

RLsite: Integrating 3D-CNN and BiLSTM for RNA-ligand binding site prediction

Yan Zou(邹艳), Lang Yang(杨浪), Yanhui Liu(刘艳辉), and Yuyu Feng(冯玉宇)

Citation: Chin. Phys. B, 2025, 34 (**8**): 088709. DOI: 10.1088/1674-1056/adea9b
http://cpb.iphy.ac.cn; http://iopscience.iop.org/cpb

What follows is a list of articles you may be interested in

Combining deep reinforcement learning with heuristics to solve the traveling salesman problem

Li Hong(洪莉), Yu Liu(刘宇), Mengqiao Xu(徐梦俏), and Wenhui Deng(邓文慧)
Chin. Phys. B, 2025, 34 (**1**): 018705. DOI: 10.1088/1674-1056/ad95f1

A large language model-powered literature review for high-angle annular dark field imaging

Wenhao Yuan(袁文浩), Cheng Peng(彭程), and Qian He(何迁)
Chin. Phys. B, 2024, 33 (**9**): 098703. DOI: 10.1088/1674-1056/ad625c

RNAGCN: RNA tertiary structure assessment with a graph convolutional network

Chengwei Deng(邓成伟), Yunxin Tang(唐蕴芯), Jian Zhang(张建), Wenfei Li(李文飞), Jun Wang(王骏), and Wei Wang(王炜)
Chin. Phys. B, 2022, 31 (**11**): 118702. DOI: 10.1088/1674-1056/ac8ce3

Relationship between manifold smoothness and adversarial vulnerability in deep learning with local errors

Zijian Jiang(蒋子健), Jianwen Zhou(周健文), and Haiping Huang(黄海平)
Chin. Phys. B, 2021, 30 (**4**): 048702. DOI: 10.1088/1674-1056/abd68e

Protein-protein docking with interface residue restraints

Hao Li(李豪) and Sheng-You Huang(黄胜友)
Chin. Phys. B, 2021, 30 (**1**): 018703. DOI: 10.1088/1674-1056/abc14e

---

SPECIAL TOPIC — A celebration of the 90th Anniversary of the Birth of Bolin Hao

RLsite: Integrating 3D-CNN and BiLSTM for RNA-ligand binding site prediction

Yan Zou(邹艳), Lang Yang(杨浪), Yanhui Liu(刘艳辉), and Yuyu Feng(冯玉宇)[†]

School of Physics, Guizhou University, Guiyang 550000, *China*

(Received 6 May 2025; revised manuscript received 22 June 2025; accepted manuscript online 2 July 2025)

Accurate identification of RNA-ligand binding sites is essential for elucidating RNA function and advancing structure-based drug discovery. Here, we present RLsite, a novel deep learning framework that integrates energy-, structure- and sequence-based features to predict nucleotide-level binding sites with high accuracy. RLsite leverages energy-based three-dimensional representations, obtained from atomic probe interactions using a pre-trained ITScore-NL potential, and models their contextual features through a 3D convolutional neural network (3D-CNN) augmented with self-attention. In parallel, structure-based features, including network properties, Laplacian norm, and solvent-accessible surface area, together with sequence-based evolutionary constraint scores, are mapped along the RNA sequence and used as sequential descriptors. These descriptors are modeled using a bidirectional long short-term memory (BiLSTM) network enhanced with multi-head self-attention. By effectively fusing these complementary modalities, RLsite achieves robust and precise binding site prediction. Extensive evaluations across four diverse RNA-ligand benchmark datasets demonstrate that RLsite consistently outperforms state-of-the-art methods in terms of precision, recall, Matthews correlation coefficient (MCC), area under the curve (AUC), and overall robustness. Notably, on a particularly challenging test set composed of RNA structures containing junctions, RLsite surpasses the second-best method by 7.3% in precision, 3.4% in recall, 7.5% in MCC, and 10.8% in AUC, highlighting its potential as a powerful tool for RNA-targeted molecular design.

Keywords: RNA-ligand, binding sites prediction, deep learning, self-attention

PACS: 87.14.gn, 87.15.A–, 87.15.B–

DOI: 10.1088/1674-1056/adea9b **CSTR:** 32038.14.CPB.adea9b

1. Introduction

RNA plays a central role in numerous cellular processes, including gene regulation,[1] protein synthesis,[2] and therapeutic intervention,[3] making it an attractive and increasingly important target for drug discovery.[4,5] As a result, substantial efforts have been directed toward understanding RNA structure and dynamics.[6–9] Recent advances in high-resolution structures of RNA-ligand complexes have significantly deepened our insights into how small molecules interact with RNA to modulate its biological functions. With the growing recognition of RNA's involvement in diseases linked to previously considered "undruggable" targets,[10–12] the rational design of RNA-targeted therapeutics has garnered considerable attention. A comprehensive understanding of RNA-ligand interactions is therefore critical for elucidating RNA function and enabling structure-based drug design.

