STRAND: Structure Refinement of RNA-Protein Complexes via Diffusion

Mohsen Al-zeqri [1] **Jörg K.H. Franke** [1] **Frederic Runge** [1]

Abstract

RNA-protein interactions play crucial roles in cellular processes, from gene regulation to viral replication. While recent advances in structure prediction have revolutionized our ability to model macromolecular complexes, achieving accurate predictions of RNA-protein binding poses remains challenging. In this work, we present *STRAND*, a diffusion-based model for monomeric RNA-protein complex refinement that builds upon the success of *DiffDock-PP* in protein-protein docking. Unlike traditional docking, we develop STRAND as a modular extension to existing RNA-Protein complex prediction tools to improve their backbone predictions. We study the effect of different transformations by training models to learn either translation, rotation, torsion, or combinations of these during the diffusion process and initialize the backward process with a complex prediction at test time. Our experiments with *AlphaFold 3* and *ProRNA3D-single* reveal that STRAND can improve the backbones of a large fraction of RNA-protein complex predictions.

1. Introduction

RNA-protein interactions are fundamental to numerous cellular processes, including transcriptional regulation, splicing, and protein synthesis (Glisovic et al., 2008). Understanding these interactions at the structural level is crucial for deciphering molecular mechanisms and developing therapeutic interventions. While experimental techniques like X-ray crystallography (Smyth & Martin, 2000), nuclear magnetic resonance spectroscopy (Hu et al., 2021), or single-particle electron microscopy (cryo-EM) (Renaud et al., 2018) provide valuable structural insights, they are often time-consuming and resource-intensive.

[1]Department of Computer Science, University of Freiburg, Germany. Correspondence to: Mohsen Al-zeqri <muhsenzaqry@hotmail.com>, Frederic Runge <runget@cs.uni-freiburg.de>.

Proceedings of the Workshop on Generative AI for Biology at the 42nd International Conference on Machine Learning, Vancouver, Canada. PMLR 267, 2025. Copyright 2025 by the author(s).

Recent breakthroughs in deep learning have revolutionized structural biology, particularly with the advent of AlphaFold 3, which enables predictions of RNA-protein complexes alongside other biological macromolecules (Abramson et al., 2024). However, while deep learning methods have driven progress in RNA secondary structure prediction (Sato et al., 2021; Franke et al., 2024) and secondary structure- (Runge et al., 2024; Patil et al., 2024) as well as 3D-based (Tan et al., 2024; Joshi et al., 2025) RNA design, RNA 3D structure prediction remains challenging (Bernard et al., 2025). This challenge is compounded by the inherent flexibility of RNA molecules (Hagerman, 1997; Franke et al., 2022) and the diverse nature of RNA-protein binding interfaces (Re et al., 2014).

The recent success of diffusion models in protein-protein docking, particularly DiffDock-PP (Ketata et al., 2023), has demonstrated the potential of these approaches in modeling molecular interactions. However, these methods are specifically designed for protein-protein interfaces and do not account for the distinct characteristics of RNA molecules.

In this work, we present STRAND (**ST**ructure **R**efinement of RN**A**-protei**N** complexes via **D**iffusion), a novel diffusion-based model that extends the capabilities of DiffDock-PP to monomeric RNA-protein complexes. Our approach leverages an RNA foundation model, RNA-FM (Chen et al., 2022), to generate RNA sequence embeddings that capture the unique properties of RNA molecules. We then combine these RNA-specific features with the powerful diffusion framework of DiffDock-PP. We train different models, applying noise either on the translation level, the rotation level, to torsion angles, or combinations of these across thousands of experimentally validated RNA-Protein complexes from the Protein Data Bank (PDB) (Burley et al., 2017). Then, at test time, we use RNA-Protein complex predictions from AlphaFold 3 (Abramson et al., 2024) or ProRNA3D-single (Roche et al., 2024) as the initial complex structure for the denoising process to refine the initial predictions. With this approach, STRAND is capable of improving many initial predictions, as indicated by improved complex RMSD (cRMSD) scores. STRAND thus represents a novel approach to RNA-protein complex refinement that leverages recent advancements in 3D RNA-protein modeling, and we propose to use STRAND as a module on top of existing RNA-protein complex prediction tools to enhance

their predictions.

Our main contributions can be summarized as follows:

- We propose STRAND, a novel RNA-protein refinement strategy that leverages the new modeling capabilities of current state-of-the-art deep learning based RNA-protein complex prediction approaches by directly improving their predictions. STRAND represents a modular extension to existing RNA-Protein complex prediction approaches.
- We extend DiffDock-PP to the prediction of RNA-Protein complexes using the embeddings of an RNA foundation model, RNA-FM (Section 3).
- Our experiments reveal improvements in complex structure prediction compared to the initial predictions of AlphaFold 3 and ProRNA3D-single across a diverse set of RNA-protein complexes from the PDB (Section 4).

We show predictions of STRAND trained with translational noise only that exemplify structural improvements together with initial predictions of AlphaFold 3, superimposed on the original PDB structure in Figure 1. Since we are reporting early results of ongoing research and due to manuscript preparations in progress, including STRAND, we currently cannot make our source code publicly available. However, we plan to publish STRAND open source in the future, including all models and training pipelines alongside our datasets.

2. Related Work

Besides exceptions (Kappel & Das, 2019; Delgado Blanco et al., 2019), traditional approaches for the prediction of RNA-protein complexes can generally be roughly divided into two groups, free docking approaches (Pérez-Cano et al., 2010; Setny & Zacharias, 2011; Huang et al., 2013; Guilhot-Gaudeffroy et al., 2014; Tuszynska et al., 2015; Iwakiri et al., 2016; Van Zundert et al., 2016; Arnautova et al., 2018) and template-based docking approaches (Zheng et al., 2016; Zhang et al., 2022). Here, we focus on deep learning based methods for RNA-protein complex predictions, but refer the interested reader to some excellent reviews of the field (Nithin et al., 2018; Bheemireddy et al., 2022; Liu et al., 2023).

Deep Learning based RNA-Protein Structure Prediction After the remarkable success of AlphaFold 2 (Jumper et al., 2021), predicting protein structures with nearly experimental accuracy, and its extension AlphaFold Multimer (Evans et al., 2021) for the prediction of protein multimers, recent deep learning approaches for the prediction of the structure of biological macromolecules extended their capabilities to other molecular entities like DNA, RNA, and small molecule ligands (Abramson et al., 2024; Krishna et al., 2024; Baek et al., 2024; Roche et al., 2024). However, while the quality of protein predictions is typically retained, particularly modeling RNAs remains challenging (Das et al., 2023; Bernard et al., 2025). This also transfers to predictions of RNA-protein interactions, and new algorithms that can improve the interaction prediction quality are highly sought after.

Diffusion Models Diffusion generative models (DGMs) have emerged as a powerful framework for modeling complex probability distributions, offering advantages over traditional likelihood-based and implicit generative approaches (Ketata et al., 2023). DGMs operate by defining a diffusion process that gradually transforms the data distribution into a tractable prior. The key insight lies in learning the score function – the gradient of the log probability density function $\nabla_{\mathbf{x}} \log p_t(\mathbf{x})^2$ – of this evolving distribution. Once learned, this score function enables sampling from the underlying probability distribution through established algorithms (Song et al., 2020). The success of DGMs has led to their widespread adoption in computational biology. These applications span diverse tasks including conformer generation (Jing et al., 2022; Xu et al., 2022; YanWang et al., 2024; Fan et al., 2024; Park & Shen), molecule generation (Hoogeboom et al., 2022), RNA secondary structure generation (Wang et al., 2025), and protein design (Trippe et al., 2022; Liu et al., 2024). Particularly noteworthy are their contributions to protein structure (Jing et al.; Watson et al., 2023; Wu et al., 2024) and backbone generation (Yim et al., 2023) as well as protein-protein docking (Ketata et al., 2023), where they have demonstrated remarkable capabilities in capturing complex structural relationships. AlphaFold 3 utilizes a diffusion-based architecture to predict the structures of various biomolecular complexes, including proteins, nucleic acids, and small molecule ligands (Abramson et al., 2024).

In this work, we employ a new strategy that leverages the recent advancements of deep learning models for structural biology by directly utilizing their predictions. However, in contrast to methods like ProRNA3D-single (Roche et al., 2024), which is trained using RNA monomer predictions from RhoFold (Shen et al., 2024) and protein monomer predictions from ESMFold (Lin et al., 2023), STRAND is trained directly on RNA-protein complexes of experimentally validated structures from the PDB. This allows us to use any RNA-protein structure as a starting point for the backward diffusion process, including predictions, making it independent of the underlying structure generation method at test time.

3. Methods

In this section, we describe STRAND, our extension of DiffDock-PP (Ketata et al., 2023) to model monomeric

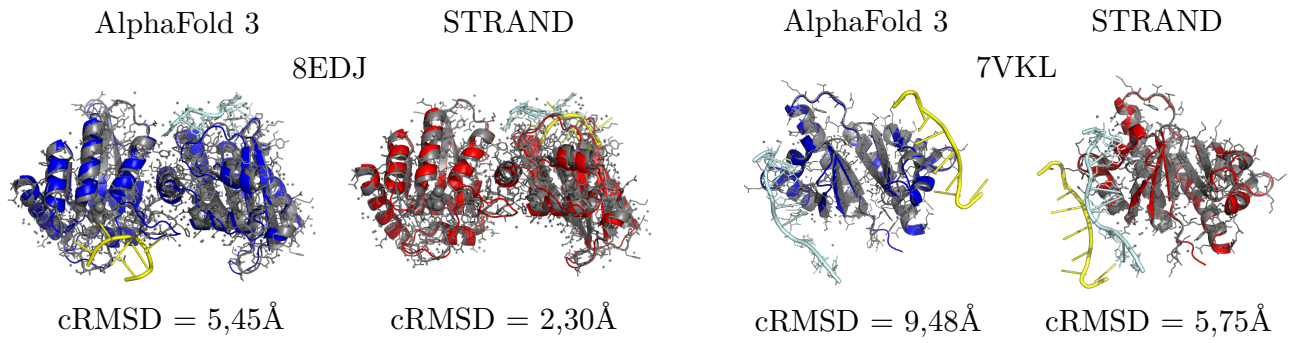

Figure 1. Example structure refinements for two AlphaFold 3 predictions for the PDB Ids 8EDJ and 7VKL using STRAND trained only on translational noise. We show the predictions of AlphaFold 3 (blue) and STRAND$_{tr}$ (red) superimposed on the true PDB structure of 8EDJ (left) and 7VKL (right). The true protein structure is shown in gray, the RNA in light blue. The RNA of the prediction is shown in yellow.

