MolBindDif: Protein-conditioned RNA structure diffusion

Anonymous Authors[1]

Abstract

Despite recent significant advancements in protein structure prediction and computational biology in general, accurate prediction of the three-dimensional structures of protein-RNA complexes remains very challenging. This is largely due to the poor quality and insufficient amount of training data, mainly RNA sequences. To overcome this obstacle, we propose a new method called MolBindDiff, which relies solely on geometric information. Inspired by the success of generative models in various domains, MolBindDiff uses a diffusion model over the SE(3) space of rigid motions. Our method introduces a novel approach to representing RNA, featuring a new model architecture that processes protein and RNA embeddings separately. MolBindDiff achieves higher accuracy than the current state-of-the-art methods and is freely available to the community.

1. Introduction

In computational biology and bioinformatics, machine learning has emerged as a powerful tool for unraveling the intricate relationships within biological macromolecules and their complexes. Particularly, the study of proteins and their interactions has benefited significantly from machine learning techniques. Some illustrative examples include the AlphaFold2 architecture (AF2) (Jumper et al., 2021), RoseTTAFold (Baek et al., 2021) and more (Lin et al., 2023). Machine-learning algorithms, such as deep learning models and ensemble methods, have also demonstrated efficacy in predicting and characterizing the structure of protein-protein complexes (Evans et al., 2021), protein-ligand interactions (Corso et al., 2022; Dong et al., 2023; Torge et al., 2023; Jing et al., 2023; Koh et al., 2023), binding site predictors, such as MaSIF (Gainza et al., 2019), ScanNet (Tubiana et al., 2022), PeSTo (Krapp et al., 2023) and others (Fang et al., 2023; Zhu et al., 2023; Crouzet et al., 2023; Mou et al., 2023).

Protein-RNA interactions regulate gene expression, post-transcriptional regulation, protein synthesis, and cellular signaling (Liu et al., 2020). Understanding the complex interplay between proteins and RNA molecules provides insights into the specificity, dynamics, and function. Furthermore, understanding the structural basis of protein-RNA complexes is pivotal for drug discovery and the design of therapeutic interventions, as numerous diseases are associated with altered protein-RNA interactions (Batista & Chang, 2013). Despite the progress in structure prediction, several RNA-specific challenges persist. One notable issue related to RNA molecules lies in the dependence of RNA structure prediction models on RNA multiple sequence alignments (MSAs) (Baek et al., 2024), which can introduce biases and limitations due to the sparsity and low quality of RNA sequence data (Das et al., 2023). An additional difficulty is the relatively low number of experimentally solved RNA and protein-RNA structures. We shall stress that compared to proteins, RNA molecules are more structurally diverse, which makes it more challenging to predict their conformations.

Recently, generative models in general and diffusion models in particular have shown significant progress in various biological domains – Martinkus et al., 2023; Alamdari et al., 2023; Zhang et al., 2023a;b; Lu et al., 2023; Yim et al., 2024; Masters et al., 2023 and others. One of the models, MMDIFF (Morehead et al., 2023), is created for a task close to ours: it jointly designs sequences and structures of nucleic acid and protein complexes. However, MMDIFF is a generative model and does not profit from the knowledge ofthe protein's 3D structure. Currently, the best 3D structure prediction results of protein-RNA complexes belong to RosettaFoldNA (Baek et al., 2024).

In this work we explore a fully geometrical approach to protein-RNA structure predictions. Precosely, we combined Invariant Point Attention (IPA) (Jumper et al., 2021), Axial Attention (Ho et al., 2019), and Riemannian score-based generative modeling approach (Yim et al., 2023) for modeling protein-RNA complexes starting from protein structure and RNA sequence. Specifically, the RNA structure is modeled in the presence of the protein by a diffusion process conditioned on the protein conformations. We constructed

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

our architecture without a MSA or language models and adopted a distinct treatment for RNAs and proteins, where each entity undergoes separate embedding creation and processing within two different segments of the model. This modular approach ensures that the unique characteristics and interactions specific to RNA and protein molecules are effectively captured and processed independently, allowing a more accurate representation of the RNA-protein interactions. Another key difference is that our approach considers the protein structure to be known. Indeed, even if the structure of the target protein is not solved experimentally, it can be predicted with high accuracy by the state-of-the-art systems like AF2. As a result, we present two novel models. The first one, binding site-aware (MolBindDif-ba), assumes the protein binding site interacting with the RNA molecule to be known. This setting is highly relevant because novel machine learning methods for predicting nucleic-acid-binding residues achieve high accuracy (Xia et al., 2021; Wei et al., 2022). The second one, binding site-blind (MolBindDif-bb), is more general and does not have such a priori knowledge. As a result, we present two novel models. The first one, binding site-aware (MolBindDif-ba), assumes the protein binding site interacting with the RNA molecule to be known. This setting is highly relevant because novel machine learning methods for predicting nucleic-acid-binding residues achieve high accuracy (Xia et al., 2021; Wei et al., 2022). The second one, binding site-blind (MolBindDif-bb), is more general and does not have such a priori knowledge. The resulting models, MolBindDif-ba and MolBindDif-bb, predict RNA's 3D structure at its binding site with a protein receptor at atomic resolution.

2. Methods

2.1. Molecular representation

Building on previous approaches (Jumper et al., 2021; Baek et al., 2024), we describe RNA and protein molecules using multiple representations: the single, the pairwise, and the spatial (i.e., three-dimensional, 3D). The primary objective of the single representation is to encapsulate information about each residue individually, while the pairwise representation aims to capture details about the interactions between each pair of residues. We construct these representations using information about the residue type, its sequence index, binding indicator, and pairwise distances. We do not use the MSA representation in our method, which is very powerful for the prediction of protein structure alone but may be too sparse in the case of RNA sequences.

We represent the spatial three-dimensional positions of nucleotides in RNAs and amino acids in proteins as rigid frames (or rigid transformations), which can be seen as a tuple consisting of a 3D translation vector and a 3D rotation matrix. For the amino acids frames, we adapted the

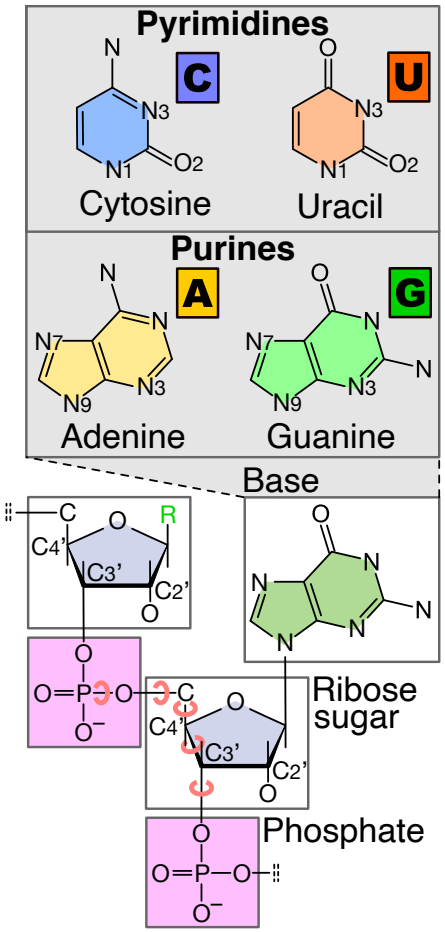

Figure 1. Schematic illustration of an RNA molecule composed of the phosphate, ribose sugar, and base groups. For the ribose representation, we built rigid frames from the C3', C2', and C4' atoms (see main text). For the base rigid frames, we use N9, N3, and N7 atoms in purines, and N1, O2, and N3 atoms in pyrimidines. The atom restoration network additionally predicts five torsion angles outlined with arcs.

AF2 strategy and built the frames accordingly, centered on the Cα atoms. Regarding the RNAs, their nucleotides are bulkier compared to amino acids. Therefore, we decided to represent each of them with two distinct frames, one corresponding to the pentose sugar and another to the nitrogenous base, as Fig. 1 shows.

The frame of the sugar centers on the C3' atoms, with Ox in the direction of C2' and Oy, such that C4' has a positive ordinate. From this frame, with the help of five additional torsion angles, outlined in Fig. 1, we can restore every other atom of the sugar and the phosphate groups. The frame of the base depends on the type of the nucleotide: adenine (A) and guanine (G) have the frame centered on N9 with Ox and Oy built from N3 and N7 directions, correspondingly; uracil (U) and cytosine (C) have the frame centered on N1

with Ox and Oy constructed from N3 and O2 directions, respectively. These frames allow us to restore other bases' atoms without additional information.

2.2. Diffusion

In our work, we decided to use a diffusional approach to network training and inference. Choosing diffusion appears intuitive, considering our objective of determining a distribution of RNA 3D structures conditioned on protein 3D structures and protein and RNA sequences. While it might seem that this distribution is highly constrained, resembling a narrow range close to the sum of delta functions – indicating that the protein unambiguously and deterministically defines the RNA structure – we suppose that diffusion remains capable of effectively reproducing such intricate relationships. Moreover, there are multiple examples of successfully trained diffusion networks published recently, where the diffusion process is conducted in a similar space (Morehead et al., 2023; Yim et al., 2023; 2024). Also, this choice is reinforced by the notion that, given its relatively small size and information volume, the network is better suited to operate with consecutive improvements rather than attempting a one-step prediction. An equivalent technique was used in the AF2 training: the same model was applied to the data multiple times, using the output of each application as input to the next one, and afterward, the gradients were backpropagated only once.

