# CRYONET.REFINE: A ONE-STEP DIFFUSION MODEL FOR RAPID REFINEMENT OF STRUCTURAL MODELS WITH CRYO-EM DENSITY MAP RESTRAINTS

**Anonymous authors**
Paper under double-blind review

## ABSTRACT

High-resolution structure determination by cryo-electron microscopy (cryo-EM) requires the accurate fitting of an atomic model into an experimental density map. Traditional refinement pipelines like *Phenix.real_space_refine* and Rosetta are computationally expensive, demand extensive manual tuning, and present a significant bottleneck for researchers. We present *CryoNet.Refine*, an end-to-end, deep learning framework that automates and accelerates molecular structure refinement. Our approach utilizes a one-step diffusion model that integrates a density-aware loss function with robust stereochemical restraints, enabling it to rapidly optimize a structure against the experimental data. *CryoNet.Refine* stands as a unified and versatile solution capable of refining not only protein complexes but also DNA/RNA-protein complexes. In benchmarks against *Phenix.real_space_refine*, *CryoNet.Refine* consistently yields substantial improvements in both model–map correlation and overall model geometric metrics. By offering a scalable, automated, and powerful alternative, *CryoNet.Refine* is poised to become an essential tool for next-generation cryo-EM structure refinement.

## 1 INTRODUCTION

Cryo-electron microscopy (cryo-EM) has emerged as a revolutionary technique in structural biology, enabling the determination of macromolecular structures, including numerous crucial biological complexes, at unprecedented resolution (Kühlbrandt, 2014; Nogales & Eva, 2016). The typical cryo-EM workflow involves several sequential stages: sample preparation, electron micrograph acquisition, particle picking, three-dimensional (3D) reconstruction, atomic model building, and final structure refinement. Despite these advancements, a persistent challenge in cryo-EM remains the inherent low signal-to-noise ratio (SNR) and the pervasive conformational dynamics of biological samples. These factors often lead to low-resolution cryo-EM density maps from 3D reconstruction, and even high-resolution maps frequently exhibit low-resolution densities at peripheral and/or flexible regions. Such limitations severely compromise the effectiveness of atomic model building tools, including traditional approaches like *Phenix.map_to_model* (Afonine et al., 2018a), *A2-Net* (Xu et al., 2019), *DeepTracer*(Pfab et al., 2021), *CryoNet*(`https://cryonet.ai`), and even more recent breakthroughs like *ModelAngelo*(Jamali et al., 2024). This can result in fragmented or incomplete structures, incorrect identification of amino acid or nucleic acid types, and in extreme cases, the inability to complete atomic model building, often necessitating integration with orthogonal data sources like RNA-seq for structure discovery(Wang et al., 2024; 2025).

Due to the limitations mentioned above, atomic model refinement becomes a crucial step to follow once an initial atomic model is built. This phase meticulously adjusts the atomic coordinates of each amino acid or nucleic acid within the model, typically employing both automated and manual interactive tools. Automated tools, exemplified by *Phenix.real_space_refine* (Afonine et al., 2018b), are highly versatile, capable of refining not only protein structures but also DNA and RNA. It incorporates a comprehensive library of structural restraints, including secondary structure, rotamer, and Ramachandran plot. By iteratively optimizing the stucture through simulated annealing and sampling from vast conformational spaces, it seeks a subset of models that best fit the cryo-EM

density map, ultimately yielding a structure with excellent geometric metrics and high model-map correlation coefficients. Concurrently, interactive tools like *Coot*(Emsley & Cowtan, 2004) provide structural biologists with user-friendly interfaces to visualize poor geometry for convenient manual adjustments. Coot further allows users to dynamically adjust the weights of local density regions with unbalanced quality, offering effective real-time constraints during manual refinement. While highly effective, both automated and manual methods often require "case-by-case" parameter tuning by experts. The process of cryo-EM refinement is in urgent need of a flexible, robust and fully automated method . Leveraging the powerful capabilities of modern AI approaches thus holds immense prospect for developing superior cryo-EM refinement tools.

The landscape of AI in structural biology has been dramatically reshaped by recent advancements in generative models, particularly diffusion models, which excel in protein generation, design, and structure prediction. Breakthroughs like *AlphaFold3*(Abramson et al., 2024), *RFDiffusion*(Watson et al., 2023), and *Chroma*(Ingraham et al., 2023) exemplify their exceptional ability to generate diverse and high-quality structures for various biomolecules, including proteins, DNA, and RNA complexes. Beyond single-chain protein prediction, where models like *AlphaFold2*(Jumper et al., 2021), *RoseTTAFold* (Baek et al., 2021), and *ESMFold*(Lin et al., 2023) have achieved remarkable accuracy, diffusion-based methods are extending to protein-protein interactions, protein-small molecule complexes, and RNA/DNA complexes (Baek et al., 2021; 2023; Wohlwend et al., 2025; Passaro et al., 2025). Notably, *AlphaFold3* has moved beyond predefined fixed bond lengths and angles from *AlphaFold2*, by learning these geometric constraints directly from PDB data using diffusion models. This includes accurately capturing geometry features like the planarity of benzene rings in amino acids. However, a significant limitation of these generative methods is that while they produce geometrically plausible structures, they often exhibit suboptimal performance on various detailed geometric metrics and, crucially, do not natively support refinement under the direct restraints of experimental data, such as cryo-EM density maps. Applying the robust generative ability of diffusion models to cryo-EM data-restraints refinement thus offers a transformative pathway to substantially improve structure quality and overcome the reliance on laborious, parameter-heavy traditional or manual refinement workflows.

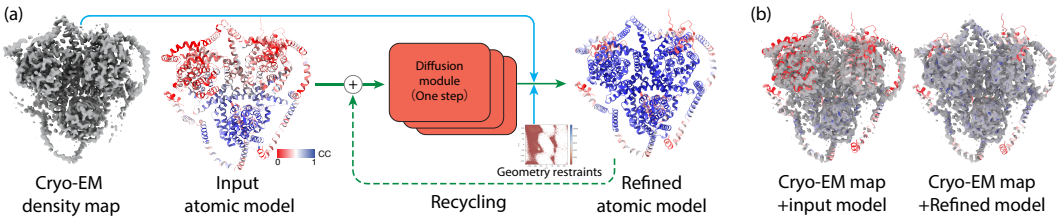



Figure 1: **(a)** Workflow of *CryoNet.Refine*. Atomic models are colored by model–map correlation coefficients (CC), with blue indicating high CC and red indicating low CC. **(b)** The input atomic model, refined atomic model within cryo-EM density map.

To address these limitations, we introduce *CryoNet.Refine*, a novel one-step diffusion model designed for cryo-EM atomic model refinement Figure 1.**(a)**. *CryoNet.Refine* pioneers the integration of advanced atomic generation principles, inspired by the capabilities seen in *AlphaFold3*, into a comprehensive refinement framework. Taking a cryo-EM density map and an atomic structure as input, Atom encoder first extracts intricate features from the structure to be refined. Concurrently, Sequence embedder encodes atomic type information derived from the input molecular sequence. These encoded representations are then fed into our one-step diffusion module, which iteratively generates the refined atomic structure. The refinement process is guided by novel designed density loss and a set of geometry losses, facilitating iterative optimization of the atomic model. *CryoNet.Refine* achieves the first successful atomic structure refinement under the direct constraints of experimental cryo-EM density maps and the guidance of standard geometry metrics within a neural network framework (Figure 1.**(b)**). Experimental benchmarkds demonstrates significant improvements over *Phenix.real_space_refine* on both protein-protein and DNA/RNA-protein complexes, showcasing its exceptional performance and broad applicability. Our work presents four main contributions:

(1) We propose the first AI-based method for cryo-EM atomic model refinement, leveraging a deep neural network-based one-step diffusion module.
(2) We develop a novel parameter-free and differentiable density generator that can produce simulated density maps from the generated atomic model. This innovation enables us to design an

effective density loss, marking the first time that density map correlation can be directly utilized as a loss function to guide neural network training.

(3) We introduce a set of differentiable geometry loss functions specifically tailored for guiding the generation of geometrically plausible macromolecular structures, which also offer valuable guidance for protein design and protein structure prediction models.

(4) We conduct extensive evaluation of *CryoNet.Refine* against *Phenix.real_space_refine* on cryo-EM datasets, showing a marked improvement in structural quality across various metrics.

## 2 RELATED WORK

In this section, we review the state-of-the-art methods in cryo-EM atomic model refinement and structure modeling, highlighting both traditional and emerging AI-driven approaches.

**Cryo-EM atomic model refinement.** Cryo-electron microscopy (cryo-EM) has revolutionized structural biology, enabling the determination of macromolecular structures at near-atomic resolution. However, the initial cryo-EM density maps often require subsequent atomic model refinement to achieve biologically meaningful and geometrically accurate structures. Traditionally, atomic model refinement methods can be broadly categorized into automated and manual approaches. Automated tools, such as *Phenix.real_space_refine*(Afonine et al., 2018b), *Rosetta*(Wang et al., 2016), and *ISOLDE*(Croll, 2018), leverage various geometric restraints (e.g., bond lengths, angles, dihedral angles) and force fields to drive atomic models to fit to density while maintaining stereochemical integrity. While these methods can be powerful and yield highly refined structures and, with expert "case-by-case" parameter tuning, their extensive parameter sets and specialized workflows often present a steep learning curve for non-expert users. In parallel, manual refinement tools like *Coot*(Emsley & Cowtan, 2004)offer powerful interactive visualization capabilities, allowing researchers to meticulously adjust individual amino acid residues or local regions directly within the density map. These tools provide unparalleled control and flexibility, enabling highly precise adjustments. Nevertheless, manual refinement is notoriously labor-intensive, time-consuming, and similarly demands significant expert knowledge, making it a bottleneck for high-throughput structure determination.