RNA-protein complexes. Similar to Ketata et al. (2023), we model proteins and RNA at the residue level in STRAND. For the protein, we exactly follow Ketata et al. (2023), using the same node features as used in DiffDock-PP that represent each residue by its type, the position of its α-carbon atom, and the processed embeddings from ESM2 (Lin et al., 2023). To enable processing of RNAs, we use similar node features: the nucleotide type, the coordinates of the phosphate atoms to represent the backbone, and the embeddings obtained from running RNA-FM (Chen et al., 2022). For the diffusion process, we treat one molecular entity as the receptor (the RNA) and apply noise to the other one (the protein). In the following, we describe STRAND in more detail with a focus on differences compared to DiffDock-PP.

3.1. STRAND

We denote $\mathbf{X}_1 \in \mathbb{R}^{3n}$ as the ligand consisting of n residues and $\mathbf{X}_2 \in \mathbb{R}^{3m}$ as the receptor with m residues. Ketata et al. (2023) assigns the ligand/receptor based on residue length. In contrast, we run the diffusion process on the protein, regardless of the relative sizes. While this increases training time due to longer molecules being diffused, it improves stability during evaluation, where test set lengths differ from those in training, thus we assign the protein $\mathbf{X}_1$ as the ligand. With $\mathbf{X}_1^* \in \mathbb{R}^{3n}$ and $\mathbf{X}_2^* \in \mathbb{R}^{3m}$ denoting the initial positions in space of both ligand and receptor respectively, the receptor is kept fixed ($\mathbf{X}_2 = \mathbf{X}_2^*$), and the task is to predict the position of the ligand with respect to the receptor. We define RNA-protein structure refinement as learning the conditional probability distribution $p(\mathbf{X}_1 \mid \mathbf{X}_2)$, which describes the possible ligand poses within a submanifold $\mathcal{M}$, given the receptor structure $\mathbf{X_2}$. Following Ketata et al. (2023), we avoid the inefficiencies arising from learning DGMs on arbitrary submanifolds (Bortoli et al., 2022), using the framework of intrinsic diffusion models (Corso et al., 2022).

Backbone Refinement via Diffusion Building on the framework of Ketata et al. (2023), which is limited to rigid-body transformations, we extend the approach by training separate models for translation, rotation, torsion, and the combinations rotation and translation, and rotation, translation, and torsion.

Formally, we introduce the 3D translation group $T(3)$, the 3D rotation group $SO(3)$, and associate changes in torsion angles at each rotatable bond with a copy of the 2D rotation group $SO(2)$.

The translation operation $A_{\text{tr}} : T(3) \times \mathbb{R}^{3n} \to \mathbb{R}^{3n}$ is naturally defined as:

$$A_{\text{tr}}(\mathbf{r}, \mathbf{x})_i = \mathbf{x}_i + \mathbf{r} \quad , \tag{1}$$

where $\mathbf{x}_i \in \mathbb{R}^3$ denotes the position of the ith backbone residue.

The rotation operation $A_{\text{rot}} : SO(3) \times \mathbb{R}^{3n} \to \mathbb{R}^{3n}$ is defined as:

$$A_{\text{rot}}(R, \mathbf{x})_i = R(\mathbf{x}_i - \bar{\mathbf{x}}) + \bar{\mathbf{x}} \quad , \tag{2}$$

where $\bar{\mathbf{x}} = \frac{1}{n}\sum_i \mathbf{x}_i$ is the (unweighted) center of mass of the ligand. This corresponds to a rotation around the center of mass.

For torsion, we follow the definition introduced in DiffDock (Corso et al., 2022), where changes in torsion angles are disentangled from global rotations and translations. In DiffDock's approach, the ambiguity of torsion changes, where the torsion angle around any bond (a_i, b_i) could be updated by rotating the a_i side, the b_i side, or both, is resolved by defining the action of elements of $SO(2)^m$ to cause minimal perturbation (in an RMSD sense) to the structure.

More precisely, let $B_k, \theta_k(\mathbf{x}) \in \mathbb{R}^{3n}$ be any valid torsion update by θ_k around the kth rotatable bond (a_k, b_k). We define $A_{\text{tor}} : SO(2)^m \times \mathbb{R}^{3n} \to \mathbb{R}^{3n}$ such that

$$A_{\text{tor}}(\boldsymbol{\theta}, \mathbf{x}) = \text{RMSDAlign}(\mathbf{x}, (B_1, \theta_1 \circ \cdots \circ B_m, \theta_m)(\mathbf{x})) \quad , \tag{3}$$

where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)$, and

$$\text{RMSDAlign}(\mathbf{x}, \mathbf{x}') = \underset{\mathbf{x}^\dagger \in \{g\mathbf{x}' \mid g \in SE(3)\}}{\arg\min} \text{RMSD}(\mathbf{x}, \mathbf{x}^\dagger) \quad . \tag{4}$$

In words, all m torsion updates are applied in any order before global RMSD alignment with the unmodified pose.

Now consider the product space $\mathbb{P} = T(3) \times SO(3) \times SO(2)^m$, and define $A : \mathbb{P} \times \mathbb{R}^{3n} \to \mathbb{R}^{3n}$ as:

$$A((\mathbf{r}, R, \boldsymbol{\theta}), \mathbf{x}) = A_{\text{tr}}\left(\mathbf{r}, A_{\text{rot}}\left(R, A_{\text{tor}}(\boldsymbol{\theta}, \mathbf{x})\right)\right) \quad . \tag{5}$$

Using this formulation, the submanifold of ligand poses can be described as:

$$\mathcal{M} = \{A\left((\mathbf{r}, \mathbf{R}, \boldsymbol{\theta}), \mathbf{X}_1\right) \mid (\mathbf{r}, \mathbf{R}, \boldsymbol{\theta}) \in \mathbb{P}\} \quad . \tag{6}$$

Following Corso et al. (2022), each mapping $A(\cdot, \mathbf{X}_1)$ is a bijection, ensuring the existence of inverse maps. This allows us to define a diffusion process over the full product space $\mathbb{P}$ to model the ligand pose distribution. In scenarios where the model is trained specifically for rotation, we set the translation vector $\mathbf{r} \in T(3)$ to the zero vector. Conversely, when training for translation only, we set the rotation matrix $\mathbf{R} \in SO(3)$ to the identity matrix $\mathbf{I}$, indicating the absence of rotation. When torsion is not modeled, we restrict the ligand to be rigid by eliminating any rotatable bonds.

Since each component of $\mathbb{P}$ forms a product manifold, we define independent forward diffusion processes (Rodola et al., 2019), where the score lies in the corresponding tangent space (Bortoli et al., 2022). The forward SDE is given by:

$$d\mathbf{x} = \sqrt{\frac{d\sigma^2(t)}{dt}}\, d\mathbf{w} \quad , \tag{7}$$

where σ^2_{tr}, σ^2_{rot}, and σ^2_{tor} govern diffusion in $T(3)$, $SO(3)$, and $SO(2)^m$, respectively. Their joint effect governs diffusion in $\mathbb{P}$. The term $d\mathbf{w}$ denotes Brownian motion on the respective manifolds.

Neural Architecture We adopt the architecture proposed by Ketata et al. (2023), originally developed for protein–protein complexes. To account for the differences in molecular composition and interaction dynamics between protein–protein and protein–RNA complexes, we introduce separate embedding layers for protein and RNA components. Furthermore, we extend the original model to additionally predict torsional scores, enabling finer-grained structural modeling.

Training Details For training, we use all RNA and protein containing PDB samples with deposition date before September 30, 2021, the training cutoff date of AlphaFold 3, downloaded on January 3, 2025. While the natural choice would be to collect monomeric RNA-protein complexes only, we found that adding RNA-protein multimers during training and using all chains of one molecular type as the receptor and all others together as the ligand is beneficial during training for all STRAND versions except STRAND$_{\text{tr+rot+tor}}$, our version using all transformations. We speculate that this process might help to directly learn the interactions of all chains in a single process. We train five different models for backbone refinement, one for translation only (STRAND$_{\text{tr}}$), one for rotation only (STRAND$_{\text{rot}}$), one for torsion only (STRAND$_{\text{tor}}$), and the combinations of translation and rotation (STRAND$_{\text{tr+rot}}$), as well as translation, rotation and torsion, and STRAND$_{\text{tr+rot+tor}}$. The models are trained on a single A40 GPU for roughly 5.75, 33, 6, 72, and 54 hours, respectively, using Adam. Please find details about hyperparameter settings in Appendix A.

4. Experiments

To assess STRAND's refinement capabilities, we initialize STRAND with the predictions of two recently proposed approaches for the modeling of monomeric RNA-protein complexes, AlphaFold 3 (Abramson et al., 2024) and ProRNA3D-single (Roche et al., 2024), and sample 40 new poses from the different models. For AlphaFold 3, we always use sample-0 predictions as only small variance is reported for the sampling from the diffusion model in AlphaFold 3 (Abramson et al., 2024). However, this aligns with our preliminary experiments, where we also did not observe strong variance across seeds nor different samples of AlphaFold 3. Predicted sequences are aligned to the ground truth via Needleman-Wunsch (Needleman & Wunsch, 1970) to handle misalignments. We report performance in terms of complex RMSD (cRMSD), determined by superimposing the ground truth and predicted complex structures via the Kabsch algorithm (Kabsch, 1976) and computing the RMSD between all Cα and P coordinates following Ganea et al. (2021).

