

Utilizing Geometric Structural Features for Functional RNA Design

Jiyue Jiang^{♡*}, Yuhan Chen^{♡*}, Xiangyu Shi[♡], Yihan Ye[♣], Han Huang[♡], Yanyu Chen[♡],
Yuheng Shan[♣], Dongchen He[♡], Jiao Yuan[◇], Yu Li[♡]

[♡] The Chinese University of Hong Kong, [♣] Hong Kong University of Science and Technology,

[♣] National University of Singapore, [◇] Guangzhou National Laboratory

{jiangjy, yuhanchen, chenyanu.cse, dche}@link.cuhk.edu.hk

sxysxym@gmail.com, yyeyby@connect.ust.hk, shan.yuheng@u.nus.edu,

hanhuang@cuhk.edu.hk, yuan_jiao@gzlab.ac.cn, liyu@cse.cuhk.edu.hk

Abstract

RNA’s conformational flexibility enables diverse biological functions but makes structure-guided design difficult. Although precise control of 3D structure is essential, existing methods often overlook the molecular surface, which critically shapes intermolecular interactions and functional outcomes. To address this, we propose gRFold, a multimodal framework that incorporates molecular surface information as a core design constraint. gRFold employs two geometric encoders: a Frame-Averaged Geometric Surface Encoder (FA-GSE) to learn E(3)-invariant surface features, and an Equivariant Backbone Graph Perceptron (EBGP) to capture backbone topology. The structural representations are integrated via hierarchical cross-attention to guide autoregressive generation; at inference, candidates are re-ranked by agreement between their predicted folds and the target base-pairing pattern. Across three RNA inverse-folding benchmarks covering RNA–protein complexes, monomeric RNAs and independently curated RNA-Puzzles targets, gRFold achieves strong sequence recovery and structural fidelity under both secondary-structure and 3D refolding evaluations, and ablations show that the surface and backbone branches contribute complementary gains. These results highlight the potential of surface-aware geometric modeling for structure-guided RNA design.

1 Introduction

As a molecule with the dual roles of carrying genetic information and executing biological functions, RNA exhibits immense potential in fields such as gene expression regulation, biocatalysis, and drug development (Isaacs et al., 2006; Yesselman et al., 2019). However, unlike proteins, RNA lacks a stable tertiary core and thus exhibits high conformational flexibility (Shen et al., 2024; Wong

et al., 2024). This flexibility enables RNA to adopt diverse spatial structures to perform complex functions, but it also poses significant challenges for precise, structure-guided design (Nori and Jin, 2024; Huang et al., 2024). Since RNA function is strongly linked to molecular conformation, improving structural control is an important objective for structure-guided RNA design (Wong et al., 2024).

Traditional RNA design methods, from early secondary-structure inverse folding (Runge et al., 2018; Zhou et al., 2023; Runge et al., 2024) to models incorporating 3D backbone topology (Huang et al., 2024; Joshi et al., 2025), largely overlook the molecular surface, a key determinant of RNA function (Song et al., 2024). RNA interactions with proteins (Corley et al., 2020), small molecules (Huang et al., 2024), and other nucleic acids (Song et al., 2024) are influenced by surface geometry and chemical properties such as charge distribution and hydrogen-bonding sites. As a result, even RNAs with plausible backbone folds may still present suboptimal interaction interfaces if critical surface sites are not properly represented (Tan et al., 2024; Huang et al., 2024).

To fill this gap, we propose gRFold, which incorporates molecular surface information as a design constraint alongside secondary structure and backbone topology (Zhou et al., 2023; Runge et al., 2024; Huang et al., 2024; Song et al., 2024). As shown in Figure 1, gRFold combines a rotation-invariant surface encoder, an equivariant backbone graph encoder, hierarchical cross-attention fusion, and a 24-layer autoregressive decoder. Across three RNA inverse-folding benchmarks, gRFold improves sequence recovery and computational structural fidelity over the compared inverse-folding methods, supporting molecular surface constraints as a useful component for computational RNA design.

Our contributions are summarized as follows: (1) We establish molecular surface geometry as a com-

*These authors are co-first authors.

putational design constraint for RNA inverse folding, addressing a key limitation of traditional methods. (2) We propose FA-GSE for $E(3)$ -invariant surface encoding and EBGp for equivariant backbone representation learning. (3) We design a hierarchical cross-attention mechanism to effectively fuse secondary structure, backbone, and surface information. (4) Extensive benchmarking demonstrates the strong computational structural performance of gRFold across three RNA inverse-folding benchmarks.

2 Related Works

RNA structure prediction and design have evolved from energy-based secondary-structure prediction relying on minimum free energy principles (Dirks et al., 2007; Lorenz et al., 2011; Zuker, 2003) to deep learning approaches for RNA inverse folding (Morandi et al., 2022; Magnus et al., 2016; Sato et al., 2021; Fu et al., 2022; Singh et al., 2019; Shen et al., 2024). While accuracy and diversity have improved, existing methods often neglect surface information, which has been reported to improve molecular interaction modeling and computational molecular design in the protein setting using mesh or point-cloud representations (Gainza et al., 2020; Sverrisson et al., 2021; Song et al., 2024). Surface information thus remains underutilized in RNA design, where methodologies predominantly focus on secondary structures (Liu et al., 2024); this study integrates surface-level features into the RNA design framework to improve structural representation. A fuller discussion is given in Appendix A.

3 Methodology: gRFold

3.1 Overview and Problem Formulation

As illustrated in Figure 1, gRFold formulates RNA inverse folding as conditional sequence generation driven by three complementary views of the target molecule: a target secondary structure written as a dot-bracket string $X = (x_1, \dots, x_N)$; a molecular-surface point cloud $C \in \mathbb{R}^{M \times 3}$ extracted from the tertiary structure, which describes the interaction interface; and a coarse-grained backbone tensor $S \in \mathbb{R}^{N \times 3 \times 3}$, which describes the global fold scaffold. Given the condition triplet $\mathcal{C} = (X, C, S)$, the goal is to generate a nucleotide sequence $Y = (y_1, \dots, y_N)$ with $y_i \in \{A, U, G, C\}$ that is simultaneously compatible with the prescribed base-pairing pattern, backbone topology, and surface geometry:

Module	Configuration	Output
FA-GSE (Surface Enc.)	$4 \times \text{EGCL} + \text{FAMHA}$	$M \times 1024$
EBGP (Structure Enc.)	$3 \times \text{GVP} + \text{message pass.}$	$N \times 512$
Fusion	$2 \times \text{cross-attn. block}$	$N \times d_{\text{model}}$
Decoder	$24 \times \text{masked Transformer}$	$N \times 4$

Table 1: Configuration of the gRFold components, where $M = 512$ is the number of surface points and N is the number of residues. Module names in parentheses correspond to the blocks of Figure 1.