**Protein Structure Refinement.** More recently, the field has seen a surge of interest in artificial intelligence (AI)-driven methods for Protein Structure Refinement. Approaches such as *DeepAccNet*(Hiranuma et al., 2021), *GNNRefine*(Jing & Xu, 2021), and *AtomRefine*(Wu et al., 2023) employ neural networks, typically 3D convolution neural networks and Graph Neural Networks (GNNs), to learn intricate geometric features of protein backbones and side-chains. These methods aim to predict corrections or refine atomic positions based on learned structural patterns. A key characteristic of these existing AI-based refinement techniques is their primary reliance on structural learning, often from large databases of known protein structures. Consequently, the refined structures are fundamentally predictions of geometrically plausible conformations. A critical limitation, however, is the general absence of direct integration with experimental cryo-EM density maps during the differentiable optimization process. This disconnect means that the final predicted structures, while potentially ideal in terms of stereochemical geometry, frequently do not optimally match experimental data. Currently, there is a notable gap in the literature for neural network-based methods that support differentiable refinement under the direct constraint of cryo-EM experimental data.

**Diffusion models for structural generation.** The advent of diffusion models has marked a significant paradigm shift in generative AI, demonstrating remarkable capabilities in complex data generation, including protein structure prediction and design. Diffusion models, exemplified by architectures like *RFDiffusion*(Watson et al., 2023) for de novo protein design and *AlphaFold3*(Abramson et al., 2024) for biomolecular interaction structure prediction and, excel at learning atom distributions and generating diverse, high-quality biomolecular structures. Their success in protein structure generation is largely attributed to their ability to capture global and local structural features, implicitly enforcing geometric characteristics like bond lengths and angles, thus generating models with high stereochemical quality. These models have shown exceptional proficiency in maintaining geometric fidelity in generated structures. Leveraging the strong structural priors and generative capabilities of diffusion models could offer a novel avenue for cryo-EM structure refinement, particularly if their powerful learning frameworks could be adapted to integrate and be driven by experimental cryo-EM density information in a differentiable manner.

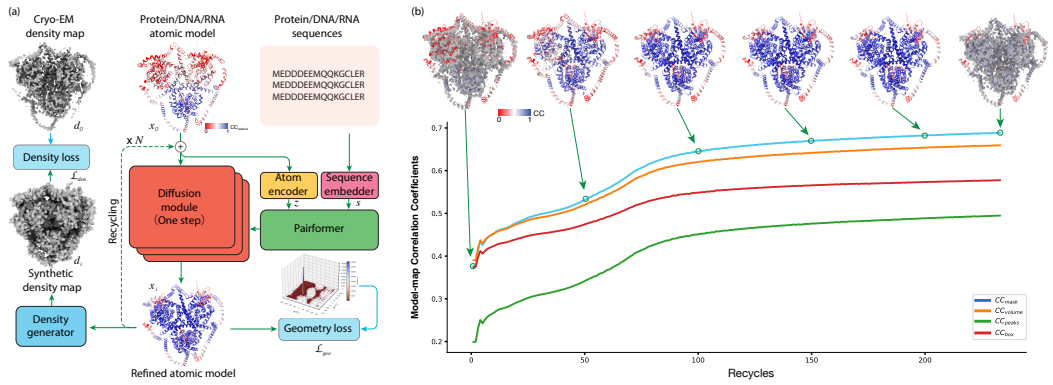



Figure 2: **(a)** Overview of the *CryoNet.Refine* framework. The *CryoNet.Refine* Network consists of four modules: Atom encoder, Sequence embedder, Diffusion module, and Density generator. The input atomic model is first processed by the encoders, and the resulting features are fed into a one-step diffusion module to generate an initial refined atomic model. Subsequently, the Density generator creates a synthetic density map, which is used to compute density loss against the input density map and geometry loss based on geometry restraints. These losses are then backpropagated to optimize the diffusion module, while the atomic model is further refined through multiple recycle steps until convergence. **(b)** Model-map correlation coefficient trajectory over 234 recycling *CryoNet.Refine* iterations on the structure of the human concentrative nucleoside transporter CNT3((Bank)). The input density map is EMD-0775, the input atomic model is predicted by *AlphaFold3*.

# 3 METHODS

## 3.1 CRYONET.REFINE FRAMEWORK

*CryoNet.Refine* is an end-to-end deep learning framework designed for the atomic model refinement of macromolecules directly against experimental cryo-EM density maps Figure 2.**(a)**. The refinement process begins by taking an experimental cryo-EM density map ($d_0$) and an initial atomic structure ($x_0$) as input. Atom encoder first processes the input structure to extract pairwise features ($z$) for each atom. Concurrently, Sequence embedder encodes atomic type information ($s$) derived from the input molecular sequence. These encoded representations ($z$,$s$) along with the initial atomic structure ($x_0$) are then fed into our one-step Diffusion Module to generate a refined atomic structure ($d_1$). Subsequently, the density generator takes the refined structure ($x_1$) as input to produce a simulated density map ($d_1$). A designed density loss ($\mathcal{L}_{\text{den}}$) is then calculated by comparing this simulated map ($d_1$) with the input map ($d_0$). Simultaneously, a Geometry loss ($\mathcal{L}_{\text{geo}}$) is computed based on the refined structure ($x_1$). These two losses are then weighted and summed to form the total loss ($\mathcal{L}_{\text{total}}$), which is backpropagated to update the parameters of the diffusion module. This process constitutes one refinement cycle. The refined structure from this cycle ($x_1$) is then used as the input for the next recycle step to generate $x_2$. After $n$ such cycles, the final refined atomic model, $x_n$, is obtained. Our network was first initialized with the pre-trained parameters of *Boltz2*(Passaro et al., 2025), a PyTorch implementation of *AlphaFold3*, followed by training with a composite loss function including two components: one measuring experimental data fidelity and the other introducing geometric restraints. The network was trained until convergence, yielding the generated atomic structures that are in high agreement with the experimental data while satisfying geometric constraints.

**One-step diffusion module.** The atomic structure generator is designed as a one-step diffusion model. Unlike the diffusion module in *AlphaFold3* that requires hundreds of sampling steps, one-step diffusion models represent a major development in generative AI, leveraging techniques like Knowledge Distillation(Meng et al., 2022) and Consistency Models(Song et al., 2023) (as explored in (Wu et al., 2025)) to compress the generation process and reduce computational bottlenecks. We have observed that the one-step diffusion module possesses a key advantage: it can easily incorporate features from experimental data and geometric restriants, into the generation process. This one-step design effectively and efficiently respects the given guidance, generating accurate results that align tightly with the restraints. Figure 2.**(b)** shows the refining atomic model trajectory colored with model-map correlation coefficients.

Concretely, the refining process is performed by training the network under a **preconditioned parameterization** following (Karras et al., 2022), and finally producing the refined coordinates $\hat{\mathbf{x}}$ as:

$$\hat{\mathbf{x}} \;=\; c_{\text{skip}}(\sigma)\,\mathbf{x}_0 \;+\; c_{\text{out}}(\sigma)\,\mathcal{F}_\theta\Big(c_{\text{in}}(\sigma)\mathbf{x}_0,\; c_{\text{noise}}(\sigma),\; \mathcal{C}\Big), \tag{1}$$

where $c_{\text{skip}}$,$c_{\text{out}}$ and $c_{\text{out}}$ denote coefficients in the preconditioned forward module(Appendix C) and $\mathcal{F}_\theta$ is a parameterized neural network in it. $\mathcal{C}$ denotes conditioned features derived from encoded structural features $s$ and $z$. This preconditioned update ensures that the one-step refinement preserves the theoretical properties of multi-step diffusion while collapsing the stochastic process into a deterministic, single-step prediction.

**Restraints of density map and structure geometry.** *CryoNet.Refine* employs a density generator and two types of loss functions to leverage these restraints. The first type of loss is a **density loss**, derived by compute the correlation between the experimental cryo-EM maps and the synthetic maps generated from refined atomic model by the density generator (Section 3.2.1). Another one is a set of geometry loss terms imposes stereochemical restraints—such as those derived from Ramachandran plot, rotamer, and bond angle distributions—to ensure standard structure geometry (Section 3.2.2). Thus the network is trained using a recycling strategy that reuses refined outputs as inputs for subsequent passes, progressively optimizing the two loss functions for density map agreement and geometry metrics. Together, these novel design choices enable CryoNet.Refine to achieve rapid, generalizable, and experimentally consistent refinement by fully respecting the restraints derived from both the cryo-EM density map and the structure geometry.

### 3.2 Loss functions

The loss function of *CryoNet.Refine* consists of density loss and geometry loss, defined as follows:

$$\begin{aligned} \mathcal{L} &= \gamma_{\text{den}} \cdot \mathcal{L}_{\text{den}} + \mathcal{L}_{\text{geo}}, \\ \mathcal{L}_{\text{geo}} &= \gamma_{\text{rama}} \cdot \mathcal{L}_{\text{rama}} + \gamma_{\text{rot}} \cdot \mathcal{L}_{\text{rot}} + \gamma_{\text{angle}} \cdot \mathcal{L}_{\text{angle}} + \gamma_{\text{C}_\beta} \cdot \mathcal{L}_{\text{C}_\beta} + \gamma_{\text{viol}} \cdot \mathcal{L}_{\text{viol}}, \end{aligned} \tag{2}$$

where $\gamma$ is the weight of each loss. See the detailed definition of these functions in Section 3.2.1, Section 3.2.2, and Appendix D. To the best of our knowledge, *CryoNet.Refine* is the first to formulate a differentiable implementation to compute density loss, ramachandran loss, rotamer loss and $C_\beta$ loss. Although these biological implications have been well-established and recognized as indispensable for protein structure prediction and atomic model building, prior methodologies have not integrated similar constraints into their nerual network based refinement processes.

