Name of student:

Derek Peng

School:

The Pingry School

City, State, Country:

Basking Ridge, New Jersey, USA

Name of instructor:

Morgan D'Ausilio

School:

The Pingry School

City, State, Country:

Basking Ridge, New Jersey, USA

Title of Research Report

A Hybrid Deep Learning Pipeline for RNA 3D Structure Reconstruction from Cryo-EM Density Maps

A Hybrid Deep Learning Pipeline for RNA 3D Structure Reconstruction from Cryo-EM Density Maps

Derek Peng[1]

[1]The Pingry School, Basking Ridge, New Jersey, USA

Abstract

Determining RNA three-dimensional structures is critical for understanding complex biological functions and developing targeted therapeutics. However, obtaining high-resolution data using traditional approaches is notoriously difficult, leading to data scarcity that severely hinders RNA research. While cryogenic electron microscopy (cryo-EM) has revolutionized structure determination by improving resolution, converting cryo-EM maps into accurate RNA all-atom models remains a major computational bottleneck requiring manual tracing. **We present RNATrace, a fully automated hybrid pipeline that reconstructs RNA all-atom models from cryo-EM maps in three stages: (i) voxel-wise segmentation, (ii) backbone construction, and (iii) all-atom reconstruction/refinement.** For segmentation, we design a Swin U-Net Transformer to perform a two-step approach that first separates RNA density from background and then classifies voxels into phosphate, ribose sugar, purine, and pyrimidine groups. For backbone construction, predicted phosphorus atoms are ordered into a single continuous Hamiltonian path, whose direction and sequence registration are then determined using sequence, secondary structure, purine/pyrimidine, and fragment anchor constraints. For reconstruction and refinement, Nucleotide templates then expand the registered trace into an all-atom model, followed by restrained geometric refinement. On a 20-map test set comprising RNAs 119–423 nt in length and cryo-EM maps at 3.0–6.0 Å resolution, RNATrace achieved a mean phosphorus-position RMSD of 2.91 Å ($p < 0.05$), a mean TM-score of 0.707 ($p < 0.0001$), and a mean GDT-TS of 63.5% ($p < 0.0001$), statistically significant improvements over current state-of-the-art models. **This hybrid pipeline enables accurate end-to-end reconstruction of continuous RNA structures, helping accelerate RNA structure determination and structure-based drug discovery.**

Keywords: RNA 3D structure reconstruction; Cryo-EM density maps; Deep learning; Swin U-Net Transformer; Hamiltonian path

Acknowledgement

My interest in working on this research project stems from my amazement at AlphaFold2's [1] ability to predict protein 3D structures from just the 1D amino acid sequence. This paper sparked my interest in the field of computational biology as a whole, seeing the potential for computational tools to make an impact on the field of healthcare/medicine. I looked to explore ways to predict 3D structures of RNA, which are also crucial to determine but the problem is much less explored (explained more in the Introduction section), eventually leading to the development of this experimental data-guided reconstruction approach described in this work.

I would like to sincerely thank my advisor, Dr. Morgan D'Ausilio, for giving feedback on my research presentations, leading to many productive conversations which helped drive my research forward. I also greatly appreciate her guidance in helping revise and edit the final manuscript.

At the time the code was written, the Gemini model available in the Antigravity IDE (Gemini 3.6 Flash Medium) was used solely for debugging purposes. The code was initially written manually by me, and Gemini was only used afterward to help identify and troubleshoot errors. Any AI-generated suggestions or modifications were carefully reviewed and verified by me before being incorporated into the final model.

Contents

1 Introduction **1**

1.1 Innovation Points 2

2 Methods **3**

2.1 Voxel-wise Segmentation 4

2.2 Identification of Phosphorus Sites 4

2.3 Hamiltonian Path Construction 4

2.4 Refine Path and Produce Coarse 3D Model 5

2.4.1 Path Orientation 5

2.4.2 Sequence Registration 5

2.4.3 Fragment Consensus and Anchor Identification 6

2.4.4 Anchor-Guided Refinement and Gap Repair 6

2.5 All-atom Reconstruction and Refinement 7

2.6 Metric Definitions 8

3 Results and Discussion **9**

3.1 Construction of Dataset 9

3.2 Test Dataset Results 10

3.3 Intermediate and Ablation Results 14

4 Future Work **18**

5 Conclusion **18**

6 Code Availability **18**

Bibliography **19**

1 Introduction

The 3D structure of RNA is closely linked to its biological function. Structured RNAs mediate catalysis, translation, splicing, ligand sensing and gene regulation [2–4]. Structural information helps identify sites that can be targeted by small molecule therapeutics [5, 6]. RNA molecules can also adopt multiple conformations, allowing their interactions and activity to change with their environment [3, 7]. This flexibility makes structure determination difficult. X-ray crystallography requires suitable crystals, while nuclear magnetic resonance (NMR) spectroscopy faces increasing challenges as RNA size and spectral complexity grow [8, 9]. Experimentally determined RNA structures remain much less abundant than protein structures, with RNA representing less than 1% of structures in the PDB database, limiting both our understanding of RNA function and the structural data available to train prediction models [10, 11].

Traditional computational RNA modeling offers a route using physics-based approaches to determine RNA 3D structure. Template-based methods, such as ModeRNA and RNABuilder, use known structures of related RNAs to guide model construction [12, 13]. When suitable templates are unavailable, de novo methods search for plausible folds using structural fragments and energy functions. Examples include the Monte Carlo method SimRNA and Rosetta fragment assembly with all-atom refinement in FARFAR and FARFAR2 [14–16]. As RNA length increases, this search must account for more possible arrangements of helices, loops, and long range interactions. Correctly predicting individual stems therefore does not necessarily recover their arrangement in the complete molecule. Blind RNA-Puzzles and CASP assessments show that global folding remains difficult with these models, particularly for large or unfamiliar RNA structures [17–19].

The success of AlphaFold2 in protein structure prediction encouraged the development of deep learning methods for RNA [1]. DeepFoldRNA, DRfold, and trRosettaRNA use learned geometric information to guide 3D prediction, while RhoFold+ incorporates a pretrained RNA language model [20–23]. RoseTTAFold All-Atom and AlphaFold 3 extend learned prediction to assemblies containing proteins, nucleic acids, and other molecules [24, 25]. These methods have improved RNA modeling, but their accuracy decreases significantly with RNA length and the information available from related sequences. Benchmarks report particular difficulties for long RNAs and for targets with shallow multiple sequence alignments, which contain few related sequences from which to infer structural constraints [26, 27]. Experimental data can help narrow the possible folds by constraining where parts of the RNA lie in space. Leveraging these measurements is a useful approach when prediction alone cannot reliably determine the global structure [28].

Cryogenic electron microscopy (cryo-EM) provides such spatial information without requiring crystallization [29]. In single-particle cryo-EM, many copies of a molecule are imaged in a thin layer of noncrystalline ice. The resulting noisy 2D images show particles in different orientations. Computational alignment and reconstruction combine these views into a 3D map [30]. This forms the density map, reflecting how electrons interact with the target [31]. Recent RNA applications include a 40-kDa SAM-IV riboswitch at 3.7 Å, a full-length *Tetrahymena* group-I intron ribozyme at 3.1 Å, and engineered RNA assemblies resolved at better than 3 Å resolution [10, 32, 33]. These examples demonstrate that cryo-EM can reveal RNA folds across different sizes and architectures. Though, a reconstructed map does not directly specify the coordinates or identities of individual atoms. Building an RNA model requires locating the backbone, following its direction, matching positions to the nucleotide sequence, and placing the remaining atoms. This process has relied heavily on manual tracing, with researchers interpreting recognizable density features to build connected RNA segments [34]. When related structures are available, rigid or flexible fitting can provide an initial model, but the result still requires manual assessment of map agreement and molecular geometry [35, 36]. At moderate or uneven local resolution, nearby features can merge and flexible regions may be poorly resolved, making the backbone path and sequence assignment uncertain [34, 37].

Current automation has attempted to reduce this manual workload. DRRAFTER combines RNA fragment assembly with experimental density, and auto-DRRAFTER extends this approach to automated modeling of RNA-only systems [28, 34]. More recent models, including CryoREAD, EM2NA, and ModelAngelo, use a primarily deep learning-based approach to identify nucleic acid features and automate major stages of atomic model building [37–39]. However, accurate local fragments do not always form a complete and correctly ordered RNA chain. Folded RNA can bring distant sequence regions close together, so their distance alone is insufficient to identify consecutive nucleotides. A continuous trace can also assign nucleotides to the wrong positions if its direction or local ordering is incorrect [37]. Thus, sequence registration, the assignment of sequence positions to traced sites, must be considered alongside

connectivity. In our single chain benchmark, these methods frequently produced disconnected RNA segments, being unable to capture the overall global structure and topology of the RNA. This weakness motivates a continuous chain building method and the incorporation of classical methods to complement the deep learning approaches. The remaining challenge is to use local density features together with sequence and secondary structure information to recover a consistent model of the full chain.

In this work, we present RNATrace, a fully automated hybrid framework that reconstructs all-atom RNA models from a cryo-EM map and nucleotide sequence. The framework combines learned density interpretation with explicit constraints on backbone connectivity and sequence assignment. Multiscale segmentation using Swin UNETR [40] identifies RNA chemical groups, which are used for phosphorus tracing and nucleotide placement. RNATrace then orders predicted phosphorus sites into one backbone path and uses predicted secondary structure, local purine/pyrimidine evidence, and fragment sequence anchors to determine its direction and improve sequence registration. Finally, nucleotide templates expand the trace into an all-atom model, followed by restrained geometric refinement. We evaluate RNATrace on 20 experimental maps spanning 119–423 nucleotides and 3.0–6.0 Å resolution, comparing local accuracy and global fold recovery with existing methods, and find that RNATrace achieves significantly higher TM-scores and GDT-TS values ($p < 0.0001$) than all three comparators. Ablation experiments were performed to assess the contributions of segmentation, sequence registration, reconstruction, and refinement.

1.1 Innovation Points

The main design innovations of RNATrace are:

1. **Multiscale group segmentation.** A two-stage voxel-wise segmentation strategy first separates RNA-associated density from background and then classifies voxels into phosphate, sugar, purine (A/G), and pyrimidine (C/U) groups. Using predictions from two patch sizes combines both local interactions and broader spatial context. These group predictions guide backbone tracing and the placement of sugars and bases.
2. **Globally constrained backbone and sequence registration.** Predicted phosphorus sites are ordered into a single open Hamiltonian path, which connects each phosphorus atom exactly once, addressing chain fragmentation. Predicted base pairing and purine/pyrimidine evidence help determine the backbone's direction and correct the order of traced sites, improving how each site is matched to a nucleotide in the input sequence.
3. **Hybrid all-atom reconstruction and refinement.** Nucleotide templates expand the registered trace into an all-atom model using predicted sugar and base locations. Refinement combines phosphorus, hydrogen bond, and density restraints, allowing local adjustment while preserving the traced backbone.