$$p_{\theta}(Y | \mathcal{C}) = \prod_{i=1}^N p_{\theta}(y_i | y_{<i}, \mathcal{C}) \quad (1)$$

The framework comprises four components shown in Figure 1: a Frame-Averaged Geometric Surface Encoder (FA-GSE, Sec. 3.2) mapping the surface point cloud to an $E(3)$ -invariant embedding H_{surf} ; an Equivariant Backbone Graph Perceptron (EBGP, Sec. 3.3) mapping the backbone graph to an orientation-invariant embedding H_{backbone} ; a 24-layer autoregressive Transformer decoder (Sec. 3.5) for nucleotide generation; and hierarchical cross-attention (Sec. 3.4) to integrate the structural representations. Maximum-likelihood training is complemented at inference by structure-aware sampling and re-ranking (Sec. 3.6). We write $[\cdot; \cdot]$ for concatenation, $\text{LN}(\cdot)$ for layer normalisation, $\text{XA}(Q, K, V)$ for multi-head cross-attention, and $\text{PW}(\cdot)$ for the position-wise residual sub-layer closing each fusion block. Table 1 summarises the components.

3.2 Molecular Surface Encoding

The geometric properties of an RNA molecule are largely reflected by the shape and chemistry of its molecular surface. We therefore first convert the raw atomic model into a compact point cloud, and then encode this point cloud into a representation that is insensitive to the arbitrary pose in which the structure was deposited.

3.2.1 Surface Extraction and Downsampling

We follow the approach of SurfPro (Song et al., 2024) to encode biomolecular surfaces. Atomic types and coordinates are obtained from PDB files (Berman et al., 2000), and the molecular surface is computed using MSMS as a dense point cloud, which is denoised via Gaussian kernel smoothing. Because the raw surface is highly redundant, we adopt the octree-based downsampling strategy of SurfPro: the 3D space is recur-

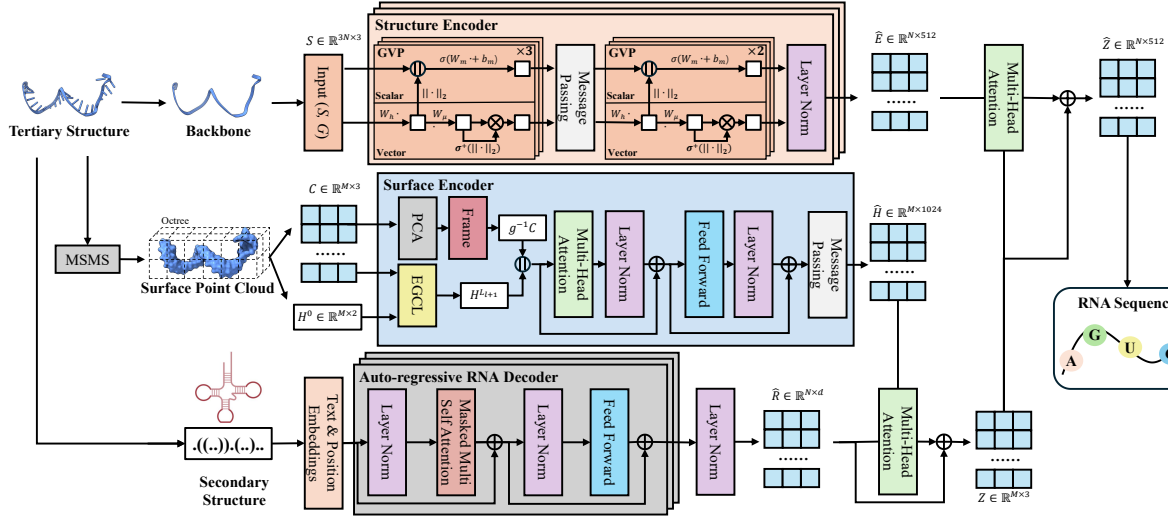


Figure 1: gRFold architecture: FA-GSE employs four equivariant graph-convolution layers followed by frame-averaged multi-head attention to yield a rotation-invariant surface embedding ($d = 1024$), while EBGP stacks three GVP layers with message passing to obtain the backbone embedding ($d = 512$). Secondary-structure representations are integrated with the surface and backbone embeddings through hierarchical cross-attention to guide autoregressive nucleotide generation. The blocks labelled *Surface Encoder*, *Structure Encoder* and *Auto-regressive RNA Decoder* correspond to Sec. 3.2, Sec. 3.3 and Sec. 3.5, and the tensor shapes match Table 1.

sively partitioned, sparse regions are collapsed into leaf nodes, and points are sampled from these leaf nodes. This yields a compressed, resolution-balanced representation $C = [x_1, \dots, x_M]^T \in \mathbb{R}^{M \times 3}$ with $M = 512$ points, keeping the encoder footprint independent of molecule size. Each retained point carries a low-dimensional scalar annotation inherited from its nearest atom, which initialises the node features of the surface encoder in Figure 1.

3.2.2 Frame-Averaged Geometric Surface Encoder (FA-GSE)

FA-GSE encodes the surface in two stages: local perspective modeling, which propagates geometric information within local neighbourhoods, and global frame averaging, which removes the dependence of the final descriptor on the orientation and the position of the molecule.

Local Perspective Modeling. This stage utilises a stack of $L_l = 4$ EGCL (Satorras et al., 2021) to iteratively update node features based on local neighbourhood information. For each point x_i we identify its $K = 16$ nearest neighbours, denoted by the set $\mathcal{N}(x_i)$, and in every layer $l \in \{0, \dots, L_l - 1\}$ a message from a neighbour j to a central node i is computed by combining the two feature vectors with their geometric relationship:

$$m'_{ij} = \phi_e([h_i^l; h_j^l; \|x_i - x_j\|^2]) \quad (2)$$

Here $h_i^l \in \mathbb{R}^{d_h}$ is the feature vector of node i at layer l and $\|x_i - x_j\|^2$ is the squared Euclidean distance (Sherali and Tuncbilek, 1992), while ϕ_e is a Multi-Layer Perceptron with a SiLU activation function. Because the coordinates enter only through pairwise distances, the messages are unaffected by rigid motions of the point cloud.

The messages are then re-weighted by a learned distribution over the neighbourhood, so that geometrically salient neighbours contribute more strongly to the update than flat or noisy surface patches:

$$\begin{aligned} z_{ij}^l &= W_s^l m'_{ij} + b_s^l, \\ \alpha_{ij}^l &= \frac{\exp(z_{ij}^l)}{\sum_{k \in \mathcal{N}(x_i)} \exp(z_{ik}^l)}, \\ m_{ij}^{l+1} &= \alpha_{ij}^l m'_{ij} \end{aligned} \quad (3)$$

where W_s^l and b_s^l are learnable parameters of the attention mechanism in layer l . The feature vector of each node is then refined by aggregating the incoming messages and applying a gated residual rule:

$$\begin{aligned} c_i^{l+1} &= \sum_{j \in \mathcal{N}(x_i)} m_{ij}^{l+1}, \\ h_i^{l+1} &= h_i^l + \phi_h(c_i^{l+1}) \end{aligned} \quad (4)$$

where $\phi_h(c) = \text{sigmoid}(c) \odot c$ is a gated non-linearity that controls the information flow. The output of this stage is a set of locally aware point features $H^{L_l} = [h_1^{L_l}, \dots, h_M^{L_l}]^\top$.