#### 3.2.1 Density loss

We compute the density loss within two steps: (i) genarate a synthetic density map for the refined atomic model; (ii) compute the overlap region between the input density and the synthetic density. Both these two steps must be implemented to be fully differentiable.

A density map $d$ can be discretized as a group of 4-dimensional vectors $[\lambda, \mu, \nu, \rho]$, where $\vec{\boldsymbol{m}} = [\lambda, \mu, \nu]$ specifies a position and $\rho$ is the density value at this point, with the origin of coordinates nearest to $\vec{\boldsymbol{m}}_{\text{min}}$ and farthest to $\vec{\boldsymbol{m}}_{\text{max}}$. For computation of the overlapped region, we first locate this region $\forall \vec{\boldsymbol{m}}_o \in \mathcal{S}_o$ between the input density map $d_0$ and the synthetic density map $d_i$ through the coordinate system:

$$\vec{\boldsymbol{m}}_{o_{\text{min}}} = \max(\vec{\boldsymbol{m}}_{0_{\text{min}}}, \vec{\boldsymbol{m}}_{i_{\text{min}}}) \quad \vec{\boldsymbol{m}}_{o_{\text{max}}} = \min(\vec{\boldsymbol{m}}_{0_{\text{max}}}, \vec{\boldsymbol{m}}_{i_{\text{max}}})$$

After locating $\mathcal{S}_o$, we can select all the density values within it from $d_0$ and $d_i$ and get their subsets $d_{0\bigcap o}$ and $d_{i\bigcap o}$. Thus, the density loss will be defined as:

$$\mathcal{L}_{\text{den}} = 1 - \frac{\boldsymbol{\rho_{0\bigcap o}} \cdot \boldsymbol{\rho_{i\bigcap o}}}{||\boldsymbol{\rho_{0\bigcap o}}||_2 \cdot ||\boldsymbol{\rho_{i\bigcap o}}||_2} \tag{3}$$

For the implementation of synthetic density map generation, we follow the *molmap* algorithm in ChimeraX(Pettersen et al., 2021), utilizing the all-atom coordinates of the refined atomic model at the $i$ recycle $\vec{\boldsymbol{x}}_i = [\boldsymbol{x}_i^0, \boldsymbol{x}_i^1, \boldsymbol{x}_i^2]$ and output a box-shaped density map $d_i = (\vec{\boldsymbol{m}}_i, \boldsymbol{\rho}_i)$, where $\vec{\boldsymbol{x}}_i \in \mathbb{R}^{3\times N_{\text{atom}}}$ and $\vec{\boldsymbol{m}}_i \in \mathbb{R}^{3\times N_{\text{atom}}\times L}$, noting that the parameter $L$ stands for the number of grid points for each atom per axis, and its value is obtained by:

$$r_{\mathcal{G}} = res/(\pi \cdot v), \quad s_{\mathcal{G}} = 4 \cdot r_{\mathcal{G}}, \quad L = 2 \cdot s_{\mathcal{G}}$$

The key idea here is to construct a Gaussian sphere for each atom, with $r_{\mathcal{G}}$ and $s_{\mathcal{G}}$ respectively meaning Gaussian radius and the radius of a Gaussian sphere. We expect that $\vec{\boldsymbol{m}}_i$ turns out to be a box as small as possible with all-atom coordinates $\vec{\boldsymbol{x}}_i$ inside it as well as a margin of Gaussian sphere around it, so:

$$\vec{m}_{i_{\text{min}}} = [\lfloor \min(\boldsymbol{x}_i^0) \rfloor, \lfloor \min(\boldsymbol{x}_i^1) \rfloor, \lfloor \min(\boldsymbol{x}_i^2) \rfloor], \quad \vec{m}_{i_{\text{max}}} = [\lceil \max(\boldsymbol{x}_i^0) \rceil, \lceil \max(\boldsymbol{x}_i^1) \rceil, \lceil \max(\boldsymbol{x}_i^2) \rceil]$$
$$\vec{m}_{i\_\text{size}} = \lceil (\vec{m}_{i_{\text{max}}} - \vec{m}_{i_{\text{min}}})/v + 2 \cdot s_{\mathcal{G}} + 1 \rceil, \quad \vec{m}_{i_{\text{center}}} = (\vec{m}_{i_{\text{min}}} + \vec{m}_{i_{\text{max}}})/(2 \cdot v)$$
$$\vec{m}_{i_{\text{min}}} = \lfloor (\vec{m}_{i_{\text{center}}} - \vec{m}_{i\_\text{size}})/2 \rfloor, \quad \vec{m}_{i_{\text{max}}} = \lfloor (\vec{m}_{i_{\text{center}}} + \vec{m}_{i\_\text{size}})/2 \rfloor$$

For each atom per axis(for example the x-axis), we can get $L$ density values around the original atom coordinate by applying Gaussian distribution:

$$\text{for } n \in (1, 2, ..., N_{\text{atom}}) \quad \boldsymbol{\rho}_\lambda^n = w_n \cdot \exp(-\frac{(x_i^0)^n + \boldsymbol{l}}{2}), \boldsymbol{l} = [-s_{\mathcal{G}}, ..., 0, -s_{\mathcal{G}}], \tag{4}$$

where $w_n$ signifies the atomic weight of the corresponding atom. To combine density values on all directions, each atom will end up with $L^3$ density values:

$$\vec{\boldsymbol{\rho}}^n = \boldsymbol{\rho}_\lambda^n \times \boldsymbol{\rho}_\mu^{n\top} \times \boldsymbol{\rho}_\mu^{n\top} \tag{5}$$

### 3.2.2 GEOMETRY LOSS

Geometry loss functions are employed to guarantee the stereochemical accuracy and structural validity of predicted proteins by enforcing conformity with established biological restraints. However, previous work has failed to incorporate geometry-related losses such as those for Ramachandran plot, rotamer, and C-beta deviation. Therefore, we innovatively implemented these differentiable geometry loss functions (ramachandran loss $\mathcal{L}_{\text{rama}}$, rotamer loss $\mathcal{L}_{\text{rot}}$, $C_\beta$ deviation loss) in this work.

**Ramachandran loss** intends to exert ramachandran plot restraints on the predicted protein structures. When computing the Ramachandran loss, all the backbone dihedral angles $\phi$ and $\psi$ will be calculated and evaluated against the Ramachandran criteria retrieved from the Top8000 dataset used by MolProbity(Hintze et al., 2016):

$$\mathcal{L}_{\text{rama}} = \sum_{i=1}^{N_{\text{res}}} \mathbf{F}_{\phi,\psi}(\vec{\boldsymbol{x}}_{i-1}, \vec{\boldsymbol{x}}_i, \vec{\boldsymbol{x}}_{i+1}, a_{i-1}, a_i, a_{i+1}), \tag{6}$$

where $a_i$ denotes the $i^{\text{th}}$ residue's amino acid type and $\mathbf{F}_{\phi,\psi}$ is an indicator function (Appendix D.1) that evaluates whether the backbone tripeptide dihedral angles $\phi_i$ and $\psi_i$ will fall into the outlier region of the Ramachandran plot.

**Rotamer loss** intends to introduce side-chain rotamer-specific restraints:

$$\mathcal{L}_{\text{rot}} = \sum_{i=1}^{N_{\text{res}}} \mathbf{F}_\chi(\chi_i, a_i), \tag{7}$$

where $a_i$ denotes the $i^{\text{th}}$ residue's amino acid type and $\mathbf{F}_\chi$ is an indicator function (Appendix D.1) that evaluates whether the side-chain structure of the $i^{\text{th}}$ residue (determined by the torsion angles $\chi_i \in (-\pi, \pi]^4$) is an outlier compared to idealized protein secondary-structure fragments from the Top8000 dataset.

The necessity to introduce $C_\beta$ **deviation loss** stems from the importance of $C_\beta$ atoms, which are side-chain carbon atoms bonded to $C_\alpha$ atoms in each amino acid (except glycine). $C_\beta$ atoms are crucial for describing side-chain orientations, indicating potential incompatibility between protein backbones and side-chains. When the actual position of $C_\beta$ atom diverges from the ideal position by over $0.25\mathring{A}$, it is thus considered as a deviation:

$$\mathcal{L}_{\text{C}_\beta} = \sum_{i=1}^{N_{\text{res}}} \mathbf{1}(|\vec{\boldsymbol{x}}_{c_\beta} - \vec{\boldsymbol{x}}'_{c_\beta}| > 0.25), \tag{8}$$

where $\boldsymbol{x}_{c_\beta} \in \mathbb{R}^3$ is the coordinates of predicted $C_\beta$ atoms while $\boldsymbol{x}'_{c_\beta} \in \mathbb{R}^3$ is the idealized $C_\beta$ position computed from positions of other 3 atoms on the same amino acid, i.e. $C_\alpha$, backbone $C$ and backbone $N$.