Small molecules have been shown to selectively bind specific RNA regions and modulate their biological activities,[13–15] underscoring RNA's promise as a therapeutic target. A key step in RNA-targeted drug discovery is the accurate identification and characterization of RNA-ligand binding sites. Such information not only guides experimental strategies, such as fragment-based screening, ligand optimization, and structure-assisted design,[16–19] but is also critical for virtual drug discovery approaches including molecular docking and high-throughput virtual screening.[20–26] In particular, the reliability of molecular docking heavily depends on the precise delineation of binding pockets. Therefore, advancing binding site prediction methodologies is essential for both experimental and computational pipelines, and is pivotal to accelerating the development of RNA-targeted therapeutics.

However, experimental characterization of RNA-ligand interactions remains laborious and resource-intensive, motivating the development of computational approaches for binding site prediction. These methods can be broadly classified into structure-based and machine learning-based strategies. Structure-based approaches such as Rsite,[27] Rsite2,[28] and RBind[29] leverage RNA structural features to predict binding sites. For instance, Rsite models RNA tertiary structures as coarse-grained assemblies of spherical units and identifies binding sites through local extrema in distance graphs. To overcome the limited availability of high-resolution tertiary structures, Rsite2 extends this strategy to secondary structure data. RBind, alternatively, represents RNA structures as graphs and computes geodesic distances, using topological features such as closeness and degree centrality to in-

[†]Corresponding author. E-mail: fengyy@gzu.edu.cn

© 2025 Chinese Physical Society and IOP Publishing Ltd. All rights, including for text and data mining, AI training, and similar technologies, are reserved.

fer potential binding regions. While effective, these approaches often underutilize sequence information, potentially constraining predictive performance. To address these limitations, machine learning-based methods have been developed that integrate both sequence and structural information. RNAsite[30] employs a random forest model trained on features including network-based metrics, solvent accessibility, and sequence-derived properties. RLBind[31] further refines these features by incorporating evolutionary conservation and biochemical descriptors. More recently, ZHmolReSTasite,[32] a deep learning-based framework, models RNA surfaces as topographical images derived from solvent-excluded surfaces. Tested across a broad range of datasets, ZHmolReSTasite has demonstrated strong performance in predicting binding sites on RNA structures with varying structural complexity.

Here, we introduce RLsite, a deep learning framework that integrates three-dimensional convolutional neural networks (3D-CNNs) and bidirectional long short-term memory networks (BiLSTMs) for RNA-ligand binding site prediction. Trained on experimentally resolved RNA-ligand complexes from the Protein Data Bank (PDB), RLsite captures structural information through energy-based 3D representations processed by a 3D-CNN enhanced with self-attention, while sequential descriptors are modeled via a BiLSTM network with multi-head self-attention. By fusing complementary structural and sequence features, RLsite enables accurate, nucleotide-level binding site identification. Extensive benchmarking across four diverse datasets demonstrates that RLsite substantially outperforms existing methods in precision, recall, Matthews correlation coefficient (MCC), and AUC, highlighting its robustness and strong generalization capabilities.

2. Materials and methods

2.1. Construction of dataset

To facilitate training and robust evaluation of RNA-ligand binding site prediction models, we curated a suite of datasets through rigorous structural and sequence-based filtering procedures. The training and validation RNA dataset (TR60) comprises 60 RNA chains selected through a rigorous multi-step filtering and clustering process. Initially, 1673 RNA chains bound to small molecules (excluding water molecules) were extracted from the PDB.[33] Chains shorter than 20 nucleotides, longer than 1500 nucleotides, or interacting only with crystallization additives were removed, yielding 712 RNA chains. Structural redundancy was reduced using RNA-align[34] with a TM-score$_{RNA}$ threshold of 0.3, retaining within each cluster the chain with the highest ratio of binding nucleotides to total sequence length. This process produced a non-redundant set of 78 RNA chains, referred to as RB78. Further sequence-based clustering using cd-hit-est[35] and BLASTclust[36] at 30% identity grouped the chains into 57 clusters, from which 60 representative RNA chains were selected to form a training set, referred to as TR60. To assess generalization, an independent test set, named TE18, was constructed from the 18 RNA chains in RB78 that are not included in TR60, ensuring strict separation between training and test data and enabling unbiased assessment of model performance. A second independent test set, denoted as RB9, was derived from the RBind benchmark dataset,[29] which originally comprised 19 RNA chains. To eliminate data overlap, 10 chains shared with TR60 were removed, resulting in a refined RB9 set for independent testing. Comprehensive details of the TR60,

Table 1. Summary of the datasets used for training and evaluating RLsite, including PDB identifiers and RNA chain lengths.