We implemented the noising and denoising diffusion processes following the approach outlined in Yim et al., 2023, separately for translations and rotations. Having the translations distribution $\mathbf{X}^{(0)}$, we can construct a time (t)–dependent process:

$$\mathrm{d}\mathbf{X}^{(t)} = f_x(t)\mathbf{X}^{(t)} + g_x(t)\mathrm{d}\mathbf{B}^{(t)}_{\mathbb{R}^3}, \tag{1}$$

where $f_x(t)$ is a drift coefficient, $f_x(t) = -\frac{1}{2}\beta(t)$ and $g_x(t)$ is a diffusion coefficient, $g_x(t) = \sqrt{\beta(t)}$ for some schedule $\beta(t)$. $\mathbf{B}^{(t)}_{\mathcal{M}}$ is the Brownian motion on manifold $\mathcal{M}$. This may be seen as a time-rescaled Ornstein–Uhlenbeck process (Song et al., 2021). Therefore, letting $G_x(t) = \int_0^t g_x(s)^2\mathrm{d}s$, we obtain a conditional probability distribution

$$p_{t|0}(\mathbf{X}^{(t)}|\mathbf{X}^{(0)}) = \\ = \mathcal{N}(\mathbf{X}^{(t)}; \exp^{-G_x(t)}\mathbf{X}^{(t)}, 1 - \exp^{-G_x(t)}), \tag{2}$$

where $\mathcal{N}$ is a normal distribution.

Similarly, for the rotations distribution $\mathbf{R}^{(t)}$,

$$\mathrm{d}\mathbf{R}^{(t)} = g_r(t)\mathrm{d}\mathbf{B}^{(t)}_{SO(3)}, \tag{3}$$

where diffusion coefficient $g_r(t) = \sqrt{\frac{d}{dt}\sigma^2(t)}$ for some noise schedule $\sigma(t)$. This gives us the following conditional distribution:

$$p_{t|0}(\mathbf{R}^{(t)}|\mathbf{R}^{(0)}) = \mathrm{IGSO}_3(\mathbf{R}^{(t)}; \mathbf{R}^{(0)}, \sigma^2(t)), \tag{4}$$

where $\mathrm{IGSO}_3(\mathbf{R}^{(t)}; \mathbf{R}^{(0)}, \sigma^2(t)) = \mathrm{IGSO}_3(\mathbf{R}^{(0)\mathrm{T}}\mathbf{R}^{(t)}, \sigma^2(t))$ is the isotropic Gaussian distribution on SO(3) (Nikolayev & Savyolov, 1997). This distribution can be parametrized in an axis-angle form, with uniformly sampled axes and rotation angle $\omega \in [0, \pi]$ with the density

$$f(\omega, t) = \sum_{l\in\mathbb{N}}(2l+1)\mathrm{e}^{-l(l+1)\sigma^2(t)/2}\frac{\sin((l+1/2)\omega)}{\sin(\omega/2)}. \tag{5}$$

The time-reversed process, corresponding to (2) and (3), is:

$$\mathrm{d}\mathbf{X}^{(t)} = (g_x(t)^2 s^x_\theta(\mathbf{X}^{(t)}, t) - f_x(t))\mathrm{d}t + \tag{6}$$

$$+\zeta g_x(t)\mathrm{d}\mathbf{B}^{(t)}_{\mathbb{R}^3}, \tag{7}$$

$$\mathrm{d}\mathbf{R}^{(t)} = g_r(t)^2 s^r_\theta(\mathbf{R}^{(t)}, t)\mathrm{d}t + \zeta g_r(t)\mathbf{B}^{(t)}_{SO(3)}, \tag{8}$$

where $s^x_\theta(\mathbf{X}^{(t)}, t)$ and $s^r_\theta(\mathbf{R}^{(t)}, t)$ are the Stein scores (Yim et al., 2023). These scores could be calculated as follows:

$$s^x_\theta(\mathbf{X}^{(t)}, t)_n = \nabla_{x_n^{(t)}}\log p_{t|0}(x_n^{(t)}|x_n^{(0)}) = \tag{9}$$

$$= -\frac{x_n^{(t)} - e^{-\frac{1}{2}\beta(t)}x_n^{(0)}}{1 - e^{\beta(t)}}, \tag{10}$$

$$s^r_\theta(\mathbf{R}^{(t)}, t)_n = \nabla_{r_n^{(t)}}\log p_{t|0}(r_n^{(t)}|r_n^{(0)}) = \tag{11}$$

$$= -\frac{r_n^{(t)}}{\omega(r_n^{(0)})}\log r_n^{(0,t)}\partial_\omega f(\omega(r_n^{(0)}), t), \tag{12}$$

where n is a number of a frame and $r_n^{(t)}$ and $x_n^{(t)}$ are frame's rotation and translation, correspondingly. Hence, with the model predicting $(r_n^{(0)}, x_n^{(0)})$ from $(r_n^{(t)}, x_n^{(t)}, t)$, it becomes possible to execute a time-reverse process, allowing the generation of samples from the distributions $\mathbf{X}^{(0)}$ and $\mathbf{R}^{(0)}$.

2.3. Model

MolBindDif comprises several main blocks – the Encoder, the RNA Block, the Protein Block, and the Atom Restoration Network, as shown in Fig. 2 and described below. The RNA Block and the Protein Block have very similar architectures (see Figs. S1 and S2 for more details). They start from the IPA blocks that produce updates for the single representations. For each type of interaction (RNA-RNA, RNA-protein, protein-RNA, and protein-protein), we have separate IPA blocks, so that the model could better catch the peculiarities of these interactions. Next, we use the Structure Transition Blocks (ResNet type of architecture), separately for RNAs and proteins, to process single representations, which are then summed with their updates and normalized. After that, we update each type of pairwise representation

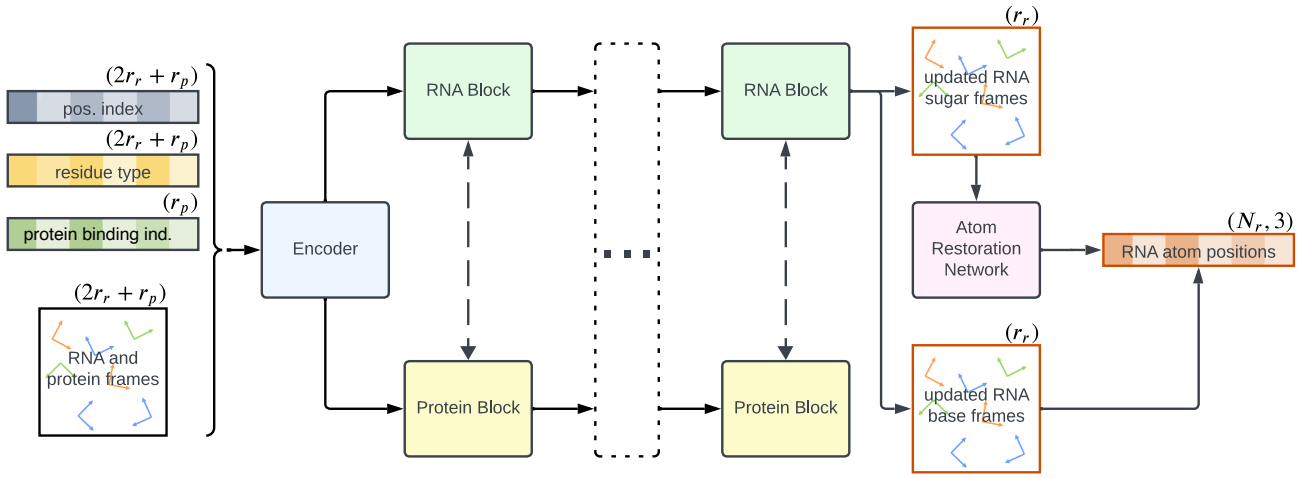

Figure 2. Schematic illustration of the MolBindDif architecture. Protein and RNA blocks are shown in more detail in Fig. S1 and Fig. S2, respectively. The atom restoration network is shown in Fig. 3. Here, r_r is the number of RNA residues, r_p is the number of protein residues, and N_r is the total number of atoms in the RNA.

with the corresponding updated single representation embeddings, using Axial Attention (separately for each type of interaction). In the RNA Block, we additionally use the Frame Update to update RNA frames. The pair of the RNA and the Protein Blocks can be repeated consequently (four times in our model). In the end, we directly restore nitrogen base atoms from the nitrogen base frames. We determine the coordinates of all other atoms from the sugar frames with the Atom Restoration Network outlined in Fig. 3.

We shall stress that the architecture of MolBindDif exhibits *equivariance* with respect to SE(3), a group of rigid rotations and translations, a property derived from the equivariance of the IPA and the Frame Update blocks. These blocks conduct all the calculations in the local frames, granting the equivariance, while the remaining blocks maintain invariance with respect to SE(3).