# 4 EXPERIMENTS

## 4.1 BENCHMARK

We curated a benchmark of 63 complexes (53 protein, 10 DNA/RNA–protein) with cryo-EM density map resolutions ranging from 2.0 Åto 4.5 Å. For each case, initial atomic models were predicted by *AlphaFold3* and subsequently refined against the experimental density map by *Phenix.real_space_refine* and *CryoNet.Refine*. Dataset statistics are provided in Figure 9 and Appendix B.

## 4.2 REFINEMENT OF PROTEIN COMPLEXES

We benchmarked *CryoNet.Refine* against the conventional *Phenix.real_space_refine*. As summarized in Table 1, *CryoNet.Refine* consistently achieves superior scores in both model–map correlation coefficients (CC) and model geometric metrics.

Table 1: Performance on protein complex (↑: higher is better, ↓: lower is better).

| Category | Metrics | *AlphaFold3* | *Phenix.real_space_refine* | *CryoNet.Refine* |
|---|---|---|---|---|
| Model-map Correlation Coefficients | $CC_{mask}$ ↑ | 0.44 | 0.61 | **0.65** |
| | $CC_{box}$ ↑ | 0.45 | 0.58 | **0.60** |
| | $CC_{mc}$ ↑ | 0.45 | 0.62 | **0.65** |
| | $CC_{sc}$ ↑ | 0.43 | 0.60 | **0.63** |
| | $CC_{peaks}$ ↑ | 0.31 | 0.48 | **0.52** |
| | $CC_{volume}$ ↑ | 0.47 | 0.61 | **0.65** |
| Model Geometric Metrics | Angle RMSD (degree)↓ | 1.56 | 0.71 | **0.48** |
| | $C_\beta$ deviations↓ | 0.05 | **0.00** | **0.00** |
| | Ramachandran favored (%)↑ | 95.34 | 96.19 | **98.82** |
| | Ramachandran outlier (%)↓ | 0.89 | **0.03** | 0.07 |
| | Rotamer favored (%)↑ | 97.08 | 83.26 | **98.65** |
| | Rotamer outlier (%)↓ | 1.05 | 1.38 | **0.48** |

Figure 3 and Figure 4 show the performance of *CryoNet.Refine* across six model-map correlation coefficients (CC) and model geometric metrics, respectively. Notably, the largest gains in $CC_{mask}$ and $CC_{mc}$ suggest the refined atomic model more accurate main-chain placement within the density map. Regarding geometry, our method consistently achieves superior stereochemistry, virtually eliminating $C_\beta$ deviations, raising ramachandran favored to nearly 99%, and reducing rotamer outliers by over 55% compared to *Phenix.real_space_refine*. Furthermore, angle RMSD is drastically reduced from 1.56° to 0.48°. These results emphatically highlight that *CryoNet.Refine* simultaneously improve the stereochemical geometry and the agreement to the cryo-EM density map.

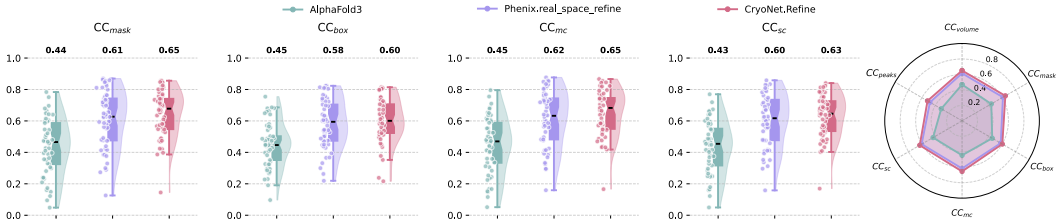



Figure 3: Model–map correlation coefficients on protein complex benchmark.

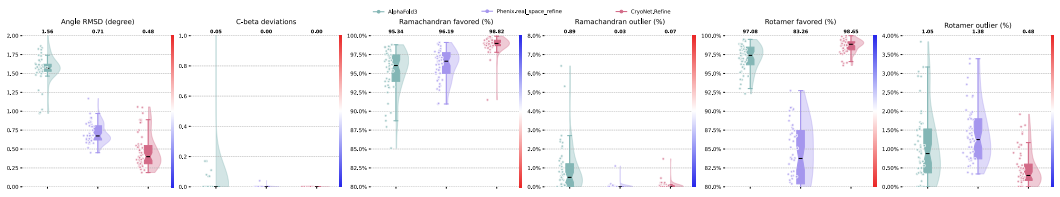



Figure 4: Model geometric metrics. (Color gradient: blue for better, red for worse)

Figure 5 shows that *CryoNet.Refine* outperforms both *AlphaFold3* and *Phenix.real_space_refine* in model-map correlation coefficients and model geometric metrics.

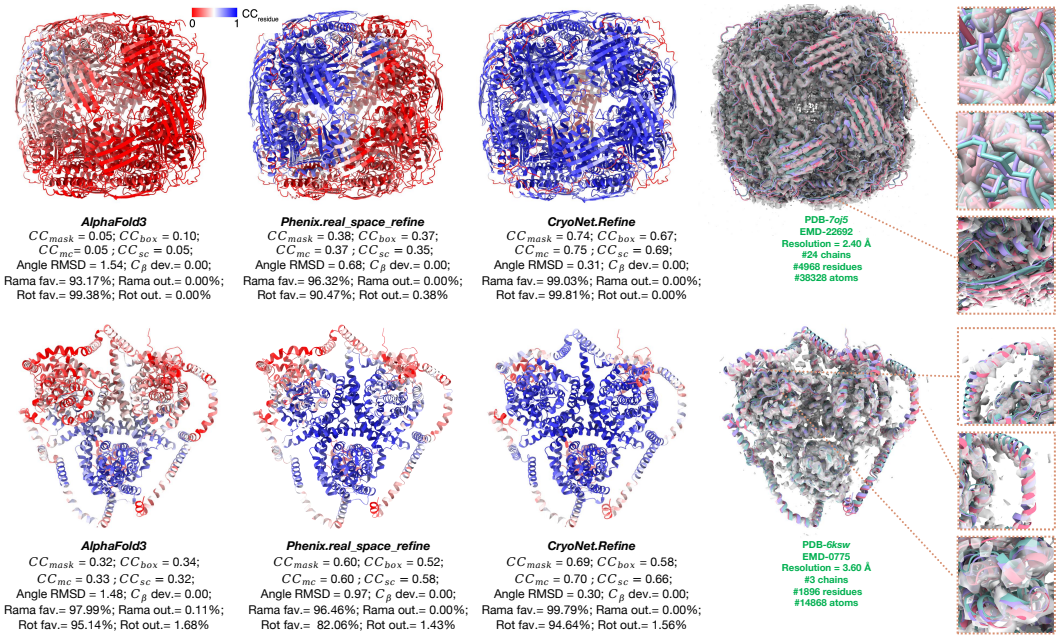



Figure 5: The input atomic models from *AlphaFold3*, the refined atomic model from *Phenix.real_space_refine* and *CryoNet.Refine* on the Medicago truncatula HISN5 protein (PDB-*7oj5*; EMD-22692) and the human concentrative nucleoside transporter CNT3 (PDB-*6ksw*, EMD-0775) complex. Inserts in the right panel show that the main-chains and side-chains generated from *CryoNet.Refine* model align well with the density map.

### 4.3 Refinement on DNA/RNA–Protein Complexes

We next evaluated *CryoNet.Refine* on DNA/RNA–protein complex. As our current implementation doesn't incorporate nucleic acid–specific stereochemical restraints, the assessment exclusively focused on model–map correlation coefficients (CC) as shown in Table 2. Across all metrics, *CryoNet.Refine* consistently outperforms both *AlphaFold3* and *Phenix.real_space_refine*, demonstrating substantially improved agreement with experimental densities.

Table 2: Performance on DNA/RNA–protein complexes.

| Category | Metrics | *AlphaFold3* | *Phenix.real_space_refine* | *CryoNet.Refine* |
|---|---|---|---|---|
| Model–map Correlation Coefficients | $\mathrm{CC}_{\mathrm{mask}}$ ↑ | 0.40 | 0.57 | **0.65** |
| | $\mathrm{CC}_{\mathrm{box}}$ ↑ | 0.49 | 0.61 | **0.67** |
| | $\mathrm{CC}_{\mathrm{mc}}$ ↑ | 0.45 | 0.62 | **0.61** |
| | $\mathrm{CC}_{\mathrm{sc}}$ ↑ | 0.42 | 0.58 | **0.67** |
| | $\mathrm{CC}_{\mathrm{peaks}}$ ↑ | 0.35 | 0.51 | **0.60** |
| | $\mathrm{CC}_{\mathrm{volume}}$ ↑ | 0.48 | 0.61 | **0.69** |

For DNA/RNA–protein complex, we show model–map correlation coefficients in Figure 6. It can be observed that *CryoNet.Refine* achieved substantial improvements in both $\mathrm{CC}_{\mathrm{mask}}$ and $\mathrm{CC}_{\mathrm{sc}}$ , strongly indicating superior performance refinement in both main-chain and side-chain. A representative case study is shown in Figure 7, where *AlphaFold3* and *Phenix.real_space_refine* achieve moderate density fitting ($\mathrm{CC}_{\mathrm{mask}} = 0.18$ and 0.36, respectively), whereas *CryoNet.Refine* attains a markedly higher $\mathrm{CC}_{\mathrm{mask}} = 0.72$, along with consistent gains across all other CC metrics.