Global Frame Averaging. To make the final surface representation insensitive to rigid transformations, we employ a frame-averaging technique. We first apply Principal Component Analysis (PCA) to the coordinates C in order to define a canonical local coordinate system, given by the three principal component vectors $\{v_1, v_2, v_3\}$ together with the centroid t of the point cloud, and we then generate a set of 8 frames by enumerating all sign patterns of the basis vectors, so that reflections are covered as well:

$$\begin{aligned} R_\alpha &= [\alpha_1 v_1, \alpha_2 v_2, \alpha_3 v_3], \\ \mathcal{F}(C) &= \{ (R_\alpha, t) : \alpha \in \{-1, +1\}^3 \} \end{aligned} \quad (5)$$

For a frame $g = (R_\alpha, t) \in \mathcal{F}(C)$, mapping the point cloud into the canonical coordinate system amounts to a centring followed by a change of basis:

$$g^{-1} \cdot C = (C - \mathbf{1} t^\top) R_\alpha \quad (6)$$

The final surface embedding is then obtained by concatenating the transformed coordinates with the learned local features, processing them with an attention layer, and averaging the outputs over the whole frame set:

$$\begin{aligned} Z_g &= \text{FAMHA}([H^{L_l}; g^{-1} \cdot C]), \\ H_{\text{surf}} &= \frac{1}{|\mathcal{F}(C)|} \sum_{g \in \mathcal{F}(C)} Z_g \end{aligned} \quad (7)$$

FAMHA (Frame-Averaged Multi-Head Attention) is a Transformer layer (Vaswani et al., 2017) over the concatenated local features and canonicalised coordinates; because the average is taken over the complete frame set, the descriptor is designed to be invariant to rotations, reflections and translations. This yields $H_{\text{surf}} \in \mathbb{R}^{M \times d_{\text{surf}}}$ with $d_{\text{surf}} = 1024$, which captures the intrinsic geometry of the RNA surface while retaining expressive power over local shape.

3.3 Backbone Structure Encoding

While the molecular surface defines where interactions take place, the three-dimensional backbone positions these interfaces in space and dictates the overall fold. We therefore introduce a graph-based encoder built on $E(3)$ equivariance, which learns structural features independent of the absolute position and orientation of the molecule.

3.3.1 Backbone Graph Construction

To construct a computationally efficient yet structurally faithful representation of the RNA backbone, we select the 3D coordinates of three key atoms from each nucleotide residue: the phosphorus atom (P), and the C1' and C3' carbon atoms of the ribose sugar. This trio of atoms characterises the local geometry and the chain progression of the backbone, so that for an RNA molecule comprising N residues the backbone structure is represented as a coordinate tensor $S \in \mathbb{R}^{N \times 3 \times 3}$.

Based on these coordinates we construct a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, in which each node $i \in \mathcal{V}$ corresponds to a single residue. The connectivity is established using a K-Nearest Neighbours algorithm: for each residue i we identify its k closest neighbours by computing the Euclidean distances between its phosphorus atom and those of all other residues, with the number of neighbours set to $k = 30$. Every edge $(i, j) \in \mathcal{E}$ carries a feature e_{ij} that summarises the relative geometry of the two residues, so that both spatial proximity and chain context are available to the network.

3.3.2 Equivariant Backbone Graph Perceptron (EBGP)

The Equivariant Backbone Graph Perceptron (EBGP) stacks $L_g = 3$ Geometric Vector Perceptron (GVP) (Jing et al., 2020) layers, which model scalar and vector features in an equivariant manner. Each node i is initialised with $h_i^{(0)} = (s_i^{(0)}, v_i^{(0)})$, where the vector part derives from intra-residue atomic coordinates relative to their centroid and the scalar part encodes local geometric descriptors. EBGP then refines these features through three operations per layer l .

Message Construction. For each node i , messages are gathered from its neighbourhood $\mathcal{N}(i)$: a GVP module g_{msg} non-linearly combines the features of a neighbour j with the corresponding edge feature, and the messages are aggregated by summation into a single permutation-invariant update:

$$\begin{aligned} h_{j \rightarrow i}^{(l)} &= g_{\text{msg}}([h_j^{(l)}; e_{ij}]), \\ \bar{h}_i^{(l)} &= \sum_{j \in \mathcal{N}(i)} h_{j \rightarrow i}^{(l)} \end{aligned} \quad (8)$$

Equivariant Update. The aggregated message updates the node feature tuple through a second GVP module g_{update} , followed by a residual connection and layer normalisation, where the normalisation is applied separately to the scalar and the vector channels so as to preserve their distinct transformation behaviour:

$$h_i^{(l+1)'} = \text{LN}(h_i^{(l)} + g_{\text{update}}(\bar{h}_i^{(l)})) \quad (9)$$

Feed-Forward Network. To further enhance the representational power of the model, the updated features are passed through a node-wise feed-forward network, which is also implemented as a GVP module g_{ffn} and is embedded in a residual block with dropout for regularisation:

$$\begin{aligned} u_i^{(l)} &= \text{Dropout}(g_{\text{ffn}}(h_i^{(l+1)'})), \\ h_i^{(l+1)} &= \text{LN}(h_i^{(l+1)'} + u_i^{(l)}) \end{aligned} \quad (10)$$

After the final layer L_g , the scalar component is extracted from each node tuple $h_i^{(L_g)} = (s_i^{(L_g)}, v_i^{(L_g)})$ and the scalars are stacked over residues, which yields an orientation-invariant description of the local and global structural environment of every nucleotide:

$$\begin{aligned} H_{\text{backbone}} &= \text{EBGP}(S, \mathcal{G}) \\ &= [s_1^{(L_g)}, \dots, s_N^{(L_g)}]^\top, \\ H_{\text{backbone}} &\in \mathbb{R}^{N \times d_{\text{backbone}}} \end{aligned} \quad (11)$$

The dimensionality of this embedding is set to $d_{\text{backbone}} = 512$, which provides a concise structural representation for the subsequent stages of our model.

3.4 Hierarchical Multimodal Fusion

The decoder must reconcile three heterogeneous signals: a discrete secondary-structure string, an unordered surface point set of length M , and a residue-indexed backbone representation of length N . Hierarchical cross-attention integrates these representations, progressively from the local interaction interface to the global scaffold.

Secondary Structure Embedding. The input dot-bracket sequence X is encoded into a secondary-structure representation $H_{\text{sec}} \in \mathbb{R}^{N \times d_{\text{model}}}$, where N is the sequence length and d_{model} is the decoder hidden dimension. This representation is used for subsequent integration with geometric features.