2.4. Encoder

For each residue (protein or RNA), we initially encode its index in the sequence using sinusoidal embeddings from (Vaswani et al., 2017) and the diffusion timestep using sinusoidal embeddings from (Ho et al., 2020). In the case of MolBindDif-ba, we then perform a one-hot encoding on binary indicators of protein residues belonging to the binding site, followed by a linear transformation. The residue types, specific to each amino acid and nucleotide's sugar and base, undergo a one-hot encoding and a subsequent linear transformation. Following these steps, we use a Multi-Layer Perceptron (MLP) to construct single representations from the concatenation of the embeddings for index, timestep, residue type, and binding site indicators (the last one only in the case of MolBindDif-ba).

We construct the pairwise representations in a similar manner - with MLPs (separate for each type of pair) and concatenate the encodings. The RNA-RNA pairwise representation incorporates a relative index and timestep encodings, while both RNA-protein and protein-RNA representations include the timestep and the binding site (for MolBindDif-ba) encodings. For the protein-protein interactions, we employ the relative sequence positional index and the Cartesian distance encodings. We binarize the latter into 22 bins (with a maximum distance of 20 Å) of relative distances. We also apply these position encodings to other pairwise representations in the case of self-conditioning.

2.5. Loss function

Having observed that the score matching loss is ineffective in training (Yim et al., 2023), we opted to implement a loss function that resembles the distogram loss in AF2,

$$L_{rr} = \sum_{i}^{N_r} \sum_{j}^{N_r} (\text{dist}(x_i, x_j) - \text{dist}(x_i^\theta, x_j^\theta))^2, \quad (13)$$

$$L_{rp} = \sum_{i}^{N_r} \sum_{j}^{N_p} (\text{dist}(x_i, x_j) - \text{dist}(x_i^\theta, x_j))^2, \quad (14)$$

$$L = L_{rr} + L_{rp}, \quad (15)$$

where L_{rr} and L_{rp} represent components of the loss associated with errors in RNA-RNA and RNA-protein relative positions, respectively. The variables x_i and x_i^θ denote the true and predicted positions of the i_{th} atom, respectively. N_r and N_p represent the total number of atoms in the RNA and the protein molecules, respectively.

2.6. Atom Restoration Network

As we cannot uniquely restore an all-atom RNA structure from the MolBindDif output, we developed an additional model for this purpose. The proposed model operates on a novel framework wherein RNA sugar frames and the nucleotide sequence serve as input, aiming to predict torsion angles required for the restoration of the positions of the remaining atoms within the sugar and phosphate groups (see Fig. 1). Leveraging the inherent structural information encoded in the sugar frames and sequence data, the model employs IPA, Axial Attention, and AngleResnet. Figure 3 schematically outlines the architecture. We trained the network on the same dataset using the same optimizer as MolBindDif. However, for the loss function, we only considered its L_{rr} component.

2.7. Training, validation and test sets

We trained MolBindDif-bb and MolBindDif-ba using protein-RNA complexes collected from the PPI3D database (Dapkunas et al., 2017). Given the potential disparity in the sequence length between full-length RNAs and their shorter protein interaction sites, we decided to truncate the RNA sequences accordingly. To do so, we selected the interacting nucleotides and their first-order and second-order neighbors according to the corresponding Voronoi graphs built with the VoroContacts software (Olechnovic & Venclovas, 2021), which describes residual interactions using Voronoi tessellation of atomic balls. These selected nucleotides constituted the *binding site nucleotides* for the model training. Restricted by the GPU memory limitations, we pruned the dataset, retaining only those complexes with a number of binding site nucleotides fewer than 100 and a protein length of fewer than 200 amino acids.

We utilized one of the PPI3D clusterisations provided to measure the homology of protein-RNA complexes. In this clusterization, proteins in one cluster have less than 40% sequence similarity with proteins in the other clusters. To ensure that our test set featured protein-RNA complexes distinct from the training data of our main baseline, RoseTTaFoldNA, we carefully selected our test data. Specifically, for the test set, we chose the complexes where the protein demonstrated less than 40% sequence identity with any entry in the Protein Data Bank (PDB) published before May 2020, which was the RoseTTaFoldNA training timestamp. We then added to the test set all other complexes from our dataset that shared the same clusters as the ones selected previously. For the training, we used the rest of the complexes from our dataset, systematically dividing them with a 90/10 split ratio during each training iteration into train and validation sets, such that complexes from one cluster are present only in one of the sets.

Ultimately, the test set consisted of 1,648 samples divided among 159 clusters; the training and validation sets comprised a total of 29,714 samples distributed across 991 clusters. The number of binding site nucleotides was evenly distributed, with a mean value of 48.2 nucleotides. The majority of proteins – over 95% had less than 100 amino acids, with a total mean value of 88.2.

To address data imbalance, where some clusters are more densely populated than others, and members within a single cluster may have identical structures (protein and/or binding site nucleotides), we adopted the random sampling strategy for each training epoch. Specifically, we randomly selected ten samples to represent each cluster (samples are repeated if the cluster size is less than 10), ensuring a more balanced data representation across different clusters.

2.8. Evaluation protocol

We compared our results with RoseTTaFoldNA, the state-of-the-art model for the prediction of protein-RNA complexes 3D structures (Baek et al., 2024). As RoseTTaFoldNA is designed to operate with complete RNA sequences, and our models only use binding site nucleotides, we created two evaluation protocols - RoseTTaFoldNA-short and RoseTTaFoldNA-long.

In both protocols, we run RoseTTaFoldNA on the same complexes from our test set. However, in the long case, we replace the binding site nucleotides with the corresponding complete RNA sequence. In our initial experiments, RoseTTaFoldNA could not process some complexes, both in the short and the long protocols. Therefore, we constructed an additional *short test set* with complexes from the test set that RoseTTaFoldNA-short successfully processed. Similarly, we constructed an additional *long test set* with complexes from the short test set that RoseTTaFoldNA-long was able to handle. In total, the short test set had 113 complexes, and the long test set had 41 complexes, with each complex belonging to a different cluster.

2.9. Training details

For the model training, we used Adam (Kingma & Ba, 2014) with the learning rate 10^{-4}, $\beta_1 = 0.9$, $\beta_1 = 0.999$, $\epsilon = 10^{-6}$ and amsgrad = True. As a learning rate scheduler, we used ExponentialLR (Li & Arora, 2019) with $\gamma = 0.9$ after each 5^{th} epoch. We trained the models on an A100 GPU using the mini-batch size of four. Each batch contained the same sample, noised with different t (t values were uniformly sampled). We train the model until both train and validation losses become stable and non-decreasing. It took 603,000 steps for MolBindDif-ba and 529,760 steps for MolBindDif-bb.

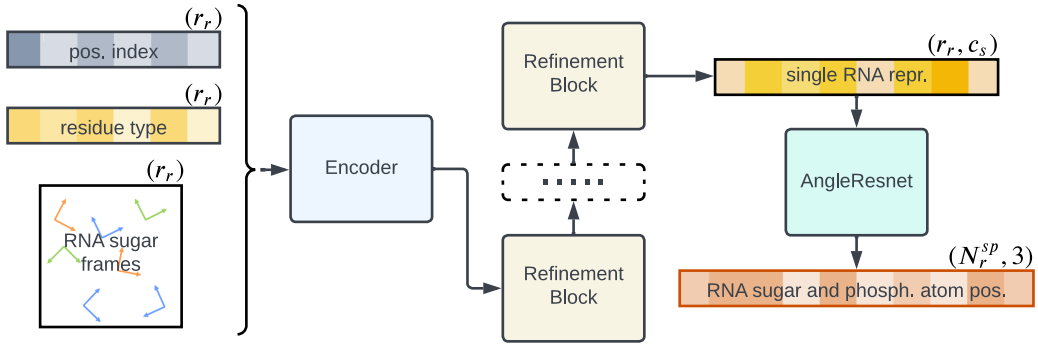

Figure 3. Schematic illustration of the Atom Restoration Network. The Refinement Block is shown in more detail in Fig. S3. Here, r_r is the number of RNA residues, c_s is the dimensionality of the single representation, and N_r^{sp} is the total number of atoms in phosphate groups and sugars of the RNA.