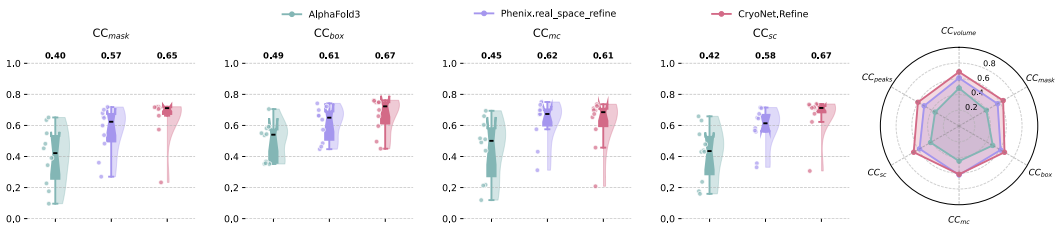



Figure 6: Model–map correlation coefficients on DNA/RNA-protein complex.

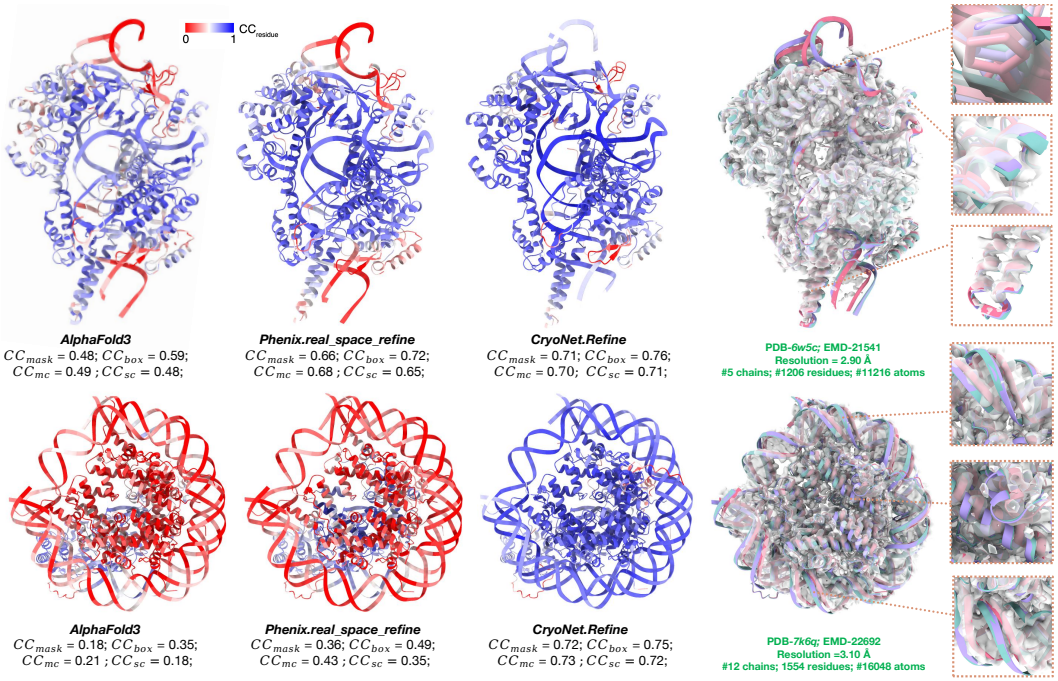



Figure 7: The input atomic models from *AlphaFold3*, the refined atomic models from *Phenix.real_space_refine* and *CryoNet.Refine* are shown on the Cas12i(E894A)–crRNA–dsDNA (PDB-*6W5C*, EMD-21541) and the active-state Dot1 bound to the H4K16ac nucleosome (PDB-*7K6Q*, EMD-22692) complex. Inserts in the right panel show that the main-chains and side-chains generated by *CryoNet.Refine* align well with the density map.

## 5 Ablation Study

Results of the ablation study can be found in Appendix E.3. The full *CryoNet.Refine* configuration achieves the best balance between model–map correlation coefficients and model geometric metrics. Removing individual loss terms leads to distinct degradations in correlation metrics or geometric metrics, confirming the complementary roles of density and geometry loss.

### 5.1 Runtime Performance

We benchmarked the runtime of *CryoNet.Refine* against *Phenix.real_space_refine*. Leveraging GPU parallelization, *CryoNet.Refine* consistently achieves highly efficient performance, whereas *Phenix.real_space_refine*ś CPU-only support incurs higher computational costs, particularly for large complexes.

Across 63 complexes, *CryoNet.Refine* ran faster than *Phenix.real_space_refine* in 40 cases (63%). These results underscore that our method combines superior accuracy with high efficiency, making it ideal for large-scale, high-throughput cryo-EM refinement.

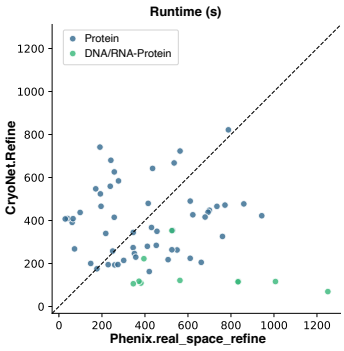



Figure 8: Runtime comparison.

## 6 Conclusion

In this work, we present *CryoNet.Refine*, a pioneering one-step diffusion module for cryo-EM atomic model refinement. It critically featuring a novel differentiable density loss and a comprehensive geometry metric loss, not only achieves unparalleled refinement but also opens new avenues: the density loss can be broadly applied to AI-based atomic model building , while the geometry loss offers a powerful guidance of standard geometry for protein structure prediction. *CryoNet.Refine* thus marks a significant advance for the cryo-EM field. Looking ahead, we aim to extend the specialized geometry loss for DNA/RNA structures, while further enriching the geometry loss family by incorporating a steric clash loss, and demonstrate superior performance on challenging low-resolution density maps, ultimately contributing more profoundly to the entire cryo-EM community.

## REFERENCES


Josh Abramson, Jonas Adler, Jack Dunger, Richard Evans, Tim Green, Alexander Pritzel, Olaf Ronneberger, Lindsay Willmore, Andrew J Ballard, Joshua Bambrick, et al. Accurate structure prediction of biomolecular interactions with alphafold 3. *Nature*, 630(8016):493–500, 2024.

Pavel V. Afonine, Ralf W. Grosse-Kunstleve, Nathaniel Echols, Jeffrey J. Headd, Nigel W. Moriarty, Marat Mustyakimov, Thomas C. Terwilliger, Alexandre Urzhumtsev, Peter H. Zwart, and Paul D. Adams. Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallographica*, 68(4):352–367, 2012.

Pavel V Afonine, Bruno P Klaholz, Nigel W Moriarty, Billy K Poon, Oleg V Sobolev, Thomas C Terwilliger, Paul D Adams, and Alexandre Urzhumtsev. New tools for the analysis and validation of cryo-em maps and atomic models. *Biological Crystallography*, 74(9):814–840, 2018a.

Pavel V. Afonine, Billy K. Poon, Randy J. Read, Oleg V. Sobolev, and Paul D. Adams. Real-space refinement in phenix for cryo-em and crystallography. *Acta Crystallographica Section D Structural Biology*, 74(6), 2018b.

Gustaf Ahdritz, Nazim Bouatta, Christina Floristean, Sachin Kadyan, Qinghui Xia, William Gerecke, Timothy J O'Donnell, Daniel Berenberg, Ian Fisk, Niccolò Zanichelli, et al. Openfold: Retraining alphafold2 yields new insights into its learning mechanisms and capacity for generalization. *Nature methods*, 21(8):1514–1524, 2024.

M. Baek, I. Anishchenko, Ian R. Humphreys, Q. Cong, D. Baker, and F. Dimaio. Efficient and accurate prediction of protein structure using rosettafold2. *bioRxiv*, 2023.

Minkyung Baek, Frank DiMaio, Ivan Anishchenko, Justas Dauparas, Sergey Ovchinnikov, Gyu Rie Lee, Jue Wang, Qian Cong, Lisa N. Kinch, R. Dustin Schaeffer, Claudia Millán, Hahnbeom Park, Carson Adams, Caleb R. Glassman, Andy DeGiovanni, Jose H. Pereira, Andria V. Rodrigues, Alberdina A. van Dijk, Ana C. Ebrecht, Diederik J. Opperman, Theo Sagmeister, Christoph Buhlheller, Tea Pavkov-Keller, Manoj K. Rathinaswamy, Udit Dalwadi, Calvin K. Yip, John E. Burke, K. Christopher Garcia, Nick V. Grishin, Paul D. Adams, Randy J. Read, and David Baker. Accurate prediction of protein structures and interactions using a three-track neural network. *Science*, 373(6557):871–876, 2021. doi: 10.1126/science.abj8754. URL https://www.science.org/doi/abs/10.1126/science.abj8754.

RCSB Protein Data Bank. RCSB PDB - 6KSW: Cryo-EM structure of the human concentrative nucleoside transporter CNT3 — rcsb.org. https://www.rcsb.org/structure/6KSW. [Accessed 25-09-2025].

Tristan Ian Croll. Isolde: a physically realistic environment for model building into low-resolution electron-density maps. *Biological Crystallography*, 74(6):519–530, 2018.

Paul Emsley and Kevin Cowtan. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr*, 60(12-1):2126–2132, 2004.

Jose Ignacio Garzon, Julio Kovacs, Ruben Abagyan, and Pablo Chacon. Adp_em: fast exhaustive multi-resolution docking for high-throughput coverage. *Bioinformatics*, 23(4):427–433, 2007.

Murray Goodman, Antonio S Verdini, Nam S Choi, and Yukio Masuda. Polypeptide stereochemistry. *Top. Stereochem*, 5:69–166, 1970.

Bradley J Hintze, Steven M Lewis, Jane S Richardson, and David C Richardson. Molprobity's ultimate rotamer-library distributions for model validation. *Proteins: Structure, Function, and Bioinformatics*, 84(9):1177–1189, 2016.