Data For all our experiments with AlphaFold 3, we use a randomly selected set of monomeric RNA-protein complexes published in PDB after the AlphaFold 3 training cutoff date to ensure independence from the training set. We split the samples based on experimental techniques into a

Table 1. RNA-protein backbone refinement results of STRAND for AlphaFold 3 predictions. We show results for the translation only model $\text{STRAND}_{\text{tr}}$, the rotation only model $\text{STRAND}_{\text{rot}}$, the torsion only model $\text{STRAND}_{\text{tor}}$, and the combined versions $\text{STRAND}_{\text{tr+rot}}$ and $\text{STRAND}_{\text{tr+rot+tor}}$. Best performance is indicated in bold, equal best performance is additionally indicated by underlined scores. Results are reported in terms of mean and median complex RMSD (cRMSD; lower is better) and percentage of samples below a certain cRMSD threshold (2Å, 5Å, 10Å; higher is better). We observe that $\text{STRAND}_{\text{tr+rot}}$ shows the best performance on average, while at least one of the STRAND models performs equally well or better than AlphaFold 3 across all metrics for both datasets. The worst performance is observed for the models trained with torsion.

Methods	**Complex RMSD (Å) ↓**									
	X-Ray					**Non X-Ray**				
	%<2Å	%<5Å	%<10Å	Median	Mean	%<2Å	%<5Å	%<10Å	Median	Mean
AlphaFold 3	34.29	85.71	**100.00**	2.20	3.09	19.57	45.65	**63.04**	7.15	10.86
$\text{STRAND}_{\text{tr}}$	34.29	82.86	**100.00**	2.30	3.06	19.57	**47.83**	60.87	**6.53**	10.71
$\text{STRAND}_{\text{rot}}$	37.14	**88.57**	97.14	2.23	3.07	**21.74**	43.48	**63.04**	6.80	10.69
$\text{STRAND}_{\text{tor}}$	28.57	85.71	**100.00**	2.71	3.29	4.35	41.30	**63.04**	7.13	11.08
$\text{STRAND}_{\text{tr+rot}}$	**42.86**	82.86	**100.00**	**2.16**	**2.97**	17.39	45.65	60.87	6.87	**10.41**
$\text{STRAND}_{\text{tr+rot+tor}}$	31.43	82.86	**100.00**	3.10	3.45	0.00	34.78	54.35	8.73	11.64

set of test structures derived from X-ray crystallography and one with samples derived by other techniques like NMR and cryo EM, and evaluate the refinement of AlphaFold 3 predictions on both splits. The datasets contain 35 and 46 samples for X-ray and non-X-ray, respectively. For the refinement of ProRNA3D-single predictions, we only use data from X-ray crystallography, but download an additional set of 28 complexes from the PDB. The X-ray dataset for evaluations with ProRNA3D-single then contains a total of 63 samples. For more details about the datasets, please see Appendix B.

4.1. RNA-protein Complex Refinement with STRAND

In this section, we evaluate different STRAND models trained for different transformations to assess their influence on performance: the translation only model $\text{STRAND}_{\text{tr}}$, the rotation only model $\text{STRAND}_{\text{rot}}$, the torsion only model $\text{STRAND}_{\text{tor}}$, and the combined versions $\text{STRAND}_{\text{tr+rot}}$ and $\text{STRAND}_{\text{tr+rot+tor}}$. As described in Section 3, we diffuse over the protein, while the RNA is treated as the receptor. We sample 40 poses for each task for each individual model. The best sample is selected manually based on the complex RMSD (cRMSD).

4.1.1. STRAND can Improve AlphaFold 3

We summarize the results for the AlphaFold 3 predictions for the X-ray and non-X-ray derived monomeric complexes from PDB and their refinement with STRAND in Table 1.

Single Transformations While AlphaFold 3 shows remarkable prediction quality on the X-ray data with a mean cRMSD of 3.09 and 100% of the predictions being below a cRMSD of 10Å, we still see some improvements when applying STRAND to these strong predictions. For the models trained for single transformations, two out of three models improve the mean performance ($\text{STRAND}_{\text{tr}}$ with 3.06Å and $\text{STRAND}_{\text{rot}}$ with 3.07Å) while the model trained with torsion achieves a slightly higher mean cRMSD of 3.29Å. In addition, we observe remarkable increases by 2.85% of the percentage of complexes predicted below a cRMSD of 2Å and by 2.86% for complexes below 5Å, with only slightly worse median performance (2.23Å and 2.20Å for $\text{STRAND}_{\text{rot}}$ and AlphaFold 3, respectively) for structure refinement done via $\text{STRAND}_{\text{rot}}$. Generally, we observe improvements for 48.5% and 40% of the X-ray modeling tasks for the individual models $\text{STRAND}_{\text{tr}}$ and $\text{STRAND}_{\text{rot}}$, respectively (see also Figures 2, 3, and 4 for results on individual tasks).

For the non-X-ray data, we generally observe worse predictions of AlphaFold 3 with only roughly 60% of the predictions being below a cRMSD of 10Å. However, the trends observed for the different STRAND versions are also confirmed for non-X-ray samples. Again, STRAND_{rot} achieves the highest percentage of predictions below 2Å (2.17% improvement) and achieves the lowest mean cRMSD of all single transformation models (10.69Å for STRAND_{rot} compared to 10.86Å for AlphaFold 3) while the torsion only model shows the highest mean cRMSD of 11.08Å. Remarkably, the model trained with translation only achieves the highest percentage of predictions below a cRMSD of 5Å, a 2.18% improvement over AlphaFold 3. Overall, the three models can improve 63%, 63% and 26% of the individual tasks for $\text{STRAND}_{\text{tr}}$, $\text{STRAND}_{\text{rot}}$, and $\text{STRAND}_{\text{tor}}$, respectively. We show visualizations for the performance on individual samples in Figures 7, 8, and 9 in Appendix C.2.

Generally, the model trained only with torsion performs

Table 2. RNA-protein backbone refinement results of STRAND for ProRNA3D-single predictions on PDB structures validated by **X-ray** crystallography. We show results for the translation only model $\text{STRAND}_{\text{tr}}$, the rotation only model $\text{STRAND}_{\text{rot}}$, the torsion only model $\text{STRAND}_{\text{tor}}$, and the combined versions $\text{STRAND}_{\text{tr+rot}}$ and $\text{STRAND}_{\text{tr+rot+tor}}$. Results are reported in terms of mean and median complex RMSD (cRMSD; lower is better) and percentage of samples below a certain cRMSD threshold (2Å, 5Å, 10Å; higher is better). We observe that all methods improve the prediction quality of ProRNA3D-single on average, with the largest improvement observed by $\text{STRAND}_{\text{tr+rot}}$. Remarkably, the models trained with torsion can improve the mean cRMSD of the ProRNA3D-single predictions.

Methods	**Complex RMSD (Å) ↓**				
	%<2Å	%<5Å	%<10Å	Median	Mean
ProRNA3D-single	4.76	36.51	57.14	8.09	12.02
$\text{STRAND}_{\text{tr}}$	**14.29**	**<u>39.68</u>**	58.73	8.22	11.52
$\text{STRAND}_{\text{rot}}$	12.70	**<u>39.68</u>**	**60.32**	**6.90**	11.30
$\text{STRAND}_{\text{tor}}$	3.17	34.92	57.14	8.18	12.01
$\text{STRAND}_{\text{tr+rot}}$	12.70	38.10	58.73	7.25	**11.19**
$\text{STRAND}_{\text{tr+rot+tor}}$	7.94	38.10	58.73	7.70	11.49

worse than the other two models. We speculate that the reason for this is that torsion only affects the internal RNA coordinates but not its general positioning with respect to the protein. Consequently, changes in overall complex RMSD through changes in the torsion angles can typically only be marginal. This can also be observed in Figure 4 and Figure 9 for the performance on individual samples, where the effect on cRMSD by changes in torsion angles is much smaller compared to the other transformations (compare e.g. with Figure 3) and further supported visually in Figure 1, where the large differences in cRMSD result from repositioning the two structures with respect to each other, while the local structures remain unchanged.

Combined Transformations We also trained two models with combinations of transformations, $\text{STRAND}_{\text{tr+rot}}$ trained with translation and rotation, and $\text{STRAND}_{\text{tr+rot+tor}}$ trained with a combination of all the individual transformations. As shown in Table 1, we observe further improvements when combining the best-performing single transformations. Specifically, $\text{STRAND}_{\text{tr+rot}}$ achieves the lowest mean cRMSDs of 2.97Å and 10.41Å for X-ray and non-X-ray samples, respectively. Furthermore, $\text{STRAND}_{\text{tr+rot}}$ improves the percentage of samples below a cRMSD of 2Å by 8.57% and achieves also the lowest median cRMSD across all methods on the X-ray dataset. In contrast, the additional training for torsion seems to hurt performance in general, as $\text{STRAND}_{\text{tr+rot+tor}}$ achieves the worst performance across all the different STRAND versions, while $\text{STRAND}_{\text{tr+rot}}$ can improve 42% of the individual AlphaFold 3 predictions on the X-ray dataset and 63% of the AlphaFold predictions on the non-X-ray dataset. Please find plots for the performance on individual tasks for the combined models for the STRAND refinements of AlphaFold 3 predictions in Appendix C.1 and C.2.

4.1.2. STRAND CAN IMPROVE PRORNA3D-SINGLE

To assess the performance of STRAND for a different predictor than AlphaFold 3, we decide to use the recently proposed ProRNA3D-single (Roche et al., 2024), which reports remarkable results for monomeric RNA-protein complex predictions. Please find an overview of the results for the X-ray dataset in Table 2. Remarkably, we observe that all versions of STRAND improve over ProRNA3D-single in terms of mean cRMSD. The general trend that training with translation, rotation, and the combination of both leads to the best results, as observed for STRAND refinement of AlphaFold 3 predictions, is also confirmed for ProRNA3D-single. However, the improvements appear more substantial: For predictions below a cRMSD of 2Å we observe an improvement of 9.53% and 7.94% for the translation only model and both the rotation and rotation+translation models, respectively. In addition, we find improvements for the predictions below 5Å and 10Å for all models except the torsion-only version of STRAND, which is on par with ProRNA3D-single for predictions below 10Å and slightly worse for the percentage of samples below a cRMSD of 5Å. For ProRNA3D-single, we also observe the strongest improvements, with decreased cRMSD scores for 84%, 77.7%, and 84% of the targets for $\text{STRAND}_{\text{tr}}$, $\text{STRAND}_{\text{rot}}$, and $\text{STRAND}_{\text{tr+rot}}$, respectively. Please find visualizations for individual modeling tasks in Appendix C.3.

For our subsequent experiments and analysis we decide to use the translation only model $\text{STRAND}_{\text{tr}}$, the rotation only model $\text{STRAND}_{\text{rot}}$ and their combination $\text{STRAND}_{\text{tr+rot}}$ as these showed remarkable refinement capabilities for both AlphaFold 3 and ProRNA3D-single.

