (vs. 1.78 for *Phenix.real_space_refine*). The simultaneous improvement in both density fit and geometry indicates that CryoNet.Refine effectively balances these objectives, as reflected in the significantly higher overall CS-scores.

The broad applicability of CryoNet.Refine is illustrated by representative cases spanning diverse molecular types and resolution regimes (Fig. 3). For protein structure, represented by the human CNT3 transporter[47] (EMD-0775, 3.60 Å, protein trimer, Fig. 3A), CryoNet.Refine achieved a CS-score of 1.00 (vs. 0.81 for *Phenix.real_space_refine*) (Fig. 3A). For nucleic-acid-containing structures, CryoNet.Refine attained CS-scores of 1.0 on both the Precleavage Cre tetrameric[48] (EMD-24470, 3.23 Å, protein–DNA, Fig. 3B) and the HIV-1 RT IC core-EFV complex[49] (EMD-22900, 2.90 Å, protein–RNA, Fig. 3C), compared with 0.81 and 0.94 for *Phenix.real_space_refine*, respectively. On cryo-ET data—exemplified by the mature RSV CA pentamer[50] (EMD-12488, 5.80 Å, cryo-ET, Fig. 3D)—CryoNet.Refine again outperformed *Phenix.real_space_refine* (CS-score: 1.00 vs. 0.56). It is worth noting that CryoNet.Refine achieves superior agreement with the density map at both the main-chain and side-chain levels (Fig. 3A-C insets). Additional examples are provided in Fig. S3, consistently demonstrating CryoNet.Refine's superiority in both map-model correlation and stereochemical quality. Together, these results establish CryoNet.Refine as a robust and generalizable refinement method that consistently outperforms current approaches across the full spectrum of cryo-EM and cryo-ET applications.

CryoNet.Refine improves expert-refined PDB-deposited models

To complement the preceding benchmark on unrefined models, we next evaluated CryoNet.Refine on PDB-deposited structures that had already undergone expert refinement and manual adjustment. Re-refining such models with a fully automated pipeline constitutes a considerably harder test, as these structures were carefully optimized by structural biology experts, leaving limited room for further improvement. Following the validation strategy proposed for *Phenix.real_space_refine*, we designed two complementary benchmarks spanning different deposition eras.

Benchmarking on pre-2018 PDB deposits.

We assembled a set of 325 experimental cryo-EM structures (resolutions 1.8–6.0 Å) deposited in the PDB before 2018, following the curation criteria of the original *Phenix.real_space_refine* study (Fig. S4A). Because these structures predate the widespread adoption of *Phenix.real_space_refine*, many were refined with earlier-generation tools, providing a baseline against which both methods can be fairly compared. Overall, both tools improved the deposited models: CryoNet.Refine raised the mean CS-score from 0.74 (original PDB deposits) to 0.90, whereas *Phenix.real_space_refine* achieved a mean CS-score of 0.80 (Fig. 4A). Specifically, CryoNet.Refine outperformed *Phenix.real_space_refine* in the majority of cases (252 out of 325, or 77.5%).

CryoNet.Refine maintained or improved map-model agreement across the evaluated density metrics. Specifically, it achieved a mean Q-score of 0.42, performing on par with *Phenix.real_space_refine* (0.42) and improving upon the original PDB deposits (0.40) (Fig. 4B). Similarly, map-model correlation metrics confirmed marginal but positive gains; for example, the mean CC_{mask} rose to 0.68 (vs. 0.67 for *Phenix.real_space_refine* and 0.65 for the original deposits).

Importantly, this competitive density fit was accompanied by substantial improvements in stereochemical quality (Fig. 4D). The mean MolProbity score decreased to 1.61, substantially outperforming both *Phenix.real_space_refine* (2.19) and the original expert-curated PDB deposits (2.31). Furthermore, CryoNet.Refine achieved a clash score of 9.19, dropping by over 40% compared with the original deposits (20.69) and markedly outperforming *Phenix.real_space_refine* (15.62). Full metric comparisons are provided in Fig. 4B-D and Fig. S4B.

Benchmarking on post-2018 PDB deposits.

We further assembled a second set of 350 structures (resolution 1.8–6.0 Å) deposited between 2018 and 2026 (Fig. S4C). This benchmark is particularly demanding, as many of these structures were likely refined with *Phenix.real_space_refine* and subsequently polished through expert manual intervention. Overall, CryoNet.Refine raised the mean CS-score from 0.84 (original expert-curated PDB deposits) to 0.93, representing a 10.7% improvement, whereas *Phenix.real_space_refine* reduced it to 0.80, corresponding to a 4.8% decline (Fig. 4E). This notable decline in

Phenix.real_space_refine performance is expected; we ran the tool in a fully automated manner, which inherently performs worse than the case-by-case parameter adjustment and follow-up manual model tuning typically employed during deposition. Specifically, CryoNet.Refine achieved higher CS-scores on 90.9% (318 out of 350) of the evaluated structures.

CryoNet.Refine maintained high-fidelity map-model agreement across the evaluated density-fit metrics. Specifically, it achieved a mean Q-score of 0.48, matching the performance of *Phenix.real_space_refine* (0.48) and surpassing the original PDB deposits (0.46; Fig. 4F).

As in the cases in the pre-2018 benchmark, CryoNet.Refine yielded pronounced improvements in stereochemical quality while maintaining this density fit. For example, the mean MolProbity score was reduced to 1.36, substantially outperforming both *Phenix.real_space_refine* (2.02) and the original PDB deposits (1.84) (Fig. 4H). Furthermore, CryoNet.Refine achieved a clash score of 6.41, outperforming the original deposits (10.24). Conversely, *Phenix.real_space_refine* exhibited a slight degradation in clash score (11.30). Full metric comparisons are provided in Fig. 4F-H and Fig. S4D.

Together, these two benchmarks—spanning different deposition eras and refinement histories—demonstrate that CryoNet.Refine consistently improves expert-refined models in a fully automated fashion.

CryoNet.Refine shows no overfitting in half-map cross-validation

A key concern with any refinement method—particularly one employing a learned potential—is whether metric improvements reflect genuine gains or overfitting to reconstruction noise. To address this, we applied the standard cryo-EM control of half-map cross-validation to 50 structures with independently reconstructed half-maps available in the EMDB (Fig. S4E). Specifically, we refined each model against the first half-map ("work map") and then evaluated against the retained second half-map ("validation map")—reasoning that a method prone to overfitting will show a clear gap between work-map and validation-map agreement, whereas one that captures genuine structural signal will perform comparably on both.

We calculated the map-model correlation for each refined model against the work map and the unseen validation map. We found that the Pearson correlation coefficient between these two values approached 1.0 for both CryoNet.Refine and *Phenix.real_space_refine* across all 50 cases (Fig. S4F), indicating that the density-fit improvements transfer almost entirely to unseen data and that CryoNet.Refine's learned energy terms do not memorize map-specific noise. Beyond density fit, we note that CryoNet.Refine's stereochemical advantages were fully retained, as expected. Models refined exclusively against the work map and subsequently evaluated on the validation map maintained substantially better geometry than those from *Phenix.real_space_refine* and the original deposits (Fig. S4G, H). In sum, CryoNet.Refine's improvements are genuine, achieved without

overfitting to noise, and persist when the model is evaluated against data it has never seen.

CryoNet.Refine benefits from all loss components in ablation studies

To dissect the contribution of each component in the CryoNet.Refine framework, we performed ablation experiments on a subset of 110 protein structures from the AlphaFold 3-predicted starting model benchmark dataset (Fig. S2D), removing individual loss terms one at a time.

We observed that removing the density loss ($\gamma_{\text{den}} = 0$)—the primary signal from experimental data—causes a dramatic collapse in map-model agreement, with mean reductions of 45–67% across correlation metrics (*e.g.*, Q-score drops from 0.43 to 0.14, CC_{mask} from 0.64 to 0.25) (Fig. 5A). Conversely, when assessing stereochemical quality, we observed an artificial "improvement" with the mean MolProbity score dropping from 1.36 to an idealized 0.91, and the mean clash score dropping from 8.47 to 1.81. This stark divergence occurs because, stripped of experimental restraints, the network over-optimizes for geometric priors at the expense of density-fit. Together, these paired observations confirm that while geometry losses ensure physical realism, the density loss is indispensable for guiding refinement toward map-model agreement.

We also ablated the components of our geometry loss suite—encompassing clash, bond and angle, Ramachandran, rotamer, and C_β deviation penalties—to evaluate their role in preventing structural overfitting (Fig. 5B and Fig. S5A, B). Across all geometric ablations, removing stereochemical restraints generally preserved or even artificially inflated map-model agreement, but at a severe cost to physical realism. For example, isolated ablation of both the clash and the non-bonded loss resulted in severe atomic overlaps throughout the structure, driving the mean clash score from a baseline of 8.47 up to an unrealistic 121.60. The most striking degradation occurred upon removing the foundational bond and angle restraints, which triggered a catastrophic collapse in structural integrity (*e.g.*, mean angle RMSD surging to 28.33°, bond RMSD to 0.80 Å, and clash score to 332.21, Fig. S5A, B). Ablating other specific terms (such as Ramachandran or rotamer losses) yielded corresponding targeted degradations in backbone and side-chain validity.