Naozumi Hiranuma, Hahnbeom Park, Minkyung Baek, Ivan Anishchenko, and David Baker. Improved protein structure refinement guided by deep learning based accuracy estimation. *Nature Communications*, 12(1), 2021.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems (NeurIPS)*, 2020. URL https://arxiv.org/abs/2006.11239.

John B. Ingraham, Max Baranov, Zak Costello, Karl W. Barber, Wujie Wang, Ahmed Ismail, Vincent Frappier, Dana M. Lord, Christopher Ng-Thow-Hing, and Erik R. Van Vlack. Illuminating protein space with a programmable generative model. *Nature*, 623(7989):15, 2023.

Kiarash Jamali, Lukas Kall, Rui Zhang, Alan Brown, Dari Kimanius, and Sjors Scheres. Automated model building and protein identification in cryo-em maps. *Nature*, 2024. doi: 10.1038/s41586-024-07215-4. URL https://www.nature.com/articles/s41586-024-07215-4.

Xiaoyang Jing and Jinbo Xu. Fast and effective protein model refinement using deep graph neural networks. *Nature Computational Science*, 1(7):462–469, 2021. doi: 10.1038/s43588-021-00098-9. URL https://doi.org/10.1038/s43588-021-00098-9.

John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, et al. Highly accurate protein structure prediction with alphafold. *nature*, 596(7873):583–589, 2021.

Tero Karras, Miika Aittala, Samuli Laine, Erik Härkönen, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Elucidating the design space of diffusion-based generative models. *Advances in Neural Information Processing Systems (NeurIPS)*, 2022. URL https://arxiv.org/abs/2206.00364.

Werner Kühlbrandt. The resolution revolution. *Science*, 343(6178):1443–1444, 2014. doi: 10.1126/science.1251652. URL https://www.science.org/doi/abs/10.1126/science.1251652.

Zeming Lin, Halil Akin, Alexander Rives, and arives@meta.com. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 379(6637):8, 2023.

Chenlin Meng, Yutong He, Yang Song, Jiaming Song, Jiahui Wu, and Stefano Ermon. On distillation of guided diffusion models. *Advances in Neural Information Processing Systems (NeurIPS)*, 2022. URL https://arxiv.org/abs/2210.03142.

Nogales and Eva. The development of cryo-em into a mainstream structural biology technique. *Nature Methods*, 13(1):24–27, 2016.

Saro Passaro, Gabriele Corso, Jeremy Wohlwend, Mateo Reveiz, Stephan Thaler, Vignesh Ram Somnath, Noah Getz, Tally Portnoi, Julien Roy, Hannes Stark, David Kwabi-Addo, Dominique Beaini, Tommi Jaakkola, and Regina Barzilay. Boltz-2: Towards accurate and efficient binding affinity prediction. *bioRxiv*, 2025. doi: 10.1101/2025.06.14.659707. URL https://www.biorxiv.org/content/early/2025/06/18/2025.06.14.659707.

Eric F. Pettersen, Thomas D. Goddard, Conrad C. Huang, Elaine C. Meng, Gregory S. Couch, Tristan I. Croll, John H. Morris, and Thomas E. Ferrin. Ucsf chimerax: Structure visualization for researchers, educators, and developers. *Protein Science*, 30(1):70–82, 2021. doi: https://doi.org/10.1002/pro.3943. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/pro.3943.

Jonas Pfab, Nhut Minh Phan, and Dong Si. Deeptracer for fast de novo cryo-em protein structure modeling and special studies on cov-related complexes. *Proceedings of the National Academy of Sciences*, 118(2):e2017525118, 2021.

Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. *International Conference on Learning Representations (ICLR)*, 2021. URL https://arxiv.org/abs/2011.13456.

Yang Song, Chenlin Meng, and Stefano Ermon. Consistency models. *Proceedings of the 40th International Conference on Machine Learning (ICML)*, 2023. URL https://arxiv.org/abs/2303.01469.

Ray Yu-Ruei Wang, Yifan Song, Benjamin A Barad, Yifan Cheng, James S Fraser, and Frank DiMaio. Automated structure refinement of macromolecular assemblies from cryo-em maps using rosetta. *elife*, 5:e17219, 2016.

Tongtong Wang, Zhangqiang Li, Kui Xu, Wenze Huang, Gaoxingyu Huang, Qiangfeng Cliff Zhang, and Nieng Yan. Cryoseek: A strategy for bioentity discovery using cryoelectron microscopy. *Proceedings of the National Academy of Sciences*, 121(42):e2417046121, 2024. doi: 10.1073/pnas.2417046121. URL https://www.pnas.org/doi/abs/10.1073/pnas.2417046121.

Tongtong Wang, Wenze Huang, Kui Xu, Yitong Sun, Qiangfeng Cliff Zhang, Chuangye Yan, Zhangqiang Li, and Nieng Yan. Cryoseek ii: Cryo-em analysis of glycofibrils from freshwater reveals well-structured glycans coating linear tetrapeptide repeats. *Proceedings of the National Academy of Sciences*, 122(1):e2423943122, 2025. doi: 10.1073/pnas.2423943122. URL https://www.pnas.org/doi/abs/10.1073/pnas.2423943122.

Joseph L. Watson, David Juergens, Nathaniel R. Bennett, Helen E. Eisenach, Woody Ahern, Andrew J. Borst, Robert J. Ragotte, Lukas F. Milles, Basile I. M. Wicky, and Nikita Hanikel. De novo design of protein structure and function with rfdiffusion. *Nature*, 620(7976):38, 2023.

Jeremy Wohlwend, Gabriele Corso, Saro Passaro, Mateo Reveiz, Ken Leidal, Wojtek Swiderski, Tally Portnoi, Itamar Chinn, Jacob Silterra, Tommi Jaakkola, and Regina Barzilay. Boltz-1 democratizing biomolecular interaction modeling. *bioRxiv*, 2025. doi: 10.1101/2024.11.19.624167. URL https://www.biorxiv.org/content/early/2025/05/06/2024.11.19.624167.

Tianqi Wu, Zhiye Guo, and Jianlin Cheng. Atomic protein structure refinement using all-atom graph representations and se(3)-equivariant graph transformer. *Bioinformatics*, 39(5):btad298, 05 2023. ISSN 1367-4811. doi: 10.1093/bioinformatics/btad298. URL https://doi.org/10.1093/bioinformatics/btad298.

Xue Wu, Jingwei Xin, Zhijun Tu, Jie Hu, Jie Li, Nannan Wang, and Xinbo Gao. One-step diffusion-based real-world image super-resolution with visual perception distillation. 2025. URL https://arxiv.org/abs/2506.02605.

Kui Xu, Zhe Wang, Jianping Shi, Hongsheng Li, and Qiangfeng Cliff Zhang. A2-net: Molecular structure estimation from cryo-em density volumes. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(01):1230–1237, Jul. 2019. doi: 10.1609/aaai.v33i01.33011230. URL https://ojs.aaai.org/index.php/AAAI/article/view/3918.

# A NOTATION

| | |
|---|---|
| | **Framework** |
| $d_0$ | Input cryo-EM density map |
| $x_0$ | All-atom coordinates of input structure |
| $s$ | Sequence Embedding |
| $z$ | Pairwise Representation |
| $x_i (i = 1, \cdots, n)$ | Refined atomic model after the $i$ recycle |
| $d_i (i = 1, \cdots, n)$ | Synthetic density map of refined atomic model $x_i$ |
| $\gamma_t$ | Weight for a loss function term $\mathcal{L}_t$ |
| | **Diffusion Module** |
| $\xi_0, \xi_{min}$ | Noise scaling hyperparameter |
| $\sigma$ | Noise scale parameter in diffusion refinement |
| $\sigma_{\text{data}}$ | Data-dependent scale constant (fixed to 16), estimated from training distribution |
| $c_{\text{skip}}(\sigma)$ | Skip coefficient in preconditioning |
| $c_{\text{out}}(\sigma)$ | Output scaling coefficient in preconditioning |
| $c_{\text{in}}(\sigma)$ | Input scaling coefficient in preconditioning |
| $c_{\text{noise}}(\sigma)$ | Noise embedding term in preconditioning, |
| $\mathcal{C}$ | Auxiliary conditioning features derived from $s$, $z$, and encoded structural features |
| | **Density Loss** |
| $w$ | Atomic weight |
| $res$ | Resolution of the input density map |
| $v$ | Voxel size of the input density map |
| $r_{\mathcal{G}}$ | Gaussian radius |
| $s_{\mathcal{G}}$ | Radius of the virtual Gaussian sphere |
| $L$ | Number of grid points for each atom per axis |
| $\vec{x} = [x^0, x^1, x^2]$ | All-atom coordinates of refined atomic models |
| $d = [\vec{m}, \rho]$ | Synthetic density map |
| $\vec{m} = [\lambda, \mu, \nu]$ | Coordinates to locate inside density maps |
| $\rho$ | density values |
| | **Geometry Loss** |
| $a$ | Amino acid type |
| $N_{\text{res}}, N_{\text{atom}}, N_{\text{bond}}$ | Number of residues, atoms and bonds in a structure |
| $\theta$ | Bond angles |
| $\chi$ | Torsion angles |

## B DATASET

Figure 9 summarizes the dataset statistics. The two panels respectively show the number of chains and the number of residues with respect to map resolution. Protein-only complexes display broader diversity in both chain counts and residue numbers, whereas DNA/RNA–protein assemblies tend to be smaller but fall within comparable resolution ranges.