Training set of 60 RNA chains (TR60)							
1aju_A(30)	1arj_N(29)	1byj_A(27)	1eht_A(33)	1ei2_A(25)	1fyp_A(27)	1hr2_A(157)	1lvj_A(31)
1nbk_A(34)	1ntb_A(22)	1qd3_A(29)	1raw_A(36)	1tob_A(27)	1uts_B(29)	1uud_B(29)	1y90_B(23)
1yls_B(38)	1yrj_A(23)	2fcy_A(23)	2g5k_A(22)	2kgp_A(25)	2ktz_A(38)	2l8h_A(29)	2lwk_A(32)
2mxs_A(27)	2nok_B(20)	2nok_C(24)	2quw_B(57)	2yie_Z(57)	3bnq_C(23)	3c44_A(23)	3fu2_A(32)
3gx3_A(94)	3mei_B(23)	3oxe_B(87)	3q3z_V(75)	3q50_A(33)	3skt_A(64)	3tzr_A(20)	3vrs_A(52)
4frg_X(84)	4k31_C(22)	4lvx_A(89)	4meg_B(121)	4mgm_B(63)	4nyb_A(80)	4pcj_A(35)	4qjh_C(26)
4qlm_A(108)	4rge_B(56)	4xw7_A(64)	5d5l_D(77)	5dh8_B(20)	5fj1_C(24)	5j02_A(414)	5o69_A(41)
5u3g_B(84)	5vci_A(22)	6fz0_A(48)	6hag_A(43)				
Test set of 18 RNA chains (TE18)							
1ddy_A(35)	1f1t_A(38)	1fmn_A(35)	1nem_A(23)	1q8n_A(38)	2juk_A(22)	2mis_A(26)	2pwt_A(21)
2tob_A(20)	364d_C(31)	379d_B(25)	430d_A(29)	4f8u_B(21)	4pqv_A(68)	4yaz_R(83)	5bjo_E(36)
5v3f_A(31)	6ez0_A(22)						
Test set of 9 RNA chains (RB9)							
1nem_A(23)	1q8n_A(38)	2tob_A(20)	1pbr_A(27)	1uui_B(29)	1y26_X(71)	2gdi_X(80)	3d2x_A(77)
3gx2_A(94)							
Test set of 10 RNA chains (JL10)							
6wtr_A(125)	6wzs_B(60)	7eaf_A(94)	7eqj_B(75)	7kga_A(89)	7kvu_G(83)	7tzr_X(80)	7uq6_B(86)
7uz0_A(87)	9bun_C(48)						
Test set of 12 RNA chains (TL12)							
7d12_A(20)	7elq_A(45)	7eog_A(48)	7l0z_G(69)	7oa3_A(52)	7rex_A(34)	7rwr_A(38)	7td7_A(78)
8f0n_A(49)	8fza_A(30)	8hb1_A(24)	8hb1_B(30)				

TE18, and RB9 datasets, including PDB IDs and RNA chain lengths, were summarized in Table 1. In addition, two benchmark datasets were curated to assess RLsite's performance on RNA structures of varying complexity. The test set JL10 targets RNA structures containing junction loops — regions with high structural complexity that pose significant challenges for binding site prediction. High-resolution RNA structures (resolution better than 4 Å) complexed exclusively with small molecules, deposited in the PDB after January 2021, were curated. Structures involving protein or DNA interactions, as well as those lacking junction loops or having fewer than 10% binding nucleotides relative to sequence length, were systematically excluded. Sequence redundancy was minimized using cd-hit at 95% identity,[37,38] resulting in a final set of 10 RNA chains for JL10. In contrast, another test set was curated to include RNA structures of low complexity. Using a similar filtering protocol but excluding RNA chains with junction loops, 12 structures were selected to form a test set with low structural complexity named TL12, providing a complementary benchmark for evaluating model robustness under simpler structural contexts. A full summary of the TR60, TE18, RB9, JL10, and TL12 datasets, including variations in PDB identifiers and RNA chain lengths, is available in Table 1.

2.2. Overview of the RLsite framework

Figure 1 illustrates the overall architecture of RLsite, a framework for nucleotide-level binding site prediction that integrates both structure- and sequence-derived features of RNA molecules.

In the energy-based feature extraction module (Fig. 1(a)), each RNA structure is discretized into a three-dimensional voxel grid with a resolution of 1 Å. Interaction energies between a multi-type atomic probe and the RNA are computed using a pre-trained scoring function, ITScore-NL.[39] To capture the influence of local spatial context, a sliding window of size $2W+1$ (with $W=5$) is centered around each nucleotide. The energy maps of the target nucleotide and its spatial neighbors are stacked to form a three-dimensional input, which is subsequently processed by a three-dimensional convolutional neural network (3D-CNN) to generate a refined two-dimensional energy representation. In parallel, sequential descriptors are extracted (Fig. 1(b)). These features include network properties and Laplacian norms, solvent-accessible surface area (SASA), and evolutionary constraint scores derived from multiple sequence alignments. A sliding window of size $2W+1$ is again utilized to incorporate local sequence context.