2 Methods

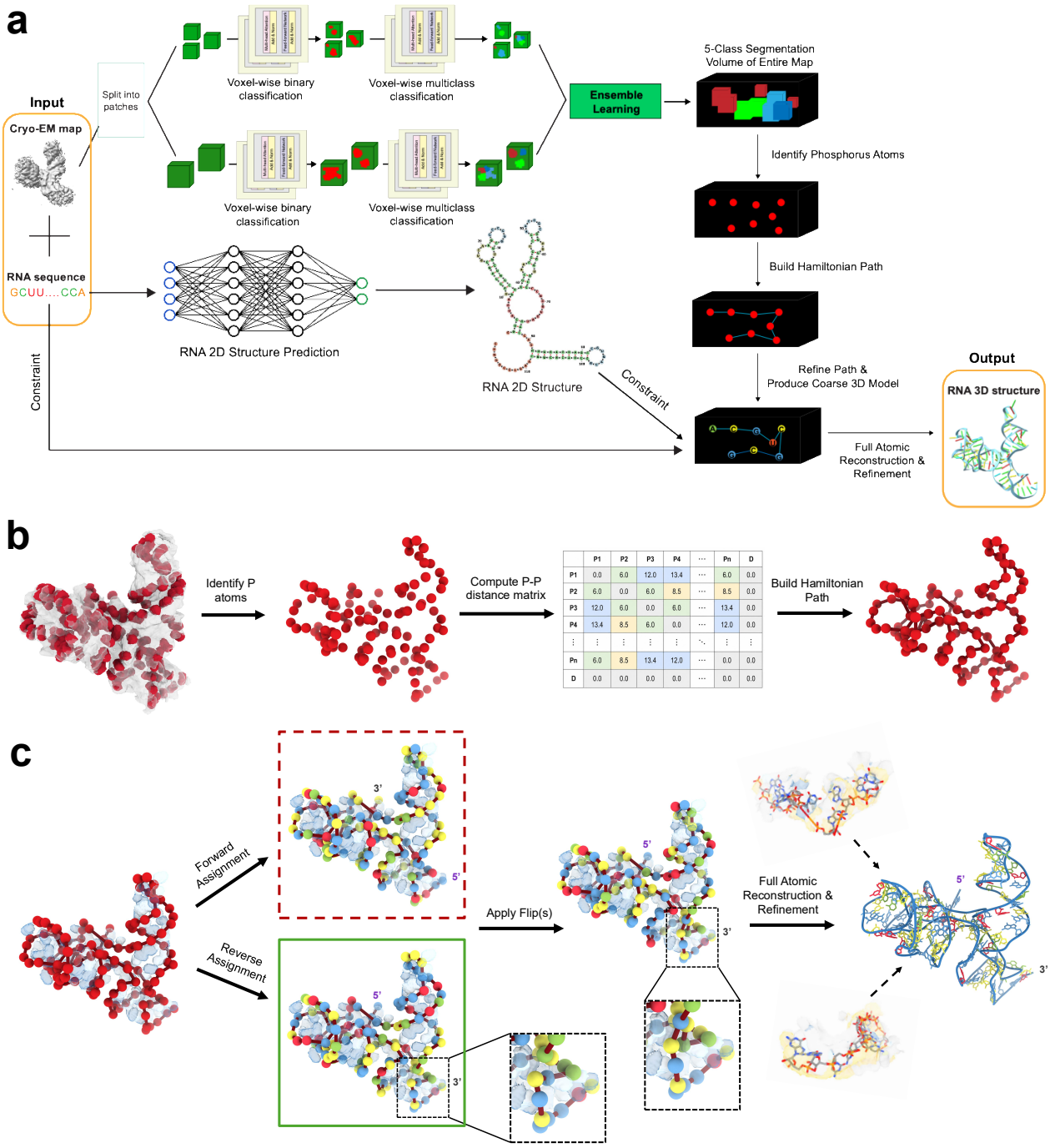


Figure 1: Overview of the RNATrace workflow. **a** The normalized and resampled cryo-EM map is divided into 64^3 and 128^3 voxel patches. Voxel-wise segmentation separates RNA from background and classifies RNA density as phosphate, sugar, purine, or pyrimidine. **b** Phosphorus sites are extracted from predicted phosphate clusters and ordered as an open path by solving a Traveling Salesman Problem with the Lin–Kernighan–Helsgaun heuristic [41]. **c** Predicted secondary structure from RNAformer [42] determines path direction and, together with purine/pyrimidine evidence and sequence anchors, guides flip-registration. Purine and pyrimidine predictions are shown in light and dark blue, respectively. Nucleotide templates, where sugar predictions are in yellow, convert the coarse 3D model, with nucleotides shown in distinct colors, to an all-atom model, followed by restrained geometric refinement.

2.1 Voxel-wise Segmentation

At each patch size, two independently trained Swin UNETR networks [40] receive the map as input (Figure 1a). The binary network has one logit output per voxel, classifying RNA foreground (RNA-associated density) versus background. The group network has four softmax outputs for phosphate, sugar, purine (A and G bases) and pyrimidine (C and U bases), and its cross-entropy loss is evaluated only at foreground voxels.

Binary networks were trained for 50 epochs with binary cross-entropy. All foreground voxels were retained, while the sampled background-to-foreground ratio increased linearly across training epochs, from 4:1 to sampling the entire patch (∼20:1). This negative sampling strategy was adapted from an approach that linearly grows the patch dimensions over training [43]. Here, it's utilized for the voxel sampling ratio, to address the class imbalance by first emphasizing positive samples and avoiding background-dominated predictions, then progressively incorporating more background as the training continued. Typed networks were trained for 100 epochs with the loss evaluated only on foreground voxels. Batch sizes were 16 for 64^3 voxel patches and 2 for 128^3 voxel patches. Both networks used AdamW with an initial learning rate of 10^{-4} and weight decay of 10^{-2}. The checkpoint with the highest validation macro-F1 was retained.

Inference used the same overlapping windows as training. Within one scale, overlapping binary predictions were combined by retaining the probability farthest from 0.5. For typed segmentation, ensemble inference was performed on ground truth foreground voxels. The complete softmax vector was retained from the patch with the largest class probability at that voxel. The combined 64^3 and 128^3 probability volumes were then averaged independently for the binary and typed tasks with equal scale weights of 0.5 and 0.5:

$$\bar{\mathbf{p}}^{(t)}(v) = 0.5\,\mathbf{p}_{64}^{(t)}(v) + 0.5\,\mathbf{p}_{128}^{(t)}(v) \tag{1}$$

for $t \in \{\text{bin}, \text{typ}\}$. A voxel was assigned to background when $\bar{p}^{(\text{bin})}(v) \leq 0.5$. Otherwise, it received the group with the largest component of $\bar{\mathbf{p}}^{(\text{typ})}(v)$.

During testing, four-class typed segmentation was only performed on voxels which were predicted as foreground in the binary classification. The predicted background voxels were masked.

2.2 Identification of Phosphorus Sites

Phosphorus sites were extracted from the predicted phosphate volume by 18-connectivity labeling (Figure 1). This stage used the input sequence to set the required number of sites, N. Components smaller than 10 voxels were discarded. For component i with volume v_i, let $V = \sum_i v_i$ and $x = V/N$. Its initial allocation was $k_i = \max(1, \text{round}(v_i/x))$. Nearby sugar-density peaks within 10 Å provided a second estimate of the phosphorus count. If at least two sugar peaks were detected and the volume-based estimate was at least twice as large, the phosphorus count was reduced to the number of sugar peaks. Allocations were then increased or decreased in order of their fractional volume residuals until $\sum_i k_i = N$.

For a component assigned one site, the density-weighted centroid defined its coordinate. A component assigned multiple sites was divided by 18-connected watershed segmentation from separated density maxima, and the density-weighted centroid of each basin defined one coordinate. Solutions containing sites closer than 3.5 Å were replaced by iterative selection of density maxima with spatial exclusion. Global non-maximum suppression prevented separations below 2.5 Å. Coordinates were transformed from voxel to Cartesian space using the retained map origin and spacing.

2.3 Hamiltonian Path Construction

The predicted sites were ordered by a Hamiltonian path that visited each site once (Figure 1). Pairwise Euclidean distances supplied the edge costs of a complete graph. A dummy node with zero-cost edges converted the open Traveling Salesman Problem (TSP) to a tour. The Lin–Kernighan–Helsgaun [41] implementation was run from 20

initializations, and the lowest-cost tour was retained. Removing the dummy node and its two incident edges produced the open path. RNATrace currently models one linear RNA chain. The path therefore represents a candidate 5' to 3' backbone trace.

2.4 Refine Path and Produce Coarse 3D Model

The Hamiltonian path constructed above gives an ordered sequence of phosphorus coordinates, but this path carries neither a biological direction (5' vs. 3') nor any correspondence to positions in the input nucleotide sequence. Producing an accurate coarse 3D model, where a nucleotide is assigned to each phosphorus atom, therefore requires designing reliable algorithms to resolve both ambiguities, while also compensating for the fact that phosphorus site detection is imperfect and can miss or introduce sites relative to the true count. This process proceeds in four stages. First, predicted secondary structure is used to determine the path's 5' to 3' direction (Section 2.4.1). Second, each phosphorus site is assigned to a sequence position through an iterative greedy registration procedure that favors placements consistent with predicted base pairing and purine/pyrimidine evidence (Section 2.4.2). Third, high confidence fragments of the path are independently aligned to the sequence to produce anchor assignments that are largely insensitive to registration errors elsewhere in the path (Section 2.4.3). Finally, when these anchors disagree with the initial registration, the registration is repeated under anchor guidance, and any remaining local gaps between the path and the sequence are resolved by inserting and removing phosphorus coordinates as needed (Section 2.4.4).

2.4.1 Path Orientation

RNAformer [42] was applied to the input sequence to predict a set $\mathcal{B}$ of canonical base pairs. For each of the two path orientations, sequence indices were assigned consecutively to phosphorus sites and the number of satisfied pairs was calculated as

$$B = \sum_{(a,b)\in\mathcal{B}} \mathbf{1}\big[15 \le \|\mathbf{p}_a - \mathbf{p}_b\|_2 \le 22 \text{ Å}\big] \tag{2}$$

where $\mathbf{p}_a$ and $\mathbf{p}_b$ represent the 3D coordinates of phosphorus atoms a and b, respectively, and $\mathbf{1}[\cdot]$ represents the indicator function. The orientation with the larger B determined the 5' end and 3' end (Figure 1c).

2.4.2 Sequence Registration

Base class (purine/pyrimidine) evidence was summarized around each phosphorus site before sequence registration. Let n_i^R and n_i^Y be the numbers of voxels labeled purine and pyrimidine within 12 Å of site i. When $n_i^R + n_i^Y \ge 5$ and the counts were unequal, the majority class was assigned with confidence

$$c_i = \frac{|n_i^R - n_i^Y|}{n_i^R + n_i^Y}. \tag{3}$$

Otherwise, the base class was treated as unknown and $c_i = 0$. For a given sequence registration, the base class agreement score was

$$C = \sum_i c_i m_i, \qquad m_i = \begin{cases} +1, & \text{sequence and local base classes agree,} \\ -1, & \text{sequence and local base classes disagree,} \\ 0, & \text{the local base class is unknown.} \end{cases} \tag{4}$$

The two agree if the sequence nucleotide label aligns with the purine/pyrimidine group prediction.