For all targets, input sequences were retrieved from the RCSB Protein Data Bank (PDB). Since *AlphaFold3* imposes a practical sequence length limit of $\leq$5000 amino acids, we adopted a chain-wise strategy: sequences were segmented at the chain level and truncated if necessary, ensuring that the combined length per case did not exceed 5000 residues. This preprocessing step guarantees compatibility with *AlphaFold3* while preserving the completeness of structural contexts across chains.

For residues with backbone correlation coefficients below 0.1, we applied the ADP-EM local refinement tool (Garzon et al., 2007) to pre-process the models and improve rigid-body docking.

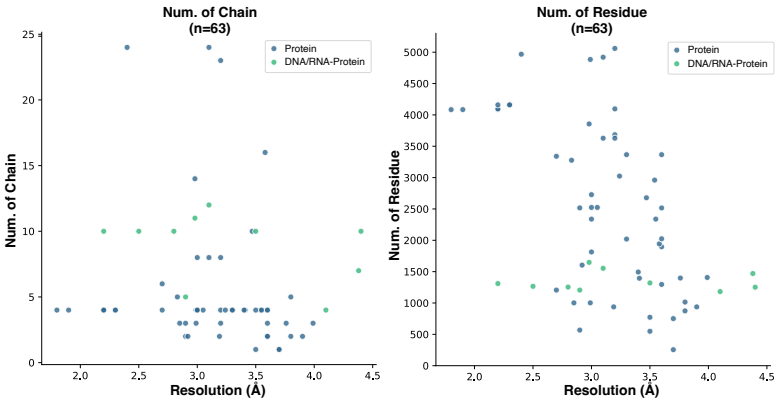



Figure 9: Dataset statistics of density map resolution, number of chains, and number of residues.

## C ONE-STEP DIFFUSION MODULE

Classical diffusion models define a forward noising process and a multi-step reverse denoising chain (Ho et al., 2020; Song et al., 2021). While powerful, such iterative schemes incur substantial computational cost. Inspired by accelerated sampling approaches such as consistency models (Song et al., 2023) and diffusion distillation

Unlike conventional diffusion that begins from Gaussian noise, we initialize directly from the the starting structure $\mathbf{x}_0$.

Refinement is then performed through a **preconditioned forward module**(Karras et al., 2022):

$$\mathbf{x}_{i+1} = c_{\text{skip}}(\sigma)\,\mathbf{x}_i \;+\; c_{\text{out}}(\sigma)\,\mathcal{F}_\theta\big(c_{\text{in}}(\sigma)\mathbf{x}_i,\; c_{\text{noise}}(\sigma),\; \mathcal{C}\big), \tag{9}$$

with auxiliary conditioning $\mathcal{C}$. $\mathcal{F}_\theta$ is a neural network responsible for the scoring mechanism. The other coefficients include:

$$c_{\text{skip}}(\sigma) = \frac{\sigma_{\text{data}}^2}{\sigma^2 + \sigma_{\text{data}}^2}, \qquad c_{\text{out}}(\sigma) = \frac{\sigma\,\sigma_{\text{data}}}{\sqrt{\sigma^2 + \sigma_{\text{data}}^2}}, \tag{10}$$

$$c_{\text{in}}(\sigma) = \frac{1}{\sqrt{\sigma^2 + \sigma_{\text{data}}^2}}, \qquad c_{\text{noise}}(\sigma) = \tfrac{1}{4}\log\Big(\tfrac{\sigma}{\sigma_{\text{data}}}\Big). \tag{11}$$

The refined structure after the $i^{th}$ recycle is obtained in a single deterministic step:

$$\mathbf{x}_{\text{i+1}} = \mathbf{x}_i. \tag{12}$$

**Algorithm 1** One-step deterministic refinement in *CryoNet.Refine*

1: $\xi_0, \xi_{\min}, \sigma$ : hyperparameter
2: $n$: number of recycles
3: **Input:** all-atom coordinates $\mathbf{x}_0$, conditioning $\mathcal{C}$
4: **for** $i$ **in** $(1, \cdots, n)$
5: $\xi \leftarrow \xi_0$ if $\sigma > \xi_{\min}$ else 0
6: $\hat{t} \leftarrow \sigma(1 + \xi)$
7: $\mathbf{x}_{i+1} \leftarrow$ *preconditionedForward*$(\mathbf{x}_i, \hat{t}, \mathcal{C})$ (see Eq. 9)
8: **return** $\{\mathbf{x}_n\}$

Compared with *AlphaFold3*, which employs multi-step stochastic denoising from Gaussian initialization (Abramson et al., 2024), our formulation collapses the diffusion chain into a single deterministic refinement. This yields (i) efficiency by eliminating iterative sampling, (ii) stability via input–output preconditioning, and (iii) broad applicability to proteins, DNA/RNAs and complexes.

# D Loss Functions

After collective parameter tuning, the weight values for each component of our loss function are listed in Table 3 .

Table 3: Loss function weights

| Loss Functions | Weight | Value |
|---|---|---|
| Density loss ($\mathcal{L}_{\text{den}}$) | $\gamma_{\text{den}}$ | 20.0 |
| Ramachandran loss ($\mathcal{L}_{\text{rama}}$) | $\gamma_{\text{rama}}$ | 500 |
| Rotation loss ($\mathcal{L}_{\text{rot}}$) | $\gamma_{\text{rot}}$ | 500 |
| Angle loss ($\mathcal{L}_{\text{angle}}$) | $\gamma_{\text{angle}}$ | 2 |
| $C_\beta$ loss ($\mathcal{L}_{C_\beta}$) | $\gamma_{C_\beta}$ | 50 |
| Violation loss ($\mathcal{L}_{\text{viol}}$) | $\gamma_{\text{viol}}$ | 1000 |

## D.1 Ramachandran and Rotamer loss

The Ramachandran plot (Goodman et al., 1970) is a 2D graph with $\phi$ angles and $\psi$ angles as x- and y-axis. By plotting the combination of these torsion angles of all residues, it provides insights into whether the conformation is sterically favored, allowed or outliered.
In order to obtain more updated and more representative protein structures for reference, MolProbity(Hintze et al., 2016), the widely used validation program for protein structures, curated the Top8000 dataset by filtering and deduplicating high-quality protein structures from the PDB Bank. The release of the Top8000 dataset came with their 3-dimensional Ramachandran grids plotting the ideal distribution for different kinds of amino acids,[1].The z-axis values act as density scores measuring the frequency for a certain combination of Ramachandran angles.
Through interpolation of the nearest grid values, we can get the density score for any combination of arbitrary Ramachandran scores and classify them as falling within favored, allowed, or outlier regions, as shown in Algorithm 2
As for $\mathcal{F}_\chi$ of rotamer loss, the procedure is highly similar to Algorithm 2 except that it requires $\chi$ angles instead of $\phi, \psi$ angles, and it performs 4-dimensional interpolation (since each density value corresponds to the combination of four $\chi$ angles)from the rotamer distribution library of Top8000 dataset.

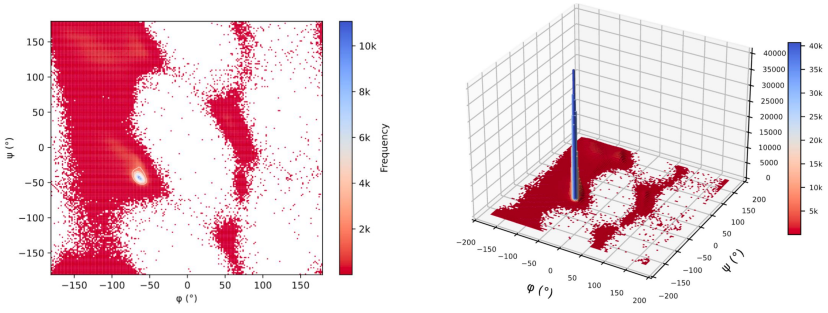
## D.2 Bond Angle loss

**Bond angle loss** applies geometry restraints on chemical bonds and forces them to approach the target values specified by the ideal geometry library. In our implementation, $\mathcal{L}_{\text{angle}}$ is essentially

[1]In fact, there are 6 distinct types of Ramachandran plots for different amino acids, including *Gly*, *Val/Ile*, *pre-Pro*, *trans-Pro*, *cis-Pro* and *Ala*. How to differentiate them and the rationale behind this categorization are beyond the scope of this paper.