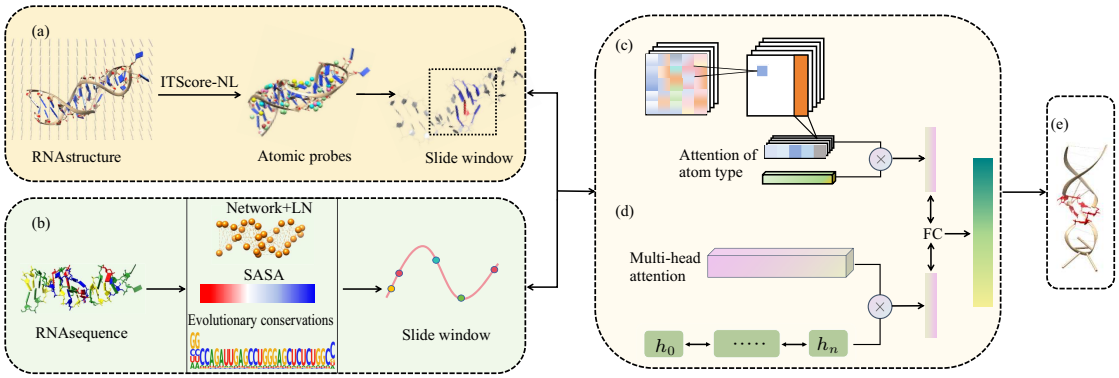


Fig. 1. Overview of RLsite. (a) Extraction of energy-based features using the pre-trained ITScore-NL scoring function. (b) Extraction of sequentially-encoded descriptors. (c) Modeling of energy-based features using a 3D-CNN with a self-attention mechanism to capture the varying importance of atomic probes. (d) Modeling of sequential descriptors using a BiLSTM network enhanced with multi-head self-attention. (e) Final prediction of RNA-ligand binding sites by RLsite.

To further enhance the energy-based feature learning, a self-attention mechanism is integrated within the 3D-CNN (Fig. 1(c)). This mechanism dynamically weighs the contributions of different atomic probes, enabling the model to focus on the most informative interaction patterns for binding site prediction. In parallel, to model the sequential descriptors, a bidirectional long short-term memory (BiLSTM) network is applied to the embedding matrix (Fig. 1(d)). Additionally, a multi-head self-attention mechanism is embedded within the BiLSTM, allowing the model to highlight key sequence regions critical for binding site formation and to enhance interpretability.

Finally, the features extracted from the 3D-CNN and BiLSTM modules are concatenated and passed through fully connected layers with nonlinear activation functions to produce the final binding site probability scores (Fig. 1(e)).

2.3. Feature construction

Our task is to predict RNA-ligand binding sites by encoding nucleotides using both structural and sequence features.

2.3.1. Conformational energy-based features

Atomic probe energy RNA structure is embedded into a three-dimensional cubic grid with a spacing of 1.0 Å. The grid dimensions are determined by extending the RNA's bounding box by 1.0 Å beyond the outermost atom in each Cartesian

direction. Recognizing that small ligands preferentially bind to concave regions of the RNA surface to maximize contact, a spherical probe with a radius of 1.4 Å, approximating the typical solvent-receptor contact distance, is used to identify solvent-accessible regions. At each grid point, the probe is translated along the six cardinal directions ($\pm x$, $\pm y$, $\pm z$) in 1.0 Å increments, up to a maximum displacement of 6.0 Å. A collision is registered when the distance between the probe center and an RNA atom is less than the sum of their radii. Grid points registering five or more collisions are preliminarily flagged as potential ligand binding regions. Since this procedure is sensitive to the initial orientation of the RNA due to the lack of rotational invariance in the grid, additional sampling is performed along the eight diagonal directions. If fewer than five collisions are detected in the cardinal scan, the probe is further displaced along the diagonals to accumulate additional collision events. A grid point is ultimately classified as part of a candidate binding pocket if the combined number of collisions across all directions reaches or exceeds eleven.

Following geometric screening, interaction energies at each grid point are computed using ITScore-NL,[39] a pre-trained statistical potential based on atomic pairwise interactions. In addition, given the frequent presence of metal ions in RNA ligands, atoms are classified into three categories: monovalent metals, divalent metals, and all other atoms. Initial interaction potentials between each pairwise combination are obtained using a training protocol aligned with ITScore-NL, resulting in 18 distinct atomic probe types. Owing to the negatively charged phosphate backbone, RNA molecules exhibit strong electrostatic fields, which pose a significant challenge for accurately modeling metal ion interactions. To address this, metal ion probes are systematically placed at every grid point surrounding the RNA, and the corresponding interaction energies are calculated. For each probe type, the ten grid points exhibiting the most favorable binding energies are selected for subsequent analysis. These energy values are then mapped to a coarse-grained nucleotide framework using an inverse distance weighting scheme.

The RNA molecule is coarse-grained by assigning the center of the six-membered ring in each nucleotide as its representative coordinate. The feature associated with each nucleotide is defined as the inverse of the distance to surrounding grid points with assigned energy values (EV), as follows:

$$f_i = \frac{\mathrm{EV}}{d_i(x,y)}. \tag{1}$$

Here $d_i(x,y)$ denotes the Euclidean distance from a grid point to the coarse-grained coordinate of the target nucleotide, while the EV quantifies the interaction strength between each selected grid point and the RNA structure. The feature vector $f_i \in \mathbb{R}^{10\times 18}$ for each nucleotide can be viewed as a distance-based, two-dimensional feature map consisting of 18 atomic probes. This map encodes the energy preferences for small molecules binding to the nucleotide at various distances. To incorporate local spatial context, a sliding window of size $2\times W+1$ (with $W=5$) is applied around each nucleotide based on its three-dimensional coordinates. The energy maps of the target nucleotide and its $2W$ nearest spatial neighbors are stacked to form a localized input tensor of shape $(2\times W+1)\times 10\times 18$. This tensor serves as the input to a three-dimensional convolutional neural network (3D-CNN), which learns spatial features reflective of RNA-ligand interaction patterns.