Starting from the oriented path, sequence registration was refined by greedy interval reversal. Every interval $[i, j)$ containing at least 20 sites was considered. A move reversed the order of the sites in that interval and hence

their sequence assignments, without moving their coordinates. For each candidate interval reversal, the change in registration score was calculated as

$$\Delta S = \Delta B + \lambda_H \Delta H + \lambda_G \Delta G + \lambda_C \Delta C + \lambda_A \Delta A, \tag{5}$$

where, for each score component X, $\Delta X = X_{\text{after}} - X_{\text{before}}$. B is defined in Equation (2). H is the number of geometrically satisfied helical steps. A helical step consists of adjacent nested base pairs (i, j) and $(i+1, j-1)$, and is satisfied only when both base pairs satisfy the phosphorus distance criterion in Equation (2). G is the reduction in the combined length of the two path edges connecting the reversal interval to the unchanged path. C is defined in Equation (4). A is the number of anchor site-to-sequence assignments, from the fragment anchors identified in Section 2.4.3, that agree with the candidate registration during evaluation. The baseline search used $\lambda_H = 0.3$, $\lambda_G = 5.0$, $\lambda_C = 2.0$, and $\lambda_A = 0$; anchor candidates used the same λ_H, λ_G, λ_C with $\lambda_A = 3$. The four coefficients were selected by a grid search over the validation maps to maximize sequence recall (Equation (17)), evaluated on the full reconstructed structures. Scores were recomputed after every accepted move, which is determined by the highest ΔS per iteration. Search ended when no available interval improved the score.

2.4.3 Fragment Consensus and Anchor Identification

Because greedy interval reversal can converge to a registration that is only locally optimal, fragment consensus was used to independently generate candidate assignments between phosphorus sites and sequence positions from short regions of the path. The most reliable of these assignments serve as anchors for correcting the registration where it disagrees with them. Windows of 24 consecutive phosphorus sites were extracted from the path at stride 2. Each path window was aligned to every 24-nucleotide sequence interval of the same length in both forward and backward orientations. Let u denote one candidate alignment and let $\mathcal{K}_u = \{i \in u : c_i > 0\}$. Using the base class agreement defined in Equation (4), the candidate alignment was scored as

$$s_u = \frac{1}{|\mathcal{K}_u|} \sum_{i \in \mathcal{K}_u} c_i m_{u,i}, \tag{6}$$

where $m_{u,i}$ is the match value m_i evaluated under the sequence assignment proposed by alignment u. Candidate alignments with $\mathcal{K}_u = \varnothing$ were discarded. Only the highest scoring sequence to phosphorus site alignment of each window was retained. Its score was converted to a nonnegative vote weight, $w_u = \max(10^{-3}, 1 + s_u)$.

If $\rho_u(i)$ denotes the sequence position assigned to phosphorus site i by the alignment u, the total vote for assigning site i to sequence position r was

$$V_i(r) = \sum_{u \in \mathcal{W}_i} w_u \, \mathbf{1}[\rho_u(i) = r], \tag{7}$$

where $\mathcal{W}_i$ is the set of retained window alignments containing site i.

For each site, let V_1 and V_2 be the largest and second-largest totals of $V_i(r)$, and let $V = \sum_r V_i(r)$. The sequence position receiving V_1 was retained when $V_1/V \geq 0.5$ and $(V_1 - V_2)/V \geq 0.3$. This means if several sites selected the same residue, only the site with the strongest consensus was retained. Confident site-to-sequence assignments were grouped into continuous blocks in which adjacent phosphorus sites mapped to adjacent sequence positions, in either the forward or reverse direction. Because assignments near block boundaries were considered less reliable, two assignments were trimmed from each end. A trimmed block containing at least eight assignments was retained as a fragment anchor.

2.4.4 Anchor-Guided Refinement and Gap Repair

The flip-registration was repeated with anchor terms only when at least 24 anchor assignments were available and either baseline anchor agreement, the fraction of anchor site-to-sequence assignments that match the baseline registration, was below 0.70 or the fraction of satisfied predicted base pairs was below 0.65. An anchor-guided result

was accepted only when anchor agreement increased by at least 0.10 and base pair satisfaction decreased by no more than 0.15.

Fragment splicing and subsequent local repairs were applied only after an anchor-guided registration was accepted. For each sequence gap g, let $\mathcal{R}_g$ be its set of unassigned sequence positions and let $\mathcal{P}_g$ be its set of candidate phosphorus sites. For an internal gap, $\mathcal{R}_g$ contains the sequence positions between two consecutive, consistently ordered anchors, and $\mathcal{P}_g$ contains the unassigned phosphorus sites between the corresponding anchor endpoints along the path. For a 5' or 3' terminal gap, $\mathcal{R}_g$ contains the sequence positions between the corresponding sequence terminus and the nearest anchor, while $\mathcal{P}_g$ contains the unassigned phosphorus sites between that anchor and the path terminus indicated by the anchor orientation. The gap sizes and their difference were defined as $N_{\mathrm{P}}^{(g)} = |\mathcal{P}_g|$, $N_{\mathrm{nt}}^{(g)} = |\mathcal{R}_g|$, and $E_g = N_{\mathrm{P}}^{(g)} - N_{\mathrm{nt}}^{(g)}$.

An $E_g > 0$ represented the number of surplus phosphorus sites, and all possible choices of E_g sites to omit were evaluated. Each assignment was scored as

$$S_{\mathrm{gap}} = -\sum_k (d_k - 6)^2 + 2C_{\mathrm{gap}}, \tag{8}$$

where d_k is the distance in Å between consecutive phosphorus sites in the proposed assignment, including the connections to the two bounding anchors, and C_{gap} is the base agreement score over the proposed gap assignments. The squared term favored phosphorus steps near 6 Å, while the second term rewarded agreement between the assigned sequence classes and the local purine/pyrimidine evidence. The candidate with the largest S_{gap} was selected. Accepted fragment assignments replaced the corresponding existing assignments.

For a gap with $E_g < 0$, let $D_g = -E_g$ be the number of phosphorus coordinates that must be derived. After orienting the candidate phosphorus sites consistently with the neighboring anchors, RNATrace evaluated each assignment of the $N_{\mathrm{P}}^{(g)}$ observed sites to $N_{\mathrm{nt}}^{(g)}$ sequence positions. For each candidate assignment, provisional phosphorus coordinates were derived for the D_g unoccupied sequence positions. Missing positions between assigned sites were linearly interpolated, whereas positions at a sequence terminus were extrapolated at 6 Å intervals. The completed assignment was scored with Equation (8) and similar to the case with $E_g > 0$, the largest S_{gap} was selected.

2.5 All-atom Reconstruction and Refinement

To place the rest of the heavy atoms, the registered phosphorus trace was expanded into an initial heavy-atom model using A, C, G and U nucleotide templates selected according to the input sequence (Figure 1c). For residue i, the local backbone tangent $\mathbf{t}_i$ and transverse axis $\mathbf{u}_i$ were defined as

$$\mathbf{t}_i = \frac{\mathbf{p}_{i+1} - \mathbf{p}_{i-1}}{\|\mathbf{p}_{i+1} - \mathbf{p}_{i-1}\|} \qquad \mathbf{u}_i = \frac{\mathbf{h}_i - (\mathbf{h}_i \cdot \mathbf{t}_i)\mathbf{t}_i}{\|\mathbf{h}_i - (\mathbf{h}_i \cdot \mathbf{t}_i)\mathbf{t}_i\|}, \tag{9}$$

and at the chain termini, the tangent was calculated from the terminal phosphoorus and its single neighboring phosphorus. Here, $\mathbf{h}_i = 0.7\hat{\mathbf{b}}_i + 0.3\hat{\mathbf{s}}_i$, where $\hat{\mathbf{b}}_i$ and $\hat{\mathbf{s}}_i$ are unit directions from phosphorus to the predicted base and sugar centroids. The third axis, $\mathbf{t}_i \times \mathbf{u}_i$, completes the local 3D frame in Equation (9). A rotation $\mathbf{R}_i$ aligned the template's P-to-O3′ axis and orthogonalized P-to-C1′ direction with $\mathbf{t}_i$ and $\mathbf{u}_i$, respectively. Template atom a was then placed at

$$\mathbf{x}_{ia} = \mathbf{p}_i + \mathbf{R}_i(\mathbf{q}_a - \mathbf{q}_{\mathrm{P}}), \tag{10}$$

where $\mathbf{q}_a$ and $\mathbf{q}_{\mathrm{P}}$ are template coordinates. Equation (10) preserves the template's internal geometry and places its phosphorus at the registered site.

After placement, optimizations are performed where the entire nucleotide template was rotated about $\mathbf{t}_i$ through $\mathbf{p}_i$ in 10° increments. The retained angle minimized

$$E_i(\theta) = w_{\mathrm{BS}}\left[D_{\mathrm{B}}(\theta) + w_{\mathrm{S}} D_{\mathrm{S}}(\theta)\right] + w_{\mathrm{WC}} D_{\mathrm{WC}}(\theta), \tag{11}$$

where θ is the rotation angle and each $D = 1 - \hat{\mathbf{d}}_{\mathrm{model}} \cdot \hat{\mathbf{d}}_{\mathrm{target}}$. The base (B) and sugar (S) terms compared phosphorus-to-centroid directions in the model template with the corresponding predicted directions. For paired

residues, those with a secondary structure pairing partner, the Watson–Crick-edge (WC) term compared the direction from phosphorus toward the centroid of WC atoms in the placed template with the direction from the same phosphorus toward the partner's phosphorus position. The weights $w_{\mathrm{BS}} = 0.75$, $w_{\mathrm{S}} = 0.5$, and $w_{\mathrm{WC}} = 0.25$ were selected by a grid search over the validation maps, evaluated on base and C4′ RMSD (Equation (12)) computed on the placed templates prior to subsequent geometric refinement.

Residues were assembled in registered sequence order and hydrogens were prepared before refinement in OpenMM [44] with the AMBER14 force field and RNA.OL3 parameters. Harmonic restraints on phosphorus coordinates preserved the traced backbone while covalent geometry and steric contacts relaxed. Hydrogen bond distance restraints and density restraints were applied. Force constants were 100,000, 100 and 500 kJ mol^{-1} nm^{-2} for phosphorus, hydrogen bond and density restraints, respectively. Restrained Langevin dynamics was then run for 2,000 steps at 300 K with a 1 fs time step and a friction coefficient of 1 ps^{-1}, followed by 250 iterations of restrained energy minimization.

2.6 Metric Definitions

Coordinate accuracy. For an atom group g, RMSD was calculated as

$$\mathrm{RMSD}_g = \sqrt{\frac{1}{N_g}\sum_{k=1}^{N_g} \|\mathbf{R}_g\mathbf{x}_k + \mathbf{t}_g - \mathbf{y}_k\|_2^2}, \tag{12}$$

where $\mathbf{x}_k$ and $\mathbf{y}_k$ are corresponding predicted and reference atom coordinates, respectively; N_g is the number of compared atom pairs; and $\mathbf{R}_g$ and $\mathbf{t}_g$ are the optimal rotation and translation, obtained with the Kabsch algorithm [45, 46]. The rotation was constrained to be proper ($\det(\mathbf{R}_g) = +1$), with no scaling or reflection.

Global structural similarity. TM-score (template modeling score) [47] measures length-normalized global fold similarity and is calculated as

$$\text{TM-score} = \frac{1}{L_{\mathrm{ref}}}\sum_{i=1}^{N_{\mathrm{aln}}} \frac{1}{1 + [d_i/d_0(L_{\mathrm{ref}})]^2}, \tag{13}$$

where d_i is the distance between corresponding C4′ atoms, d_0 is its reference-length-based distance normalization factor computed by RNA-align [48], N_{aln} is the number of aligned pairs, and L_{ref} is the full reference length.