Stage 1: Structure-to-Surface Attention. The secondary-structure representation attends to the surface features, allowing the model to learn associations between sequence positions and surface patches. The result is merged through a residual connection and layer normalisation, followed by the position-wise sub-layer $\text{PW}(\cdot)$ of the fusion block in Figure 1:

$$\begin{aligned} A^{(1)} &= \text{XA}(H_{\text{sec}}, H_{\text{surf}}, H_{\text{surf}}), \\ \tilde{H}^{(1)} &= \text{LN}(H_{\text{sec}} + A^{(1)}), \\ H^{(1)} &= \text{PW}(\tilde{H}^{(1)}). \end{aligned} \quad (12)$$

Stage 2: Surface-to-Backbone Attention. The resulting surface-aware representation then attends to the backbone features, which anchors the interface information onto the global backbone scaffold:

$$\begin{aligned} A^{(2)} &= \text{XA}(H^{(1)}, H_{\text{backbone}}, H_{\text{backbone}}), \\ \tilde{H}^{(2)} &= \text{LN}(H^{(1)} + A^{(2)}), \\ H_{\text{final}} &= \text{PW}(\tilde{H}^{(2)}). \end{aligned} \quad (13)$$

Arranging the two stages in this order lets fine-grained interface cues be resolved before they are placed in the global structural context, and the contribution of each stream is reported by the ablation rows in Table 2. The resulting tensor $H_{\text{final}} \in \mathbb{R}^{N \times d_{\text{model}}}$ integrates all three modalities to guide autoregressive nucleotide generation.

3.5 Autoregressive Decoding and Training

The autoregressive Transformer decoder has $L_d = 24$ layers (Vaswani et al., 2017) and uses the fused structural representation to guide nucleotide generation. The output states are projected by a linear head onto the four nucleotides ($N \times 4$), as drawn in Figure 1. At each step i , the prediction is conditioned on $y_{<i}$ and the complete structural condition $\mathcal{C}$, and the model is trained end-to-end on paired structure–sequence data by minimising the negative log-likelihood:

$$\mathcal{L}_{\text{MLE}}(\theta) = - \sum_{i=1}^N \log p_\theta(y_i | y_{<i}, \mathcal{C}) \quad (14)$$

Algorithm 1 gRFold: geometric encoding, hierarchical fusion and structure-aware re-ranking

Require: Atomic model $\mathcal{A}$; dot-bracket target X ; sample size n

Ensure: Designed RNA sequence Y_{best}

```

1: // Surface preprocessing
2:  $P \leftarrow \text{GaussianSmooth}(\text{MSMS}(\mathcal{A}))$ 
3:  $C \leftarrow \text{OctreeDownsample}(P, M)$ 
4: // FA-GSE: local perspective modeling
5: Build  $K$ -NN graph on  $C$ ; init  $\{h_i^0\}_{i=1}^M$ 
6: for  $l = 0$  to  $L_l - 1$  do
7:    $m'_{ij} \leftarrow \phi_e([h_i^l; h_j^l; \|x_i - x_j\|^2])$ 
8:    $c_i^{l+1} \leftarrow \sum_j \alpha_{ij} m'_{ij}$ 
9:    $h_i^{l+1} \leftarrow h_i^l + \phi_h(c_i^{l+1})$ 
10: end for
11: // FA-GSE: global frame averaging
12:  $\{v_1, v_2, v_3\}, t \leftarrow \text{PCA}(C)$ ; form  $\mathcal{F}(C)$ 
13:  $H_{\text{surf}} \leftarrow \frac{1}{|\mathcal{F}(C)|} \sum_g \text{FAMHA}([H^{L_l}; g^{-1}C])$ 
14: // EBG: backbone encoding
15:  $S \leftarrow \text{Extract}(\mathcal{A})$ ;  $\mathcal{G} \leftarrow \text{KNNGraph}(S, k)$ 
16: Init  $h_i^{(0)} \leftarrow (s_i^{(0)}, v_i^{(0)})$ 
17: for  $l = 0$  to  $L_g - 1$  do
18:    $\tilde{h}_i^{(l)} \leftarrow \sum_j g_{\text{msg}}([h_j^{(l)}; e_{ij}])$ 
19:    $h_i^{(l)} \leftarrow \text{LN}(h_i^{(l)} + g_{\text{update}}(\tilde{h}_i^{(l)}))$ 
20:    $h_i^{(l+1)} \leftarrow \text{LN}(h_i^{(l)} + g_{\text{fn}}(h_i^{(l)}))$ 
21: end for
22:  $H_{\text{backbone}} \leftarrow [s_1^{(L_g)}, \dots, s_N^{(L_g)}]^\top$ 
23: // Hierarchical multimodal fusion
24:  $H_{\text{sec}} \leftarrow \text{EncodeSecondaryStructure}(X)$ 
25:  $\tilde{H}^{(1)} \leftarrow \text{LN}(H_{\text{sec}} + \text{XA}(H_{\text{sec}}, H_{\text{surf}}, H_{\text{surf}}))$ 
26:  $H^{(1)} \leftarrow \text{PW}(\tilde{H}^{(1)})$ 
27:  $\tilde{H}^{(2)} \leftarrow \text{LN}(H^{(1)} + \text{XA}(H^{(1)}, H_{\text{backbone}}, H_{\text{backbone}}))$ 
28:  $H_{\text{final}} \leftarrow \text{PW}(\tilde{H}^{(2)})$ 
29: // Sampling and re-ranking
30:  $s_{\text{best}} \leftarrow -\infty$ 
31: for  $t = 1$  to  $n$  do
32:    $\hat{Y} \leftarrow \text{NucleusSample}(p_\theta(\cdot | H_{\text{final}}))$ 
33:    $s \leftarrow \text{F1}(X, \text{RNAfold}(\hat{Y}))$ 
34:   if  $s > s_{\text{best}}$  then
35:      $s_{\text{best}} \leftarrow s$ ;  $Y_{\text{best}} \leftarrow \hat{Y}$ 
36:   end if
37: end for
38: return  $Y_{\text{best}}$ 

```

Here θ collects the parameters of the two geometric encoders, the fusion module and the decoder, optimised jointly. Since H_{surf} and H_{backbone} are both designed to be invariant to rigid transformations of the input structure, the induced conditional distribution is expected to satisfy

$$p_\theta(Y | g \cdot \mathcal{C}) = p_\theta(Y | \mathcal{C}), \quad g \in E(3) \quad (15)$$

so that it is designed to be insensitive to rotated, reflected or translated copies of the same target, removing the need for pose augmentation. In terms of cost, the surface branch scales as $\mathcal{O}(L_l M K d_h)$ plus a constant factor $|\mathcal{F}(C)| = 8$

for frame-averaged attention, the backbone branch as $\mathcal{O}(L_g N k)$, and the decoder retains the usual quadratic dependence on length; since M is fixed by the octree downsampling, the geometric encoders add only bounded overhead.