2.3.2. Structure-based features

Network properties RNA structures are represented as nucleotide interaction networks, where each nucleotide is a node and edges denote noncovalent interactions between nucleotides.[29] In these networks, two key parameters, closeness (CL) and degree (DG), are computed for each nucleotide. Closeness measures the average shortest path from a given nucleotide to all others, while degree quantifies the number of direct interactions each nucleotide has within the network. Closeness (CL) and degree (DG), are calculated for each nucleotide as follows:

$$C(x) = \frac{n-1}{\sum d(x,y)}. \tag{2}$$

The distance $d(x,y)$ represents the shortest path between a node x and any other node y.[40] The shortest paths between all pairs of nodes are computed using the Floyd–Warshall algorithm. The centrality (CL) of a nucleotide is defined by its average shortest distance to all other nodes, reflecting its global importance within the network, while the degree (DG) captures its local connectivity by counting the number of edges directly linked to it. These two parameters are encoded as a 2D vector to characterize the network's topological properties. Similar network metrics have been employed in RLBind to assess nucleotide interactions.[31]

Laplacian norm The Laplace norm (LN) has been applied in RNA and protein structure analysis, as well as in functional prediction. This operator provides a measure of the local geometric features of biomolecules by quantifying the convexity and concavity of structural surfaces.[41] In this study, we use LN to characterize RNA structures. The LN of each nucleotide is defined as the distance between the target nucleotide and the weighted center of its neighboring nucleotides. To compute LN, a discrete Laplace operator is calculated using a Gaussian kernel, as defined below:

$$\Omega_{ij}(\sigma) = \begin{cases} \mathrm{e}^{-\|p_i-p_j\|^2/\sigma^2}, & \text{if } |i-j|>1, \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

As defined in RNAsite,[30] p_i and p_j are the coordinates of the 'C3' atoms of nucleotide i and nucleotide j, respec-

tively. The parameter σ acts as a scale factor in the calculation of the Laplace operator. This operator could capture the influence of surrounding nucleotides on the target nucleotide with different values of σ. A lower value of σ gives greater weight to adjacent nucleotides, while a higher value of σ includes more distant nucleotides. To account for both nearby and distant nucleotides, distances corresponding to the 0, 1/4, 1/2, 3/4, and 1 quantiles of the distribution are considered. As a result, the LN for the target nucleotide is represented by a 1×5 vector. LN values under different σ scales are then generated as follows:

$$\mathrm{LN}_i(\sigma)=\left|\left|p_i-\frac{\sum_j^{|i-j|>1}[p_i\times\Omega_{ij}(\sigma)]}{\sum_j^{|i-j|>1}\Omega_{ij}(\sigma)}\right|\right|. \tag{4}$$

The value of LN can reflect the geometrical features of surface convexity/concavity. A high value of LN means the nucleotide position is convex in the RNA structure, while a low LN implies a concave position. Thus, one 5D vector was used to characterize RNA as a multidimensional sequence of quantity of deformation at different scales of the local space in which the nucleotide is embedded.

Solvent accessible surface area RNAsol is a specialized algorithm developed to predict the solvent accessibility of RNA nucleotides. Built upon improved sequence profiles derived from covariance models, RNAsol is trained using long short-term memory (LSTM) neural networks.[42] This approach offers an efficient means of estimating nucleotide exposure to the solvent. Nucleotides with higher accessible surface areas (ASAs) are typically exposed on the RNA surface, while those with lower ASAs are often buried within the interior, such as in binding cavities. We utilized the solvent accessible surface area (SASA) predictions from RNAsol in our analysis.

2.3.3. Sequence-based features

Evolutionary constraint score To characterize RNA nucleotides from a sequence-based perspective, evolutionary conservation is employed as a key descriptor. RNA multiple sequence alignments (MSAs) are first generated using BLASTN[43] with an E-value threshold of 0.001, searching against the MARS database.[44] To extract meaningful conservation signals, direct coupling analysis (DCA) is applied to the MSAs. This approach builds a global statistical model to distinguish direct from indirect nucleotide correlations, thereby capturing underlying co-evolutionary constraints.[45] Subsequently, the plmDCA algorithm is used to represent each nucleotide as a 1×5 feature vector, where each element corresponds to a conservation score calculated under a distinct cutoff threshold.[32,46] For any pair of sites i and j in the sequence, the evolutionary constraint score is computed as follows:

$$\mathrm{DI}_{ij}=\sum_{\sigma,\omega=1}^{q}p_{ij}^{(\mathrm{dir})}(\sigma,\omega)\ln\frac{p_{ij}^{(\mathrm{dir})}(\sigma,\omega)}{f_i(\sigma)f_j(\omega)}. \tag{5}$$