Phosphorus GDT-TS (GDT-TS) [49] is calculated as

$$\text{GDT-TS} = \frac{100}{4N_{P,\mathrm{ref}}} \sum_{\delta\in\{1,2,4,8\}\,\text{Å}} \max_{\mathbf{R}_\delta,\mathbf{t}_\delta} \sum_{i=1}^{N_{P,\mathrm{aln}}} \mathbf{1}(\|\mathbf{R}_\delta\mathbf{p}_i + \mathbf{t}_\delta - \mathbf{q}_i\|_2 \le \delta), \tag{14}$$

where $\mathbf{p}_i$ and $\mathbf{q}_i$ are corresponding predicted and reference phosphorus coordinates, respectively; $N_{P,\mathrm{ref}}$ is the full reference phosphorus count; and $\mathbf{1}(\cdot)$ is the indicator function. The proper rigid superposition $(\mathbf{R}_\delta, \mathbf{t}_\delta)$ is optimized independently at each distance threshold δ.

Voxel-wise segmentation accuracy. For a class c, the F_1 score was calculated from the numbers of true positives (TP_c), false positives (FP_c), and false negatives (FN_c) as $F_{1,c} = \frac{2\mathrm{TP}_c}{2\mathrm{TP}_c+\mathrm{FP}_c+\mathrm{FN}_c}$.

Macro-F_1 is the unweighted mean of each class' scores,

$$F_{1,\text{macro}} = \frac{1}{C}\sum_{c=1}^{C} F_{1,c}, \tag{15}$$

where C is the number of evaluated classes.

Sequence assignment accuracy. Let N_{exact} be the number of aligned residue pairs for which the canonical nucleotide identities (A, C, G, or U) are identical. Define:

$$\text{Nucleotide assignment accuracy } (\%) = 100 \times \frac{N_{\text{exact}}}{N_{\text{aln}}}, \tag{16}$$

where N_{aln} is the number of aligned C4′ residue pairs.

Let L_{ref} denote the total number of nucleotides in the complete reference RNA sequence. Define:

$$\text{Sequence recall } (\%) = 100 \times \frac{N_{\text{exact}}}{L_{\text{ref}}}. \tag{17}$$

3 Results and Discussion

We first describe the dataset construction, preprocessing, and division into training, validation, and test sets used to evaluate RNATrace. We then compare reconstructed RNA structures with those produced by existing methods on the same 20 test maps, assessing coordinate accuracy, structural completeness, and global similarity. Next, we evaluate voxel-wise segmentation on the validation set to assess the intermediate predictions that guide structure reconstruction. Finally, ablation analyses examine how individual pipeline components contribute to reconstruction quality and sequence assignment.

3.1 Construction of Dataset

We prepared a dataset for training, validation, and testing by extracting cryo-EM maps from the EMDB database and the corresponding RNA 3D structures from the PDB database. We used maps from a wide range of resolutions (2–7 Å). To prevent RNAs with similar sequences from appearing across different dataset splits, we clustered sequences using Cd-hit [50] at an 80% sequence similarity threshold, retaining one map per cluster before splitting into training, validation, and test sets. We were left with 201 single-chain RNA maps, where we did an approximate 80%-10%-10% train-val-test split with 161 maps used for training, 20 for validation, and 20 test maps. Each test map corresponded to a single RNA chain.

For each map, data preprocessing was performed where each were denoised by removing surface dust and resampled to an isotropic voxel size of 0.5 Å with ChimeraX [51]. Values above the contour were clipped at their 95th percentile, shifted so that the contour was zero and scaled to $[0, 1]$. The values below the contour were set to zero.

Each map was divided into 64^3 and 128^3 patches with 50% overlap. Binary training labels marked voxels within 2 Å of any RNA atom. Group labels assigned foreground voxels to phosphate (P, OP1 and OP2), sugar, purine base, or pyrimidine base according to the nearest RNA chemical group. The other voxels were background. Background-only patches were not extracted.

The 161 training and 20 validation maps yielded 51,831 and 6,208 patches at 64^3, and 8,522 and 1,106 at 128^3, respectively.

3.2 Test Dataset Results

We benchmarked RNATrace against current state-of-the-art methods that are able to predict all-atom RNA structures from the cryo-EM map: CryoREAD [38], EM2NA [37], and ModelAngelo [39]. All four methods are evaluated on the same 20-map test set using the same ground truth reference structures and metric calculations. We first compare RMSD metrics in Figure 2. For each target, the predicted model was structurally aligned to the ground truth reference using RNA-align [48]. Phosphorus and C4′ RMSD values were calculated using Equation (12). Backbone and base RMSDs included canonical phosphate/ribose and base heavy atoms, respectively.

In Figure 2, we can see that across the test set, RNATrace achieved mean phosphorus, C4′, backbone, and base RMSDs of 2.91, 3.36, 3.68, and 5.34 Å, respectively.

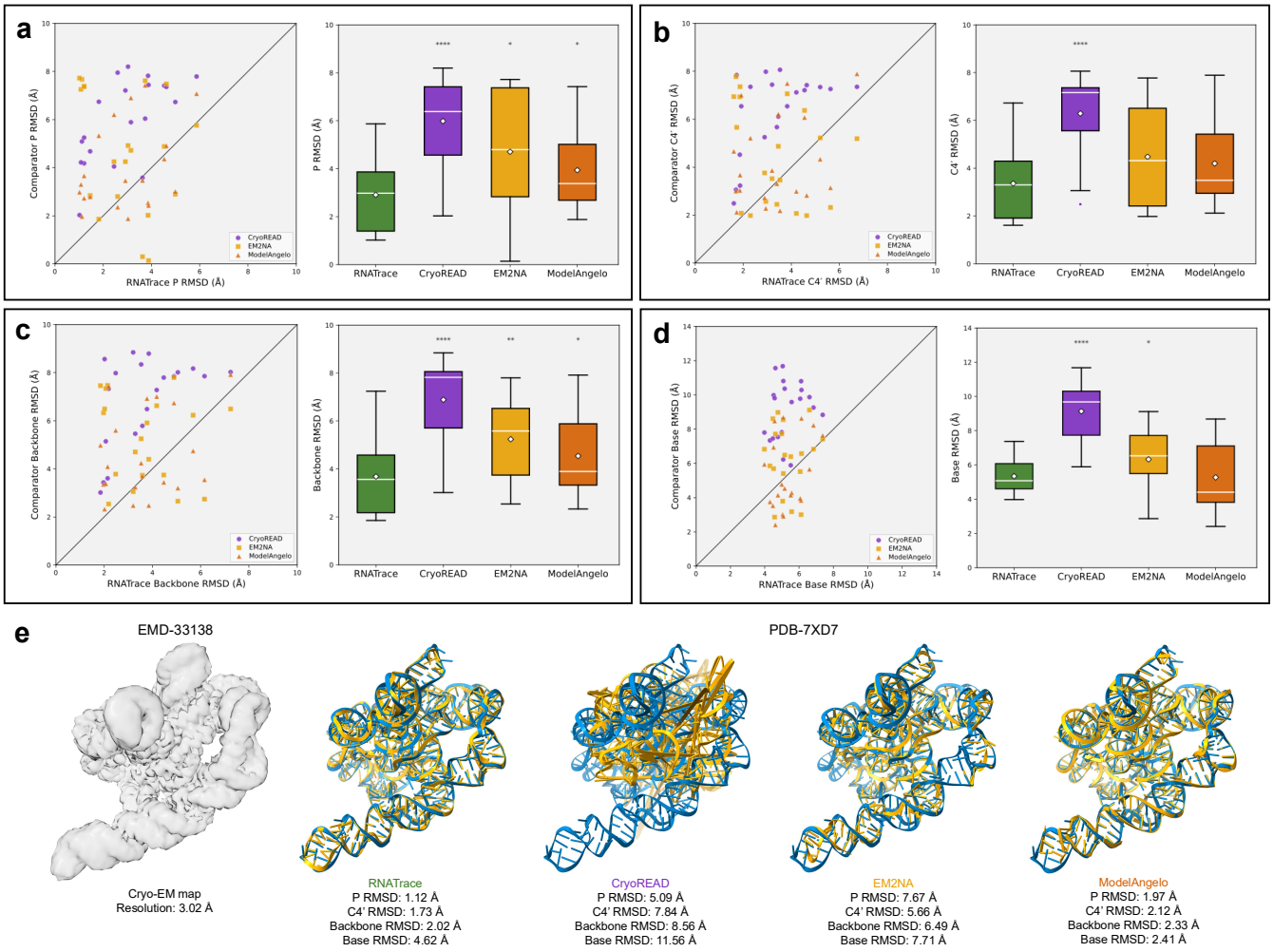


Figure 2: Comparison of RNATrace with CryoREAD, EM2NA, and ModelAngelo for RMSD metrics from a test set of 20 cryo-EM maps. **a**–**d** Scatter plots (left) and box plots (right) comparing the built models for phosphorus position RMSD, C4′ RMSD, backbone RMSD, and base RMSD, respectively. For the box plots, the box indicates the 25th to 75th percentiles (IQR), white line indicates the median, white diamond indicates the mean, and the whiskers indicate the furthest observations within 1.5 × IQR below Q1 and above Q3. Outliers are indicated as a dot beyond the whiskers. Paired Wilcoxon signed-rank tests were performed. *: $p < 0.05$, **: $p < 0.01$, ****: $p < 0.0001$. **e** An example of the built structures (in gold) overlayed with the ground truth structures (in blue) and the cryo-EM map for the pre-Tet-C state of the wild-type *Tetrahymena* group I intron (EMD-33138, PDB-7XD7) are shown, as well as the corresponding RMSD values for each model.

Compared with CryoREAD, RNATrace achieved significantly lower RMSD values for all four metrics ($p < 0.0001$; Figure 2a–d). Compared with EM2NA, RNATrace achieved lower phosphorus, C4′, and base RMSD with $p < 0.05$, and lower backbone RMSD with $p < 0.01$ (Figure 2a–d). Compared with ModelAngelo, RNATrace achieved lower phosphorus, C4′, and backbone RMSD with $p < 0.05$, while the base RMSD difference was not significant (Figure 2a–d).

The scatter plots show that RNATrace outperformed the majority of targets for every comparator and RMSD metric except base RMSD against ModelAngelo, where ModelAngelo achieved lower RMSD on 13 of the 20 targets (Figure 5). This distinction is also visible for EMD-33138, where RNATrace achieved lower phosphorus, C4′, and backbone RMSD, while ModelAngelo achieved a lower base RMSD of 2.41 Å compared with 4.62 Å for RNATrace (Figure 2e). The larger base RMSD relative to the RNATrace backbone metrics suggests that base placement remains the main source of error.

Since RMSD is calculated only based on the aligned residues with the ground truth, these RMSD metrics focus on the coordinate accuracy within the aligned subset and do not measure how completely the prediction is representative of the full structure. There are two EM2NA points that have phosphorus RMSD below 0.25 Å, but fewer than 10 nucleotides were aligned in each case (Figure 2a). Therefore, a small aligned subset can produce a deceptively small RMSD while excluding most of the RNA chain, showing that RMSD should be interpreted alongside aligned resiude percentage.