Table 3. RNA-protein backbone refinement results of STRAND for AlphaFold 3 **non-X-ray** predictions using sample selection with a trained confidence model. We show results for $\text{STRAND}_{\text{tr}}$, $\text{STRAND}_{\text{rot}}$ and the combined version $\text{STRAND}_{\text{tr+rot}}$. Best performance is indicated in bold, equal best performance is indicated by underlined scores. Results are reported in terms of mean and median cRMSD (lower is better) after sample selection with the confidence model for STRAND compared to the cRMSD achieved by the initial AlphaFold 3 prediction.

Methods	**Complex RMSD (Å) ↓**							
	Top 1		**Top 5**		**Top 10**		**Top 20**	
	Median	Mean	Median	Mean	Median	Mean	Median	Mean
AlphaFold 3	7.15	**10.86**	–	–	–	–	–	–
$\text{STRAND}_{\text{tr}}$	**6.81**	11.59	**6.78**	10.75	**6.78**	10.67	**6.58**	10.58
$\text{STRAND}_{\text{rot}}$	7.24	11.56	7.03	**10.73**	7.03	**10.51**	6.80	**10.27**
$\text{STRAND}_{\text{tr+rot}}$	7.15	12.81	6.95	11.15	6.95	10.94	6.94	10.48

4.2. Sample Selection for STRAND Predictions

In the previous section, we observe remarkable refinement capabilities of STRAND, improving AlphaFold 3 and ProRNA3D-single predictions across all datasets and metrics. However, the sample selection in our initial study was done manually based on the cRMSD across 40 samples drawn from the individual STRAND models. Selecting the best sample from a set of predictions is a challenging task. We decide to follow Ketata et al. (2023) and train a confidence model based on the predictions of $\text{STRAND}_{\text{tr+rot}}$. However, in contrast to Ketata et al. (2023), we do not train a classifier to select samples below a certain ligand RMSD threshold (5Å for Ketata et al. (2023)) but directly train a regression model for the prediction of ligand RMSDs. In addition, we cannot train our confidence model at the same scale as Ketata et al. (2023) due to compute limitations; we only use 3% of the training samples compared to Ketata et al. (2023) and train the confidence model for 60 hours on a single A40 GPU. One advantage of STRAND is that we can always fall back to the initial structure predictions of AlphaFold 3 or ProRNA3D-single in case STRAND cannot improve the prediction, since we are only doing refinement. Therefore, we include the initial predictions in the set of samples of the different STRAND versions, essentially ranking 41 instead of 40 samples per complex prediction task with the confidence model. Following Corso et al. (2022), we evaluate the confidence model in terms of top1, top5, as well as additional top10, and top20 performance, where topN refers to selecting the most accurate pose out of the N highest ranked predictions by the confidence model. In the following, we discuss the sample selection results for the three best models $\text{STRAND}_{\text{tr}}$, $\text{STRAND}_{\text{rot}}$, and $\text{STRAND}_{\text{tr+rot}}$.

4.2.1. Selection for AlphaFold 3 Refinements

We show the results for the sample selection for STRAND using the confidence model for the refined predictions of AlphaFold 3 on non-X-ray data in Table 3. The results for X-ray data can be found in Table 10 in Appendix C.1. Generally, we observe that the confidence model can select strong samples well for the non-X-ray dataset (see Table 3). $\text{STRAND}_{\text{tr}}$ already improves the AlphaFold 3 median performance for the top1 selected sample (6.81Å and 7.15Å for $\text{STRAND}_{\text{tr}}$ and AlphaFold 3, respectively), while both other STRAND versions perform equally well or only slightly worse in terms of median cRMSD. From top5 onwards, all STRAND versions consistently show lower median cRMSDs, while $\text{STRAND}_{\text{tr}}$ and $\text{STRAND}_{\text{rot}}$ also outperform AlphaFold 3 in terms of mean performance. However, sample selection on the X-ray dataset appears more challenging. None of the STRAND versions can improve AlphaFold 3 performance for top1 and top5 selected samples, while $\text{STRAND}_{\text{rot}}$ is the only model that can improve AlphaFold 3 mean performance for top10 selection with equal median cRMSD. However, for top20 selection, we still observe improvements with STRAND across all models for mean and median cRMSDs, except for $\text{STRAND}_{\text{tr}}$ with slightly higher mean cRMSD but the lowest median cRMSD.

We attribute the difference in sample selection performance between the non-X-ray and the X-ray dataset to the overall high prediction quality of AlphaFold 3 and the STRAND refinements across all samples of the X-ray dataset. Distinguishing very good from good predictions seems to be much harder than selection for the non-X-ray examples, where we observed higher cRMSDs for the initial AlphaFold 3 predictions in general.

4.2.2. Selection for ProRNA3D-single Refinements

The results for the sample selection for STRAND refinements of ProRNA3D-single predictions are shown in Table 4. Generally, both $\text{STRAND}_{\text{tr+rot}}$ and $\text{STRAND}_{\text{rot}}$ achieve similar performance, with $\text{STRAND}_{\text{tr+rot}}$ performing best, achieving the lowest mean and median cRMSDs

Table 4. RNA-protein backbone refinement results of STRAND for ProRNA3D-single predictions for **X-ray** data using sample selection with a trained confidence model. We show results for $\text{STRAND}_{\text{tr}}$, $\text{STRAND}_{\text{rot}}$ and the combined version $\text{STRAND}_{\text{tr+rot}}$. Best performance is indicated in bold. Results are reported in terms of mean and median cRMSD (lower is better) after sample selection with the confidence model for STRAND compared to the cRMSD achieved by the initial ProRNA3D-single prediction.

Methods	**Complex RMSD (Å) ↓**							
	Top 1		**Top 5**		**Top 10**		**Top 20**	
	Median	Mean	Median	Mean	Median	Mean	Median	Mean
ProRNA3D-single	**8.09**	**12.02**	–	–	–	–	–	–
$\text{STRAND}_{\text{tr}}$	8.95	12.37	8.77	11.88	8.77	11.75	8.10	11.42
$\text{STRAND}_{\text{rot}}$	8.34	12.41	8.02	11.78	8.02	11.66	7.68	11.34
$\text{STRAND}_{\text{tr+rot}}$	8.38	12.08	**7.11**	**11.74**	**6.88**	**11.32**	**6.88**	**11.22**

from top5 selection onwards (7.11Å and 8.09Å median, 11.74Å and 12.02Å mean cRMSD for $\text{STRAND}_{\text{tr+rot}}$ top5 selection and ProRNA3D-single, respectively), while being only slightly worse than ProRNA3D-single for the top1 selected sample. While $\text{STRAND}_{\text{tr}}$ can similarly reduce the mean cRMSD compared to ProRNA3D-single from top5 selection onwards, it cannot achieve a lower median cRMSD.

5. Discussion, Limitations & Future Work

In this work, we introduce Structure refinement of RNA-protein complexes via diffusion (STRAND), a novel approach for monomeric RNA-protein complex refinement. We leverage the capabilities of current deep learning approaches for structural biology that can model different molecule types and complexes of these and combine it with the powerful diffusion framework of DiffDock-PP. Our new method successfully refines monomeric RNA-protein complex predictions by enhancing docking positions, showing remarkable improvements in structure quality when applied to AlphaFold 3 and ProRNA3D-single predictions.

With our analysis using different individual transformations and their combinations, we show that simple rotation and translation of the initial predictions can often already lead to improvements, indicating that the 3D structure predictors might often generally misplace the protein with respect to the RNA when modeling their interaction. This is further supported by our finding that the local refinement via torsion angles does not lead to substantial improvements, although torsion angle refinement generally comes with its own additional challenges, and the training of the models, including torsion, might require further development, specifically regarding noise schedules. However, we think that our results provide new insights that could help to develop stronger 3D RNA-protein structure prediction tools in the future, which would also be beneficial for new refinement methods.

Our experiments for sample selection show that it is generally possible to select strong samples by training a confidence model for ligand RMSD prediction. We think that direct prediction of ligand RMSD could be beneficial over training a classifier to select samples below a certain ligand RMSD threshold, as implemented in DiffDock-PP. However, we follow Ketata et al. (2023) and optimize the ligand RMSD, which might not be optimal. It would be interesting to analyze the results when optimizing e.g. for cRMSD directly. Nevertheless, our selection model already shows remarkable results, although we train only on a fraction of the data used by Ketata et al. (2023) due to compute limitations.

Limitations However, despite notable improvements in performance, our approach also has limitations. Our refinement currently only involves backbone atoms similar to DiffDock-PP. However, it would be generally preferable to refine full-atom structures. Additionally, we are bound by the length limitations of the foundation models for obtaining sequence embeddings. We currently limit the length to 1022 residues. Furthermore, selecting the best sample remains challenging. We note, however, that our confidence model is currently trained on a relatively small set of complexes, and we expect improved performance when scaling the training to larger amounts of data.

Future Work For the future, we mainly plan to address the current limitations. Most importantly, we would like to further analyze the inclusion of torsion angle refinement and scale the data for training the confidence model for sample selection. Also, optimizing the confidence model for cRMSD instead of ligand RMSD would be a natural next step. In addition, we are currently developing a full-atom version of STRAND with diffusion over the RNA instead of the protein.

We believe that our novel refinement approach bears large potential for the community, potentially enabling high prediction quality of monomeric RNA-protein complexes in the future.

Impact Statement

Our paper presents work whose goal is to advance the field of 3D RNA-protein complex predictions. There are many potential scientific and societal consequences of our work, including potential advances in drug discovery and general understanding of biological systems. However, we feel that none of these potential consequences must be specifically highlighted here.

ACKNOWLEDGMENTS

The authors acknowledge funding by the German Research Foundation (DFG) under SFB 1597 (SmallData), grant no. 499552394, and through grant no. 417962828 as well as support by the state of Baden-Württemberg through bwHPC and the German Research Foundation (DFG) through grant no INST 39/963-1 FUGG (bwForCluster NEMO) and grant INST 35/1597-1 FUGG (bwForCluster Helix). This research was funded by the European Union (via ERC Consolidator Grant DeepLearning 2.0, grant no. 101045765). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them.