A two-site model is constructed for each pair of sites i and j as follows:

$$P_{ij}^{\mathrm{dir}}(\sigma,\omega)=\frac{\exp\left(e_{ij}(\sigma,\omega)+\tilde{h}_i(\sigma)+\tilde{h}_j(\omega)\right)}{Z_{ij}}, \tag{6}$$

where $\tilde{h}_i(\sigma)$ and $\tilde{h}_j(\omega)$ are parameters assigned based on the empirical single-nucleotide frequency counts and are required to satisfy the constraints

$$f_i(\sigma)=\sum_{\omega=1}^{q}p_{ij}^{(\mathrm{dir})}(\sigma,\omega)),\quad f_i(\omega)=\sum_{\sigma=1}^{q}p_{ij}^{(\mathrm{dir})}(\sigma,\omega)). \tag{7}$$

Here, $f_i(\sigma)$ denotes the empirical frequency of nucleotide σ at position i, and $p_{ij}^{(\mathrm{dir})}(\sigma,\omega)$ represents the empirical joint frequency of nucleotides σ and ω at positions i and j, respectively, in the multiple sequence alignment (MSA). The partition function Z_{ij} is defined to satisfy

$$\sum_{\sigma,\omega=1}^{q}p_{ij}^{(\mathrm{dir})}(\sigma,\omega)=1. \tag{8}$$

The coupling coefficient $e_{ij}(\sigma,\omega)$ is estimated in plmDCA using the method of pseudolikelihood maximization. Maximum-likelihood inference is a way to evaluate the parameters of a model.[47]

In summary, structure-based and sequence-based features are integrated as sequential descriptors and fed into the BiLSTM model, where each RNA nucleotide is encoded by a total of 1×13 feature vector. To capture the sequential dependencies among nucleotides, a sliding window of size W is applied across the RNA sequence. For each target nucleotide at position i, the window includes W upstream and W downstream nucleotides, thereby providing contextual information reflective of its local sequence environment. Consequently, each nucleotide is represented by a total of $(2\times W+1)\times 13$ sequential descriptors. For nucleotides near the sequence termini where the window extends beyond the sequence boundaries, zero-padding is employed to maintain dimensional consistency.

To comprehensively characterize the energetic, structural, and evolutionary properties of RNA molecules, we constructed a multi-view feature set encompassing three categories: energy-based, structure-based, and sequence-based descriptors. Each feature type encodes distinct biological signals relevant to RNA-ligand interactions. Energy-based features are derived from an atomic probing scheme, wherein 18 atomic probe types are systematically placed around the RNA, and the ten most favorable interaction energies per type are selected based on the ITScore-NL[39] statistical potential. Structure-based features capture geometric and topological characteristics, including network properties, Laplacian norms, and solvent accessibility, thereby describing both the spatial organization and surface exposure of the RNA.

Table 2. Overview of feature types incorporated in the model. L denotes the number of nucleotides in the RNA sequence. Atomic probe features capture the spatial and energetic properties of potential ligand binding sites. Structure-based features represent topological and geometric characteristics via network descriptors, Laplacian norms, and solvent accessibility. Evolutionary constraint scores encode residue-level covariation signals from multiple sequence alignments.

Feature type	Description	Input shape	Source / tool used
Energy-based	atomic probe features	$10 \times 18 \times L$	ITScore-NL[39]
Structure-based	network proterties	$2 \times L$	RBind[29]
	Laplacian norm	$5 \times L$	RLsite source code
	solvent accessible surface area	$1 \times L$	RNAsol[42]
Sequence-based	evolutionary constraint score	$5 \times L$	plmDCA[46]

These features help identify physicochemical microenvironments favorable for ligand binding. Sequence-based features encode covariation signals inferred from multiple sequence alignments using plmDCA,[49] reflecting residue-level evolutionary constraints. All features are formatted into fixed size matrices with dimensionality determined by the RNA length L, as detailed in Table 2.

2.4. Deep learning framework

In this section, we will introduce the framework of our RNA-ligand binding site prediction model RLsite.

Attentive 3D CNN In our model, a 3D convolutional neural network (CNN) is employed to encode RNA conformational energy maps into fixed-size vector representations. The CNN module consists of stacked residual blocks followed by a sequential self-attention block. Each residual block, borrowed from ResNet, comprises two 3×3 convolutional layers. To preserve the spatial dimensions of the input throughout the residual blocks, pooling layers are omitted, and zero-padding is applied on both sides of the input. Specifically, given a nucleotide conformational energy map $P_{\text{input}} \in \mathbb{R}^{1\times(2\times W+1)\times 10\times 18}$, the output of the last residual block retains the dimensions $N_{\text{filters}} \times (2 \times W + 1) \times 10 \times 18$, where N_{filters} denotes the number of filters in the final convolutional layer. We then apply average pooling to compress the feature map and minimum pooling to capture the optimal interaction between the atomic probes and the nucleotide. This process results in a convolutional nucleotide feature map $P_{\text{conv}} \in \mathbb{R}^{18\times N_{\text{filters}}}$, which is used for downstream processing. The feature map P_{conv} can be interpreted as the interaction between each nucleotide and various atomic probes at different distances, where 18 is the number of atomic probes used, and N_{filters} represents the spatial features of each probe. Certain atomic probes, such as those involved in stacking within binding pockets through benzene ring formation or establishing electrostatic, hydrogen-bond, and hydrophobic interactions, are critical for nucleotide recognition. To highlight the importance of these probes, we incorporate a self-attention mechanism that takes P_{conv} as input and outputs a vector of attention weights, att^{p}, for each probe.