**Algorithm 2** Classification of Ramachandran Angle Outliers

$\vec{x}_i$: The i-th residue coordinates
$a_i$: The amino acid type of the i-th residue
$N_{\text{res}}$: total number of residues
$\mathcal{S}_{\text{rama\_type}} = \{\textit{Gly, Val/Ile, pre-Pro, trans-Pro, cis-Pro, Ala}\}$
$r_i \in \mathcal{S}_{\text{rama\_type}}$: categorization Ramachandran angles between $n_{i-1}, n_i, n_{i+1}$,
$\tau_r$: density value threshold for Ramachandran category $r$
$o$: count of Ramachandran outliers
**Input** $\vec{\boldsymbol{x}} = [\vec{x}_1, \vec{x_2}, \cdots, \vec{x}_{N_{\text{res}}}], [a_1, a_2, \cdots, a_{N_{\text{res}}}]$
**Function** *getDensity*$(\phi, \psi, r_i)$:
  $\mathcal{R} \leftarrow$ Ramachandran distribution grids for $\text{r}_i$
  density_score$\leftarrow$ *2DInterpolate*$(\text{R}(\lfloor\phi\rfloor, \lfloor\psi\rfloor), \mathcal{R}(\lceil\phi\rceil\lceil\psi\rceil))$
**return** density_score
**for** $i = 2$ to $N_{\text{res}} - 1$
$\phi_i, \psi_i \leftarrow$ *calculateDihedrals*$(\vec{x}_{i-1}, \vec{x}_i, \vec{x}_{i+1})$
$r_i \leftarrow$ *RamaType*$(\text{a}_{i-1}, a_i, a_{i+1})$
density_score$_i \leftarrow$ *getDensity*$(\phi_i, \psi_i, r_i)$
**if** density_score$_i < \tau_r$ **then** $o \leftarrow$o+1
**return** $o/N_{\text{res}}$




(a) 2D Ramachandran plot (b) 3D Ramachandran distribution

Figure 10: Visualization of Ramachandran distributions

equivalent to the structure's bond angle RMSD value:

$$\mathcal{L}_{\text{anlge}} = \sqrt{\frac{1}{N_{\text{bond}}} \sum_{i=1}^{N_{\text{bond}}} \Delta\theta_i^2}, \tag{13}$$

where $N_{\text{bond}}$ counts the number of all chemical bonds inside the whole protein structure, and $\Delta\theta_i \in (-\pi, \pi]$ calculates the minimal angular difference between the predicted angle value $\theta_i$ and the ideal angle value $\theta_i'$:

$$\Delta\theta_i = \arctan(\sin(\theta_i - \theta_i')/\cos(\theta_i - \theta_i')) \tag{14}$$

To identify chemical bonds given all-atom coordinates, we follow the algorithm as well as ideal bond angle values given by *phenix.pdb_interpretation*(Afonine et al., 2012) and rewrite the implementation using PyTorch.

### D.3 Violation Loss

**Violation loss** penalizes steric clashes between nonbonded atoms to prevent potential clashes. We borrow the OpenFold implementation (Ahdritz et al., 2024) and follow the definition of (Jumper et al., 2021):

$$\mathcal{L}_{\text{viol}} = \sum_{i=1}^{N_{\text{nbpairs}}} \max(s_{lit}^i - \tau - s_{pred}^i, 0), \tag{15}$$

where $s^i_{pred}$ is the distance of two non-bonded atoms in the predicted structure and $s^i_{lit}$ is the "clashing distance" of these two atoms according to their literature Van der Waals radii. $N_{\text{nbpairs}}$ is the number of all nonbonded atom pairs in this structure. The tolerance $\tau$ is set to 1.5 Å.

# E RESULTS

## E.1 RECYCLE STRATEGY

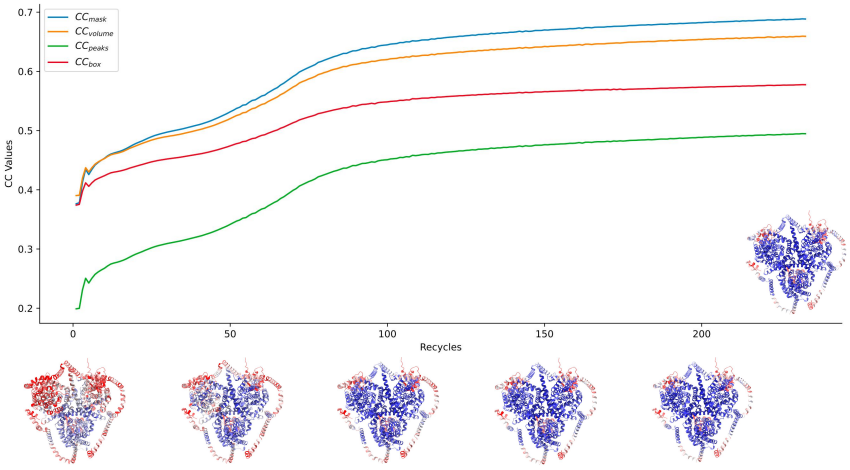



Figure 11: Model-map Correlation Coefficients (CC) analysis of protein complexes across different recycling numbers (We provide a trajectory video over the entire 234 refinement recyclings in supplementary material).

The Recycling strategy of *CryoNet.Refine* is setting 300 iterations as maximum with an early-stop mechanism. To demonstrate the rationale behind this, we picked a protein structure from our dataset (PDB-*6ksw*) and trace its model-map correlation coefficients (CC) values after each recycle iteration, as shown in Figure 11.
These 4 CC values basically follow a similar pattern. The early-stop mechanism effectively spots a good termination where the curves converge after the uphill slope.

## E.2 DIFFUSION SAMPLING STEPS

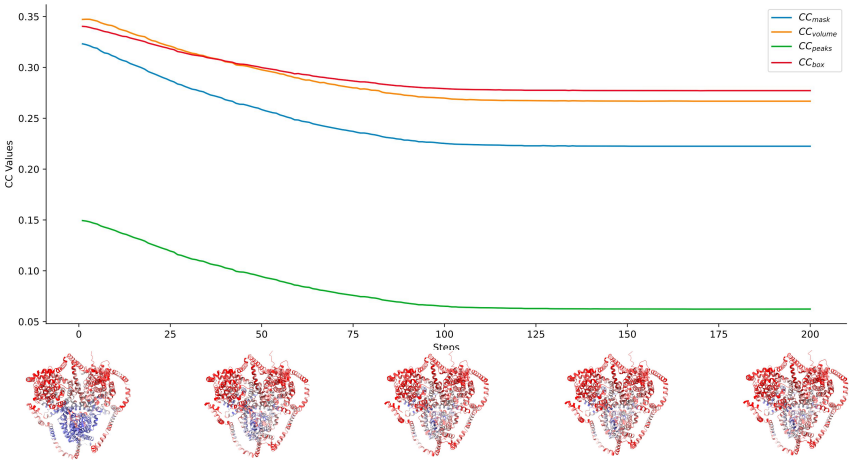



Figure 12: Model-map Correlation Coefficients (CC) analysis of DNA/RNA–protein complex across different diffusion sampling steps.

To show that our one-step diffusion module is an improvement over previous diffusion models with multiple sampling steps, we picked a protein structure (PDB-*6ksw*) and run it with 200 steps. The model-map correlation coefficients (CC) values after each step are shown in Figure 12. Each curve has its optimal value at the beginning step and steadily falls when sampling more steps, which demonstrates the efficacy of our one-step diffusion module.

### E.3 Ablation Study

We performed an ablation study on 27 protein complexes (Table 4). The full *CryoNet.Refine* configuration achieves the best overall performance between model-map correlation coefficients and model geometric metrics, underscoring the necessity of integrating density- and geometry-aware objectives.

Table 4: Ablation study: how loss functions influence model-map correlation coefficients and model geometric metrics

| Category | Metrics | $\gamma_{\text{den}} = 0$ | $\gamma_{\text{rama}} = 0$ | $\gamma_{\text{rot}} = 0$ | *CryoNet.Refine* |
|---|---|---|---|---|---|
| Model-map Correlation Coefficients | $\text{CC}_{\text{mask}} \uparrow$ | 0.41 | **0.65** | 0.64 | **0.65** |
| | $\text{CC}_{\text{box}} \uparrow$ | 0.42 | **0.58** | 0.57 | **0.58** |
| | $\text{CC}_{\text{mc}} \uparrow$ | 0.42 | **0.66** | 0.65 | **0.66** |
| | $\text{CC}_{\text{sc}} \uparrow$ | 0.41 | **0.64** | **0.64** | 0.63 |
| | $\text{CC}_{\text{peaks}} \uparrow$ | 0.24 | **0.48** | 0.47 | 0.47 |
| | $\text{CC}_{\text{volume}} \uparrow$ | 0.43 | **0.65** | 0.64 | **0.65** |
| Model Geometric Metrics | Angle RMSD (°)↓ | 0.45 | **0.41** | 0.44 | 0.54 |
| | C-beta deviations↓ | 0.00 | 0.00 | 0.00 | 0.00 |
| | Ramachandran favored (%)↑ | 99.09 | 90.75 | **99.22** | 98.80 |
| | Ramachandran outlier (%)↓ | 0.06 | 2.27 | **0.03** | 0.10 |
| | Rotamer favored (%)↑ | **98.67** | 98.64 | 94.48 | 98.58 |
| | Rotamer outlier (%)↓ | 0.54 | 2.11 | 1.38 | **0.51** |

Removing the density loss ($\gamma_{\text{den}} = 0$) leads to a pronounced drop across all CC metrics, with $\text{CC}_{\text{mask}}$ and $\text{CC}_{\text{mc}}$ reduced by over 35%, confirming that the density term is indispensable for guiding accurate model–map fitting. In contrast, omitting the Ramachandran prior ($\gamma_{\text{rama}} = 0$) preserves correlation coefficients but severely compromises geometry, reducing favored residues from 98.80% to 90.75% and inflating outliers more than 20-fold. This highlights the critical role of stereochemical priors in constraining backbone conformations.

The effect of removing the rotamer prior ($\gamma_{\text{rot}} = 0$) is more nuanced: CC values remain competitive, and Ramachandran statistics are slightly improved, but side-chain packing deteriorates, with favored rotamers dropping to 94.48%. This indicates that local side-chain restraints complement backbone priors in ensuring chemically realistic conformations.

Together, these results establish three key principles: (i) density loss is essential for achieving high model–map correlation, (ii) stereochemical priors, particularly Ramachandran constraints, safeguard backbone geometry, and (iii) rotamer priors contribute to realistic side-chain packing. The synergy of these components underpins *CryoNet.Refine*'s ability to deliver models that are both density-consistent and stereochemically robust.