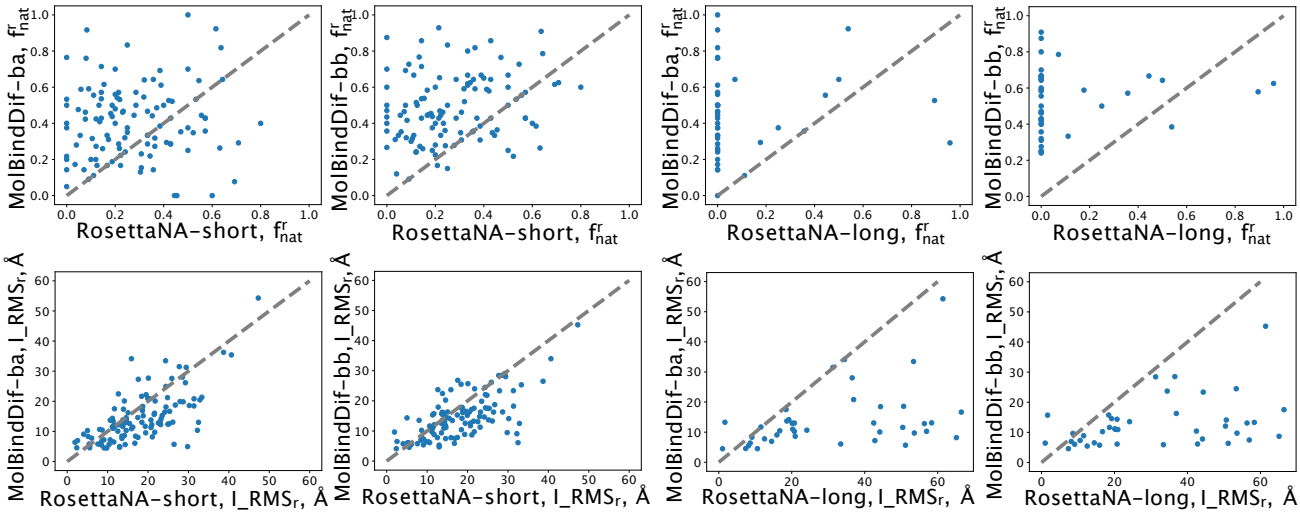

Figure 4. Scatter plots of MolBindDif versus RoseTTaFoldNA performance on multiple metrics. Top row: f^r_{nat} metrics. Bottom row: l_RMS$_r$ metrics. The leftmost two columns show MolBindDif-ba vs RoseTTaFoldNA-short and MolBindDif-bb vs RoseTTaFoldNA-short comparison. The rightmost two columns show MolBindDif-ba vs RoseTTaFoldNA-long and MolBindDif-bb vs RoseTTaFoldNA-long comparison.

3. Results

For the models' evaluation, we used multiple metrics. We can split them into three categories: superposition-free 3D structure evaluation, superposition-based 3D structure evaluation, and contacts evaluation.

For the superposition-free 3D structure evaluation, we designed lDDT-inspired metrics rRMSD$_{rr}$, rRMSD$_{rp}$, lDDT$_{rr}$, and lDDT$_{rp}$ (see Table 1) (Mariani et al., 2013). rRMSD is a mean error in the prediction of pairwise distances between atoms of different RNA nucleotides (rRMSD$_{rr}$) or between atoms of protein amino acids and RNA nucleotides (rRMSD$_{rp}$). lDDT is the average of preserved distances fractions under given thresholds – 0.5 Å, 1.0 Å, 2.0 Å, 4.0 Å. Unlike the original metric, we did not use the maximum distance cut-off of 15 Å. Practically, we calculate lDDT as rRMSD for two kinds of distances (rr and rp) described above.

For the superposition-based 3D structure evaluation, we used the Critical Assessment of PRediction of Interactions (CAPRI) metrics L_RMS and I_RMS (see Table 1) (Méndez et al., 2003). L_RMS is the root-mean-square deviation (RMSD) of RNA backbone atoms from the ground truth positions with proteins backbones being superposed. I_RMS is the RMSD of superposed backbone atoms at the protein-RNA interface. In contrast to the original CAPRI definition, which we found task-specific, we redefined amino acid-nucleotide contacts with the VoroContacts software, as those sharing common Voronoi faces. We also introduced a variation of I_RMS – I_RMS$_r$, which is calculated only

on the RNA part of the protein-RNA interface. This metric is practical if one is only interested in the prediction of the geometry of the RNA binding site alone.

For the contacts evaluation, we used the CAPRI-inspired f_{nat} and $f_{non-nat}$ metrics (Méndez et al., 2003). We define the interface as a set of nucleotide-amino-acid pairs that have a common Voronoi face, as defined above. f_{nat} is a fraction of native (correct) contacts in the predicted complex to the number of contacts in the ground-truth complex. $f_{non-nat}$ is the number of non-native (incorrect) contacts in the predicted complex divided by the total number of contacts in that complex. However, we introduced two new variations of these metrics – $f^r_{nat/non-nat}$ and $f^p_{nat/non-nat}$. $f^r_{nat/non-nat}$ are computed on only RNA nucleotides in contact with any protein amino acid and $f^p_{nat/non-nat}$ vice versa (see Table 2). We shall specifically note that in case of a small number of predicted contacts, all variations of $f_{non-nat}$ and I_RMS are not representative.

Figure S4 in SI shows several examples of predictions with the corresponding f^r_{nat} and f^p_{nat} scores for the four methods, MolBindDif-ba, MolBindDif-bb, RoseTTaFoldNA-short, and RoseTTaFoldNA-long. For better rendering in PyMol, we relaxed MolBindDif-ba/bb predictions with OpenMM (Eastman et al., 2017) using the Amber14 force field to improve the local geometry of nucleotides. We computed all the scores for the unrelaxed structures. We can see in this figure that sometimes RoseTTaFoldNA-long produced highly accurate models, but in 5 out of 8 cases, it completely misplaced the binding site nucleotides ($f^r_{nat} = 0$). RoseTTaFoldNA-short produced structures with non-zero f^r_{nat} and f^p_{nat} scores more often than RoseTTaFoldNA-long. However, for 6 out of 8 given example tasks, higher scores were achieved by MolBindDif-ba/bb. MolBindDif-bb always had higher f^r_{nat} scores than MolBindDif-ba, but it also produced visibly more clashes between the protein and the RNA residues.

Tables 1 and 2 list the comparison between MolBindDif-ba, MolBindDif-bb, and the RoseTTaFoldNA-short protocol on the short test set. In its turn, Table 3 compares four methods, MolBindDif-ba, MolBindDif-bb, RoseTTaFoldNA-short, and RoseTTaFoldNA-long, on the long test set.

We may see that MolBindDif-ba, our model informed by the locations of the protein binding residues, performs the best on nearly every metric. The blind interface version, MolBindDif-bb, has the highest f^r_{nat} on both test sets. However, its f^r_{nonnat} is worse than the one of MolBindDif-ba and RoseTTaFoldNA-long. Compared to RoseTTaFoldNA-short, which has no information about interface contacts, MolBindDif-bb performs better on every metric.

Results in Table 3 suggest that RoseTTaFoldNA-long has, on average, lower performance than RoseTTaFoldNA-short on I_RMS, f^r_{nat} and f^p_{nat} metrics. Its $f^r_{non-nat}$ and $f^p_{non-nat}$ values are relatively low because the model predicts fewer contacts in total. RoseTTaFoldNA-long demonstrates I_RMS scores on the same level (even slightly better) with MolBindDif-ba and MolBindDif-bb ones.

As the mean values in the tables above may not be very illustrative, it is also very useful to compare the performance of different models on each example individually from the test sets. Figure 4 shows scatter plots of f^r_{nat} and I_RMS$_r$ scores of MolBindDif-ba and MolBindDif-bb plotted against those of RoseTTaFoldNA-short and RoseTTaFoldNA-long. We can see that the f^r_{nat} values of the MolBindDif model lie above the diagonal (so they are better), while I_RMS$_r$ values are more frequently located below the diagonal (so they are better). Supplementary figures S5-S19 provide comparison of other score metrics.

4. Ablation studies

4.1. New schedulers

As the noising schedules, described by Yim et al., 2023, have demonstrated imperfect results for the diffusion on some toy distributions, we have introduced a new noise schedule for the diffusion in $\mathbb{R}_3$: $\beta(t) = 64t^{15} + 4t^3 + 2t$. We also changed the coefficients from (Yim et al., 2023) in the SE$_3$ noise schedule: $\sigma(t) = \log(te^{1.0} - (1-t)e^{0.01})$. These new schedules have demonstrated significant improvements in the diffusion results on the toy examples and led to more stable training of our main models. Figure 5 shows the comparison between the original and improved schedules.

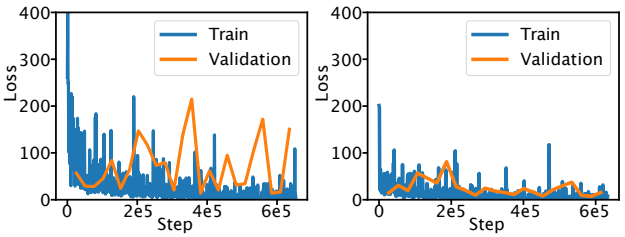

Figure 5. Loss of MolBindDif-ba for the old (left) and new (right) schedule versions. The values of the loss function (y-axis) as a function of the training step (x-axis) is shown.

4.2. Architecture

We have tested several other possible architectures. The first was with RNAs and proteins processed together. The second architecture was with protein representations not updated after the Encoder. The third and fourth architectures used all pairwise representations united and processed with the Axial Attention or the pair stack from AF2, respectively. All of these models have demonstrated weak or

Model	rRMSD$_{rr}$, Å	rRMSD$_{rp}$, Å	lDDT$_{rr}$	lDDT$_{rp}$	L_RMS, Å	I_RMS, Å	I_RMS$_r$, Å
MolBindDif-ba	**10.2 ± 6.2**	**9.9 ± 7.4**	**0.18 ± 0.07**	**0.16 ± 0.07**	**25.1 ± 12.1**	**15.8 ± 8.9**	**14.7 ± 8.3**
MolBindDif-bb	10.7 ± 5.5	14.1 ± 7.8	**0.17 ± 0.06**	0.10 ± 0.05	36.4 ± 11.1	**16.5 ± 7.1**	14.6 ± 7.3
RoseTTaFoldNA-short	13.3 ± 6.3	17.6 ± 7.8	**0.19 ± 0.10**	0.08 ± 0.04	43.2 ± 15.5	20.5 ± 7.9	18.1 ± 8.6

Table 1. Comparison of MolBindDif-ba, MolBindDif-bb, and RoseTTaFoldNA-short on 113 samples with less than 40% protein sequence identity from the short test set. Superposition-free and superposition-based metrics are shown. Best values are highlighted in bold. For all metrics measured in Å, the lower values are better, and vice versa for the lDDT metrics.