$$\text{att}^{\text{p}} = \text{softmax}(w_{\text{p2}} \tanh(W_{\text{p1}} P_{\text{conv}}^{\text{T}})), \tag{9}$$

$$\sum_{i=1}^{l} \left(\text{att}_i^{\text{p}}\right) = 1, \quad \forall i,\ 1 \le i \le l, \tag{10}$$

where $W_{\text{p1}} \in \mathbb{R}^{d_{\text{p}}\times N_{\text{filter}}}$ is a learnable weight matrix and $w_{\text{p2}} \in \mathbb{R}^{d_{\text{p}}}$ is a learnable parameter vector. Here, d_{p} is an adjustable hyperparameter controlling the hidden dimension. While this vector representation captures contributions from atomic probes, RNA binding pockets typically interact with multiple ligand atoms in three-dimensional space. To better model this complexity, we extend the w_{p2} into a matrix $W_{\text{p2}} \in \mathbb{R}^{r_{\text{p}}\times d_{\text{p}}}$, thereby generating a multi-head attention matrix $\text{ATT}^{\text{p}} \in \mathbb{R}^{r_{\text{p}}\times l}$:

$$\text{ATT}^{\text{p}} = \text{softmax}(W_{\text{p2}} \tanh(W_{\text{p1}} P_{\text{conv}}^{\text{T}})). \tag{11}$$

Equation (11) can be interpreted as a two-layer multilayer perceptron (MLP) without bias terms, where d_{p} denotes the hidden dimension and $W_{\text{p1}}, W_{\text{p2}}$ are the learnable parameters. Subsequently, the attention matrix ATT^{p} is applied to the probe feature map P_{conv} to produce an attention-refined feature map P_{att}:

$$P_{\text{att}} = \text{ATT}^{\text{p}} P_{\text{conv}}, \tag{12}$$

where P_{att} is an attentional feature map containing the latent relationship between contribution of probes on the interaction. The size of P_{att} is $r_{\text{p}} \times N_{\text{filters}}$, where r_{p} is an adjustable hyperparameter representing the number of attention vectors. where $P_{\text{att}} \in \mathbb{R}^{r_{\text{p}}\times N_{\text{filters}}}$ encodes the latent relationships between atomic probes and their contributions to ligand binding. Here, r_{p} is a tunable hyperparameter that defines the number of attention vectors.

Self-attentive BiLSTM Each nucleotide, along with its adjacent nucleotides, is encoded into a 2D embedding matrix $M \in \mathbb{R}^{n\times d}$, where n corresponds to the number of nucleotides within a local region, determined by a sliding window of size $2W + 1$. This window encompasses both the target nucleotide and its adjacent nucleotides. Each nucleotide is embedded as a vector of length d, which represents its sequence-specific features. This encoding captures both the individual characteristics of the nucleotide and the contextual relationships between the target and its neighboring nucleotides. To model the dependencies between the target nucleotide and its neighbors,

we employ a bidirectional long short-term memory (BiLSTM) network. The forward pass of the BiLSTM is expressed as

$$\overrightarrow{h_i} = \overrightarrow{\text{LSTM}}(e_i, \overrightarrow{h_{i-1}}) \tag{13}$$

and the backward pass as

$$\overleftarrow{h_i} = \overleftarrow{\text{LSTM}}(e_i, \overleftarrow{h_{i+1}}). \tag{14}$$

The forward hidden state $\overrightarrow{h_i}$ is concatenated with the backward hidden state $\overleftarrow{h_i}$, resulting in a unified hidden state h_i that replaces the nucleotide embedding e_i. This process produces an information-enriched vector h_i that encodes the dependencies between adjacent nucleotides in the RNA sequence. We denote the collection of all hidden states across the sequence as H:

$$h_i = (\overrightarrow{h_i}, \overleftarrow{h_i}), \tag{15}$$

$$H = (h_0, h_i, \ldots, h_n). \tag{16}$$

If the hidden unit number for each unidirectional LSTM is denoted as u, the resulting matrix H has a shape of $n \times 2u$.

To further capture the contribution of each nucleotide and its sequential neighbors, we incorporate a multi-head self-attention mechanism. The attention mechanism takes the LSTM hidden states H as input and outputs a weight vector ATT^{s}, representing the attention matrix for the sequence, as

$$\text{ATT}^{\text{s}} = \text{softmax}(\text{MLP}(H^{\text{T}})). \tag{17}$$

Here, the multi-layer perceptron (MLP) consists of two layers, with hidden units and parameters denoted as d_{s}, W_{s1}, and W_{s2}. The self-attentive nucleotide feature map is computed by multiplying the attention matrix ATT^{s} with the LSTM hidden states H:

$$\text{Seq}_{\text{att}} = \text{ATT}^{\text{s}} H. \tag{18}$$

The resulting matrix Seq_{att} represents the latent relationships between the target nucleotide and its adjacent nucleotides that contribute to RNA-ligand binding. The size of Seq_{att} is $r_{\text{s}} \times 2u$. Here, r_{s} is a tunable hyperparameter that defines the number of attention vectors.