Collectively, these geometric ablations demonstrate a clear algorithmic trade-off: without explicit stereochemical enforcement, the network exploits unconstrained degrees of freedom to overfit the experimental density map. This underscores that density optimization alone is insufficient, and a comprehensive suite of geometric priors is essential to maintain physically realistic atomic models.

DISCUSSION

CryoNet.Refine addresses a remaining bottleneck in cryo-EM structure determination: automated refinement that improves map–model agreement while preserving stereochemical correctness without extensive manual tuning. By casting atomic model refinement as a physics-informed,

differentiable, and unrolled optimization process, CryoNet.Refine refines models under the guidance of a composite loss encoding experimental density agreement and stereochemical validity. Across 225 AlphaFold 3-predicted starting models and 675 expert-refined PDB deposits, CryoNet.Refine consistently outperforms *Phenix.real_space_refine* in both map-model correlation and geometric quality. Ultimately, by replacing iterative manual correction and optimization cycles with this deep learning framework, CryoNet.Refine effectively removes this bottleneck. We note that CryoBoltz[51] was recently developed for cryo-EM density-guided protein structure prediction. While highly effective for its intended goal, it is not specifically tailored for fine-grained model refinement. Nevertheless, because it does possess the capability to fit models with density maps, we performed a direct comparison to CryoNet.Refine. Notably, CryoBoltz successfully processed only 24 cases in our test set, and within this small subset, CryoNet.Refine achieved higher performance on both density fit and geometry metrics (Fig. S6A, B). We do not wish to overstate this performance difference, as it is expected and directly attributable to the distinct design objectives of the two frameworks. Ultimately, these tools are complementary: CryoNet.Refine serves as an ideal downstream optimizer for CryoBoltz predictions.

The ability of CryoNet.Refine to improve upon post-2018 PDB deposits—structures already rigorously optimized in Phenix and polished by human experts—is highly significant. It demonstrates that the network's learned structural priors capture complex, high-dimensional stereochemical principles that elude both classical heuristics and manual human intuition. However, outperforming expert-curated models inherently raises a critical danger: the risk that the neural network might be "gaming" the metrics by overfitting to experimental noise or introducing chemically unrealistic micro-conformations to artificially satisfy the loss function. To rigorously check against this, we relied on half-map cross-validation. This analysis confirms that the observed gains reflect genuine structural signal rather than noise fitting, as evidenced by the lack of a systematic gap between work-map and validation-map agreement (Fig. S4F).

Several conceptual insights emerge from this work. First, we demonstrate how pretrained generative models can be repurposed for constrained optimization in structural biology. Rather than using these networks for de novo prediction, CryoNet.Refine leverages their rich, learned structural priors to iteratively guide models into agreement with experimental restraints, effectively acting as a highly sophisticated regularizer. Beyond this architectural insight, the differentiable density generator and the suite of differentiable geometry penalties introduced here are not limited to refinement: they can serve as drop-in modules for any end-to-end differentiable pipeline in cryo-EM or protein structure prediction, as demonstrated by the incorporation of these geometry losses into the fine-tuning of Boltz-2-like models, which substantially improved their stereochemical validity (Fig. S5C, D).

Despite its strong performance, CryoNet.Refine has limitations that define its current scope. First, regarding its pulling range, CryoNet.Refine is designed to polish and optimize rather than perform structure prediction or modeling from scratch. Consequently, applying it to initial models with massive structural deviations from the target density will exceed its pulling range. Second, regarding molecular composition, while it accommodates standard amino acids and nucleic acids, it does not yet support small-molecule ligands or post-translational modifications (*e.g.*, phosphorylation, glycosylation). Extending the coordinate encoder and the loss function to handle these diverse chemical entities remains an important direction for future work. Finally, regarding conformational and structural completeness, the current framework optimizes a single set of coordinates against a single consensus map. As a result, it does not explicitly model regions of conformational flexibility or partial occupancy, which can lead to suboptimal refinement in locally disordered regions where the density reflects a mixture of states. Furthermore, refinement performance may decrease for structures with many missing or unresolved residues, as incomplete backbone connectivity limits the effectiveness of density-guided geometric restraints.

Taken together, CryoNet.Refine demonstrates that tightly coupling learned structural priors with experimental restraints through differentiable optimization can yield refinement quality surpassing expert-guided pipelines. Meanwhile, its modular components—the differentiable density generator and geometry loss suite—are readily transferable to other tasks in AI-driven structural biology. Looking ahead, because refinement proceeds in an automated way with no case-specific parameter tuning, the framework is well positioned to scale to the hundreds or thousands of independently reconstructed states generated by time-resolved or conformational-sorting experiments—where manual intervention per structure is impractical. More broadly, CryoNet.Refine establishes a foundation for cryo-EM to mature into an enterprise for high-throughput, high-accuracy structural determination—from single-particle analysis to the time-resolved and conformational frontiers.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Qiangfeng Cliff Zhang (qczhang@tsinghua.edu.cn).

Data and code availability

All data used in this study are publicly available. The density maps are from EMDB (https://ftp.ebi.ac.uk/pub/databases/emdb/), the structural models are from PDB (https://www.wwpdb.org/ftp/pdb-ftp-sites) and AlphaFold structure database, (https://alphafold.ebi.ac.uk/). All the CryoNet.Refine refined models are available from figshare at https://doi.org/10.6084/m9.figshare.31883266. The source code is available at https://github.com/kuixu/cryonet.refine. The CryoNet.Refine web server is freely available for academic use at https://cryonet.ai/refine.

ACKNOWLEDGMENTS

Q.C.Z. was supported by National Natural Science Foundation of China (32125007, 32230018), Science & Technology Fundamental Resources Investigation Program (Grant No. 2025FY100300), the Xplorer Prize (2024), and Tsinghua University Dushi Program and the TsienTang Life Science Development Fund at Tsinghua University. K.X. was supported by the Young Scientists Fund of the National Natural Science Foundation of China (32400548).

AUTHOR CONTRIBUTIONS

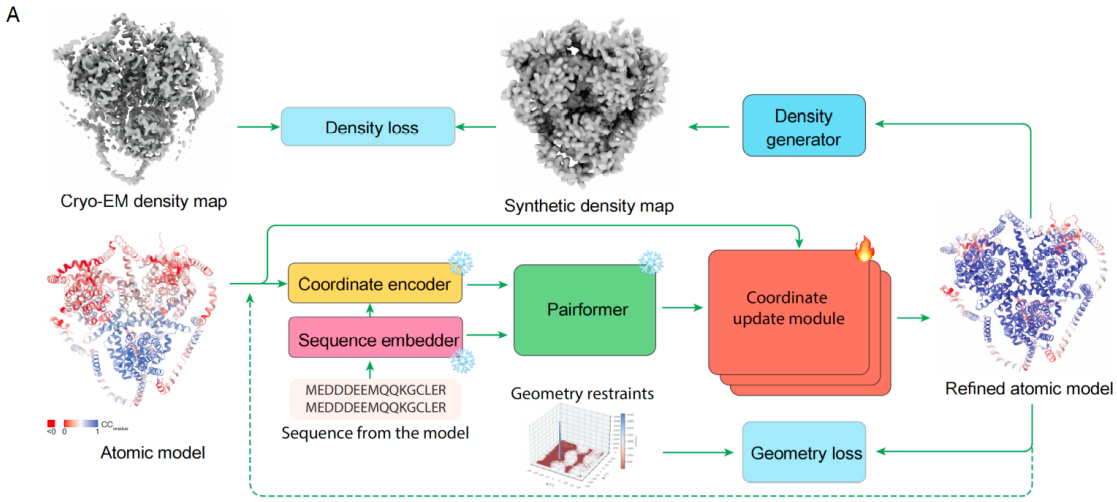
Q.C.Z. supervised and K.X. conceived the project. K.X., F.H., and X.Y. designed and implemented the CryoNet.Refine model. F.H., K.X., K.M., and R.W. validated the CryoNet.Refine model and analyzed the results. Q.C.Z., K.X., F.H., and K.M. wrote the manuscript, with inputs from all the authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