In Figure 3a, each comparator point lies below the $y = x$ line, and in Figure 3b, RNATrace achieves a much higher aligned residue percentage than each of the other three models ($p < 0.0001$ for all cases). This observation shows that RNATrace traced a larger fraction of the RNA chain for every paired comparison in the test set. This result also reflects a specific failure mode observed in the comparator predictions: CryoREAD, EM2NA, and ModelAngelo frequently produced models split across multiple disconnected RNA chains. RNATrace instead uses explicit Hamiltonian path construction together with sequence, secondary-structure, and purine/pyrimidine constraints to guide a single ordered 5′ to 3′ trace. The resulting improvement in aligned-residue percentage indicates that enforcing global backbone continuity helps recover a larger fraction of the complete RNA structure.

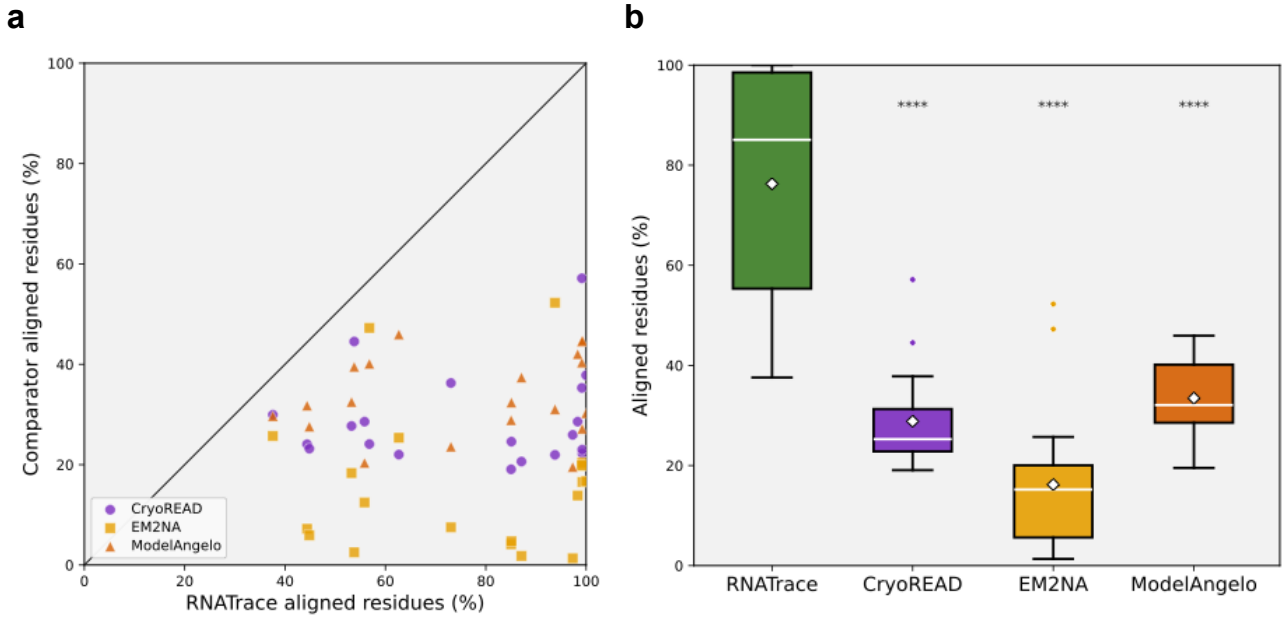


Figure 3: **a** Scatter plot and **b** box plot comparing the percentage of aligned residues between predicted RNATrace structures and CryoREAD, EM2NA, and ModelAngelo's predicted structures. The same box plot format as Figure 2 is used. ****: $p < 0.0001$.

For metrics that compare the global topology and overall structure of the RNA, we use TM-score and GDT-TS,

which are calculated based on Equation (13) and Equation (14), respectively. RNATrace achieved a mean TM-score of 0.707, compared with 0.195 for CryoREAD, 0.147 for EM2NA, and 0.284 for ModelAngelo (Figure 4a). Its mean GDT-TS was 63.5, compared with 11.9, 14.2, and 23.2, respectively (Figure 4b).

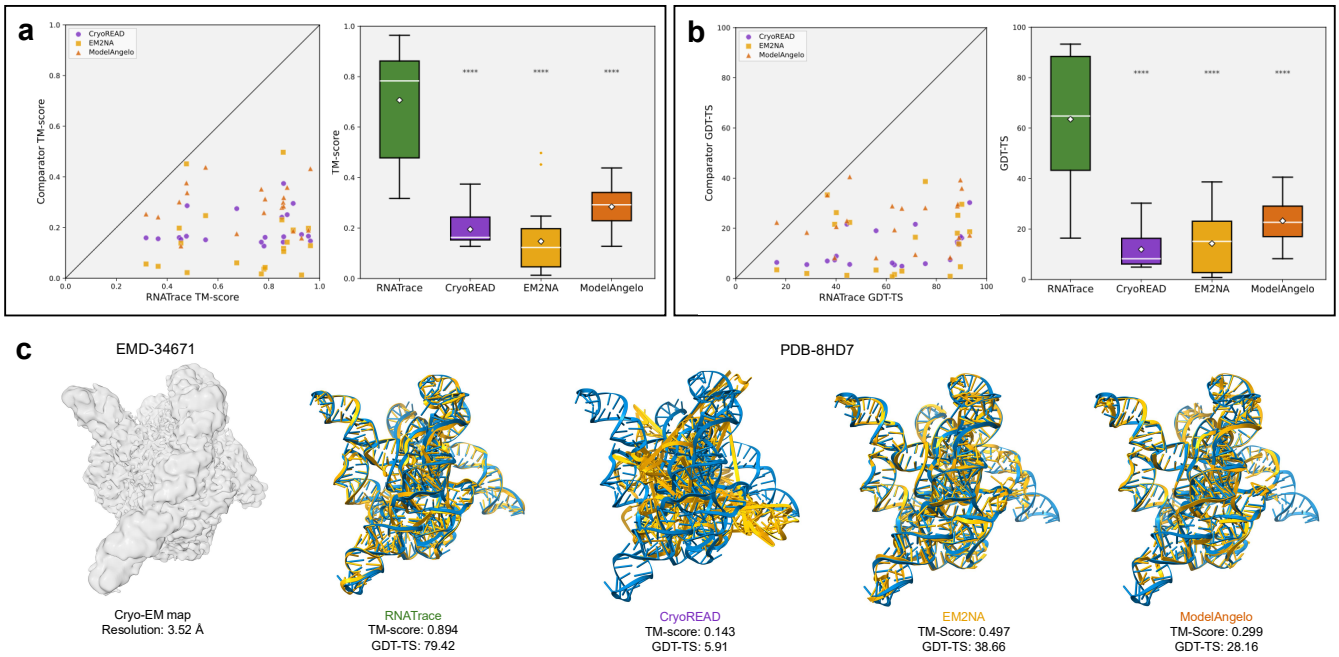


Figure 4: Comparison of RNATrace with CryoREAD, EM2NA, and ModelAngelo for TM-scores and GDT-TS. **a**–**b** Scatter plots (left) and box plots (right) comparing the built models for TM-score and GDT-TS, respectively. The same box plot format as Figure 2 is used. ****: $p < 0.0001$. **c** An example of the built structures (in gold) overlayed with the ground truth structures (in blue) and the cryo-EM map for the intermediate pre-Tet-S1 state of the G264A mutated *Tetrahymena* group I intron (EMD-34671, PDB-8HD7) are shown, as well as the corresponding TM-scores and GDT-TS for each model.

For each model and metric, RNATrace achieves a statistically significant ($p < 0.0001$) increase in TM-score and GDT-TS compared to all of the other three models (Figure 4a–b). For TM-score, neither CryoREAD, EM2NA, or ModelAngelo prediction ever exceeded RNATrace for any map (Figure 5). For GDT-TS, EM2NA and CryoREAD never exceeded RNATrace, while ModelAngelo did so for one map (Figure 5). In the EMD-34671 example, RNATrace achieved a TM-score of 0.894 and GDT-TS of 79.42, compared with maximum comparator values of 0.497 and 38.66, respectively (Figure 4c). Together with the aligned residue results, these global metrics emphasize RNATrace's significant advantage across a larger portion of each RNA and is not limited to a small, accurately aligned fragment.

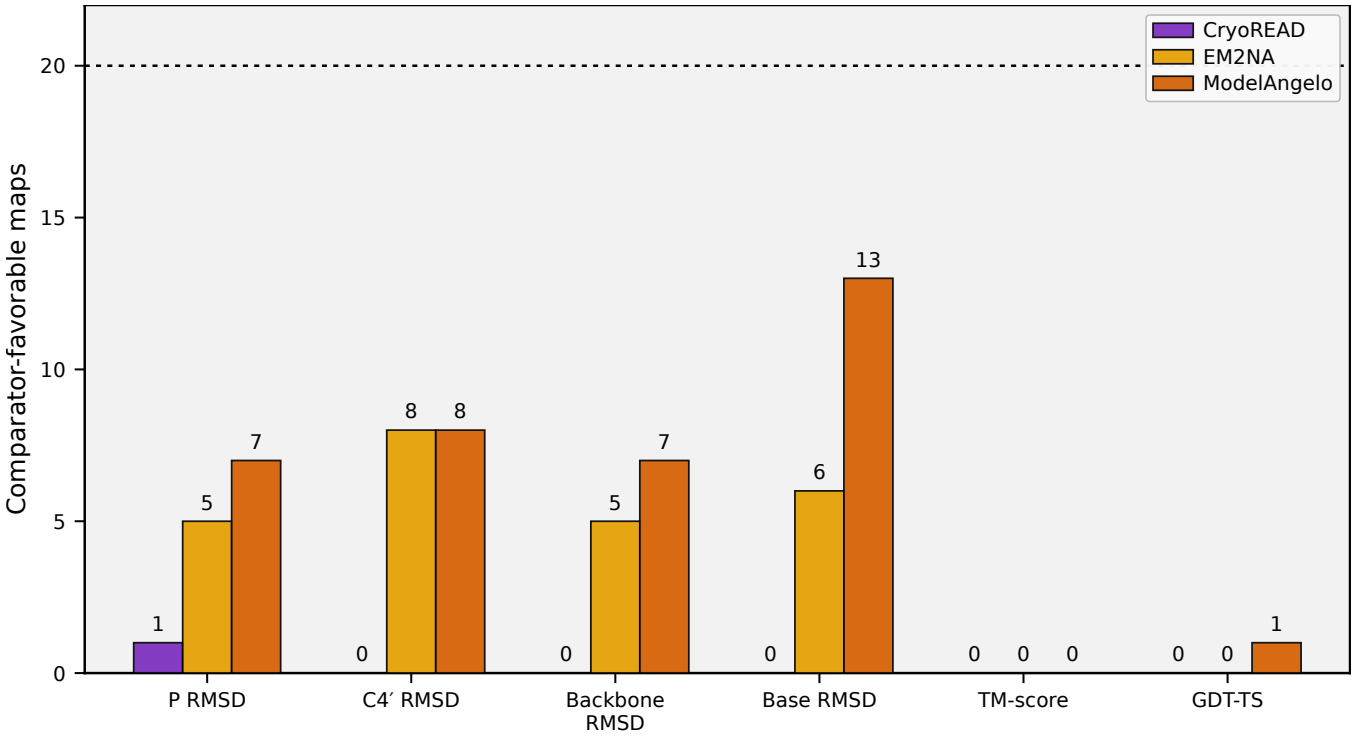


Figure 5: Number of comparator against RNATrace scatter plot points on the comparator-favorable side of the $y = x$ line for the 20 maps in the test set. For RMSD metrics, each bar counts points below $y = x$; for TM-score and GDT-TS, bars count points above $y = x$. Numbers above the bars give the exact counts, and the dotted horizontal line marks all 20 maps.