Model	f_{nat}	f^r_{nat}	f^p_{nat}	$f_{non-nat}$	$f^r_{non-nat}$	$f^p_{non-nat}$
MolBindDif-ba	**0.10 ± 0.11**	0.41 ± 0.20	**0.59 ± 0.18**	**0.83 ± 0.19**	**0.44 ± 0.23**	**0.15 ± 0.12**
MolBindDif-bb	0.04 ± 0.07	**0.49 ± 0.18**	0.31 ± 0.23	0.96 ± 0.08	0.48 ± 0.17	0.68 ± 0.26
RoseTTaFoldNA-short	0.01 ± 0.02	0.27 ± 0.19	0.20 ± 0.21	0.99 ± 0.03	0.63 ± 0.20	0.75 ± 0.26

Table 2. Comparison of MolBindDif-ba, MolBindDif-bb, and RoseTTaFoldNA-short on contact-based metrics computed with 113 samples with less than 40% protein sequence identity from the short test set. Best values are highlighted in bold. For the nat metrics, the higher values are better; for the $non - nat$ metrics, the lower values are better.

Model	I_RMS, Å	I_RMS$_r$, Å	f^r_{nat}	$f^r_{non-nat}$	f^p_{nat}	$f^p_{non-nat}$
MolBindDif-ba	**15.2 ± 10.6**	13.9 ± 9.9	0.45 ± 0.23	0.41 ± 0.22	**0.55 ± 0.18**	**0.16 ± 0.13**
MolBindDif-bb	**15.3 ± 8.2**	**12.8 ± 8.1**	**0.53 ± 0.17**	0.48 ± 0.18	0.26 ± 0.24	0.75 ± 0.25
RoseTTaFoldNA-short	18.7 ± 9.1	15.9 ± 9.7	0.33 ± 0.22	0.61 ± 0.22	0.25 ± 0.22	0.71 ± 0.27
RoseTTaFoldNA-long	**14.9 ± 11.1**	31.0 ± 19.1	0.10 ± 0.23	**0.39 ± 0.44**	0.19 ± 0.29	0.32 ± 0.37

Table 3. Comparison between MolBindDif-ba, MolBindDif-bb, RoseTTaFoldNA-short, and RoseTTaFoldNA-long on superposition-free and contact-based metrics computed with 41 samples with less than 40% protein sequence identity from the long test set. Best values are highlighted in bold. For the RMS metrics, the lower values are better. For the contact nat metrics, the higher values are better. For the contact $non - nat$ metrics, the lower values are better.

absent convergence during training.

5. Code availability

The model, along with the pretrained parameters, is available at https://github.com/icml2024submission/MolBindDif. We also provide scripts to preprocess input data.

6. Conclusion

This paper presents a novel method to predict three-dimensional structures of protein-RNA complexes. Our model, MolBindDiff, surpasses current state-of-the-art methods across various metrics. A key feature of our approach is its independence from Multiple Sequence Alignment data and any evolutionary information in general, such that it solely relies on protein geometry and sequences of the partners. This aspect enables the application of our model in scenarios with the absence of homologous sequences.

7. Broader impact

Applications of this method may have social and industrial benefits. Potential applications include in-silico vaccine design, the development of novel antigens, and some other pharmaceutical tasks.

References

Alamdari, S., Thakkar, N., Van Den Berg, R., Lu, A. X., Fusi, N., Amini, A. P., and Yang, K. K. Protein generation with evolutionary diffusion: sequence is all you need. preprint, Bioengineering, September 2023. URL http://biorxiv.org/lookup/doi/10.1101/2023.09.11.556673.

Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G. R., Wang, J., Cong, Q., Kinch, L. N., Schaeffer, R. D., Millán, C., Park, H., Adams, C., Glassman, C. R., DeGiovanni, A., Pereira, J. H., Rodrigues, A. V., Van Dijk, A. A., Ebrecht, A. C., Opperman, D. J., Sagmeister, T., Buhlheller, C., Pavkov-Keller, T., Rathinaswamy, M. K., Dalwadi, U., Yip, C. K., Burke, J. E., Garcia, K. C., Grishin, N. V., Adams, P. D., Read, R. J., and Baker, D. Accurate prediction of protein structures and interactions using a three-track neural network. *Science*, 373(6557):871–876, August 2021. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.abj8754. URL https://www.science.org/doi/10.1126/science.abj8754.

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, January 2024. ISSN 1548-7091, 1548-7105. doi: 10.1038/s41592-023-02086-5. URL https://www.nature.com/articles/s41592-023-02086-5.

Batista, P. and Chang, H. Long Noncoding RNAs: Cellular Address Codes in Development and Disease. *Cell*, 152(6):1298–1307, March 2013. ISSN 00928674. doi: 10.1016/j.cell.2013.02.012. URL https://linkinghub.elsevier.com/retrieve/pii/S0092867413002018.

Corso, G., Stärk, H., Jing, B., Barzilay, R., and Jaakkola, T. DiffDock: Diffusion Steps, Twists, and Turns for Molecular Docking. 2022. doi: 10.48550/ARXIV.2210.01776. URL https://arxiv.org/abs/2210.01776. Publisher: arXiv Version Number: 2.

Crouzet, S. J., Lieberherr, A. M., Atz, K., Nilsson, T., Sach-Peltason, L., Müller, A. T., Peraro, M. D., and Zhang, J. D. G- *PLIP* : Knowledge graph neural network for structure-free protein-ligand bioactivity prediction. preprint, Bioinformatics, September 2023. URL http://biorxiv.org/lookup/doi/10.1101/2023.09.01.555977.

Dapkunas, J., Timinskas, A., Olechnovic, K., Margelevicius, M., Diciūnas, R., and Venclovas, C. The PPI3D web server for searching, analyzing and modeling protein–protein interactions in the context of 3D structures. *Bioinformatics*, 33(6):935–937, March 2017. ISSN 1367-4803, 1367-4811. doi: 10.1093/bioinformatics/btw756. URL https://academic.oup.com/bioinformatics/article/33/6/935/2585028.

Das, R., Kretsch, R. C., Simpkin, A. J., Mulvaney, T., Pham, P., Rangan, R., Bu, F., Keegan, R. M., Topf, M., Rigden, D. J., Miao, Z., and Westhof, E. Assessment of three-dimensional RNA structure prediction in CASP15. preprint, Biophysics, April 2023. URL http://biorxiv.org/lookup/doi/10.1101/2023.04.25.538330.

Dong, T., Yang, Z., Zhou, J., and Chen, C. Y.-C. Equivariant Flexible Modeling of the Protein–Ligand Binding Pose with Geometric Deep Learning. *Journal of Chemical Theory and Computation*, 19(22):8446–8459, November 2023. ISSN 1549-9618, 1549-9626. doi: 10.1021/acs.jctc.3c00273. URL https://pubs.acs.org/doi/10.1021/acs.jctc.3c00273.

Eastman, P., Swails, J., Chodera, J. D., McGibbon, R. T., Zhao, Y., Beauchamp, K. A., Wang, L.-P., Simmonett, A. C., Harrigan, M. P., Stern, C. D., Wiewiora, R. P., Brooks, B. R., and Pande, V. S. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLOS Computational Biology*, 13(7):e1005659, July 2017. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1005659. URL https://dx.plos.org/10.1371/journal.pcbi.1005659.

Evans, R., O'Neill, M., Pritzel, A., Antropova, N., Senior, A., Green, T., Žídek, A., Bates, R., Blackwell, S., Yim, J., Ronneberger, O., Bodenstein, S., Zielinski, M., Bridgland, A., Potapenko, A., Cowie, A., Tunyasuvunakool, K., Jain, R., Clancy, E., Kohli, P., Jumper, J., and Hassabis, D. Protein complex prediction with AlphaFold-Multimer. preprint, Bioinformatics, October 2021. URL http://biorxiv.org/lookup/doi/10.1101/2021.10.04.463034.

Fang, Y., Jiang, Y., Wei, L., Ma, Q., Ren, Z., Yuan, Q., and Wei, D.-Q. DeepProSite: structure-aware protein binding site prediction using ESMFold and pretrained language model. *Bioinformatics*, 39(12):btad718, December 2023. ISSN 1367-4811. doi: 10.1093/bioinformatics/btad718. URL https://academic.oup.com/bioinformatics/article/doi/10.1093/bioinformatics/btad718/7453375.

Gainza, P., Sverrisson, F., Monti, F., Rodola, E., Bronstein, M. M., and Correia, B. E. MaSIF - Deciphering interaction fingerprints from protein molecular surfaces. April 2019. doi: 10.5281/ZENODO.2625420. URL https://zenodo.org/record/2625420. Publisher: Zenodo.

Ho, J., Kalchbrenner, N., Weissenborn, D., and Salimans, T. Axial Attention in Multidimensional Transformers. 2019. doi: 10.48550/ARXIV.1912.12180. URL https://arxiv.org/abs/1912.12180. Publisher: arXiv Version Number: 1.