Funded by the European Union

References

Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A. J., Bambrick, J., et al. Accurate structure prediction of biomolecular interactions with alphafold 3. *Nature*, pp. 1–3, 2024.

Arnautova, Y. A., Abagyan, R., and Totrov, M. Protein-rna docking using icm. *Journal of Chemical Theory and Computation*, 14(9):4971–4984, 2018.

Baek, M., McHugh, R., Anishchenko, I., Jiang, H., Baker, D., and DiMaio, F. Accurate prediction of protein–nucleic acid complexes using rosettafoldna. *Nature methods*, 21 (1):117–121, 2024.

Bernard, C., Postic, G., Ghannay, S., and Tahi, F. Has alphafold3 achieved success for rna? *Biological Crystallography*, 81(2), 2025.

Bheemireddy, S., Sandhya, S., Srinivasan, N., and Sowdhamini, R. Computational tools to study rna-protein complexes. *Frontiers in molecular biosciences*, 9:954926, 2022.

Bortoli, V. D., Mathieu, E., Hutchinson, M., Thornton, J., Teh, Y. W., and Doucet, A. Riemannian score-based generative modeling. *arXiv preprint arXiv:2202.02763*, 2022.

Burley, S. K., Berman, H. M., Kleywegt, G. J., Markley, J. L., Nakamura, H., and Velankar, S. Protein data bank (pdb): the single global macromolecular structure archive. *Protein crystallography: methods and protocols*, pp. 627–641, 2017.

Chen, J., Hu, Z., Sun, S., Tan, Q., Wang, Y., Yu, Q., Zong, L., Hong, L., Xiao, J., Shen, T., et al. Interpretable rna foundation model from unannotated data for highly accurate rna structure and function predictions. *arXiv preprint arXiv:2204.00300*, 2022.

Corso, G., Stärk, H., Jing, B., Barzilay, R., and Jaakkola, T. Diffdock: Diffusion steps, twists, and turns for molecular docking. *arXiv preprint arXiv:2210.01776*, 2022.

Das, R., Kretsch, R. C., Simpkin, A. J., Mulvaney, T., Pham, P., Rangan, R., Bu, F., Keegan, R. M., Topf, M., Rigden, D. J., et al. Assessment of three-dimensional rna structure prediction in casp15. *Proteins: Structure, Function, and Bioinformatics*, 91(12):1747–1770, 2023.

Delgado Blanco, J., Radusky, L. G., Cianferoni, D., and Serrano, L. Protein-assisted rna fragment docking (rnax) for modeling rna–protein interactions using modelx. *Proceedings of the National Academy of Sciences*, 116(49): 24568–24573, 2019.

Evans, R., O'Neill, M., Pritzel, A., Antropova, N., Senior, A., Green, T., Žídek, A., Bates, R., Blackwell, S., Yim, J., et al. Protein complex prediction with alphafold-multimer. *biorxiv*, pp. 2021–10, 2021.

Fan, J., Li, Z., Alcaide, E., Ke, G., Huang, H., and E, W. Accurate conformation sampling via protein structural diffusion. *Journal of Chemical Information and Modeling*, 64(22):8414–8426, 2024.

Franke, J., Runge, F., and Hutter, F. Probabilistic transformer: Modelling ambiguities and distributions for RNA folding and molecule design. *Advances in Neural Information Processing Systems*, 35:26856–26873, 2022.

Franke, J. K., Runge, F., Köksal, R., Backofen, R., and Hutter, F. Rnaformer: A simple yet effective deep learning model for rna secondary structure prediction. *bioRxiv*, pp. 2024–02, 2024.

Ganea, O.-E., Huang, X., Bunne, C., Bian, Y., Barzilay, R., Jaakkola, T., and Krause, A. Independent se(3)-equivariant models for end-to-end rigid protein docking. *arXiv preprint*, arXiv:2111.07786, 2021. doi: 10.48550/arXiv.2111.07786.

Glisovic, T., Bachorik, J. L., Yong, J., and Dreyfuss, G. Rna-binding proteins and post-transcriptional gene regulation. *FEBS letters*, 582(14):1977–1986, 2008.

Guilhot-Gaudeffroy, A., Froidevaux, C., Azé, J., and Bernauer, J. Protein-rna complexes and efficient automatic docking: expanding rosettadock possibilities. *PloS one*, 9(9):e108928, 2014.

Hagerman, P. J. Flexibility of rna. *Annual review of biophysics and biomolecular structure*, 26(1):139–156, 1997.

Hoogeboom, E., Satorras, V. G., Vignac, C., and Welling, M. Equivariant diffusion for molecule generation in 3d. In *International Conference on Machine Learning (ICML)*, pp. 8867–8887. PMLR, 2022.

Hu, Y., Cheng, K., He, L., Zhang, X., Jiang, B., Jiang, L., Li, C., Guan, W., Yang, Y., and Liu, M. Nmr-based methods for protein analysis. *Analytical Chemistry*, XXXX, 01 2021. doi: 10.1021/acs.analchem.0c03830.

Huang, Y., Liu, S., Guo, D., Li, L., and Xiao, Y. A novel protocol for three-dimensional structure prediction of rna-protein complexes. *Scientific reports*, 3(1):1887, 2013.

Iwakiri, J., Hamada, M., Asai, K., and Kameda, T. Improved accuracy in rna-protein rigid body docking by incorporating force field for molecular dynamics simulation into the scoring function. *Journal of Chemical Theory and Computation*, 12(9):4688–4697, 2016. doi: 10.1021/acs.jctc.6b00254. URL https://doi.org/10.1021/acs.jctc.6b00254. Epub 2016 Aug 17.

Jing, B., Erives, E., Pao-Huang, P., Corso, G., Berger, B., and Jaakkola, T. S. Eigenfold: Generative protein structure prediction with diffusion models. In *ICLR 2023-Machine Learning for Drug Discovery workshop*.

Jing, B., Corso, G., Chang, J., Barzilay, R., and Jaakkola, T. Torsional diffusion for molecular conformer generation. *arXiv preprint arXiv:2206.01729*, 2022.

Joshi, C. K., Jamasb, A. R., Viñas, R., Harris, C., Mathis, S. V., Morehead, A., Anand, R., and Liò, P. grnade: Geometric deep learning for 3d rna inverse design. *bioRxiv*, pp. 2024–03, 2025.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. Highly accurate protein structure prediction with alphafold. *nature*, 596(7873):583–589, 2021.

Kabsch, W. A solution for the best rotation to relate two sets of vectors. *Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography*, 32(5):922–923, 1976.

Kappel, K. and Das, R. Sampling native-like structures of rna-protein complexes through rosetta folding and docking. *Structure*, 27(1):140–151, 2019.

Ketata, M. A., Laue, C., Mammadov, R., Stärk, H., Wu, M., Corso, G., Marquet, C., Barzilay, R., and Jaakkola, T. S. Diffdock-pp: Rigid protein-protein docking with diffusion models. 2023. URL https://arxiv.org/abs/2304.04573.

Krishna, R., Wang, J., Ahern, W., Sturmfels, P., Venkatesh, P., Kalvet, I., Lee, G. R., Morey-Burrows, F. S., Anishchenko, I., Humphreys, I. R., et al. Generalized biomolecular modeling and design with rosettafold all-atom. *Science*, 384(6693):eadl2528, 2024.

Lin, Z., Akin, H., Rao, R., Hie, B., Zhu, Z., Lu, W., Smetanin, N., Verkuil, R., Kabeli, O., Shmueli, Y., et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 379(6637): 1123–1130, 2023.

Liu, X., Duan, Y., Hong, X., Xie, J., and Liu, S. Challenges in structural modeling of rna-protein interactions. *Current Opinion in Structural Biology*, 81:102623, 2023.

Liu, Y., Wang, S., Dong, J., Chen, L., Wang, X., Wang, L., Li, F., Wang, C., Zhang, J., Wang, Y., et al. De novo protein design with a denoising diffusion network independent of pretrained structure prediction models. *Nature Methods*, pp. 1–10, 2024.

Needleman, S. B. and Wunsch, C. D. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology*, 48 (3):443–453, 1970. doi: 10.1016/0022-2836(70)90057-4.

Nithin, C., Ghosh, P., and Bujnicki, J. M. Bioinformatics tools and benchmarks for computational docking and 3d structure prediction of rna-protein complexes. *Genes*, 9 (9):432, 2018.

Park, J. and Shen, Y. Equivariant blurring diffusion for hierarchical molecular conformer generation. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*.

Patil, S., Runge, F., Franke, J. K., and Hutter, F. Towards generative rna design with tertiary interactions. In *ICLR 2024 Workshop on Generative and Experimental Perspectives for Biomolecular Design*, 2024.

Pérez-Cano, L., Solernou, A., Pons, C., and Fernández-Recio, J. Structural prediction of protein-rna interaction by computational docking with propensity-based statistical potentials. In *Biocomputing 2010*, pp. 293–301. World Scientific, 2010.

Re, A., Joshi, T., Kulberkyte, E., Morris, Q., and Workman, C. T. Rna–protein interactions: an overview. *RNA sequence, structure, and function: Computational and bioinformatic methods*, pp. 491–521, 2014.

Renaud, J.-P., Chari, A., Ciferri, C., Liu, W.-t., Rémigy, H.-W., Stark, H., and Wiesmann, C. Cryo-em in drug discovery: achievements, limitations and prospects. *Nature reviews Drug discovery*, 17(7):471–492, 2018.

Roche, R., Tarafder, S., and Bhattacharya, D. Single-sequence protein-rna complex structure prediction by geometric attention-enabled pairing of biological language models. *bioRxiv*, 2024.

Rodola, E., Lähner, Z., Bronstein, A. M., Bronstein, M. M., and Solomon, J. Functional maps representation on product manifolds. *Computer Graphics Forum*, 38:678–689, 2019.

Runge, F., Franke, J., Fertmann, D., Backofen, R., and Hutter, F. Partial rna design. *Bioinformatics*, 40 (Supplement_1):i437–i445, 2024.

Sato, K., Akiyama, M., and Sakakibara, Y. RNA secondary structure prediction using deep learning with thermodynamic integration. *Nature communications*, 12(1):1–9, 2021.

Setny, P. and Zacharias, M. A coarse-grained force field for protein–rna docking. *Nucleic acids research*, 39(21): 9118–9129, 2011.