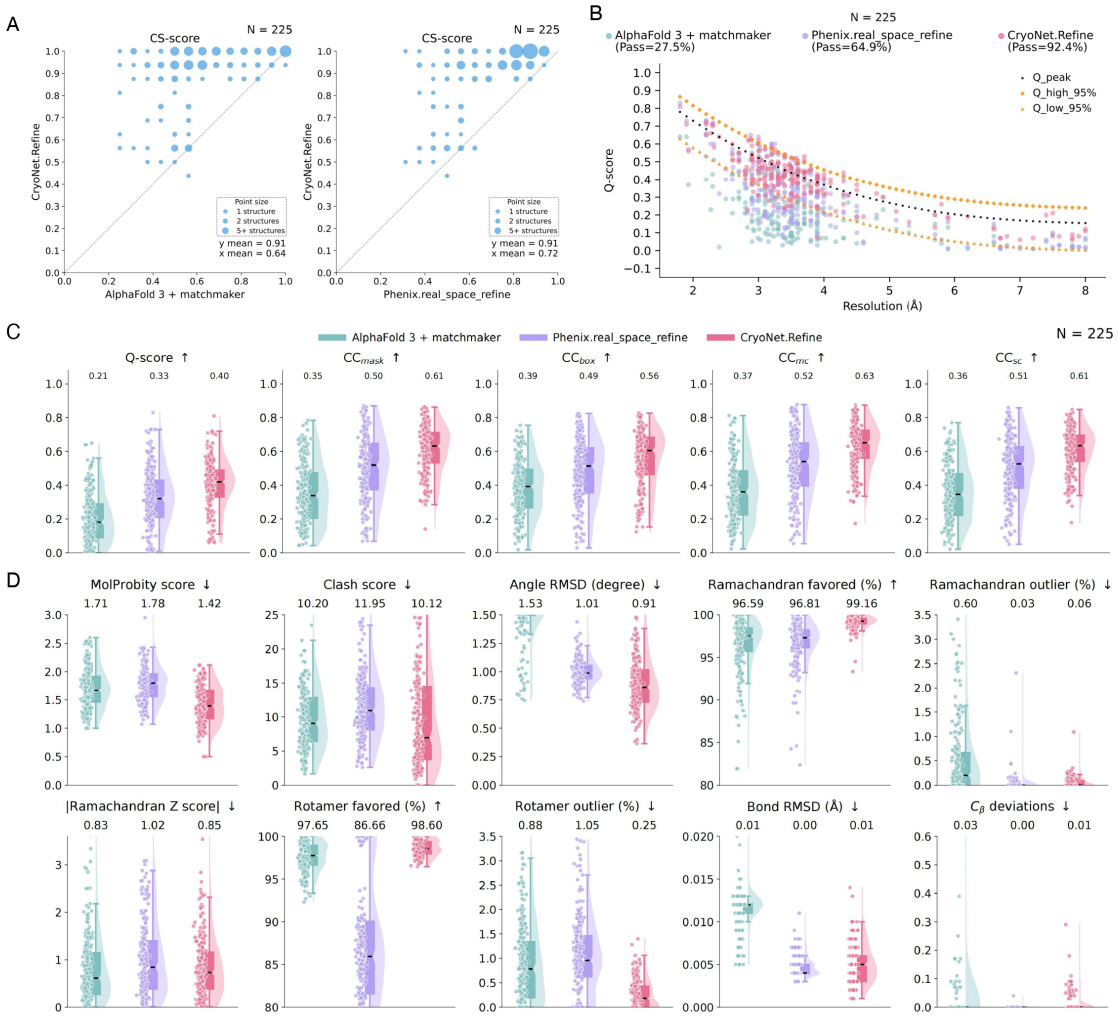
FIGURE LEGENDS

Figure 1. Architecture of CryoNet.Refine.



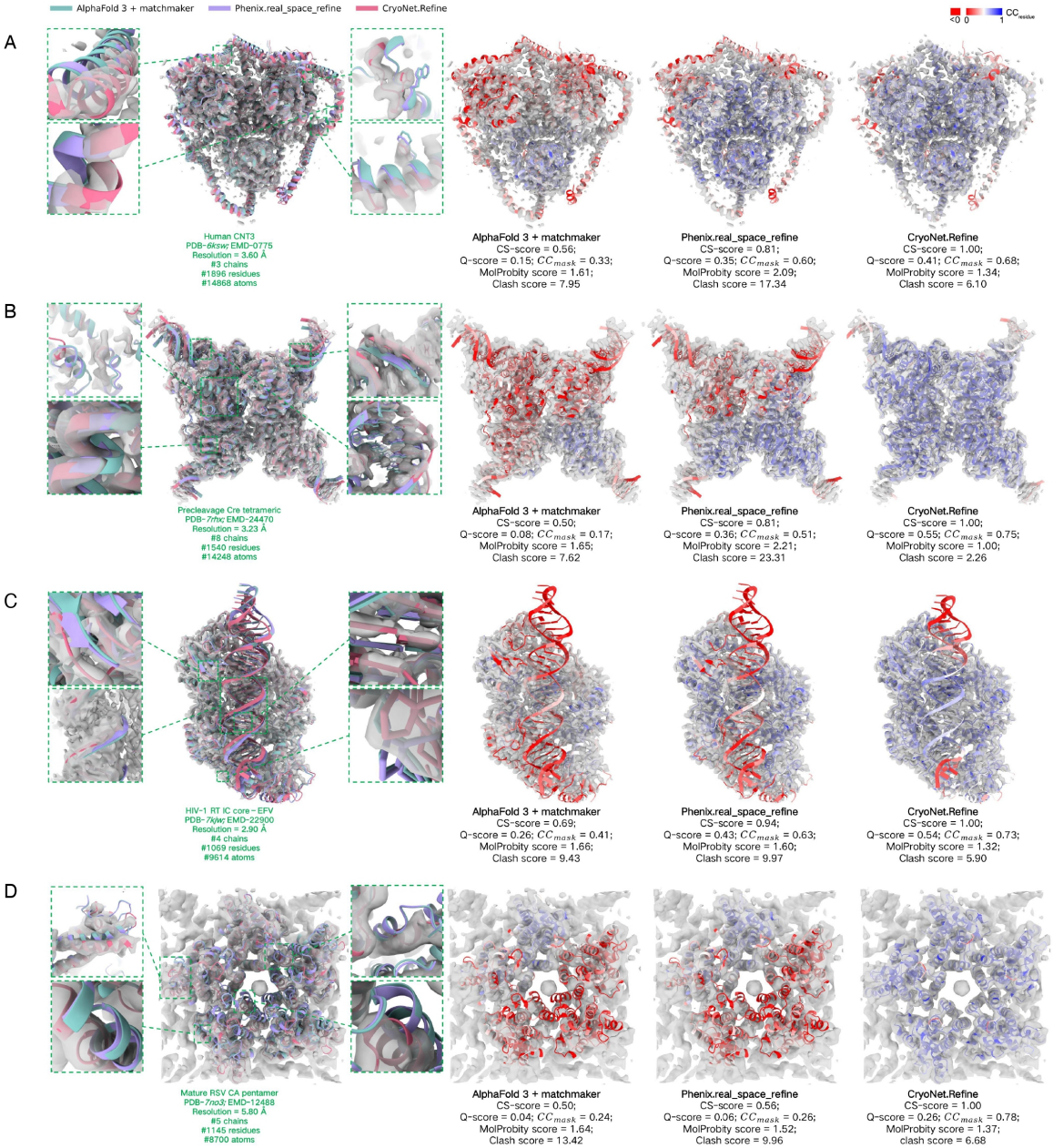
(**A**) Overview of the physics-informed framework integrating geometry priors with experimental density restraints. The pipeline accepts an initial atomic model and a target cryo-EM density map as input. Coordinate and sequence features are extracted via a coordinate encoder and a sequence embedder, respectively. These features are subsequently processed by a pairformer module to capture inter-residue dependencies. To preserve pre-trained knowledge, the weights of these three modules remain frozen (indicated by snowflake icons in the figure). A coordinate update module (trainable weights denoted by the flame icon) deterministically predicts refined atomic coordinates. To enable gradient-based optimization directly against experimental data, a differentiable density generator converts the refined atomic model into a synthetic density map for density loss computation. The framework undergoes instance-specific optimization guided by a composite objective function, which comprises a density loss and a suite of geometry losses. This process is performed iteratively via a recycling mechanism (green dashed line) until convergence.

Figure 2. Benchmarking on AlphaFold 3-predicted starting models.