We next examined the associations of phosphorus RMSD, TM-score, and GDT-TS with flip count and reported map resolution across the 20 test maps. These analyses assessed whether more extensive path correction or lower map resolution was associated with reduced reconstruction accuracy.

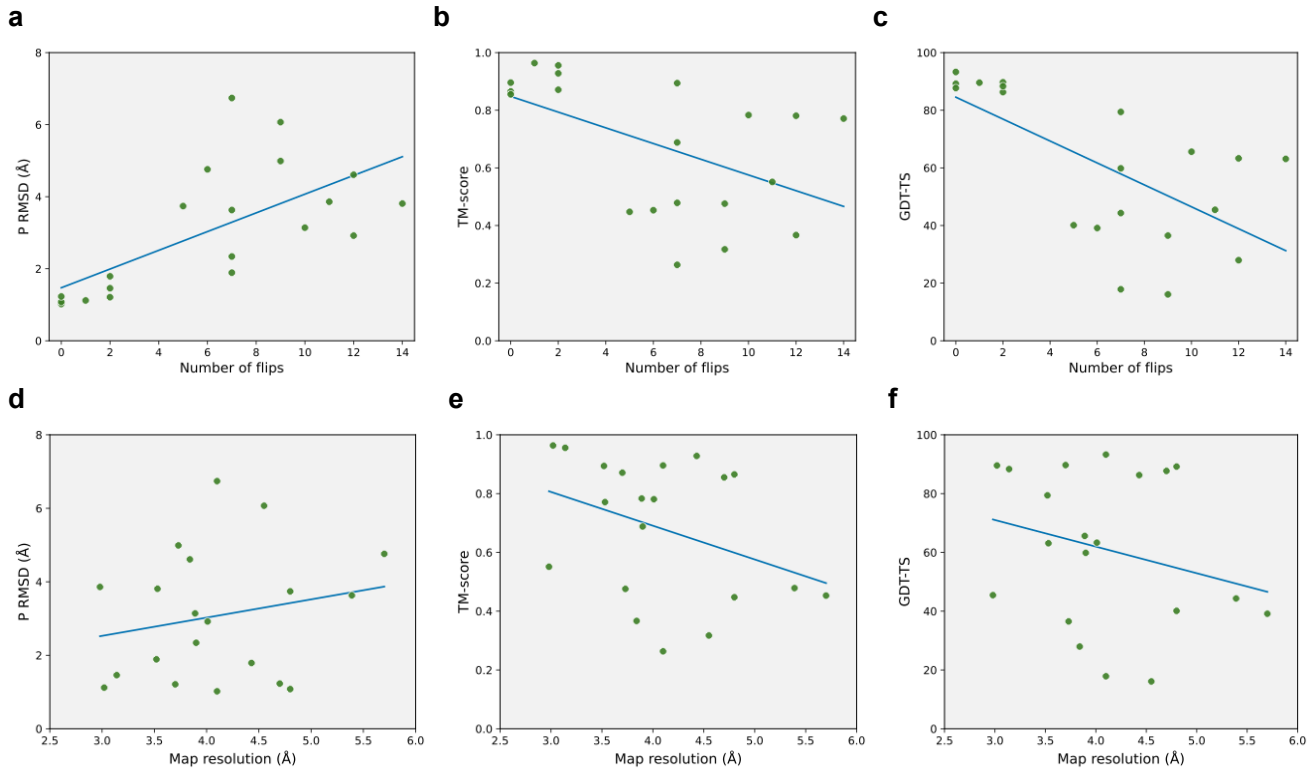


Figure 6: P RMSD, TM-score, and GDT-TS, respectively, are plotted against the number of flips per structure (**a**–**c**) and against the map resolution (**d**–**f**) for all 20 RNATrace predictions in the test set. The blue line for each graph represents the line of best fit using least-squared regression. A two-sided t-test of the least-squares regression slope is performed. **a**: $p = 0.0014$, **b**: $p = 0.0189$, **c**: $p = 0.0018$. **d**–**f**: not significant ($p \geq 0.05$).

The number of flips was significantly positively associated with phosphorus RMSD and significantly negatively associated with TM-score and GDT-TS (Figure 6a–c). Counterintuitively, this association does not indicate that flips reduce accuracy as we can see in Table 2 that removing flips deteriorated all measured metrics. Flips are applied when sequence and secondary structure evidence identifies local inconsistencies in the baseline LKH path, so a high flip count is more likely a marker of a difficult initial trace. Flips therefore provide a corrective benefit, although targets requiring many flips remain more difficult.

No significant association was detected between reported map resolution and phosphorus RMSD, TM-score, or GDT-TS (Figure 6d–f). However, all three fitted slopes point toward poorer performance as map resolution worsens. With only 20 maps and relatively few targets at the poorest resolutions, the absence of significance does not establish that performance is independent of resolution. These results show that RNATrace can reconstruct structures across the tested resolution range, but additional maps are needed to determine how density quality concretely affects performance.

3.3 Intermediate and Ablation Results

Equal-weight inference is defined in Section 2.1. For the weighted-inference comparison, the scale-specific voxel probabilities were combined using weights proportional to the validation macro-F1 scores of the selected checkpoints:

$$w_s^{(t)} = \frac{F_s^{(t)}}{F_{64}^{(t)} + F_{128}^{(t)}}, \qquad \bar{\mathbf{p}}_{\mathrm{wt}}^{(t)}(v) = w_{64}^{(t)}\mathbf{p}_{64}^{(t)}(v) + w_{128}^{(t)}\mathbf{p}_{128}^{(t)}(v),$$

where $s \in \{64, 128\}$ and $t \in \{\text{bin}, \text{typ}\}$. Here, $F_s^{(t)}$ is the validation macro-F1 of the selected checkpoint for task t and scale s, and $w_s^{(t)}$ is its normalized ensemble weight.

Let $F_{\text{eq}}^{(t)}$ denote the macro-F1 obtained from the equal-weight ensemble probability volume defined in Equation (1), and let $F_{\text{wt}}^{(t)}$ denote the macro-F1 obtained from $\bar{\mathbf{p}}_{\text{wt}}^{(t)}$.

Table 1 reports the validation macro-F1 scores for the voxel-wise segmentation step (Section 2.1).

Table 1: Validation macro-F1 scores of the selected single-scale checkpoints for each patch and the resulting ensemble predictions across each map in the validation set. Four-class typed scores are calculated over phosphate, sugar, purine, and pyrimidine voxels; binary is calculated over atom and background noise voxels (Equation (15)). The vertical line separates single-scale checkpoint performance from ensemble performance. **Bold** indicates the best result within each section of a row.

			Equal-weight inference	Weighted inference
Task t	$F_{64}^{(t)}$	$F_{128}^{(t)}$	$F_{\text{eq}}^{(t)}$	$F_{\text{wt}}^{(t)}$
Binary	**0.978**	0.957	**0.985**	0.979
Four-class typed	0.934	**0.949**	**0.952**	0.942

At the single-scale level, the 64^3 model achieved a higher binary macro-F1, while the 128^3 model achieved a higher four-class typed macro-F1. This difference suggests that the two patch sizes provide complementary information for segmentation. Across the validation maps, equal-weight inference achieved higher macro-F1 than weighted inference for both tasks, reaching 0.985 for binary segmentation and 0.952 for typed segmentation. Equal-weight inference was therefore used in the full RNATrace model.

Model variants and ablation conditions. The evaluated variants were defined as follows:

- 128^3 only and 64^3 only use only the corresponding trained patch size for inference instead of the equal-weight ensemble used in the full RNATrace model.
- No flips removes flip-registration, defined in Section 2.4, while no anchors removes fragment-consensus anchor guidance but retains baseline flip-registration.
- ARENA reconstruction uses ARENA [52], with the same coarse 3D model as input, for full atomic reconstruction instead of the reconstruction method defined in Section 2.5.
- No geometric refinement removes the restrained AMBER14 refinement entirely.

Table 2: Ablation study on the fixed 20-RNA cohort. Values are arithmetic means. **Bold** indicates best score per column.

Model	P RMSD (Å)	C4′ RMSD (Å)	Backbone RMSD (Å)	Base RMSD (Å)	TM-score	GDT-TS (%)
RNATrace	**2.91**	**3.36**	**3.68**	5.34	**0.707**	63.5
Weighted inference	3.07	3.48	3.82	5.49	0.680	61.1
128^3 only	3.23	3.58	3.93	5.65	0.661	59.6
64^3 only	3.26	3.54	3.94	5.59	0.654	60.1
No flips	3.47	3.88	4.21	6.31	0.592	52.0
No anchors	3.55	3.81	4.16	5.96	0.647	57.8
ARENA reconstruction	**2.91**	3.50	3.83	**5.15**	0.684	62.0
No geometric refinement	**2.91**	3.81	4.00	6.25	0.666	**64.7**

Full RNATrace achieved the best mean values for phosphorus, C4′, and backbone RMSD and TM-score. ARENA reconstruction achieved a slightly lower base RMSD of 5.15 Å compared with 5.34 Å for RNATrace, but RNATrace

outperformed or matched ARENA on every other metric. The model without geometric refinement achieved a slightly higher GDT-TS of 64.7 compared with 63.5, but RNATrace outperformed or matched it on every other metric (Table 2).

Both single-scale variants performed worse than the two ensemble variants across every tested metric. Equal-weight inference also outperformed weighted inference across all six metrics. Together with the segmentation results in Table 1, this indicates that combining the two patch scales improves downstream reconstruction, while weighting them according to validation macro-F1 provides no additional benefit over equal averaging.

The phosphorus RMSD remained 2.91 Å for RNATrace, ARENA reconstruction, and the model without geometric refinement. For ARENA reconstruction, this is expected because the same coarse phosphorus trace is provided as input to both reconstruction methods. During geometric refinement, strong harmonic position restraints similarly limit displacement of the phosphorus atoms. Despite the unchanged phosphorus RMSD, refinement reduced C4′ RMSD from 3.81 to 3.36 Å, backbone RMSD from 4.00 to 3.68 Å, and base RMSD from 6.25 to 5.34 Å, while increasing TM-score from 0.666 to 0.707. These changes suggest that refinement primarily improves the geometry surrounding the restrained phosphorus trace.

The slightly higher GDT-TS without geometric refinement does not follow the trend of the other structural metrics. Because GDT-TS counts phosphorus atoms within fixed distance thresholds, small restrained coordinate changes can alter the score even when mean phosphorus RMSD remains unchanged at the reported precision. Furthermore, the coordinate metrics do not directly measure steric clashes, which are also targeted during geometric refinement.

ARENA reconstruction achieved a slightly lower base RMSD, but RNATrace achieved lower C4′ and backbone RMSD and higher TM-score and GDT-TS. This result is consistent with the comparison in Section 3.2, where base placement remained the largest local source of error, and suggests that improving nucleotide base placement is an important direction for the all-atom reconstruction stage.