3.6 Inference: Sampling with Structure-Aware Re-ranking

Because sequence-to-structure mapping is many-to-one, a greedy decode may miss the target topology. We therefore draw $n = 50$ candidates $\mathcal{Y}_n$ by nucleus sampling (Holtzman et al., 2019), fold each with RNAfold (Lorenz et al., 2011), and score its base-pair agreement with the target. With $\mathcal{P}(\cdot)$ denoting the base-pair set of a dot-bracket string, the score and selected design are

$$\begin{aligned} \text{F1}(X, \hat{X}) &= \frac{2|\mathcal{P}(X) \cap \mathcal{P}(\hat{X})|}{|\mathcal{P}(X)| + |\mathcal{P}(\hat{X})|}, \\ Y_{\text{best}} &= \arg \max_{\hat{Y} \in \mathcal{Y}_n} \text{F1}(X, \text{RNAfold}(\hat{Y})) \end{aligned} \quad (16)$$

Re-ranking provides a thermodynamic quality-control filter, with the n independent folding calls treated as lightweight post-processing. Algorithm 1 summarises the complete pipeline.

4 Experiments and Analysis

We evaluate gRFold on held-out RNAsolo and RNA-Puzzles test sets and ablate both geometric components to quantify their individual contributions to performance.

4.1 Experiment Settings

Implementation Details All models are implemented with PyTorch (Paszke et al., 2019) on eight NVIDIA A100-80G GPUs. Training is orchestrated using the AdamW optimizer (Loshchilov and Hutter, 2017) with decoupled weight decay ($\lambda = 0.01$). We employ a *Linear Warmup with Cosine Annealing Learning Rate Schedule* (Vaswani et al., 2017). Batch size of 16 is utilized. The total duration for pre-training and fine-tuning is six weeks. During inference, a temperature of 1.0, a top- k value of 800, and a top- p value of 0.7 are set for stochastic decoding.

Datasets We first pre-train the model on RNA sequences from RNACentral (rna, 2019). For supervised training and subsequent fine-tuning, we use the RNAsolo dataset (Adamczyk et al., 2022); after preprocessing and quality filtering,

Secondary structure-based inverse folding						Tertiary structure-based inverse folding							
Methods	Recovery (%)	RhoFold RMSD ↓	TM ↑	Boltz-2 RMSD ↓	TM ↑	Methods	Recovery (%)	F1 ↑ RNAfold	BPfold	RhoFold RMSD ↓	TM ↑	Boltz-2 RMSD ↓	TM ↑
RNASOLO RBP-COMPLEX BENCHMARK													
MCTS-RNA	28.604	3.497	0.203	3.119	0.281	gRNAde	40.240	<u>0.479</u>	0.525	<u>3.092</u>	0.299	3.267	0.268
RNAinverse	23.888	<u>3.362</u>	0.190	2.681	0.251	RDesign	40.317	0.146	0.172	<u>3.407</u>	0.213	3.386	0.219
SAMFEO(ned)	30.744	3.522	0.215	3.182	0.271	RiboDiffusion	36.851	0.460	<u>0.507</u>	3.333	0.242	<u>3.002</u>	<u>0.303</u>
SAMFEO(pd)	30.265	3.456	0.216	3.195	0.268								
w/o backbone	41.013	3.490	0.232	3.301	0.273	w/o backbone	41.013	0.337	0.373	3.490	0.232	3.301	0.273
w/o surface	<u>49.175</u>	3.401	<u>0.255</u>	3.073	<u>0.293</u>	w/o surface	<u>49.175</u>	0.464	0.491	3.401	0.255	3.073	0.293
gRFold (Ours)	49.650	3.031	0.297	<u>2.700</u>	0.336	gRFold (Ours)	49.650	0.497	0.525	3.031	<u>0.297</u>	2.700	0.336
RNASOLO MONOMERIC BENCHMARK													
MCTS-RNA	27.828	3.452	0.221	3.165	0.281	gRNAde	33.729	0.378	0.380	<u>3.364</u>	0.222	<u>2.962</u>	<u>0.311</u>
RNAinverse	23.478	<u>3.188</u>	0.207	2.437	0.258	RDesign	36.949	0.187	0.208	3.657	0.183	3.117	0.262
SAMFEO(ned)	30.754	3.552	0.216	3.230	<u>0.283</u>	RiboDiffusion	34.948	0.432	0.500	3.551	0.219	3.034	0.285
SAMFEO(pd)	29.975	3.500	0.222	3.237	0.276								
w/o backbone	42.306	3.632	0.218	3.461	0.254	w/o backbone	42.306	0.355	0.373	3.632	0.218	3.461	0.254
w/o surface	<u>50.726</u>	3.568	<u>0.236</u>	3.175	<u>0.283</u>	w/o surface	<u>50.726</u>	<u>0.451</u>	0.440	3.568	<u>0.236</u>	3.175	0.283
gRFold (Ours)	51.671	3.165	0.277	<u>2.789</u>	0.323	gRFold (Ours)	51.671	0.486	<u>0.485</u>	3.165	0.277	2.789	0.323
RNA PUZZLES BENCHMARK													
MCTS-RNA	27.045	4.283	0.226	3.887	0.315	gRNAde	34.314	0.320	0.302	4.018	0.247	3.870	0.301
RNAinverse	24.415	3.883	0.213	2.810	0.256	RDesign	37.333	0.020	0.023	4.710	0.155	3.703	0.260
SAMFEO(ned)	36.690	4.299	0.254	3.865	0.330	RiboDiffusion	37.211	0.457	<u>0.497</u>	4.361	0.225	3.781	0.297
SAMFEO(pd)	32.858	4.257	0.248	3.922	0.304								
w/o backbone	40.090	4.651	0.197	3.968	0.269	w/o backbone	40.090	0.415	0.396	4.651	0.197	3.968	0.269
w/o surface	<u>44.725</u>	<u>3.841</u>	<u>0.308</u>	3.358	<u>0.355</u>	w/o surface	<u>44.725</u>	0.626	0.580	<u>3.841</u>	<u>0.308</u>	<u>3.358</u>	<u>0.355</u>
gRFold (Ours)	45.004	3.820	0.310	<u>3.332</u>	0.370	gRFold (Ours)	45.004	0.626	0.580	3.820	0.310	3.332	0.370

Table 2: Comparison with secondary- and tertiary-structure inverse folding methods across three RNA benchmarks. We report sequence recovery, secondary-structure F1 from **RNAfold** and **BPfold**, and 3D fidelity after refolding with **RhoFold** and **Boltz-2**. RMSD (Å) and TM-score (reported as TM in the table) are computed against native structures, while ablation rows (w/o surface/backbone) assess geometric contributions. Best values are bold; second-best values are underlined to support a consistent cross-dataset comparison.

the training split contains 13,769 RNA sequence–tertiary-structure pairs. We evaluate the resulting model on three held-out benchmarks: the RNAsolo RBP-Complex Benchmark (283 sequences), the RNAsolo Monomeric Benchmark (278 sequences), and the RNA Puzzles Benchmark (21 sequences). These benchmarks cover RNA–protein complexes, monomeric RNA structures, and independently curated RNA puzzle targets, respectively, probing both an interface-driven regime, where surface geometry is expected to be informative, and a free-standing structural-design regime.