Ho, J., Jain, A., and Abbeel, P. Denoising Diffusion Probabilistic Models. 2020. doi: 10.48550/ARXIV.2006.11239. URL https://arxiv.org/abs/2006.11239. Publisher: arXiv Version Number: 2.

Jing, B., Erives, E., Pao-Huang, P., Corso, G., Berger, B., and Jaakkola, T. EigenFold: Generative Protein Structure Prediction with Diffusion Models, April 2023. URL http://arxiv.org/abs/2304.02198. arXiv:2304.02198 [physics, q-bio].

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver,

D., Vinyals, O., Senior, A. W., Kavukcuoglu, K., Kohli, P., and Hassabis, D. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873):583–589, August 2021. ISSN 0028-0836, 1476-4687. doi: 10.1038/s41586-021-03819-2. URL https://www.nature.com/articles/s41586-021-03819-2.

Kingma, D. P. and Ba, J. Adam: A Method for Stochastic Optimization. 2014. doi: 10.48550/ARXIV.1412.6980. URL https://arxiv.org/abs/1412.6980. Publisher: arXiv Version Number: 9.

Koh, H. Y., Nguyen, A. T., Pan, S., May, L. T., and Webb, G. I. PSICHIC: physicochemical graph neural network for learning protein-ligand interaction fingerprints from sequence data. preprint, Bioinformatics, September 2023. URL http://biorxiv.org/lookup/doi/10.1101/2023.09.17.558145.

Krapp, L. F., Abriata, L. A., Cortés Rodriguez, F., and Dal Peraro, M. PeSTo: parameter-free geometric deep learning for accurate prediction of protein binding interfaces. *Nature Communications*, 14(1): 2175, April 2023. ISSN 2041-1723. doi: 10.1038/s41467-023-37701-8. URL https://www.nature.com/articles/s41467-023-37701-8.

Li, Z. and Arora, S. An Exponential Learning Rate Schedule for Deep Learning. 2019. doi: 10.48550/ARXIV.1910.07454. URL https://arxiv.org/abs/1910.07454. Publisher: arXiv Version Number: 3.

Lin, Z., Akin, H., Rao, R., Hie, B., Zhu, Z., Lu, W., Smetanin, N., Verkuil, R., Kabeli, O., Shmueli, Y., Dos Santos Costa, A., Fazel-Zarandi, M., Sercu, T., Candido, S., and Rives, A. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 379(6637):1123–1130, March 2023. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.ade2574. URL https://www.science.org/doi/10.1126/science.ade2574.

Liu, S., Li, B., Liang, Q., Liu, A., Qu, L., and Yang, J. Classification and function of RNA –protein interactions. *WIREs RNA*, 11(6):e1601, November 2020. ISSN 1757-7004, 1757-7012. doi: 10.1002/wrna.1601. URL https://wires.onlinelibrary.wiley.com/doi/10.1002/wrna.1601.

Lu, J., Zhong, B., and Tang, J. Score-based Enhanced Sampling for Protein Molecular Dynamics, June 2023. URL http://arxiv.org/abs/2306.03117. arXiv:2306.03117 [cs, q-bio].

Mariani, V., Biasini, M., Barbato, A., and Schwede, T. lDDT: a local superposition-free score for comparing protein structures and models using distance difference tests. *Bioinformatics*, 29(21):2722–2728, November 2013. ISSN 1367-4803, 1367-4811. doi: 10.1093/bioinformatics/btt473. URL https://academic.oup.com/bioinformatics/article/29/21/2722/195896.

Martinkus, K., Ludwiczak, J., Cho, K., Liang, W.-C., Lafrance-Vanasse, J., Hotzel, I., Rajpal, A., Wu, Y., Bonneau, R., Gligorijevic, V., and Loukas, A. AbDiffuser: Full-Atom Generation of In-Vitro Functioning Antibodies, July 2023. URL http://arxiv.org/abs/2308.05027. arXiv:2308.05027 [cs, q-bio, stat].

Masters, M. R., Mahmoud, A. H., Wei, Y., and Lill, M. A. Deep Learning Model for Efficient Protein–Ligand Docking with Implicit Side-Chain Flexibility. *Journal of Chemical Information and Modeling*, 63(6):1695–1707, March 2023. ISSN 1549-9596, 1549-960X. doi: 10.1021/acs.jcim.2c01436. URL https://pubs.acs.org/doi/10.1021/acs.jcim.2c01436.

Morehead, A., Ruffolo, J., Bhatnagar, A., and Madani, A. Towards Joint Sequence-Structure Generation of Nucleic Acid and Protein Complexes with SE(3)-Discrete Diffusion, December 2023. URL http://arxiv.org/abs/2401.06151. arXiv:2401.06151 [cs, q-bio].

Mou, M., Pan, Z., Zhou, Z., Zheng, L., Zhang, H., Shi, S., Li, F., Sun, X., and Zhu, F. A Transformer-Based Ensemble Framework for the Prediction of Protein–Protein Interaction Sites. *Research*, 6:0240, January 2023. ISSN 2639-5274. doi: 10.34133/research.0240. URL https://spj.science.org/doi/10.34133/research.0240.

Méndez, R., Leplae, R., De Maria, L., and Wodak, S. J. Assessment of blind predictions of protein–protein interactions: Current status of docking methods. *Proteins: Structure, Function, and Bioinformatics*, 52(1): 51–67, July 2003. ISSN 0887-3585, 1097-0134. doi: 10.1002/prot.10393. URL https://onlinelibrary.wiley.com/doi/10.1002/prot.10393.

Nikolayev, D. I. and Savyolov, T. I. Normal Distribution on the Rotation Group So(3). *Textures and Microstructures*, 29(3-4):201–233, January 1997. ISSN 0730-3300, 1029-4961. doi: 10.1155/TSM.29.201. URL https://www.hindawi.com/journals/tsm/1997/173236/abs/.

Olechnovic, K. and Venclovas, C. VoroContacts: a tool for the analysis of interatomic contacts in macromolecular structures. *Bioinformatics*, 37(24):4873–4875, December 2021. ISSN 1367-4803, 1367-4811.

doi: 10.1093/bioinformatics/btab448. URL https://academic.oup.com/bioinformatics/article/37/24/4873/6300513.

Song, Y., Sohl-Dickstein, J., Kingma, D. P., Kumar, A., Ermon, S., and Poole, B. Score-Based Generative Modeling through Stochastic Differential Equations, February 2021. URL http://arxiv.org/abs/2011.13456. arXiv:2011.13456 [cs, stat].

Torge, J., Harris, C., Mathis, S. V., and Lio, P. DiffHopp: A Graph Diffusion Model for Novel Drug Design via Scaffold Hopping, August 2023. URL http://arxiv.org/abs/2308.07416. arXiv:2308.07416 [q-bio].

Tubiana, J., Schneidman-Duhovny, D., and Wolfson, H. J. ScanNet: an interpretable geometric deep learning model for structure-based protein binding site prediction. *Nature Methods*, 19(6):730–739, June 2022. ISSN 1548-7091, 1548-7105. doi: 10.1038/s41592-022-01490-7. URL https://www.nature.com/articles/s41592-022-01490-7.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. Attention Is All You Need. 2017. doi: 10.48550/ARXIV.1706.03762. URL https://arxiv.org/abs/1706.03762. Publisher: arXiv Version Number: 7.

Wei, J., Chen, S., Zong, L., Gao, X., and Li, Y. Protein–RNA interaction prediction with deep learning: structure matters. *Briefings in Bioinformatics*, 23(1):bbab540, January 2022. ISSN 1467-5463, 1477-4054. doi: 10.1093/bib/bbab540. URL https://academic.oup.com/bib/article/doi/10.1093/bib/bbab540/6470965.

Xia, Y., Xia, C.-Q., Pan, X., and Shen, H.-B. GraphBind: protein structural context embedded rules learned by hierarchical graph neural networks for recognizing nucleic-acid-binding residues. *Nucleic Acids Research*, 49(9):e51–e51, May 2021. ISSN 0305-1048, 1362-4962. doi: 10.1093/nar/gkab044. URL https://academic.oup.com/nar/article/49/9/e51/6134185.

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, May 2023. URL http://arxiv.org/abs/2302.02277. arXiv:2302.02277 [cs, q-bio, stat].

Yim, J., Campbell, A., Mathieu, E., Foong, A. Y. K., Gastegger, M., Jiménez-Luna, J., Lewis, S., Satorras, V. G., Veeling, B. S., Noé, F., Barzilay, R., and Jaakkola, T. S. Improved motif-scaffolding with SE(3) flow matching, January 2024. URL http://arxiv.org/abs/2401.04082. arXiv:2401.04082 [cs, q-bio, stat].

Zhang, C., Leach, A., Makkink, T., Arbesú, M., Kadri, I., Luo, D., Mizrahi, L., Krichen, S., Lang, M., Tovchigrechko, A., Lopez Carranza, N., Şahin, U., Beguir, K., Rooney, M., and Fu, Y. FrameDiPT: SE(3) Diffusion Model for Protein Structure Inpainting. preprint, Immunology, November 2023a. URL http://biorxiv.org/lookup/doi/10.1101/2023.11.21.568057.