Removing either flips or fragment anchors worsened every mean metric relative to the full RNATrace model. The no-anchors variant retained better C4′, backbone, and base RMSD, TM-score, and GDT-TS than the no-flips variant, showing the importance of baseline flip-registration for correcting the initial Hamiltonian path. Adding fragment anchors further improved every mean metric relative to no anchors, indicating that local sequence evidence effectively provides crucial additional guidance to determine the overall sequence alignment.

To further examine the importance of anchors and flips in sequence registration, we analyzed nucleotide assignment accuracy and sequence recall, defined by Equations (16) and (17), respectively. Nucleotide assignment accuracy measures the correctness of nucleotide identities among aligned residues, while sequence recall also accounts for how much of the complete reference sequence is correctly recovered.

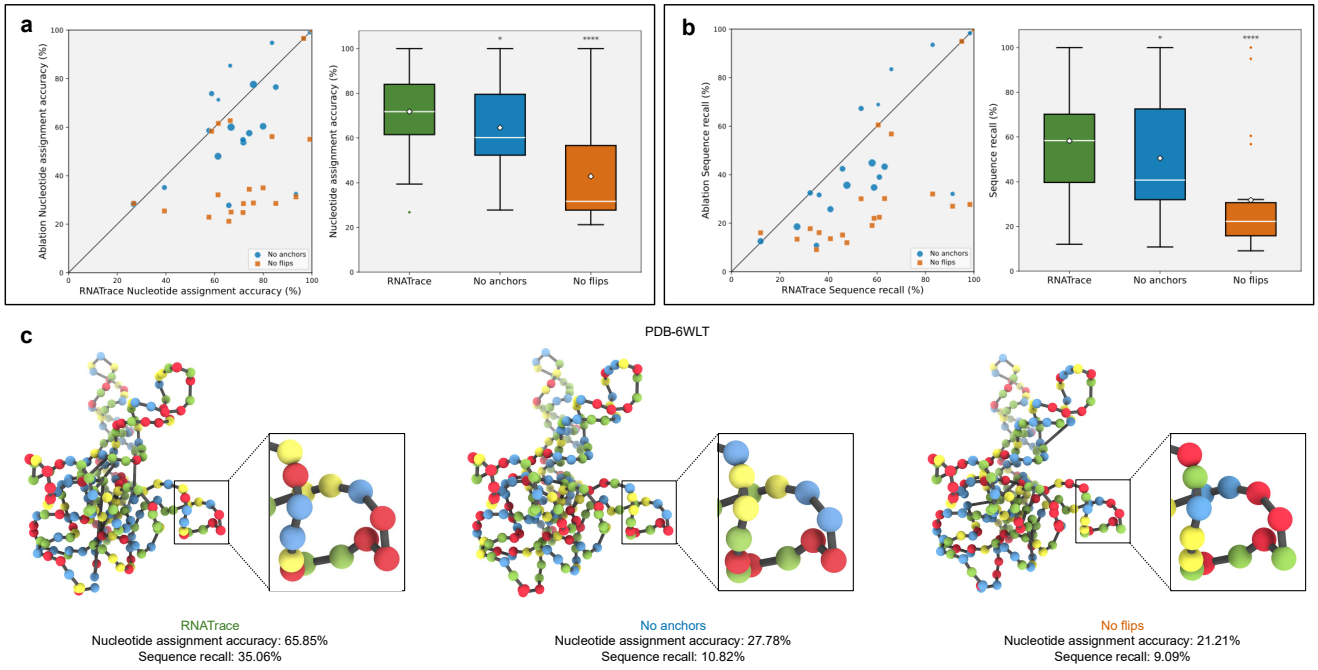


Figure 7: **a**–**b** Scatter plots (left) and box plots (right) comparing the nucleotide assignment accuracy and sequence recall, respectively, of the full RNATrace model against without the fragment-anchor guidance, and without flips entirely. For the scattter plots, the larger the dot in the no anchors case, the higher number of flips performed. The same box plot format as Figure 2 is used. *: $p < 0.05$, ****: $p < 0.0001$. **c** Predicted coarse 3D models for the 4.8 Å apo *Vibrio cholerae* glycine riboswitch (EMD-21841, PDB-6WLT) produced by RNATrace, without fragment-anchor guidance, and without flips, with nucleotides shown in distinct colors.

For both nucleotide assignment accuracy and sequence recall, RNATrace achieved significant improvements compared with no anchors ($p < 0.05$) and no flips ($p < 0.0001$; Figure 7a–b). Across the test set, the no anchors variant exceeded RNATrace on seven targets for nucleotide assignment accuracy and five for sequence recall (Figure 7a–b). The no flips variant exceeded RNATrace on only one target for each metric (Figure 7a–b). These results further reinforce the idea that flip-registration provides a substantial improvement in sequence assignment, while fragment anchors provide additional guidance.

We can also see in the scatter plots of Figures 7a and 7b that the no anchors structures below the $y = x$ line have, on average, more flips than those above. For nucleotide assignment accuracy, there was an average of 5.29 flips above and 8.10 flips below the line; for sequence recall, there was an average of 3.40 flips above and 8.64 flips below the line (Figure 7a–b). This pattern suggests that fragment anchors are particularly beneficial for structures requiring more extensive flip-registration. A larger number of flips likely reflects greater ambiguity in the initial Hamiltonian path, making the sequence assignment more susceptible to incorrect rearrangements without anchor guidance. The anchors therefore appear to provide additional sequence constraints in these more difficult cases, while having less influence when few flips are required. Because flip count is also associated with target difficulty, this relationship should be interpreted as an observed association rather than a direct causal effect.

For the PDB-6WLT example, nucleotide assignment accuracy increased from 27.78% without anchors and 21.21% without flips to 65.85% for RNATrace (Figure 7c). Sequence recall similarly increased from 10.82% and 9.09% to 35.06%, respectively (Figure 7c). For this structure, removing anchors changed 162 of the 230 residue assignments relative to full RNATrace, while removing flips changed 199 assignments (Figure 7c). The large number of changed assignments shows that anchors and flips can alter the overall sequence alignment rather than only correcting isolated nucleotides. However, the sequence recall of 35.06% for the full RNATrace model and that 25% of the test maps have a sequence recall below 40% highlight how difficult registration cases remain unresolved.

4 Future Work

RNATrace currently assumes a single linear RNA chain and constructs one open Hamiltonian path through a determined number of phosphorus sites. Extending it to multi-chain targets will require tracing multiple Hamiltonian paths that jointly partition sites, order each chain, and match each trace to an input sequence.

Sequence registration remains a central limitation. Although flips and fragment anchors improved every mean structural metric, 25% of the test maps had sequence recall below 40%, and high flip counts marked difficult initial traces (Figure 7a–b). Future work could incorporate secondary structure and base-class evidence into initial LKH backbone tracing and replace or augment greedy interval reversal with beam search optimization. Retaining multiple candidate edges and registrations with calibrated confidence could reduce large interval rearrangements. Because of the improvements in the nucleotide assignment accuracy and sequence recall for some maps in the no anchors approach, specifically when there's a low number of flips, the anchor selection approach seems to be optimizable as well based on the number of flips (Figure 7a–b).

Robustness to low or missing density and all-atom accuracy should also be improved. A confidence-aware formulation could permit optional phosphorus sites and model missing-density states. Because base RMSD remained the largest local error and ARENA achieved a slightly lower base RMSD, density-conditioned nucleotide conformers, explicit stacking and hydrogen bond geometry, and refinement against the original map should be evaluated (Table 2).

5 Conclusion

We present RNATrace, a fully automated hybrid framework that converts cryo-EM density maps and RNA sequences into continuous all-atom models. The pipeline combines multiscale Swin UNETR segmentation, phosphorus-site identification and Hamiltonian path tracing, sequence- and secondary-structure-guided flip-registration with fragment anchors, and restrained all-atom reconstruction and refinement. On a 20-map single-chain test set spanning 119–423 nucleotides and 3.0–6.0 Å resolution, RNATrace achieved mean phosphorus, C4′, backbone, and base RMSDs of 2.91, 3.36, 3.68, and 5.34 Å, respectively, together with a mean TM-score of 0.707 and GDT-TS of 63.5.

Compared with CryoREAD, EM2NA, and ModelAngelo, RNATrace achieved lower phosphorus, C4′, and backbone RMSDs and significantly higher aligned-residue percentages, TM-scores, and GDT-TS values. The comparator methods frequently produced multiple disconnected RNA chains, while RNATrace's global backbone construction is designed to recover a continuous trace. Ablation experiments showed that multiscale segmentation and flip-registration each improved reconstruction quality. **Together, these results show that combining learned density interpretation with explicit global connectivity, sequence registration constraints (secondary structure and purine/pyrimidine evidence), and geometric refinement can reduce chain fragmentation while improving local backbone accuracy and global fold recovery.**

6 Code Availability

RNATrace was implemented in Python 3.12. Neural network training and inference used PyTorch 2.9.1 with CUDA 12.8, PyTorch Lightning 2.6.0, and MONAI 1.5.1. Numerical and data pre-processing and predicting phosphorus atoms used CD-HIT 4.8.1, NumPy 2.3.5, and SciPy 1.16.3. Cryo-EM maps and molecular structures were handled with mrcfile 1.5.4, Gemmi 0.7.4, and Biopython 1.86. Hamiltonian path construction used elkai 2.0.1. Secondary-structure predictions used the RNAformer 32M BP-RNA model. All-atom refinement used OpenMM 8.4.0.

The RNATrace source code, trained weights, environment specifications, and scripts for running the complete pipeline are available at `https://github.com/littlebullGit/RNATrace`.

Bibliography

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

[2] Matthew W L Lau and Adrian R Ferré-D'Amaré. Many activities, one structure: Functional plasticity of ribozyme folds. *Molecules (Basel, Switzerland)*, 21(11), 2016.

[3] Laura R. Ganser, Megan L. Kelly, Daniel Herschlag, and Hashim M. Al-Hashimi. The roles of structural dynamics in the cellular functions of RNAs. *Nature Reviews Molecular Cell Biology*, 20(8):474–489, 2019.

[4] Jason R Stagno and Yun-Xing Wang. Riboswitch mechanisms for regulation of p1 helix stability. *International journal of molecular sciences*, 25(19), 2024.

[5] Jennifer A. Doudna and Thomas R. Cech. The chemical repertoire of natural ribozymes. *Nature*, 418(6894):222–228, 2002.

[6] Jessica L. Childs-Disney, Xueyi Yang, Quentin M. R. Gibaut, Yuquan Tong, Robert T. Batey, and Matthew D. Disney. Targeting RNA structures with small molecules. *Nature Reviews Drug Discovery*, 21(10):736–762, 2022.

[7] José Almeida Cruz and Eric Westhof. The dynamic landscapes of RNA architecture. *Cell*, 136(4):604–609, 2009.

[8] Ailong Ke and Jennifer A. Doudna. Crystallization of RNA and RNA-protein complexes. *Methods*, 34(3):408–414, 2004.

[9] Boris Fürtig, Christian Richter, Jens Wöhnert, and Harald Schwalbe. NMR spectroscopy of RNA. *ChemBioChem*, 4(10):936–962, 2003.

[10] Di Liu, François A. Thélot, Joseph A. Piccirilli, Maofu Liao, and Peng Yin. Sub-3-å cryo-EM structure of RNA enabled by engineered homomeric self-assembly. *Nature Methods*, 19(5):576–585, 2022.