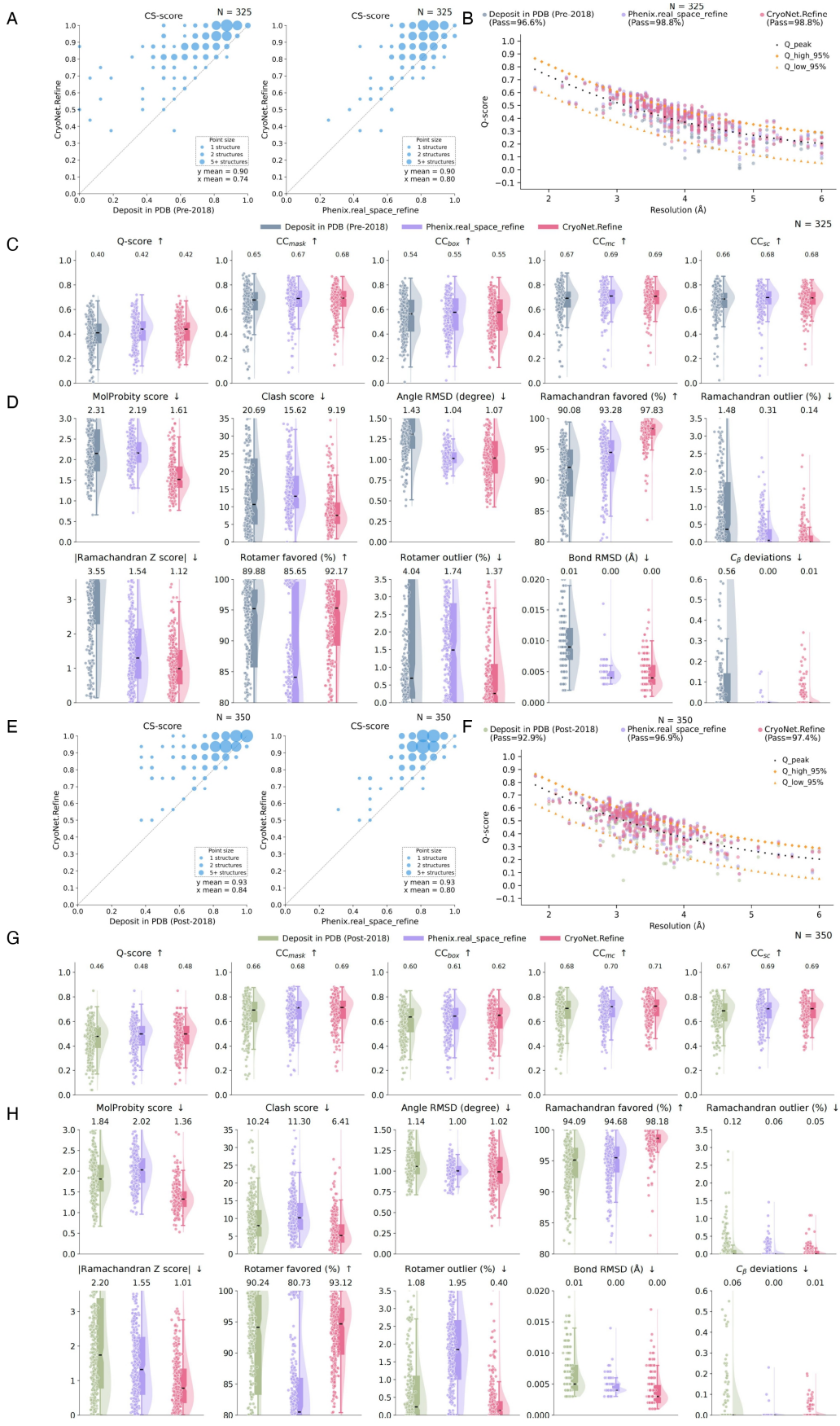
Evaluation was performed on 225 AlphaFold 3-predicted cryo-EM targets (1.8–8.0 Å), including protein and protein-nucleic acid structures. **(A)** CS-score comparisons of CryoNet.Refine against AlphaFold 3 starting models (rigidly aligned to corresponding PDB deposits via *matchmaker*; left, mean CS-score: 0.91 vs. 0.64) and those optimized by *Phenix.real_space_refine* (right, mean CS-score: 0.91 vs. 0.72). Point sizes indicate structure frequency. (**B**) Q-score versus resolution. Dotted lines denote theoretical resolvability thresholds (Q_peak, Q_high_95%, and Q_low_95%). CryoNet.Refine achieves a 92.4% pass rate (Q-score ≥ Q_low_95%), representing a substantial improvement over AlphaFold 3 starting models (27.5%), and outperforming *Phenix.real_space_refine* (64.9%). (**C**) Distributions of Q-score and map-model correlations (CC_{mask}, CC_{box}, CC_{mc}, and CC_{sc}), showing CryoNet.Refine consistently yields comparable or superior performance on density-related metrics. The mean values are indicated. (**D**) Comprehensive stereochemical assessment across ten metrics. CryoNet.Refine improves or maintains stereochemical quality relative to *Phenix.real_space_refine* across seven of the ten evaluated metrics. The mean values are indicated. (↑: higher is better, ↓: lower is better).

Figure 3. Performance on diverse biomolecular complexes and cryo-ET targets.



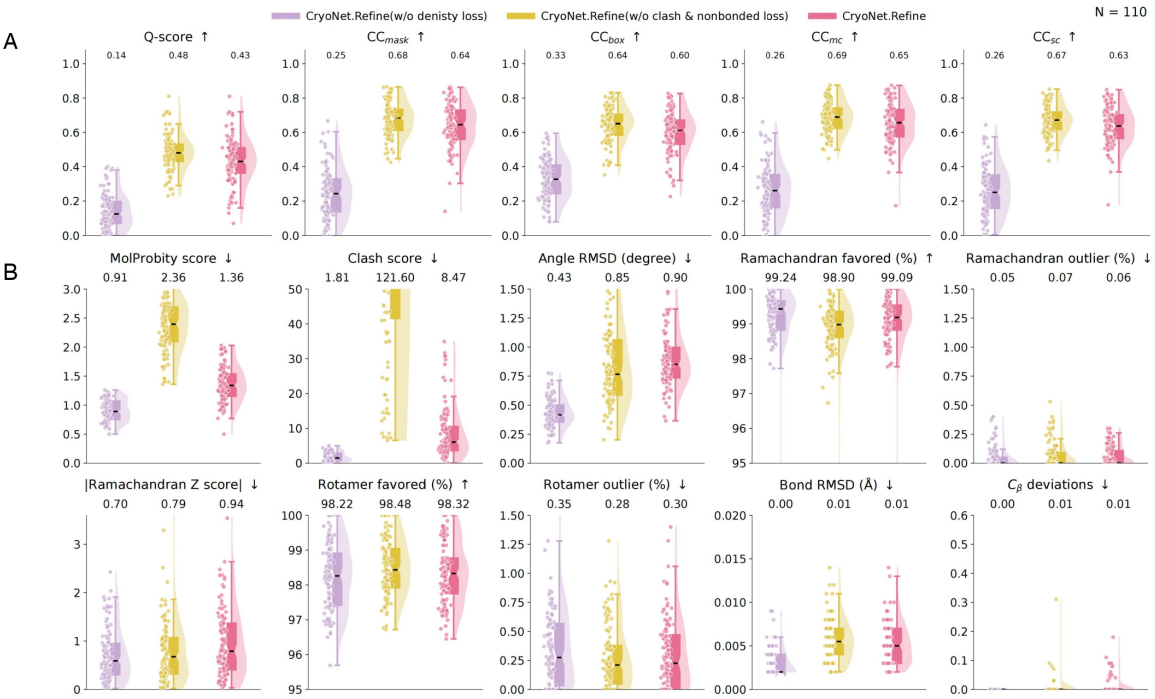
Representative refinement results compared to AlphaFold 3 starting models and models refined by *Phenix.real_space_refine*: **(A)** human CNT3 (EMD-0775, 3.60 Å, with 3 protein chains); (**B**) precleavage Cre tetrameric complex (EMD-24470, 3.23 Å, with 4 protein chains and 4 DNA chains); (**C**) HIV-1 RT IC core-EFV complex (EMD-22900, 2.90 Å, with 2 protein chains and 2 RNA chains); and (**D**) the mature RSV CA pentamer from *in situ* cryo-ET (EMD-12488, 5.80 Å, with 5 protein chains). Across all cases, CryoNet.Refine models exhibit superior map-model correlation (blue regions: high $CC_{residue}$), significantly lower MolProbity and clash scores, while adhering to Ramachandran and rotamer favored regions. Inserts highlight accurate backbone and side-chain alignments with the experimental density maps.

Figure 4. Large-scale evaluation on PDB-deposited structures.