Shen, T., Hu, Z., Sun, S., Liu, D., Wong, F., Wang, J., Chen, J., Wang, Y., Hong, L., Xiao, J., et al. Accurate rna 3d structure prediction using a language model-based deep learning approach. *Nature Methods*, pp. 1–12, 2024.

Smyth, M. S. and Martin, J. H. X-ray crystallography. *Molecular Pathology*, 53(1):8–14, 2000. doi: 10.1136/mp.53.1.8.

Song, Y., Sohl-Dickstein, J., Kingma, D. P., Kumar, A., Ermon, S., and Poole, B. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*, 2020.

Tan, C., Zhang, Y., Gao, Z., Hu, B., Li, S., Liu, Z., and Li, S. Z. Rdesign: Hierarchical data-efficient representation learning for tertiary structure-based rna design. In *The Twelfth International Conference on Learning Representations*, 2024.

Trippe, B. L., Yim, J., Tischer, D., Broderick, T., Baker, D., Barzilay, R., and Jaakkola, T. Diffusion probabilistic modeling of protein backbones in 3d for the motif-scaffolding problem. *arXiv preprint arXiv:2206.04119*, 2022.

Tuszynska, I., Magnus, M., Jonak, K., Dawson, W., and Bujnicki, J. M. Npdock: a web server for protein–nucleic acid docking. *Nucleic acids research*, 43(W1):W425–W430, 2015.

Van Zundert, G., Rodrigues, J., Trellet, M., Schmitz, C., Kastritis, P., Karaca, E., Melquiond, A., van Dijk, M., De Vries, S., and Bonvin, A. The haddock2. 2 web server: user-friendly integrative modeling of biomolecular complexes. *Journal of molecular biology*, 428(4): 720–725, 2016.

Wang, Z., Feng, Y., Tian, Q., Liu, Z., Yan, P., and Li, X. Rnadifffold: generative rna secondary structure prediction using discrete diffusion models. *Briefings in Bioinformatics*, 26(1):bbae618, 2025.

Watson, J. L., Juergens, D., Bennett, N. R., Trippe, B. L., Yim, J., Eisenach, H. E., Ahern, W., Borst, A. J., Ragotte, R. J., Milles, L. F., et al. De novo design of protein structure and function with rfdiffusion. *Nature*, 620(7976): 1089–1100, 2023.

Wu, K. E., Yang, K. K., van den Berg, R., Alamdari, S., Zou, J. Y., Lu, A. X., and Amini, A. P. Protein structure generation via folding diffusion. *Nature communications*, 15(1):1059, 2024.

Xu, M., Yu, L., Song, Y., Shi, C., Ermon, S., and Tang, J. Geodiff: A geometric diffusion model for molecular conformation generation. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=PzcvxEMzvQC.

YanWang, Wang, L., Shen, Y., Wang, Y., Yuan, H., Wu, Y., and Gu, Q. Protein conformation generation via force-guided SE(3) diffusion models. In *Forty-first International Conference on Machine Learning*, 2024. URL https://openreview.net/forum?id=aCl LSa4nXs.

Yim, J., Trippe, B. L., De Bortoli, V., Mathieu, E., Doucet, A., Barzilay, R., and Jaakkola, T. Se (3) diffusion model with application to protein backbone generation. In *International Conference on Machine Learning*, pp. 40001–40039. PMLR, 2023.

Zhang, C., Shine, M., Pyle, A. M., and Zhang, Y. Us-align: universal structure alignments of proteins, nucleic acids, and macromolecular complexes. *Nature methods*, 19(9): 1109–1115, 2022.

Zheng, J., Kundrotas, P. J., Vakser, I. A., and Liu, S. Template-based modeling of protein-rna interactions. *PLoS computational biology*, 12(9):e1005120, 2016.

A. Hyperparameters

Table 5. Hyperparameters of the Tensor Product Score Model.

Parameter	Value
Max Ligand Radius	5
Receptor Max Radius	30
Cross Max Distance	80
Dynamic Max Cross	True
Cross Cutoff Weight	3
Cross Cutoff Bias	40
Center Max Distance	30
Spherical Harmonics (lmax)	2
Node Features (ns)	16
Vector Features (nv)	4
Scale by Sigma	True
Number of Conv Layers	4
Dropout Rate	0
Batch Normalization	False
Edge Features	4
Distance Embedding Dim	32
Cross Distance Embedding Dim	32
Sigma Embedding Dim	32
LM Embedding Dim	640
Hidden Features	3 * ns
Gaussian Smearing Start	0.0
Gaussian Smearing Stop	5.0
Sinusoidal Max Positions	1e4
Embedding Scale	10000

Table 6. Other Hyperparameters.

Parameter	Value
Batch Size $\text{STRAND}_{\text{tr}}$	4
Batch Size $\text{STRAND}_{\text{rot}}$	24
Batch Size $\text{STRAND}_{\text{tor}}$	12
Batch Size $\text{STRAND}_{\text{tr+rot}}$	12
Batch Size $\text{STRAND}_{\text{tr+rot+tor}}$	12
Number of Nearest Neighbors	30
Optimizer	Adam

B. Data Insights

B.1. Statistics of the three sets used in the evaluation of structure refinement.

Table 7. RNA and Protein Residue Statistics for Different Evaluation Sets.

	X-ray	**Non-X-ray**	**ProRNA3D-single**
Num samples	35	46	63
		RNA	
Max Length	84.0	828.0	270.0
Min Length	3.0	5.0	3.0
Median Length	6.0	25.5	10.0
Mean length	12.9	96.0	35.0
		Protein	
Max length	816.0	1017.0	850.0
Min Length	75.0	100.0	82.0
Median Length	167.0	669.5	388.0
Mean Length	289.7	572.3	416.3

B.2. Statistics train, validation, and test splits used in training and evaluating the diffusion model.

Table 8. RNA and Protein Residue Statistics for Train, Validation, and Test Sets (all three combined).

	Train	**Val**	**Test**
	RNA		
Max Length	829.0	706	240
Min Length	1.0	2	14
Median Length	22.0	21.0	68.0
Mean Length	59.0	60.88	79.0
	Protein		
Max Length	1022	1007	999
Min Length	14	61	75
Median Length	411.0	416.0	374.0
Mean Length	449.5	484.0	423.6

Table 9. RNA and Protein Residue Statistics for Train, Validation, and Test Sets used in $\text{STRAND}_{\text{tr+rot+tor}}$.

	Train	**Val**	**Test**
RNA			
Max Length	1012	924	950
Min Length	2.0	2	2
Median Length	118	119.5	114.5
Mean Length	150.86	148.76	118.68
Protein			
Max Length	1015	823	476
Min Length	2	50	51
Median Length	155	126.5	152.5
Mean Length	449.65	196.26	178.32Í

C. Additional Results

C.1. Improving AlphaFold 3 Prediction for X-Ray Structures

Table 10. RNA-protein backbone refinement results of STRAND for AlphaFold 3 **X-ray** predictions using sample selection with a trained confidence model. We show results for $\text{STRAND}_{\text{tr}}$, $\text{STRAND}_{\text{rot}}$ and the combined version $\text{STRAND}_{\text{tr+rot}}$. Best performance is indicated in bold, equal best performance is indicated by underlined scores. Results are reported in terms of mean and median cRMSD (lower is better) after sample selection with the confidence model for STRAND compared to the cRMSD achieved by the initial AlphaFold 3 prediction.

Methods	**Complex RMSD (Å) ↓**							
	Top 1		**Top 5**		**Top 10**		**Top 20**	
	Median	Mean	Median	Mean	Median	Mean	Median	Mean
AlphaFold 3	**2.20**	**3.09**	–	–	–	–	–	–
$\text{STRAND}_{\text{tr}}$	2.78	3.94	2.39	3.57	2.30	3.47	**2.11**	3.15
$\text{STRAND}_{\text{rot}}$	3.64	4.46	2.22	3.30	<u>2.20</u>	**2.90**	<u>2.20</u>	**2.83**
$\text{STRAND}_{\text{tr+rot}}$	2.84	4.48	2.45	3.60	2.29	3.34	2.12	2.90

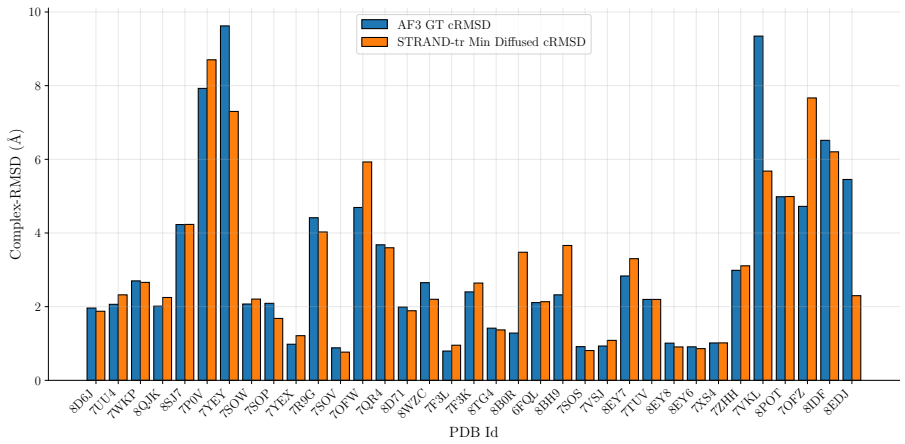

Figure 2. Per sample cRMSD of AlphaFold 3 predictions on X-ray data and the corresponding prediction of STRAND. We show results of the translation-only model $\text{STRAND}_{\text{tr}}$.

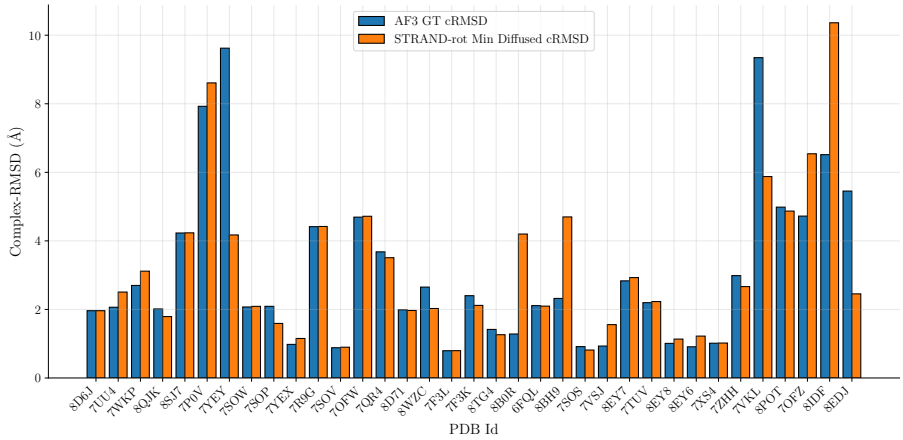

Figure 3. Per sample cRMSD of AlphaFold 3 predictions on X-ray data and the corresponding prediction of STRAND. We show results of the rotation-only model $\text{STRAND}_{\text{rot}}$.