Evaluation For sequence-level evaluation, we report Recovery Rate. For secondary-structure consistency, we report the F1 score after refolding the designed sequence with RNAfold (Lorenz et al., 2011) and BPfold. For 3D structural fidelity, we report RMSD and TM-score against the native structure after refolding with RhoFold and Boltz-2. Reporting two independent predictors at each structural level guards against an apparent improvement being an artefact of a single evaluation tool.

Baselines We compare gRFold with four secondary-structure inverse folding baselines: **MCTS-RNA** (Yang et al., 2017), **RNAinverse** (Lorenz et al., 2011), and two variants of **SAMFEO**, **SAMFEO(ned)** and **SAMFEO(pd)** (Zhou et al., 2023). For tertiary-structure inverse folding, we include **gRNAde** (Joshi et al., 2025), **RDesign** (Tan et al., 2024), and **RiboDiffusion** (Huang et al., 2024). We additionally report two ablations of our own model: **w/o backbone** removes the EBG branch and therefore the term $A^{(2)}$, while **w/o surface** removes the FA-GSE branch and therefore the term $A^{(1)}$, which isolates the individual contribution of each geometric modality.

4.2 Results and Analysis

4.2.1 Main Results

Overall Structural Performance. Across all three benchmarks, gRFold achieves the highest sequence recovery among all compared methods, reaching 49.65%, 51.67%, and 45.00% on the RNAsolo

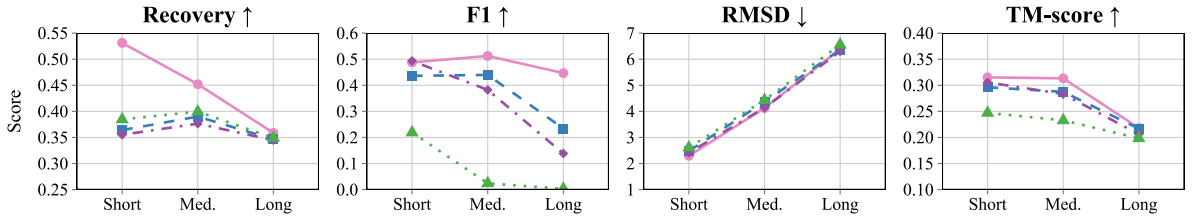


Figure 2: Performance scaling analysis of gRFold versus baselines across different RNA sequence lengths (Short: <80 nt, Med.: 80–200 nt, Long: >200 nt). Subplots report (a) **Recovery**, (b) **Secondary Structure F1 evaluated by RNAfold**, (c) **RMSD** and (d) **TM-score evaluated by Boltz-2**. Methods are color-coded as: **gRFold** (pink), **gRNAde** (blue), **RiboDiffusion** (purple), and **RDesign** (green).

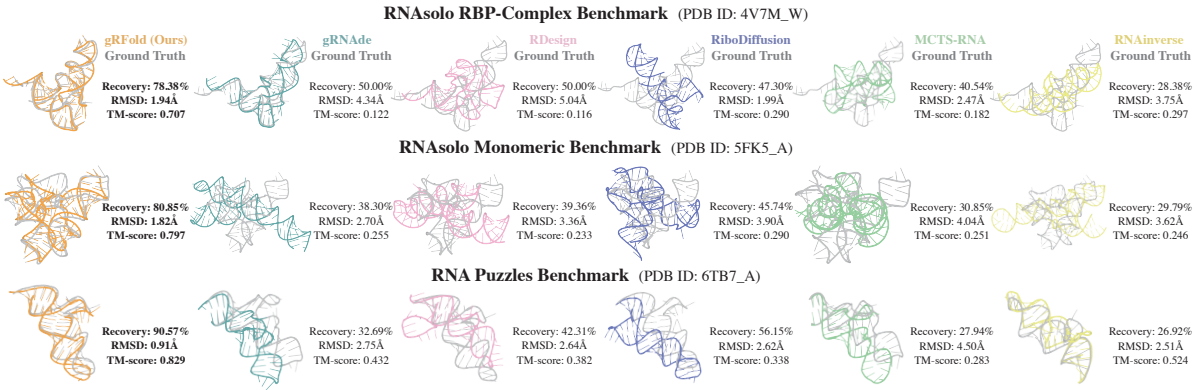


Figure 3: Structural alignment of ground-truth RNA structures with **Boltz-2** predictions generated from sequences designed by **gRFold** and baseline methods across three benchmarks. Ground-truth structures are shown in gray, predictions are colored by method, and each panel reports recovery, RMSD, and TM-score; **gRFold** generally achieves higher TM-scores, lower RMSDs, and stronger recovery.

RBP-Complex, RNAsolo Monomeric, and RNA Puzzles benchmarks (Table 2). It also preserves the target structural constraints more accurately, obtaining the best or near-best secondary-structure F1 under both RNAfold and BPfold while generally achieving lower RMSD and higher TM-scores than the compared tertiary-structure baselines under RhoFold and Boltz-2. The improvement is clearest on RNA Puzzles, where gRFold reaches RNAfold and BPfold F1 scores of 0.626 and 0.580 and a Boltz-2 TM-score of 0.370. The agreement between the secondary- and tertiary-structure evaluations indicates that the gains are not limited to sequence similarity or a single evaluation tool. One pattern deserves comment: RNAinverse attains the lowest Boltz-2 RMSD on all three benchmarks (2.681, 2.437, and 2.810 Å) while holding the lowest or near-lowest TM-scores and recoveries below 25%. RMSD alone is therefore not sufficient, since it can be dominated by short, locally well-superimposed fragments, whereas TM-score is length-normalized and reflects global fold agreement. Overall, gRFold provides a more reliable

balance between recovery, secondary-structure consistency, and preservation of the global RNA fold.

Ablation of the Geometric Modalities. The two ablation rows separate the roles of the two geometric streams. Removing the backbone branch is the more damaging intervention: recovery drops from 49.650% to 41.013%, from 51.671% to 42.306%, and from 45.004% to 40.090% on the three benchmarks, with the RNAfold F1 falling correspondingly from 0.497 to 0.337, from 0.486 to 0.355, and from 0.626 to 0.415. Removing the surface branch has a smaller effect on recovery but a consistent effect on 3D fidelity, with the RhoFold TM-score decreasing from 0.297 to 0.255 and from 0.277 to 0.236 on the two RNAsolo benchmarks. The surface contribution is largest on the RBP-Complex benchmark and smallest on RNA Puzzles, where the F1 scores of the two variants coincide at 0.626 and 0.580, which is consistent with surface geometry being most informative when the target RNA is engaged in an intermolecular interface.