Zhang, Y., Ma, Z., and Gong, H. TopoDiff: Improving Protein Backbone Generation with Topology-aware Latent Encoding. preprint, Bioengineering, December 2023b. URL http://biorxiv.org/lookup/doi/10.1101/2023.12.13.571602.

Zhu, C., Zhang, C., Shang, T., Zhang, C., Zhai, S., Su, Z., and Duan, H. GAPS: Geometric Attention-based Networks for Peptide Binding Sites Identification by the Transfer Learning Approach. preprint, Bioinformatics, December 2023. URL http://biorxiv.org/lookup/doi/10.1101/2023.12.26.573336.

A. Supplementary Information

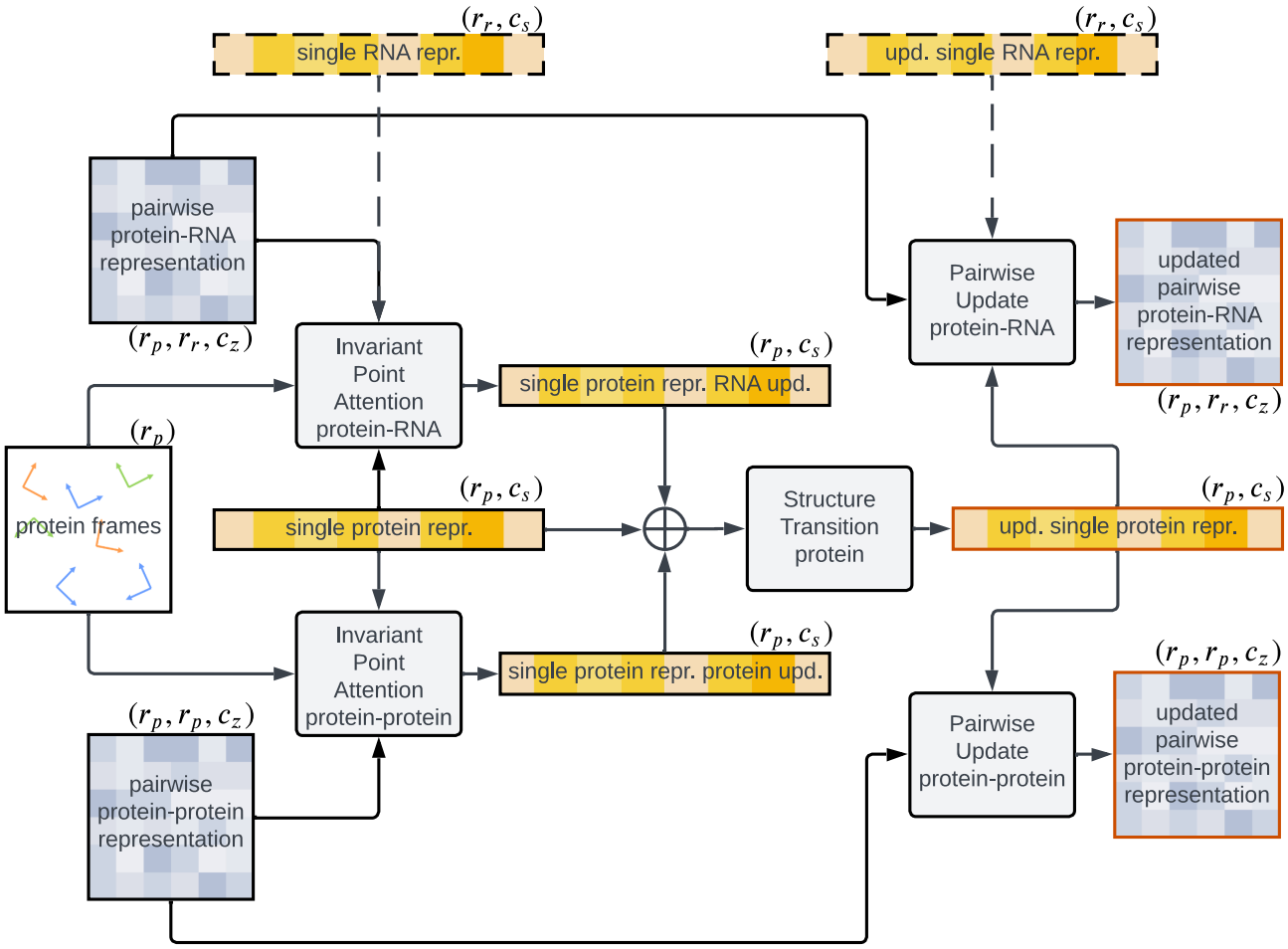

Figure S1. Architecture of the Protein Block. Here, r_r is the number of RNA residues, r_p is the number of protein residues, c_s is the dimensionality of the single representation, and c_s is the dimensionality of the pair representation.

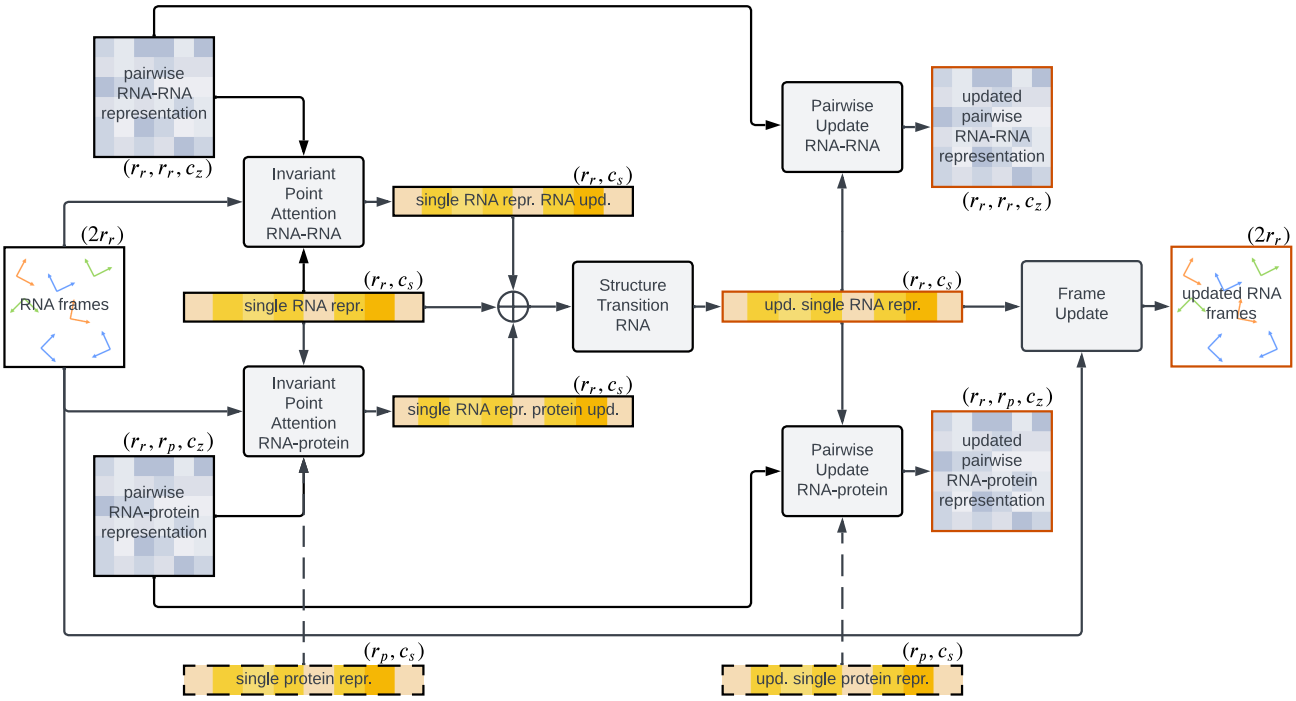

Figure S2. Architecture of the RNA Block. Here, r_r is the number of RNA residues, r_p is the number of protein residues, c_s is the dimensionality of the single representation, and c_s is the dimensionality of the pair representation.

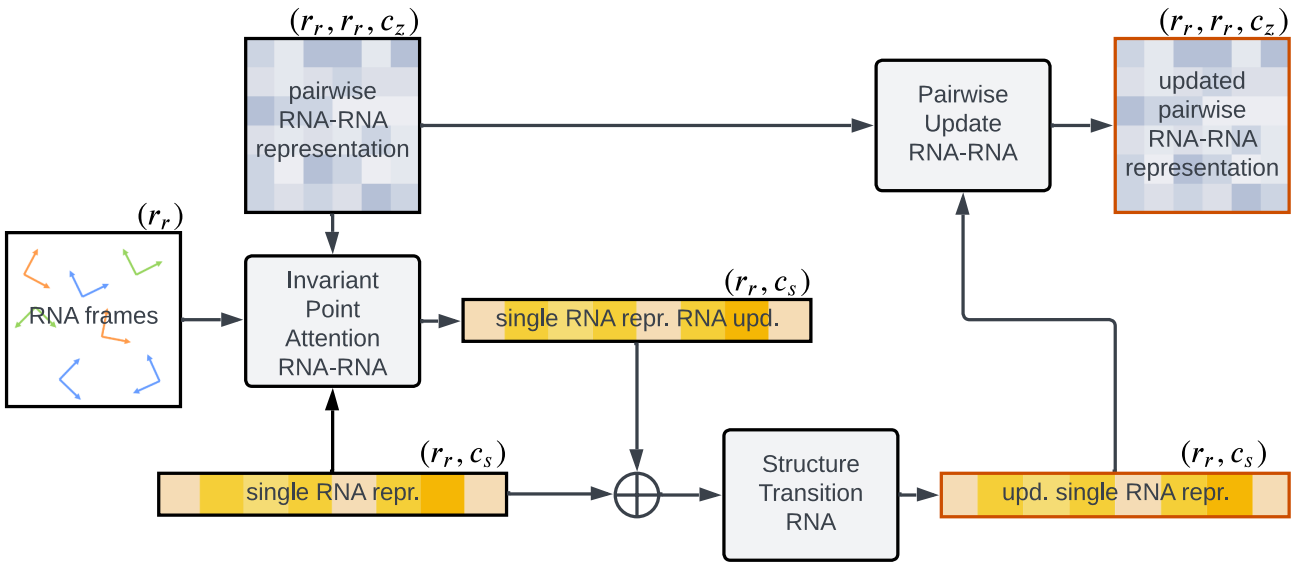

Figure S3. Architecture of the Refinement Block. Here, r_r is the number of RNA residues, c_s is the dimensionality of the single representation, c_s is the dimensionality of the pair representation.