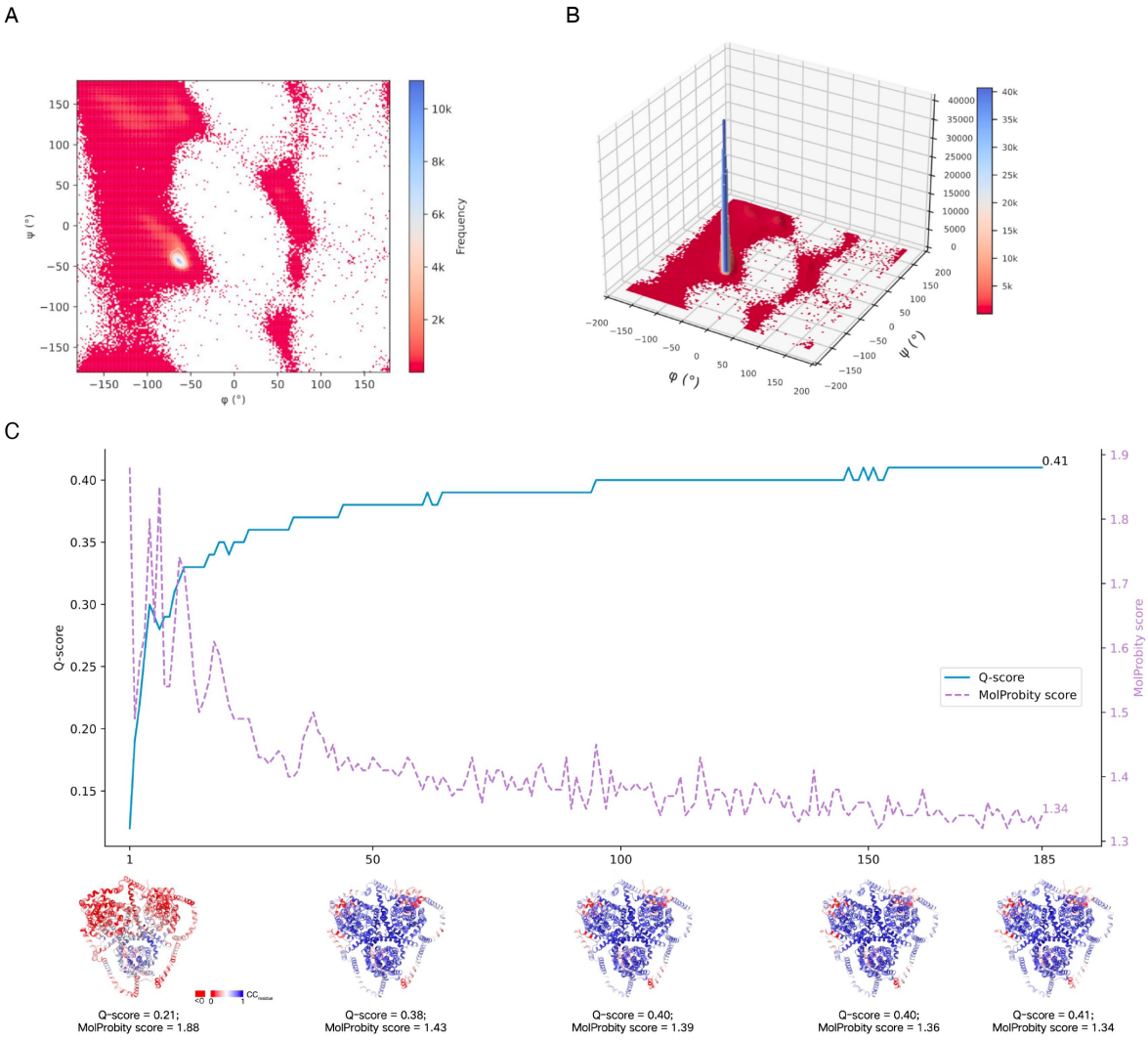
Evaluation on 325 pre-2018 (**A**-**D**) and 350 post-2018 (**E**-**H**) PDB deposits (1.8–6.0 Å). **(A, E)** CS-score comparisons against original PDB-deposits (left, mean CS-scores: 0.90 vs. 0.74 for pre-2018, and 0.93 vs. 0.84 for post-2018) and *Phenix.real_space_refine* (right, mean CS-scores: 0.90 vs. 0.80 for pre-2018, and 0.93 vs. 0.80 for post-2018). **(B, F)** Q-score versus resolution. CryoNet.Refine achieves pass rates (Q-score ≥ Q_low_95%) of 98.8% and 97.4% on *pre-2018* and *post-2018* deposits, respectively. (**C, G**) Distributions of Q-score and correlation metrics (CC_{mask}, CC_{box}, CC_{mc}, and CC_{sc}). The mean values are indicated. (**D, H**) Stereochemical assessment across ten metrics. The mean values are indicated. (↑: higher is better, ↓: lower is better).

Figure 5. Ablation study on loss components.



Evaluations performed on 110 AlphaFold 3-predicted starting protein structure models. (**A**) Distributions of Q-score and correlation metrics (CC_{mask}, CC_{box}, CC_{mc}, and CC_{sc}). Omitting density loss drastically reduces all correlation metrics, confirming its essential role in guiding the coordinate update module to conform to the experimental cryo-EM map. (**B**) Distributions of stereochemical quality metrics. Excluding the clash and non-bonded loss results in severe steric degradation (mean clash score surging to 121.60), demonstrating that the network requires this explicit penalty to resolve atomic overlaps during density fitting. The mean values are indicated. (↑: higher is better, ↓: lower is better).

464 **Figure S1. Ramachandran restraints and the coordinate update strategy.**

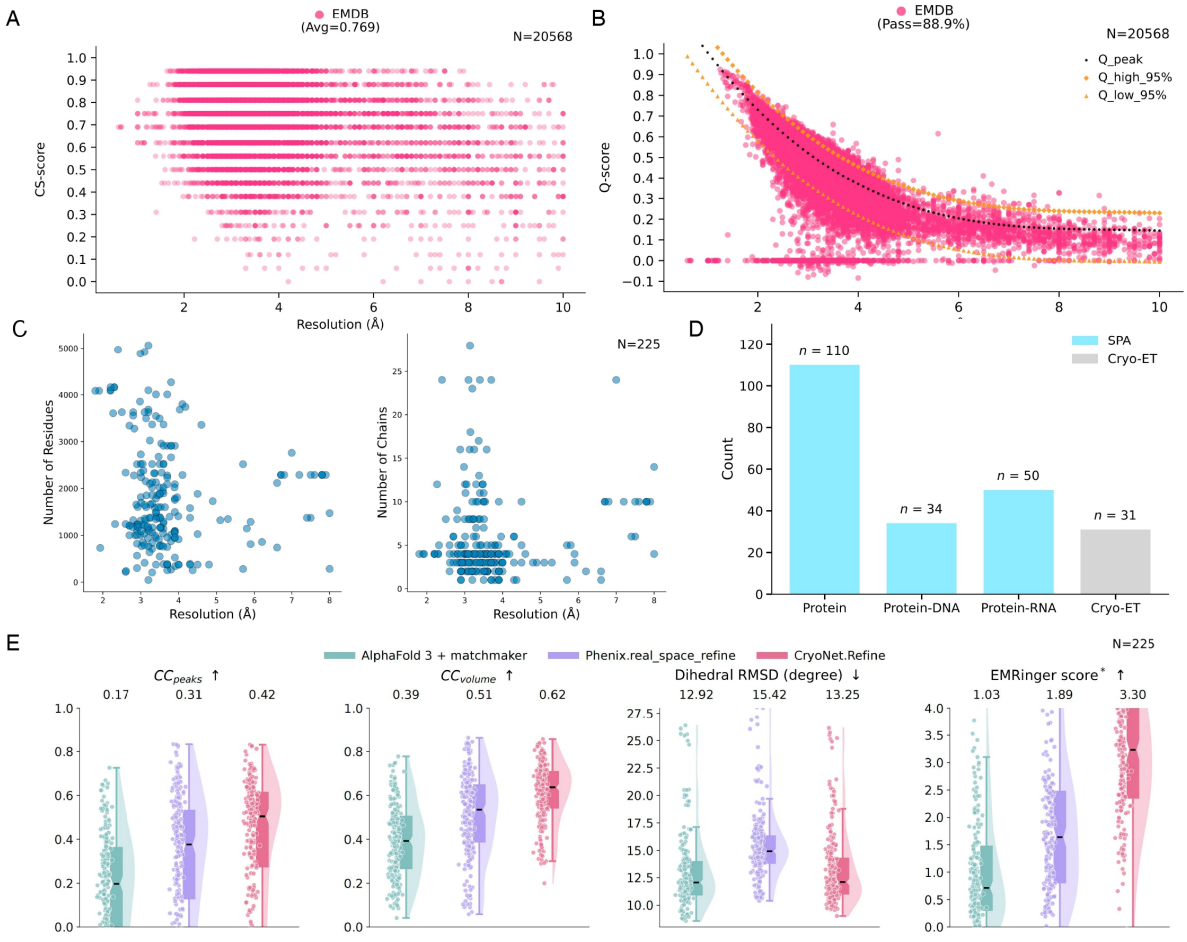


465

466 **(A, B)** 2D and 3D empirical backbone dihedral angle (ϕ, ψ) distributions from the Top8000 dataset,
467 defining the statistical basis for CryoNet.Refine's Ramachandran loss. Outliers are penalized to
468 enforce stereochemical quality. **(C)** Convergence of an AlphaFold 3-predicted starting model (PDB-
469 6ksw) over 185 refinement iterations. Q-score rapidly increase and converge, coupled concurrently
470 with a continuous decline in the MolProbity score. An early-stopping mechanism halts recycling
471 upon plateauing to prevent redundant computation.