Scalability Across Sequence Lengths. Figure 2

demonstrates the performance of gRFold and three tertiary-structure baselines across Short (<80 nt), Medium (80–200 nt), and Long (>200 nt) RNA sequences. gRFold achieves the best recovery in the Short and Medium bins and maintains high F1 scores across all sequence lengths. As sequence length increases, RMSD rises and TM-score decreases for all methods, while gRFold remains among the most stable models, with the lowest or near-lowest RMSD and competitive TM-scores across the three length ranges. The degradation on long sequences is shared by every compared method, which suggests that it may reflect the difficulty of the underlying folding problem rather than a specific weakness of the geometric encoders.

4.2.2 Qualitative Case Studies

Figure 3 presents representative inverse-folding cases evaluated with **Boltz-2** from three benchmarks: RNAsolo RBP-Complex (PDB ID: 4V7M_W), RNAsolo Monomeric (PDB ID: 5FK5_A), and RNA Puzzles (PDB ID: 6TB7_A). Across all three, **gRFold** achieves the highest recovery (78.38%, 80.85%, and 90.57%), the lowest RMSD (1.94, 1.82, and 0.91 Å), and the highest TM-score (0.707, 0.797, and 0.829), indicating that its designed sequences preserve native folds more accurately than baseline-generated sequences in these cases under Boltz-2-based refolding. These values remain computational structural metrics rather than experimental validation.

5 Conclusion

In this work, we show that integrating molecular surface features as a computational design constraint improves selected RNA inverse-folding metrics under the evaluated benchmarks. The proposed gRFold framework combines secondary structure, backbone topology, and surface geometry, and achieves strong structural fidelity compared with existing methods across three benchmarks. These results support structure-conditioned RNA sequence design under computational evaluation, while biochemical function remains to be established by further validation.

Limitations

Despite its promising performance, gRFold has several limitations: (1) **Data Scarcity.** The model relies on high-quality 3D RNA structures for training. Since available RNA tertiary structures are

limited, its generalization to rare RNA families or unusual motifs may be constrained. (2) **Dependence on Target Structure.** gRFold presupposes an experimentally determined or reliably predicted tertiary structure, which relocates rather than removes part of the design difficulty. (3) **Evaluation Scope.** The current evaluation is based on computational benchmarks and structural metrics, and the reported RMSD and TM-score inherit refolding-predictor error, so they may not fully capture all factors affecting RNA behavior in broader settings. (4) **Long Sequences.** Performance degrades on sequences beyond 200 nucleotides, although this degradation is shared by compared methods.

Ethics Statement

This work presents a computational method for RNA inverse folding and structure-guided sequence design. The model is intended for research and benchmark evaluation. Generated sequences should be interpreted as computational candidates and require expert review before downstream use.

AI-assisted tools were used only for language polishing. All research ideas, methods, analyses, and conclusions were developed by the authors.

Acknowledgement

This work was supported by Shenzhen Medical Research Fund (Grant No. A2503002 to Yu Li), the Chinese University of Hong Kong (CUHK; award numbers 4937025, 4937026, 5501517 and 5501329 to Yu Li), the IdeaBooster Fund (IDBF23ENG05 and IDBF24ENG06 to Yu Li), partially supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region (Hong Kong SAR), China (project no. CUHK 24204023 to Yu Li), a grant from the Innovation and Technology Commission of the Hong Kong SAR, China (project no. GHP/065/21SZ and ITS/247/23FP to Yu Li), and the Research Matching Grant Scheme at CUHK (award numbers 8601603 and 8601663 to Yu Li) from the Research Grants Council, Hong Kong SAR, China. The Major Project of Guangzhou National Laboratory (Grant No. GZNL2024A01003, GZNL2023A02007, GZNL2025C02028 to Jiao Yuan), National Natural Science Foundation of China (Grant No. 32400547 to Jiao Yuan), Pearl River Talent Recruitment Program (2023QN10Y296 to Jiao Yuan), Guangzhou Young Top Talent Program, National Key R&D Program of China (2023YFF1204701 to Jiao Yuan).