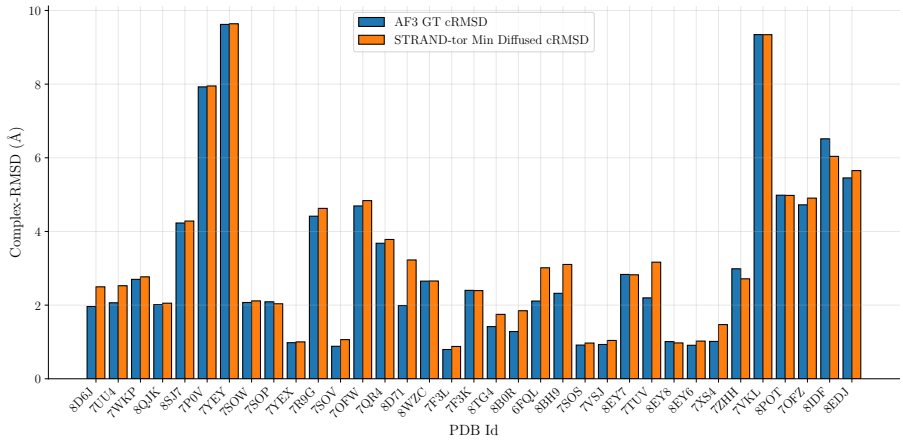

Figure 4. Per sample cRMSD of AlphaFold 3 predictions on X-ray data and the corresponding prediction of STRAND. We show results of the torsion-only model $\text{STRAND}_{\text{tor}}$.

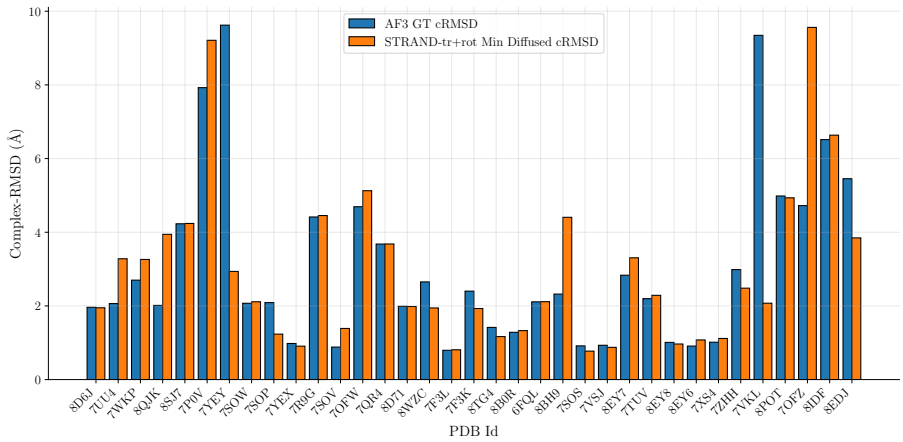

Figure 5. Per sample cRMSD of AlphaFold 3 predictions on X-ray data and the corresponding prediction of STRAND. We show results of the combined model of translation and rotation $\text{STRAND}_{\text{tr+rot}}$.

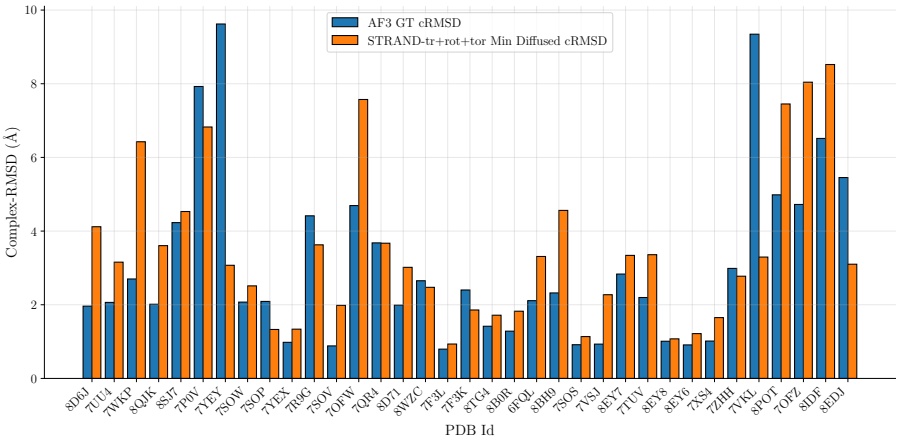

Figure 6. Per sample cRMSD of AlphaFold 3 predictions on X-ray data and the corresponding prediction of STRAND. We show results of the combined model for translation, rotation, and torsion $\text{STRAND}_{\text{tr+rot+tor}}$.

C.2. Improving AlphaFold 3 Prediction for Non-X-Ray Structures

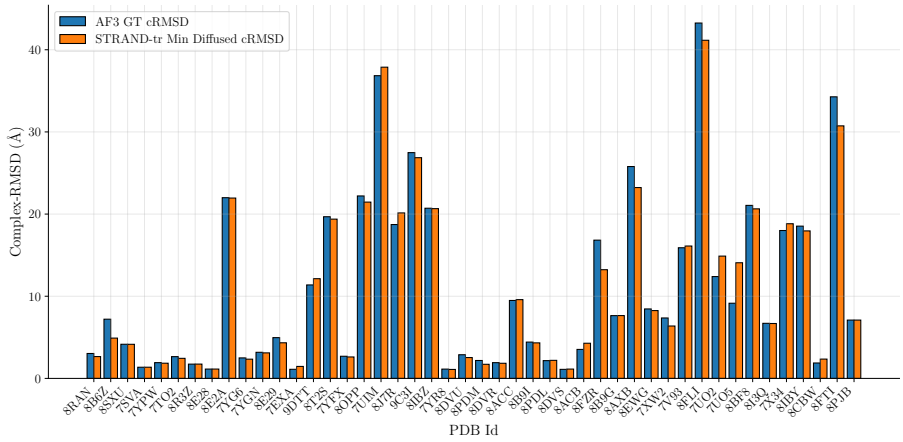

Figure 7. Per sample cRMSD of AlphaFold 3 predictions on non-X-ray data and the corresponding prediction of STRAND. We show results of the translation-only model $\text{STRAND}_{\text{tr}}$.

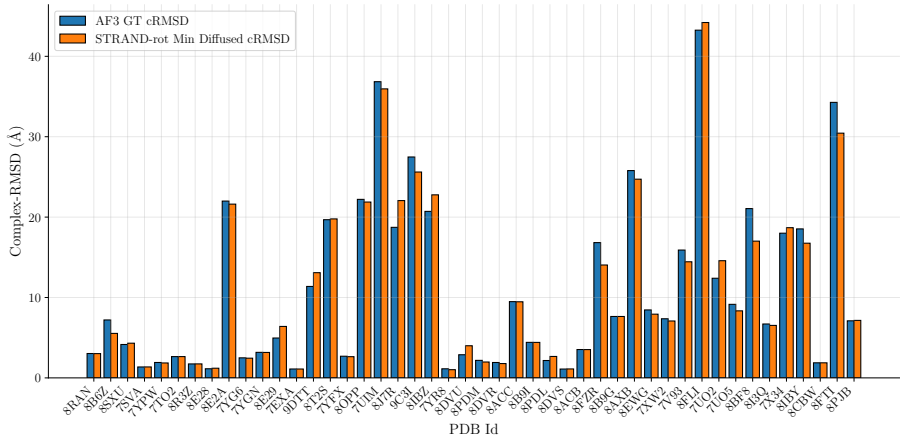

Figure 8. Per sample cRMSD of AlphaFold 3 predictions on non-X-ray data and the corresponding prediction of STRAND. We show results of the rotation-only model $\text{STRAND}_{\text{rot}}$.

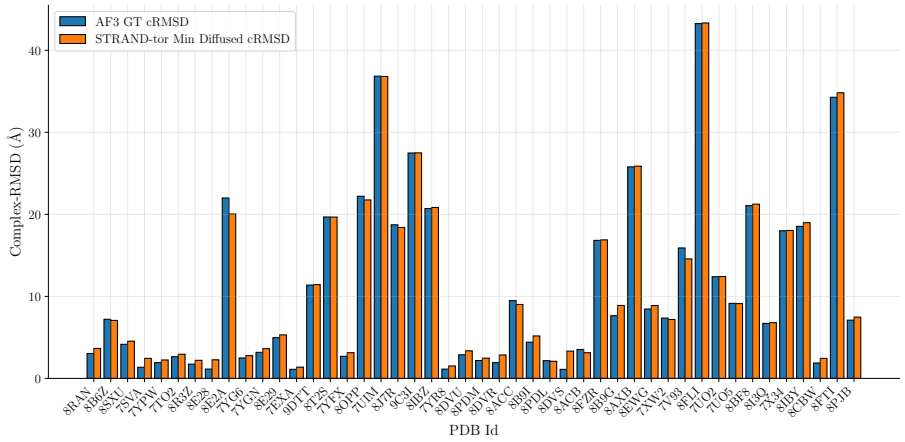

Figure 9. Per sample cRMSD of AlphaFold 3 predictions on non-X-ray data and the corresponding prediction of STRAND. We show results of the torsion only model $\text{STRAND}_{\text{tor}}$.