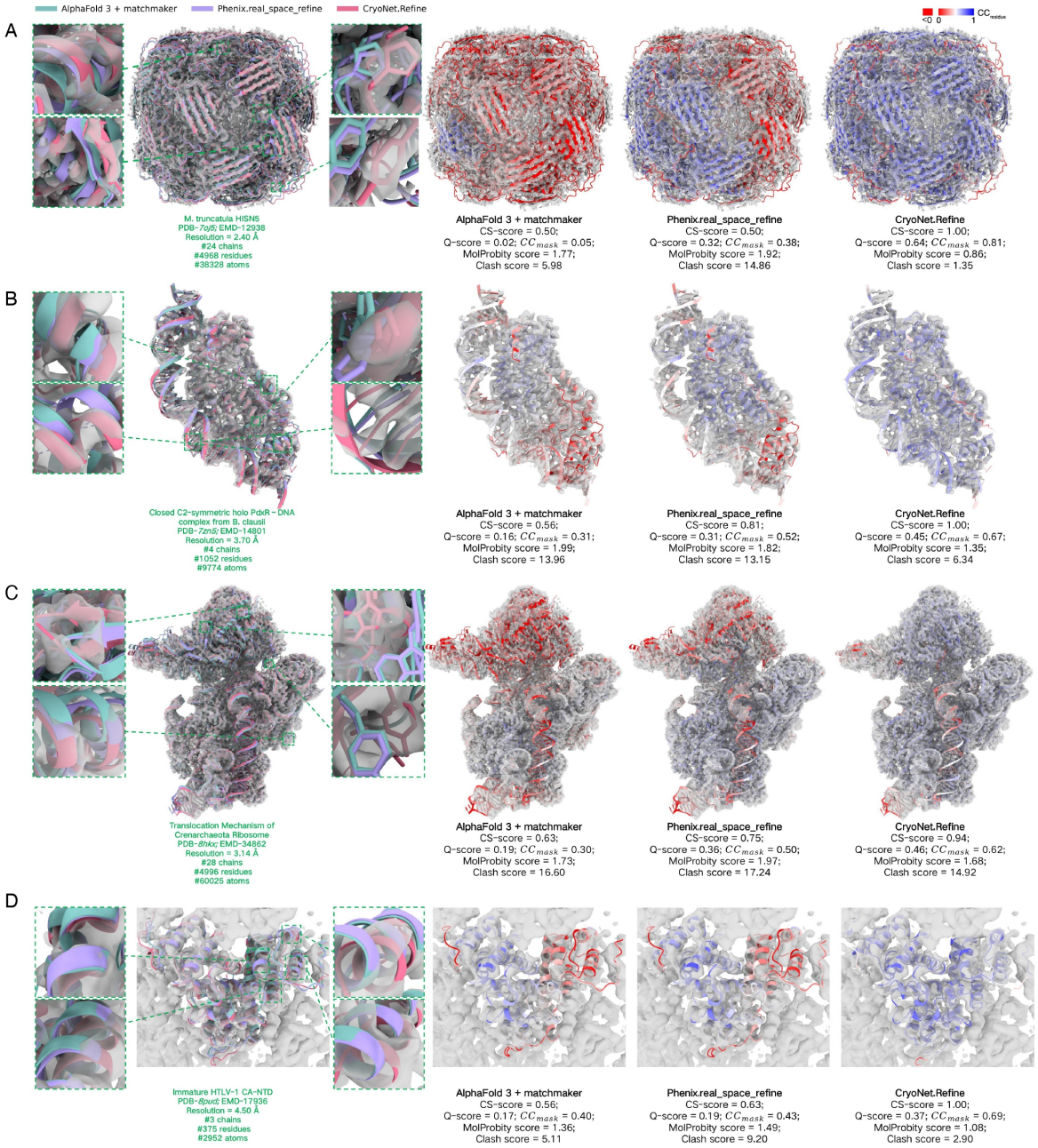
472

Figure S2. Statistical distributions and extended benchmark metrics.



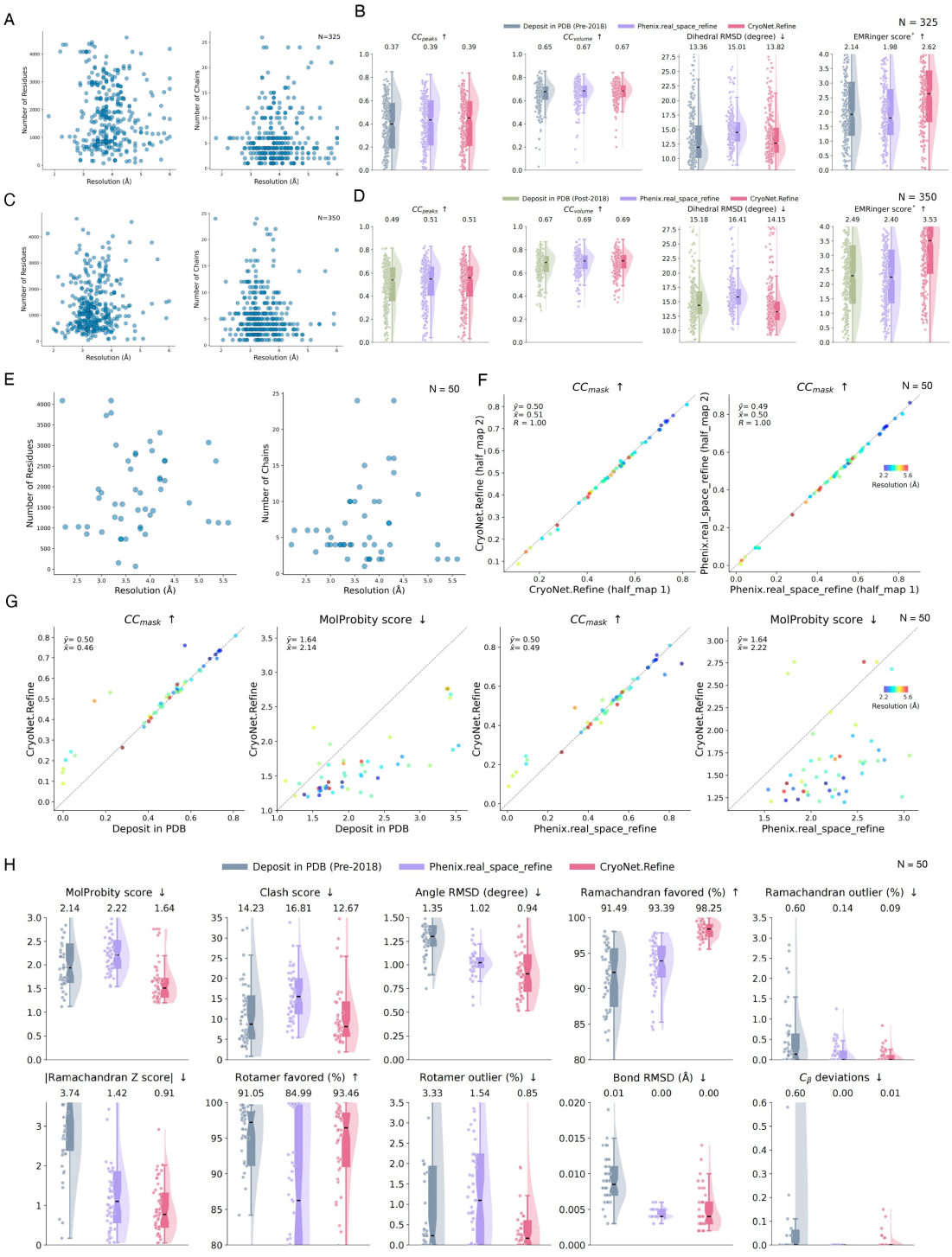
(A) CS-score versus resolution for 20,568 PDB/EMDB structures (0.6–10 Å; mean 0.769), establishing a global baseline. (**B**) Q-score versus resolution for 20,568 structures (0.6–10 Å). The *Q*_low_95% curve defines the resolution-dependent reliability; 88.9% of deposits pass this threshold. (**C, D**) Dataset composition of AlphaFold 3-predicted starting models (N = 225, 1.8–8.0 Å). **(C)** Detailed residue/chain counts and **(D)** composition of biomolecular types (110 protein, 34 protein-DNA, and 50 protein-RNA structures) and experimental methods (194 single-particle cryo-EM targets and 31 *in situ* cryo-electron tomography targets). (**E**) Extended metrics (CC_{peaks}, CC_{volume}, dihedral RMSD, and EMRinger score) on AlphaFold 3-predicted starting models. CryoNet.Refine yields higher mean correlations and EMRinger scores while preserving better backbone geometry (lower Dihedral RMSD). The mean values are indicated. *EMRinger evaluated only for maps with resolution ⩽4.5 Å (N = 197).

Figure S3. Performance on diverse complexes and cryo-ET targets.



Representative refinement results compared to AlphaFold 3-predicted starting models and models refined by *Phenix.real_space_refine*: **(A)** *Medicago truncatula* HISN5 (EMD-12938, 2.40 Å, with 24 protein chains); **(B)** the closed C2-symmetric holo PdxR–DNA complex from *B. clausii* (EMD-14801, 3.70 Å, with 2 protein chains and 2 DNA chains); **(C)** Crenarchaeota ribosome (EMD-34862, 3.14 Å, with 27 protein chains and 1 RNA chain); and **(D)** immature HTLV-1 CA-NTD from cryo-ET (EMD-17936, 4.50 Å, with 3 protein chains). CryoNet.Refine yields superior map-model correlation (blue regions: high $CC_{residue}$), lower MolProbity and clash scores. Inserts confirm accurate backbone and side-chain density alignments.