References

2019. Rnacentral: a hub of information for non-coding rna sequences. *Nucleic acids research*, 47(D1):D221–D229.
- Bartosz Adamczyk, Maciej Antczak, and Marta Szachniuk. 2022. Rnasolo: a repository of cleaned pdb-derived rna 3d structures. *Bioinformatics*, 38(14):3668–3670.
- Helen M Berman, John Westbrook, Zukang Feng, Gary Gilliland, Talapady N Bhat, Helge Weissig, Ilya N Shindyalov, and Philip E Bourne. 2000. The protein data bank. *Nucleic acids research*, 28(1):235–242.
- Meredith Corley, Margaret C Burns, and Gene W Yeo. 2020. How rna-binding proteins interact with rna: molecules and mechanisms. *Molecular cell*, 78(1):9–29.
- Robert M Dirks, Justin S Bois, Joseph M Schaeffer, Erik Winfree, and Niles A Pierce. 2007. Thermodynamic analysis of interacting nucleic acid strands. *SIAM review*, 49(1):65–88.
- Laiyi Fu, Yingxin Cao, Jie Wu, Qinke Peng, Qing Nie, and Xiaohui Xie. 2022. Ufold: fast and accurate rna secondary structure prediction with deep learning. *Nucleic acids research*, 50(3):e14–e14.
- Pablo Gainza, Freyr Sverrisson, Federico Monti, Emanuele Rodola, Davide Boscaini, Michael M Bronstein, and Bruno E Correia. 2020. Deciphering interaction fingerprints from protein molecular surfaces using geometric deep learning. *Nature methods*, 17(2):184–192.
- Ari Holtzman, Jan Buys, Li Du, Maxwell Forbes, and Yejin Choi. 2019. The curious case of neural text degeneration. *arXiv preprint arXiv:1904.09751*.
- Han Huang, Ziqian Lin, Dongchen He, Liang Hong, and Yu Li. 2024. Ribodiffusion: tertiary structure-based rna inverse folding with generative diffusion models. *Bioinformatics*, 40(Supplement_1):i347–i356.
- Farren J Isaacs, Daniel J Dwyer, and James J Collins. 2006. Rna synthetic biology. *Nature biotechnology*, 24(5):545–554.
- Bowen Jing, Stephan Eismann, Patricia Suriana, Raphael JL Townshend, and Ron Dror. 2020. Learning from protein structure with geometric vector perceptrons. *arXiv preprint arXiv:2009.01411*.
- Chaitanya Joshi, Arian Jamasb, Ramon Viñas, Charles Harris, Simon Mathis, Alex Morehead, Rishabh Anand, and Pietro Liò. 2025. grnade: Geometric deep learning for 3d rna inverse design. In *International Conference on Learning Representations*, volume 2025, pages 10166–10191.
- Haoquan Liu, Chen Zhuo, Jiaming Gao, Chengwei Zeng, and Yunjie Zhao. 2024. Ai-integrated network for rna complex structure and dynamic prediction. *Biophysics Reviews*, 5(4).
- Ronny Lorenz, Stephan H Bernhart, Christian Höner zu Siederdisen, Hakim Tafer, Christoph Flamm, Peter F Stadler, and Ivo L Hofacker. 2011. Viennarna package 2.0. *Algorithms for molecular biology*, 6(1):26.
- Ilya Loshchilov and Frank Hutter. 2017. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*.
- Marcin Magnus, Michał J Boniecki, Wayne Dawson, and Janusz M Bujnicki. 2016. Simrnaweb: a web server for rna 3d structure modeling with optional restraints. *Nucleic acids research*, 44(W1):W315–W319.
- Edoardo Morandi, Martijn J van Hemert, and Danny Incarnato. 2022. Shape-guided rna structure homology search and motif discovery. *Nature Communications*, 13(1):1722.
- Divya Nori and Wengong Jin. 2024. Rnaflow: Rna structure & sequence design via inverse folding-based flow matching. *arXiv preprint arXiv:2405.18768*.
- Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, and 1 others. 2019. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32.
- Frederic Runge, Jörg Franke, Daniel Fertmann, Rolf Backofen, and Frank Hutter. 2024. Partial rna design. *Bioinformatics*, 40(Supplement_1):i437–i445.
- Frederic Runge, Danny Stoll, Stefan Falkner, and Frank Hutter. 2018. Learning to design rna. *arXiv preprint arXiv:1812.11951*.
- Kengo Sato, Manato Akiyama, and Yasubumi Sakakibara. 2021. Rna secondary structure prediction using deep learning with thermodynamic integration. *Nature communications*, 12(1):941.
- Victor Garcia Satorras, Emiel Hoogetboom, and Max Welling. 2021. E (n) equivariant graph neural networks. In *International conference on machine learning*, pages 9323–9332. PMLR.
- Tao Shen, Zhihang Hu, Siqi Sun, Di Liu, Felix Wong, Jiuming Wang, Jiayang Chen, Yixuan Wang, Liang Hong, Jin Xiao, and 1 others. 2024. Accurate rna 3d structure prediction using a language model-based deep learning approach. *Nature Methods*, 21(12):2287–2298.
- Hanif D Sherali and Cihan H Tuncbilek. 1992. A squared-euclidean distance location-allocation problem. *Naval Research Logistics (NRL)*, 39(4):447–469.
- Jaswinder Singh, Jack Hanson, Kuldip Paliwal, and Yaoqi Zhou. 2019. Rna secondary structure prediction using an ensemble of two-dimensional deep neural networks and transfer learning. *Nature communications*, 10(1):5407.

- Zhenqiao Song, Tinglin Huang, Lei Li, and Wengong Jin. 2024. Surfpro: Functional protein design based on continuous surface. *arXiv preprint arXiv:2405.06693*.
- Freyr Sverrisson, Jean Feydy, Bruno E Correia, and Michael M Bronstein. 2021. Fast end-to-end learning on protein surfaces. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 15267–15276. IEEE.
- Cheng Tan, Yijie Zhang, Zhangyang Gao, Bozhen Hu, Siyuan Li, Zicheng Liu, and Stan Z Li. 2024. Rde-sign: Hierarchical data-efficient representation learning for tertiary structure-based rna design. In *International Conference on Learning Representations*, volume 2024, pages 10304–10333.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.
- Felix Wong, Dongchen He, Aarti Krishnan, Liang Hong, Alexander Z Wang, Jiuming Wang, Zhihang Hu, Sotaka Omori, Alicia Li, Jiahua Rao, and 1 others. 2024. Deep generative design of rna aptamers using structural predictions. *Nature Computational Science*, 4(11):829–839.
- Xiufeng Yang, Kazuki Yoshizoe, Akito Taneda, and Koji Tsuda. 2017. Rna inverse folding using monte carlo tree search. *BMC bioinformatics*, 18(1):468.
- Joseph D Yesselman, Daniel Eiler, Erik D Carlson, Michael R Gotrik, Anne E d’Aquino, Alexandra N Ooms, Wipapat Kladwang, Paul D Carlson, Xuesong Shi, David A Costantino, and 1 others. 2019. Computational design of three-dimensional rna structure and function. *Nature nanotechnology*, 14(9):866–873.
- Tianshuo Zhou, Ning Dai, Sizhen Li, Max Ward, David H Mathews, and Liang Huang. 2023. Rna design via structure-aware multifrontier ensemble optimization. *Bioinformatics*, 39(Supplement_1):i563–i571.
- Michael Zuker. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic acids research*, 31(13):3406–3415.

A Appendix

A.1 Related Work

RNA Structure Prediction and Design. RNA structure prediction and design are fundamental topics in RNA research due to the critical role of RNA structure in determining its function. Traditional methods primarily focus on predicting secondary structures using energy-based models(Dirks et al., 2007). These approaches rely on principles such as minimum free energy (MFE) to determine

the most stable folding configurations(Lorenz et al., 2011; Zuker, 2003). However, with the growing interest in RNA applications across synthetic biology and therapeutic fields, research has increasingly shifted toward RNA inverse folding, which seeks to generate RNA sequences that adopt target structures(Morandi et al., 2022; Magnus et al., 2016). Modern advancements, particularly those leveraging deep learning, have shown promising results in enhancing prediction accuracy and sequence diversity(Sato et al., 2021; Fu et al., 2022; Singh et al., 2019; Shen et al., 2024). Nevertheless, existing methods often face challenges in balancing structural fidelity and functional adaptability, highlighting the need for novel approaches to address these limitations.

Surface Information in Molecular Design.

Surface information represents a critical paradigm in molecular design, particularly in elucidating the geometric and functional characteristics of biomolecules (Gainza et al., 2020). Previous investigations have reported that surface feature integration can improve molecular interaction modeling and computational molecular design under specific benchmark settings. Pioneering work such as MaSIF (Sverrisson et al., 2021) leverages molecular surfaces represented as meshes through pre-calculated physical and geometrical descriptors, while subsequent work including dMaSIF and Surf-Pro (Song et al., 2024) reports competitive performance of point-cloud representations with atom-category annotations. Despite the established utility of surface information in protein design, its application in RNA design remains conspicuously limited (Liu et al., 2024). Conventional RNA design methodologies predominantly concentrate on secondary structural configurations, neglecting the nuanced three-dimensional geometric context crucial for functional determination. By integrating surface-level features into RNA design frameworks, this study aims to address these methodological constraints and improve structural representation through a computational method.