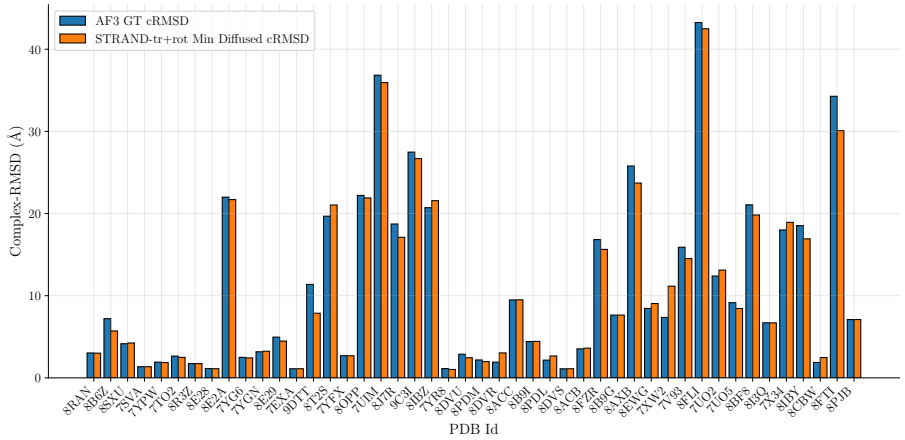

Figure 10. Per sample cRMSD of AlphaFold 3 predictions on non-X-ray data and the corresponding prediction of STRAND. We show results of the combined model of translation and rotation $\text{STRAND}_{\text{tr+rot}}$.

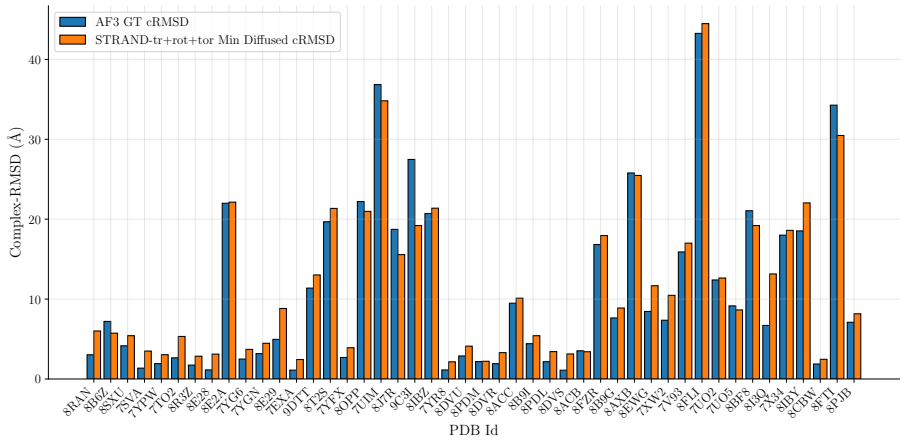

Figure 11. Per sample cRMSD of AlphaFold 3 predictions on non-X-ray data and the corresponding prediction of STRAND. We show results of the combined model of translation, rotation, and torsion $\text{STRAND}_{\text{tr+rot+tor}}$.

C.3. Improving ProRNA3D-single Predictions

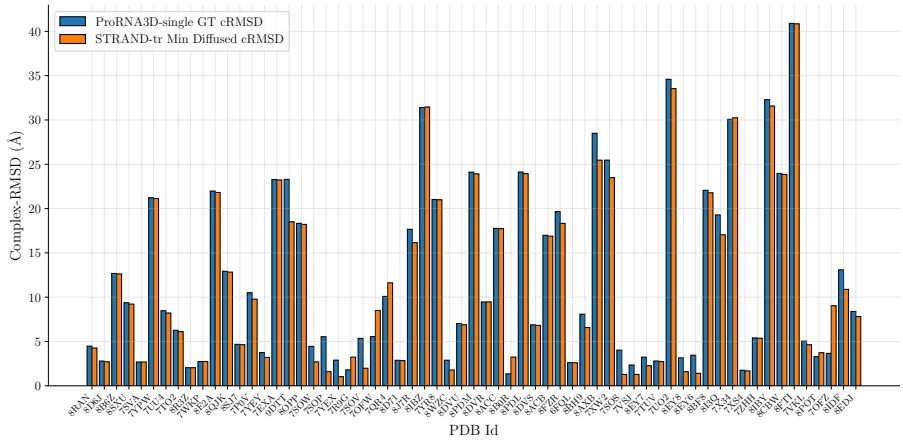

Figure 12. Per sample cRMSD of ProRNA3D-single predictions on X-ray data and the corresponding prediction of STRAND. We show results of the translation-only model $\text{STRAND}_{\text{tr}}$.

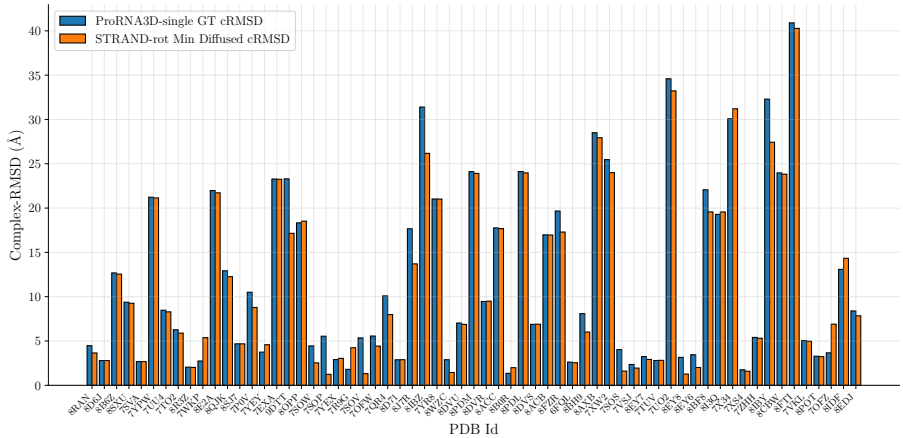

Figure 13. Per sample cRMSD of ProRNA3D-single predictions on X-ray data and the corresponding prediction of STRAND. We show results of the rotation-only model $\text{STRAND}_{\text{rot}}$.

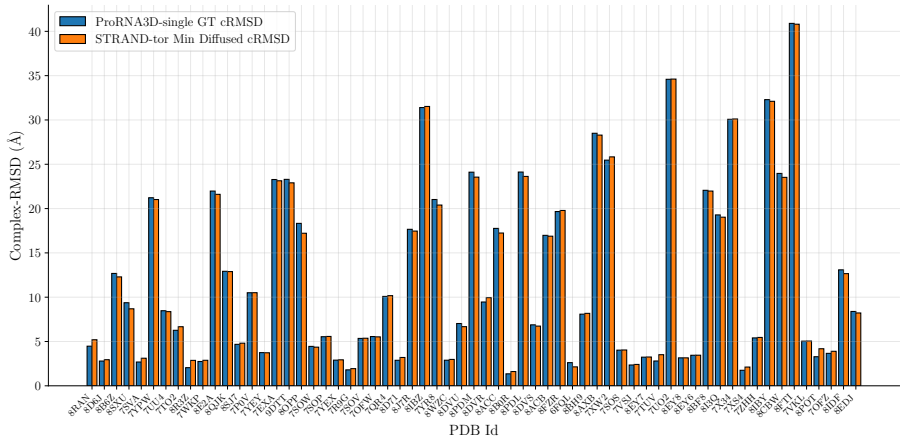

Figure 14. Per sample cRMSD of ProRNA3D-single predictions on X-ray data and the corresponding prediction of STRAND. We show results of the torsion-only model $\text{STRAND}_{\text{tor}}$.

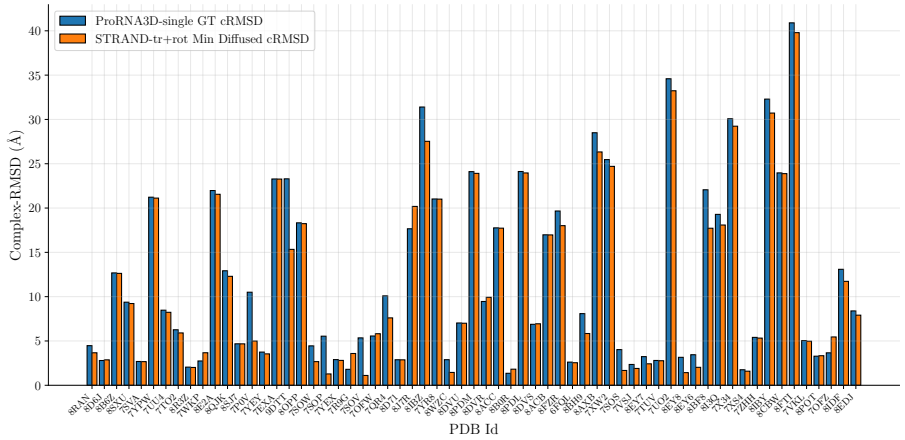

Figure 15. Per sample cRMSD of ProRNA3D-single predictions on X-ray data and the corresponding prediction of STRAND. We show results of the combined model for translation and rotation $\text{STRAND}_{\text{tr+rot}}$.

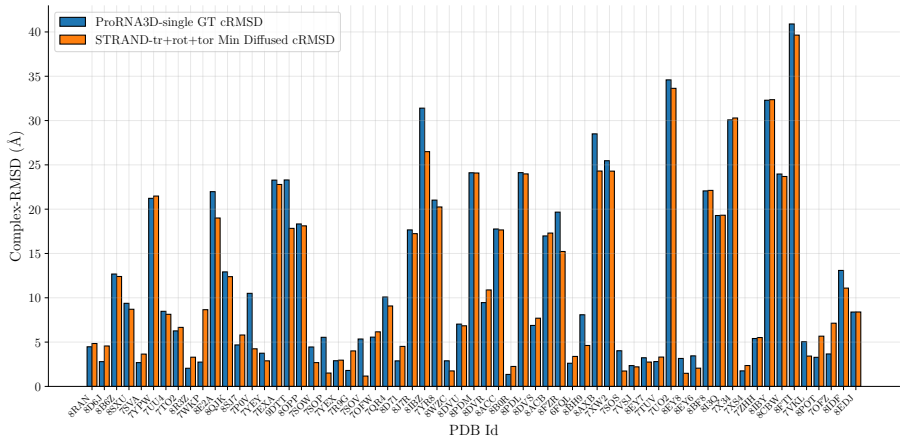

Figure 16. Per sample cRMSD of ProRNA3D-single predictions on X-ray data and the corresponding prediction of STRAND. We show results of the combined model for translation, rotation, and torsion $\text{STRAND}_{\text{tr+rot+tor}}$.