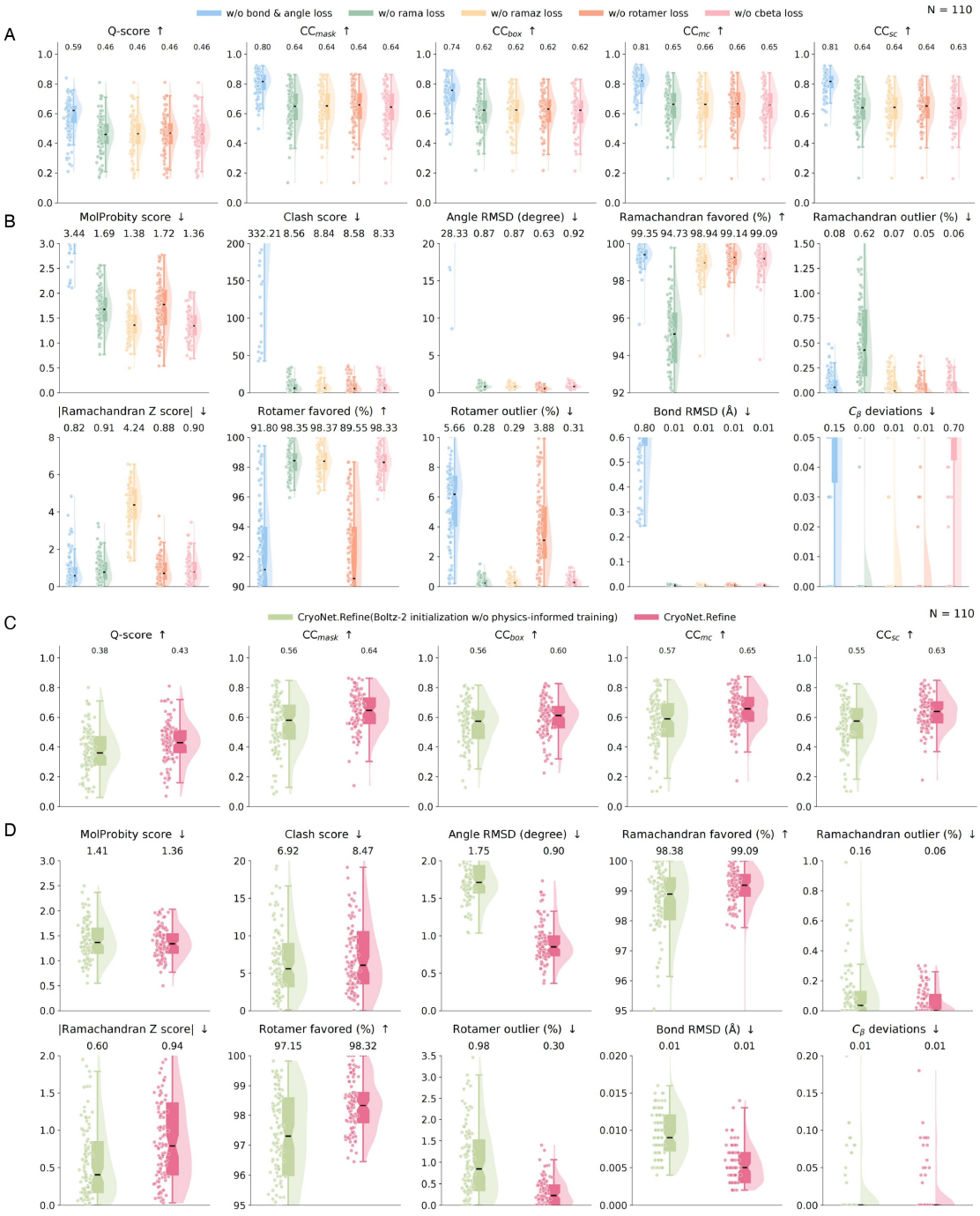
Figure S4. Extended evaluation on PDB-deposits and half-map cross-validation.



(**A, C**) Residue/chain count distributions on 325 pre-2018 (**A**) and 350 post-2018 (**C**) PDB deposits. **(B, D)** Extended metrics on 325 pre-2018 (**B**) and 350 post-2018 (**D**) PDB deposits. Violin plots confirm CryoNet.Refine yields comparable or superior CC_{peaks}, CC_{volume}, and EMRinger scores while maintaining lower dihedral RMSDs compared to baselines. The mean values are indicated. *EMRinger evaluated only for maps with resolution ⩽4.5 Å (pre-2018: N = 271; post-2018: N =

329). (**E-H**) Half-map cross-validation (N = 50, 2.20–5.60 Å). (**E**) Residue/chain count distributions. (**F**) Work map vs. validation map CC_{mask}. Points lie tightly along the diagonal parity line (y = x, $R = 1.00$). (**G**) Validation map evaluations. CryoNet.Refine outperforms PDB deposits in CC_{mask} and MolProbity score and surpasses *Phenix.real_space_refine* in geometric quality while maintaining comparable correlation, indicating minimal generalization gap and absence of overfitting to noise. (**H**) Comprehensive stereochemical assessment on the unseen validation map confirms CryoNet.Refine improves or maintains stereochemistry across nine of ten metrics relative to *Phenix.real_space_refine*. The mean values are indicated. (↑: higher is better, ↓: lower is better).

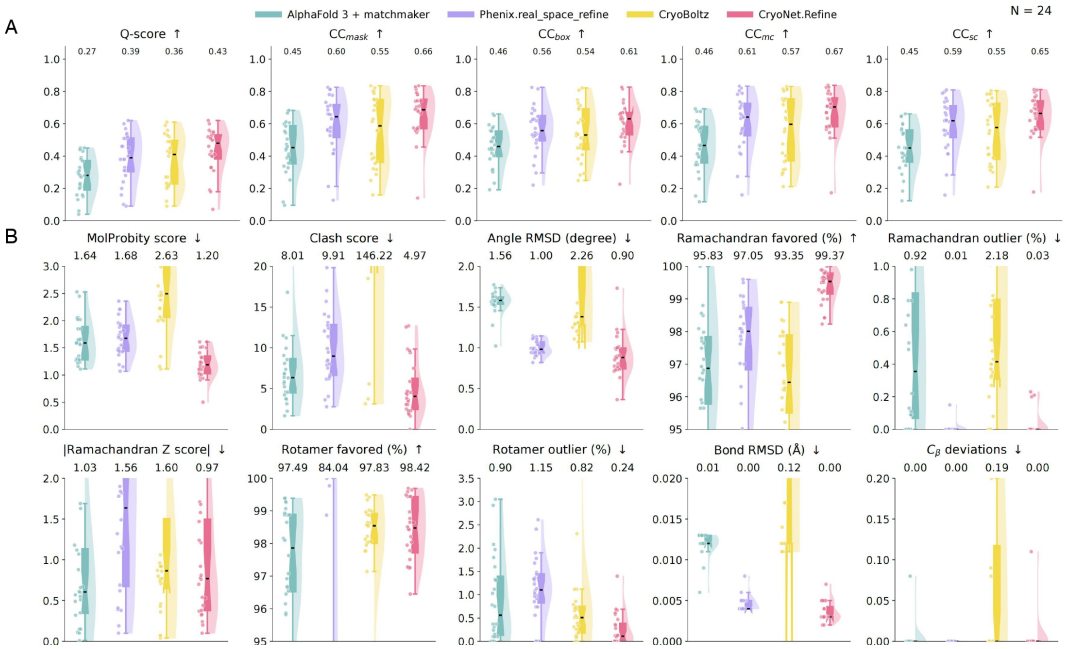
Figure S5. Ablation of extended geometry losses and physics-informed training.



Evaluations on 110 AlphaFold 3-predicted starting models. (**A, B**) Impact of omitting individual stereochemical restraints. Removing bond and angle restraints artificially inflates density-fit metrics (**A**) but catastrophically degrades geometry (**B**). Conversely, removing Ramachandran, rotamer, or C_β deviations causes only moderate density-fit reductions, indicating they act primarily as regularizers for geometric plausibility. The mean values are indicated. (**C, D**) Impact of physics-informed training. Compared to the full model, a baseline variant without physics-informed training

yields consistently lower Q-score and map-model correlations (**C**) and exhibits poorer stereochemistry (*e.g.*, mean MolProbity score of 1.41 vs. 1.36) (**D**). The mean values are indicated. (↑: higher is better, ↓: lower is better).

Figure S6. Benchmarking on AlphaFold 3-predicted starting models with CryoBoltz.



(**A, B**) Performance comparison including CryoBoltz on 24 AlphaFold 3-predicted starting protein starting models. CryoNet.Refine consistently outperforms baselines in correlation (**A**) and maintains/improves stereochemistry **(B)** relative to CryoBoltz. The mean values are indicated. (↑: higher is better, ↓: lower is better).

Reference

1 Kühlbrandt, W. The resolution revolution. *Science* **343**, 1443–1444 (2014).

2 Nakane, T. *et al.* Single-particle cryo-EM at atomic resolution. *Nature* **587**, 152–+ (2020). https://doi.org/10.1038/s41586-020-2829-0

3 Yip, K. M., Fischer, N., Paknia, E., Chari, A. & Stark, H. Atomic-resolution protein structure determination by cryo-EM. *Nature* **587**, 157–+ (2020). https://doi.org/10.1038/s41586-020-2833-4

4 Pfab, J., Phan, N. M. & Si, D. DeepTracer for fast de novo cryo-EM protein structure modeling and special studies on CoV-related complexes. *Proc Natl Acad Sci U S A* **118** (2021). https://doi.org/10.1073/pnas.2017525118

5 Xu, K., Wang, Z., Shi, J. P., Li, H. S. & Zhang, Q. C. A(2)-Net: Molecular Structure Estimation from Cryo-EM Density Volumes. *Thirty-Third Aaai Conference on Artificial Intelligence* **33**, 1230–1237 (2019).