[11] Clément Bernard, Guillaume Postic, Sahar Ghannay, and Fariza Tahi. State-of-the-RNArt: benchmarking current methods for RNA 3D structure prediction. *NAR Genomics and Bioinformatics*, 6(2):lqae048, 2024.

[12] Magdalena Rother, Kristian Rother, Tomasz Puton, and Janusz M. Bujnicki. ModeRNA: a tool for comparative modeling of RNA 3D structure. *Nucleic Acids Research*, 39(10):4007–4022, 2011.

[13] Samuel C. Flores, Yaqi Wan, Rick Russell, and Russ B. Altman. Predicting RNA structure by multiple template homology modeling. In *Pacific Symposium on Biocomputing 2010*, pages 216–227, 2010.

[14] Michal J. Boniecki, Grzegorz Lach, Wayne K. Dawson, Konrad Tomala, Pawel Lukasz, Tomasz Soltysinski, Kristian M. Rother, and Janusz M. Bujnicki. SimRNA: a coarse-grained method for RNA folding simulations and 3D structure prediction. *Nucleic Acids Research*, 44(7):e63, 2016.

[15] Rhiju Das, John Karanicolas, and David Baker. Atomic accuracy in predicting and designing noncanonical RNA structure. *Nature Methods*, 7(4):291–294, 2010.

[16] Andrew Martin Watkins, Ramya Rangan, and Rhiju Das. FARFAR2: improved de novo Rosetta prediction of complex global RNA folds. *Structure*, 28(8):963–976.e6, 2020.

[17] José Almeida Cruz, Marie-Françoise Blanchet, Michal Boniecki, Janusz M. Bujnicki, Shi-Jie Chen, Song Cao, Rhiju Das, Feng Ding, Nikolay V. Dokholyan, Samuel C. Flores, et al. RNA-Puzzles: a CASP-like evaluation of RNA three-dimensional structure prediction. *RNA*, 18(4):610–625, 2012.

[18] Rhiju Das, Rachael C. Kretsch, Adam J. Simpkin, et al. Assessment of three-dimensional RNA structure prediction in CASP15. *Proteins: Structure, Function, and Bioinformatics*, 91(12):1747–1770, 2023.

[19] Fan Bu et al. RNA-Puzzles round V: blind predictions of 23 RNA structures. *Nature Methods*, 22(2):399–411, 2025.

[20] Robin Pearce, Gilbert S. Omenn, and Yang Zhang. De novo RNA tertiary structure prediction at atomic resolution using geometric potentials from deep learning. *bioRxiv*, page 2022.05.15.491755, 2022.

[21] Yang Li, Chengxin Zhang, Chenjie Feng, Robin Pearce, P. Lydia Freddolino, and Yang Zhang. Integrating end-to-end learning with deep geometrical potentials for ab initio RNA structure prediction. *Nature Communications*, 14:5745, 2023.

[22] Wenkai Wang, Chenjie Feng, Renmin Han, Ziyi Wang, Lisha Ye, Zongyang Du, Hong Wei, Fa Zhang, Zhenling Peng, and Jianyi Yang. trRosettaRNA: automated prediction of RNA 3D structure with transformer network. *Nature Communications*, 14:7266, 2023.

[23] Tao Shen, Zhihang Hu, Siqi Sun, Di Liu, Felix Wong, Jiuming Wang, Jiayang Chen, Yixuan Wang, Liang Hong, Jin Xiao, Liangzhen Zheng, Tejas Krishnamoorthi, Irwin King, Sheng Wang, Peng Yin, James J. Collins, and Yu Li. Accurate RNA 3D structure prediction using a language model-based deep learning approach. *Nature Methods*, 21(12):2287–2298, 2024.

[24] Rohith Krishna, Jue Wang, Woody Ahern, Pascal Sturmfels, Preetham Venkatesh, Indrek Kalvet, Gyu Rie Lee, F. S. Morey-Burrows, Ivan Anishchenko, Ian R. Humphreys, et al. Generalized biomolecular modeling and design with RoseTTAFold All-Atom. *Science*, 384(6693):eadl2528, 2024.

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

[26] Akash Bahai, Chee Keong Kwoh, Yuguang Mu, and Yinghui Li. Systematic benchmarking of deep-learning methods for tertiary RNA structure prediction. *PLOS Computational Biology*, 20(12):e1012715, 2024.

[27] Chandran Nithin, Sebastian Kmiecik, Roman Błaszczyk, Julita Nowicka, and Irina Tuszyńska. Comparative analysis of RNA 3D structure prediction methods: towards enhanced modeling of RNA–ligand interactions. *Nucleic Acids Research*, 52(13):7465–7486, 2024.

[28] Kalli Kappel, Kaiming Zhang, Zhaoming Su, Andrew M. Watkins, Wipapat Kladwang, Shanshan Li, Grigore Pintilie, Ved V. Topkar, Ramya Rangan, Ivan N. Zheludev, Joseph D. Yesselman, Wah Chiu, and Rhiju Das. Accelerated cryo-EM-guided determination of three-dimensional RNA-only structures. *Nature Methods*, 17(7):699–707, 2020.

[29] Haiyun Ma, Xinyu Jia, Kaiming Zhang, and Zhaoming Su. Cryo-EM advances in RNA structure determination. *Signal Transduction and Targeted Therapy*, 7:58, 2022.

[30] Amit Singer and Fred J. Sigworth. Computational methods for single-particle electron cryomicroscopy. *Annual Review of Biomedical Data Science*, 3:163–190, 2020.

[31] Alok Bharadwaj and Arjen J. Jakobi. Electron scattering properties of biological macromolecules and their use for cryo-EM map sharpening. *Faraday Discussions*, 240:168–183, 2022.

[32] Kaiming Zhang, Shanshan Li, Kalli Kappel, Grigore Pintilie, Zhaoming Su, Tom C. Mou, Michael F. Schmid, Rhiju Das, and Wah Chiu. Cryo-EM structure of a 40 kDa SAM-IV riboswitch RNA at 3.7 å resolution. *Nature Communications*, 10:5511, 2019.

[33] Zhaoming Su, Kaiming Zhang, Kalli Kappel, Shanshan Li, Michael Z. Palo, Grigore D. Pintilie, Ramya Rangan, Bingnan Luo, Yuquan Wei, Rhiju Das, and Wah Chiu. Cryo-EM structures of full-length Tetrahymena ribozyme at 3.1 å resolution. *Nature*, 596(7873):603–607, 2021.

[34] Kalli Kappel, Shiheng Liu, Kevin P. Larsen, Georgios Skiniotis, Elisabetta Viani Puglisi, Joseph D. Puglisi, Z. Hong Zhou, Rui Zhao, and Rhiju Das. De novo computational RNA modeling into cryo-EM maps of large ribonucleoprotein complexes. *Nature Methods*, 15(11):947–954, 2018.

[35] Doo Nam Kim and Karissa Y. Sanbonmatsu. Tools for the cryo-EM gold rush: going from the cryo-EM map to the atomistic model. *Bioscience Reports*, 37(6):BSR20170072, 2017.

[36] Niels Volkmann. The joys and perils of flexible fitting. In Ke-li Han, Xin Zhang, and Ming-jun Yang, editors, *Protein Conformational Dynamics*, volume 805 of *Advances in Experimental Medicine and Biology*, pages 137–155. Springer, Cham, 2014.

[37] Tao Li, Hong Cao, Jiahua He, and Sheng-You Huang. Automated detection and de novo structure modeling of nucleic acids from cryo-EM maps. *Nature Communications*, 15:9367, 2024.

[38] Xiao Wang, Genki Terashi, and Daisuke Kihara. CryoREAD: de novo structure modeling for nucleic acids in cryo-EM maps using deep learning. *Nature Methods*, 20(11):1739–1747, 2023.

[39] Kiarash Jamali, Lukas Käll, Rui Zhang, Alan Brown, Dari Kimanius, and Sjors H. W. Scheres. Automated model building and protein identification in cryo-EM maps. *Nature*, 628(8007):450–457, 2024.

[40] Ali Hatamizadeh, Vishwesh Nath, Yucheng Tang, Dong Yang, Holger Roth, and Daguang Xu. Swin UNETR: Swin transformers for semantic segmentation of brain tumors in MRI images, 2022.

[41] Keld Helsgaun. An effective implementation of the lin–kernighan traveling salesman heuristic. *European Journal of Operational Research*, 126(1):106–130, 2000.

[42] Jörg K.H. Franke, Frederic Runge, Ryan Köksal, Dominika Matus, Rolf Backofen, and Frank Hutter. Rnaformer: A simple yet effective model for homology-aware rna secondary structure prediction. *bioRxiv*, 2024.

[43] Stefan M. Fischer, Lina Felsner, Richard Osuala, Johannes Kiechle, Daniel M. Lang, Jan C. Peeken, and Julia A. Schnabel. Progressive Growing of Patch Size: Resource-Efficient Curriculum Learning for Dense Prediction Tasks . In *proceedings of Medical Image Computing and Computer Assisted Intervention – MICCAI 2024*, volume LNCS 15009. Springer Nature Switzerland, October 2024.

[44] Peter Eastman, Raimondas Galvelis, Raúl P Peláez, Charlles R A Abreu, Stephen E Farr, Emilio Gallicchio, Anton Gorenko, Michael M Henry, Frank Hu, Jing Huang, Andreas Krämer, Julien Michel, Joshua A Mitchell, Vijay S Pande, João Pglm Rodrigues, Jaime Rodriguez-Guerra, Andrew C Simmonett, Sukrit Singh, Jason Swails, Philip Turner, Yuanqing Wang, Ivy Zhang, John D Chodera, Gianni De Fabritiis, and Thomas E Markland. Openmm 8: Molecular dynamics simulation with machine learning potentials. *The journal of physical chemistry. B*, 128(1):109–116, 2024.

[45] W. Kabsch. A solution for the best rotation to relate two sets of vectors. *Acta Crystallographica Section A*, 32(5):922–923, 1976.

[46] W. Kabsch. A discussion of the solution for the best rotation to relate two sets of vectors. *Acta Crystallographica Section A*, 34(5):827–828, 1978.

[47] Yang Zhang and Jeffrey Skolnick. Scoring function for automated assessment of protein structure template quality. *Proteins*, 57(4):702–710, 2004.

[48] Sha Gong, Chengxin Zhang, and Yang Zhang. Rna-align: quick and accurate alignment of rna 3d structures based on size-independent tm-scorerna. *Bioinformatics*, 35(21):4459–4461, 11 2019.

[49] Adam Zemla. Lga: a method for finding 3d similarities in protein structures. *Nucleic Acids Research*, 31(13):3370–3374, 07 2003.

[50] Weizhong Li and Adam Godzik. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, 22(13):1658–1659, 07 2006.

[51] Eric F Pettersen, Thomas D Goddard, Conrad C Huang, Elaine C Meng, Gregory S Couch, Tristan I Croll, John H Morris, and Thomas E Ferrin. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein science : a publication of the Protein Society*, 30(1):70–82, 2021.

[52] Zion R. Perry, Anna Marie Pyle, and Chengxin Zhang. Arena: Rapid and accurate reconstruction of full atomic rna structures from coarse-grained models. *Journal of Molecular Biology*, 435(18):168210, 2023.