Figure S4. Examples of MolBindDif-ba, MolBindDif-bb, RoseTTaFoldNA-short, and RoseTTaFoldNA-long predictions with the corresponding f_{nat}^{r} and f_{nat}^{p} scores. Protein chains are colored in green, RNA — in magenta for the bases and riboses, in orange for the cartoon trace. The examples were chosen to illustrate how MolBindDif-ba and MolBindDif-bb perform in cases where RoseTTaFoldNA-long or RoseTTaFoldNA-short produce either high-accuracy or completely incorrect predictions. The rows are ordered by the size of the ground truth structure, from smaller to larger. 3D visualizations were rendered using PyMol. To avoid breaking the cartoon representations, we relaxed MolBindDif-ba/bb predictions with OpenMM using the Amber14 force field to improve the local geometry of nucleotides.

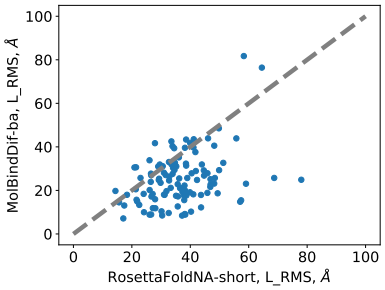

Figure S5. L_RMS, MolBindDif-ba vs RoseTTaFoldNA.

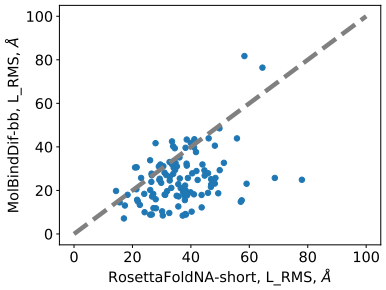

Figure S6. L_RMS, MolBindDif-bb vs RoseTTaFoldNA.

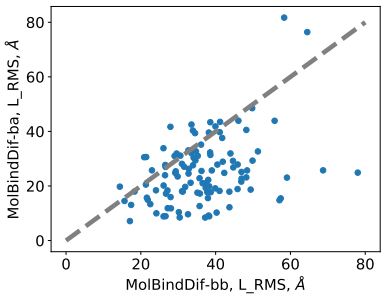

Figure S7. L_RMS, MolBindDif-ba vs MolBindDif-bb

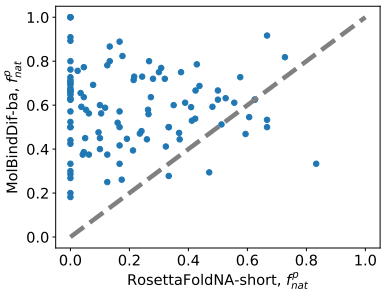

Figure S8. f^p_{nat}, MolBindDif-ba vs RoseTTaFoldNA.

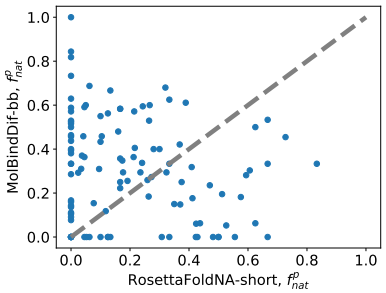

Figure S9. f^p_{nat}, MolBindDif-bb vs RoseTTaFoldNA.

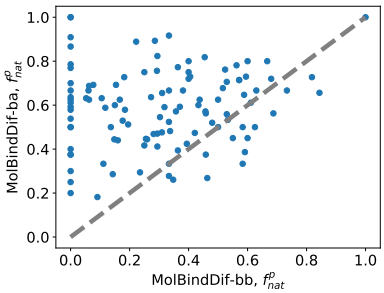

Figure S10. f^p_{nat}, MolBindDif-ba vs MolBindDif-bb

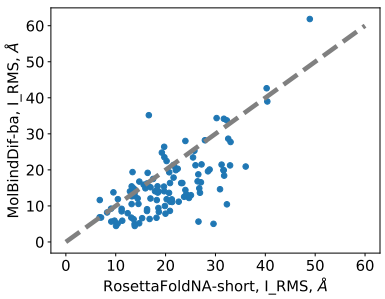

Figure S11. I_RMS, MolBindDif-ba vs RoseTTaFoldNA.

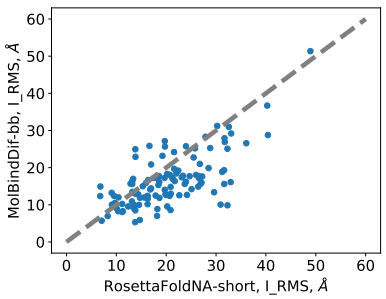

Figure S12. I_RMS, MolBindDif-bb vs RoseTTaFoldNA.

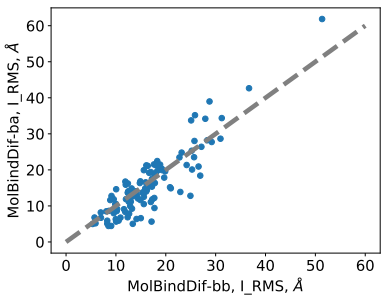

Figure S13. I_RMS, MolBindDif-ba vs MolBindDif-bb

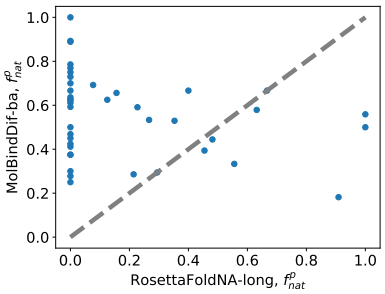

Figure S14. f^p_{nat}, MolBindDif-ba vs RoseTTaFoldNA-long.

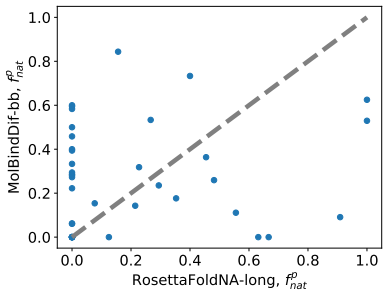

Figure S15. f^p_{nat}, MolBindDif-bb vs RoseTTaFoldNA-long.

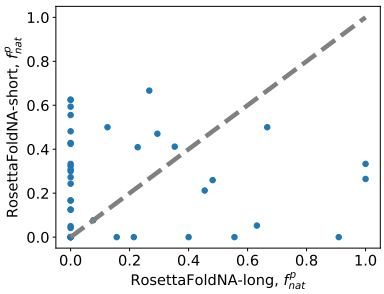

Figure S16. f^p_{nat}, RoseTTaFoldNA vs RoseTTaFoldNA-long

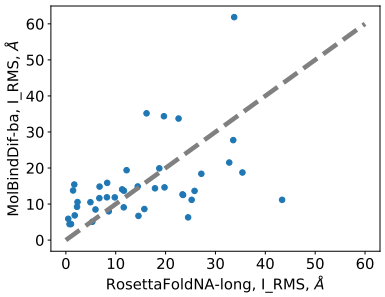

Figure S17. I_RMS, MolBindDif-ba vs RoseTTaFoldNA-long.

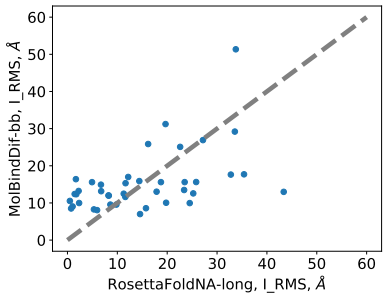

Figure S18. I_RMS, MolBindDif-bb vs RoseTTaFoldNA-long.

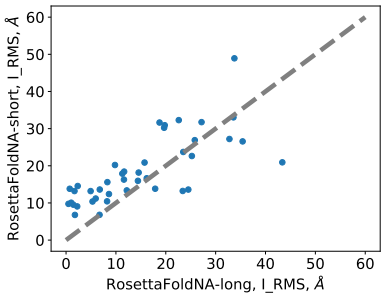

Figure S19. I_RMS, RoseTTaFoldNA vs RoseTTaFoldNA-long