6 Li, S., Terashi, G., Zhang, Z. & Kihara, D. Advancing structure modeling from cryo-EM maps with deep learning. *Biochem Soc Trans* **53**, 259–265 (2025). https://doi.org/10.1042/BST20240784

7 Wang, T. *et al.* CryoSeek: A strategy for bioentity discovery using cryoelectron microscopy. *Proc Natl Acad Sci U S A* **121**, e2417046121 (2024). https://doi.org/10.1073/pnas.2417046121

8 Terwilliger, T. C., Adams, P. D., Afonine, P. V. & Sobolev, O. V. A fully automatic method yielding initial models from high-resolution cryo-electron microscopy maps. *Nat Methods* **15**, 905–908 (2018). https://doi.org/10.1038/s41592-018-0173-1

9 Jamali, K. *et al.* Automated model building and protein identification in cryo-EM maps. *Nature* **628**, 450–457 (2024). https://doi.org/10.1038/s41586-024-07215-4

10 Afonine, P. V. *et al.* Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr D* **74**, 531–544 (2018). https://doi.org/10.1107/S2059798318006551

11 Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr* **67**, 355–367 (2011). https://doi.org/10.1107/S0907444911001314

12 Engh, R. A. & Huber, R. Accurate Bond and Angle Parameters for X-Ray Protein-Structure Refinement. *Acta Crystallogr A* **47**, 392–400 (1991). https://doi.org/Doi 10.1107/S0108767391001071

13 Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486–501 (2010). https://doi.org/10.1107/S0907444910007493

14 Croll, T. I. ISOLDE: a physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallogr D Struct Biol* **74**, 519–530 (2018). https://doi.org/10.1107/S2059798318002425

15 Wang, R. Y. *et al.* Automated structure refinement of macromolecular assemblies from cryo-EM maps using Rosetta. *Elife* **5** (2016). https://doi.org/10.7554/eLife.17219

16 Wang, F. *et al.* DeepPicker: A deep learning approach for fully automated particle picking in cryo-EM. *J Struct Biol* **195**, 325–336 (2016). https://doi.org/10.1016/j.jsb.2016.07.006

17 Bepler, T. *et al.* Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nat Methods* **16**, 1153–1160 (2019). https://doi.org/10.1038/s41592-019-0575-8

18 Levy, A. *et al.* CryoAI: Amortized Inference of Poses for Ab Initio Reconstruction of 3D Molecular Volumes from Real Cryo-EM Images. *Comput Vis ECCV* **13681**, 540–557 (2022). https://doi.org/10.1007/978-3-031-19803-8_32

19 Sanchez-Garcia, R. *et al.* DeepEMhancer: a deep learning solution for cryo-EM volume post-processing. *Commun Biol* **4**, 874 (2021). https://doi.org/10.1038/s42003-021-02399-1

20 He, J., Li, T. & Huang, S. Y. Improvement of cryo-EM maps by simultaneous local and non-local deep learning. *Nat Commun* **14**, 3217 (2023). https://doi.org/10.1038/s41467-023-39031-1

21 He, J., Lin, P., Chen, J., Cao, H. & Huang, S. Y. Model building of protein complexes from intermediate-resolution cryo-EM maps with deep learning-guided automatic assembly. *Nat Commun* **13**, 4066 (2022). https://doi.org/10.1038/s41467-022-31748-9

22 Li, T. *et al.* All-atom RNA structure determination from cryo-EM maps. *Nat Biotechnol* **43**, 97–105 (2025). https://doi.org/10.1038/s41587-024-02149-8

23 Su, B., Huang, K., Peng, Z., Amunts, A. & Yang, J. CryoAtom improves model building for cryo-EM. *Nat Struct Mol Biol* **33**, 351–361 (2026). https://doi.org/10.1038/s41594-025-01713-3

24 Wang, X., Zhu, H., Terashi, G., Taluja, M. & Kihara, D. DiffModeler: large macromolecular structure modeling for cryo-EM maps using a diffusion model. *Nat Methods* **21**, 2307–2317 (2024). https://doi.org/10.1038/s41592-024-02479-0

25 Karniadakis, G. E. *et al.* Physics-informed machine learning. *Nature Reviews Physics* **3**, 422–440 (2021). https://doi.org/10.1038/s42254-021-00314-5

26 Raissi, M., Perdikaris, P. & Karniadakis, G. E. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J Comput Phys* **378**, 686–707 (2019). https://doi.org/10.1016/j.jcp.2018.10.045

27 Wang, S. F., Teng, Y. J. & Perdikaris, P. Understanding and Mitigating Gradient Flow Pathologies in Physics-Informed Neural Networks. *Siam J Sci Comput* **43**, A3055–A3081 (2021). https://doi.org/10.1137/20m1318043

28 Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021). https://doi.org/10.1038/s41586-021-03819-2

29 Baek, M. *et al.* Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373**, 871–+ (2021). https://doi.org/10.1126/science.abj8754

30 Sohl-Dickstein, J., Weiss, E. A., Maheswaranathan, N. & Ganguli, S. Deep Unsupervised Learning using Nonequilibrium Thermodynamics. *arXiv e-prints*, arXiv:1503.03585 (2015). https://doi.org/10.48550/arXiv.1503.03585

31 Ho, J., Jain, A. & Abbeel, P. Denoising Diffusion Probabilistic Models. *arXiv e-prints*, arXiv:2006.11239 (2020). https://doi.org/10.48550/arXiv.2006.11239

32 Song, Y. *et al.* Score-Based Generative Modeling through Stochastic Differential Equations. *arXiv e-prints*, arXiv:2011.13456 (2020). https://doi.org/10.48550/arXiv.2011.13456

33 Watson, J. L. *et al.* De novo design of protein structure and function with RFdiffusion. *Nature* **620**, 1089–1100 (2023). https://doi.org/10.1038/s41586-023-06415-8

34 Ingraham, J. B. *et al.* Illuminating protein space with a programmable generative model. *Nature* **623**, 1070–1078 (2023). https://doi.org/10.1038/s41586-023-06728-8

35 Abramson, J. *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500 (2024). https://doi.org/10.1038/s41586-024-07487-w

36 Passaro, S. *et al.* Boltz-2: Towards Accurate and Efficient Binding Affinity Prediction. *bioRxiv* (2025). https://doi.org/10.1101/2025.06.14.659707

37 Meng, E. C. *et al.* UCSF ChimeraX: Tools for structure building and analysis. *Protein Sci* **32**, e4792 (2023). https://doi.org/10.1002/pro.4792

38 Goddard, T. D. *et al.* UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci* **27**, 14–25 (2018). https://doi.org/10.1002/pro.3235

621 39 Paszke, A. *et al.* PyTorch: An Imperative Style, High-Performance Deep Learning Library. *arXiv*
622 *e-prints*, arXiv:1912.01703 (2019). https://doi.org/10.48550/arXiv.1912.01703

623 40 Lawson, C. L. *et al.* EMDataBank unified data resource for 3DEM. *Nucleic Acids Res* **44**, D396–
624 D403 (2016). https://doi.org/10.1093/nar/gkv1126

625 41 Fleming, J. *et al.* AlphaFold Protein Structure Database and 3D-Beacons: New Data and
626 Capabilities. *J Mol Biol* **437**, 168967 (2025). https://doi.org/10.1016/j.jmb.2025.168967

627 42 Berman, H., Henrick, K. & Nakamura, H. Announcing the worldwide Protein Data Bank. *Nat*
628 *Struct Biol* **10**, 980 (2003). https://doi.org/10.1038/nsb1203-980

629 43 Pintilie, G. *et al.* Measurement of atom resolvability in cryo-EM maps with Q-scores. *Nature*
630 *Methods* **17**, 328–334 (2020). https://doi.org/10.1038/s41592-020-0731-1

631 44 Williams, C. J. *et al.* MolProbity: More and better reference data for improved all-atom structure
632 validation. *Protein Sci* **27**, 293–315 (2018). https://doi.org/10.1002/pro.3330

633 45 Barad, B. A. *et al.* EMRinger: side chain-directed model and map validation for 3D cryo-electron
634 microscopy. *Nat Methods* **12**, 943–946 (2015). https://doi.org/10.1038/nmeth.3541

635 46 Afonine, P. V. *et al.* New tools for the analysis and validation of cryo-EM maps and atomic models.
636 *Acta Crystallogr D Struct Biol* **74**, 814–840 (2018). https://doi.org/10.1107/S2059798318009324

637 47 Zhou, Y. *et al.* Cryo-EM structure of the human concentrative nucleoside transporter CNT3. *PLoS*
638 *Biol* **18**, e3000790 (2020). https://doi.org/10.1371/journal.pbio.3000790

639 48 Stachowski, K., Norris, A. S., Potter, D., Wysocki, V. H. & Foster, M. P. Mechanisms of Cre
640 recombinase synaptic complex assembly and activation illuminated by Cryo-EM. *Nucleic Acids*
641 *Res* **50**, 1753–1769 (2022). https://doi.org/10.1093/nar/gkac032

642 49 Ha, B. *et al.* High-resolution view of HIV-1 reverse transcriptase initiation complexes and
643 inhibition by NNRTI drugs. *Nat Commun* **12**, 2500 (2021). https://doi.org/10.1038/s41467-021-
644 22628-9

645 50 Obr, M. *et al.* Structure of the mature Rous sarcoma virus lattice reveals a role for IP6 in the
646 formation of the capsid hexamer. *Nat Commun* **12**, 3226 (2021). https://doi.org/10.1038/s41467-
647 021-23506-0

648 51 Raghu, R., Levy, A., Wetzstein, G. & Zhong, E. D. Multiscale guidance of protein structure
649 prediction with heterogeneous cryo-EM data. *arXiv e-prints*, arXiv:2506.04490 (2025).
650 https://doi.org/10.48550/arXiv.2506.04490

651