Classifier For the classification task, we aggregate the attention vectors across all nucleotides and normalize the resulting weight vector to sum to 1. This process yields two enriched 1D vectors $\widehat{P_{\text{att}}}$ and $\widehat{\text{Seq}_{\text{att}}}$, which are then fed into the classification layer. The final output o is obtained by concatenating $\widehat{P_{\text{att}}}$ and $\widehat{\text{Seq}_{\text{att}}}$, as follows:

$$o = W_o[\widehat{P_{\text{att}}}; \widehat{\text{Seq}_{\text{att}}}] + b_o, \tag{19}$$

where $W_o \in \mathbb{R}^{N_{\text{filters}}+2u}$ is the weight matrix and $b_o \in \mathbb{R}$ is the bias. Given a dataset $D = \{(N_i, y_i, \widehat{y_i})\}$, the parameters are updated iteratively based on the loss of each nucleotide (with batch size equal to 8). The training objective is to minimize the cross-entropy loss

$$\text{Loss} = \sum_{i=1}^{N} -y_i \cdot \log(\widehat{y_i}) - (1 - y_i) \cdot \log(1 - \widehat{y_i}), \tag{20}$$

where N is the total number of nucleotides in the training dataset, and y_i and $\widehat{y_i}$ are the true and predicted probability scores for nucleotide i, respectively.

2.5. The training protocol

The RLsite model was implemented using PyTorch v2.0.1 and trained on an NVIDIA GeForce RTX 3080Ti GPU. Training was conducted on the TR60 dataset for up to 20 epochs using the AdamW optimizer,[48] with an initial learning rate of 0.001, a weight decay of 0.001, and a batch size of 8. Model parameters were initialized using He initialization.[49] The network architecture comprises eight residual blocks, each containing 32 filters. The hidden state size of the bidirectional long short-term memory (BiLSTM) network was set to 64 (denoted as $2u$ in the self-attentive BiLSTM module), and a dropout rate of 0.2 was applied. Both the CNN and BiLSTM components incorporated attention-based multilayer perceptrons (MLPs), each with a hidden layer size of 32 (denoted as d_{p} and d_{s}, respectively). The embedding matrices for RNA conformational energy-based and sequential descriptors were configured with 10 rows, denoted as r_{p} and r_{s}. Model optimization was performed via backpropagation using the same learning rate, with parameter updates occurring at each epoch. The binary cross-entropy (BCE) loss function was used to compute prediction loss for each training batch. To promote convergence, the learning rate was adaptively reduced to 10% of its previous value if the Matthews correlation coefficient (MCC) on the validation set did not improve over two consecutive epochs. A five-fold cross-validation strategy was adopted, and the model yielding the best performance on the validation set was selected for final evaluation.

2.6. Performance evaluation

We evaluated the model's performance across a range of characteristics. During training, 20% of the training data was randomly withheld for validation. Model performance was assessed using four standard metrics: precision, recall, Matthews correlation coefficient (MCC), and the area under the receiver operating characteristic curve (AUC). These metrics are defined as follows:

$$\text{Precision} = \frac{|\text{TP}|}{(|\text{TP}| + |\text{FP}|)},$$

$$\text{Recall} = \frac{|\text{TP}|}{(|\text{TP}| + |\text{FN}|)},$$

$$\text{MCC} = \frac{\text{TP} \times \text{TN} - \text{FP} \times \text{FN}}{\sqrt{(\text{TP}+\text{FP})(\text{TP}+\text{FN})(\text{TN}+\text{FP})(\text{TN}+\text{FN})}}.$$

True positives (TP) denote correctly predicted binding nucleotides, while false positives (FP) refer to incorrectly predicted binding sites. True negatives (TN) correspond to correctly identified nonbinding nucleotides, whereas false negatives (FN) represent binding nucleotides misclassified as nonbinding.

3. Results

3.1. Performance in easy tasks

RLsite was evaluated on two independent test sets, TE18, containing 18 RNA chains, and RB9, comprising 9 RNA chains. These datasets primarily consist of RNAs with simple secondary structures — combinations of hairpin loops, internal loops, bulges, and stems, and are widely used as test datasets in the methods of ZeSTa, RNAsite, and RBind. The performance of RLsite was compared with these competing methods using precision, recall, Matthews correlation coefficient (MCC), and area under the ROC curve (AUC).

As shown in Fig. 2, in the TE18 dataset, RLsite outperformed in precision (0.712), recall (0.392), MCC (0.335) and AUC (0.740) compared with ZeSTa's 0.379, 0.327 and 0.709, but obtained slightly lower precision than ZeSTa's 0.729. Similarly, RLsite achieved a significantly better performance in precision, recall and MCC compared to RNAsite's 0.675, 0.263, 0.253, respectively with a lower AUC compared with RNAsite's 0.776. Compared with RLBind, RLsite absolutely achieves a better performance on the four criterions. These improvements underscore RLsite's enhanced ability to comprehensively capture true binding sites in RNAs with low structural complexity.

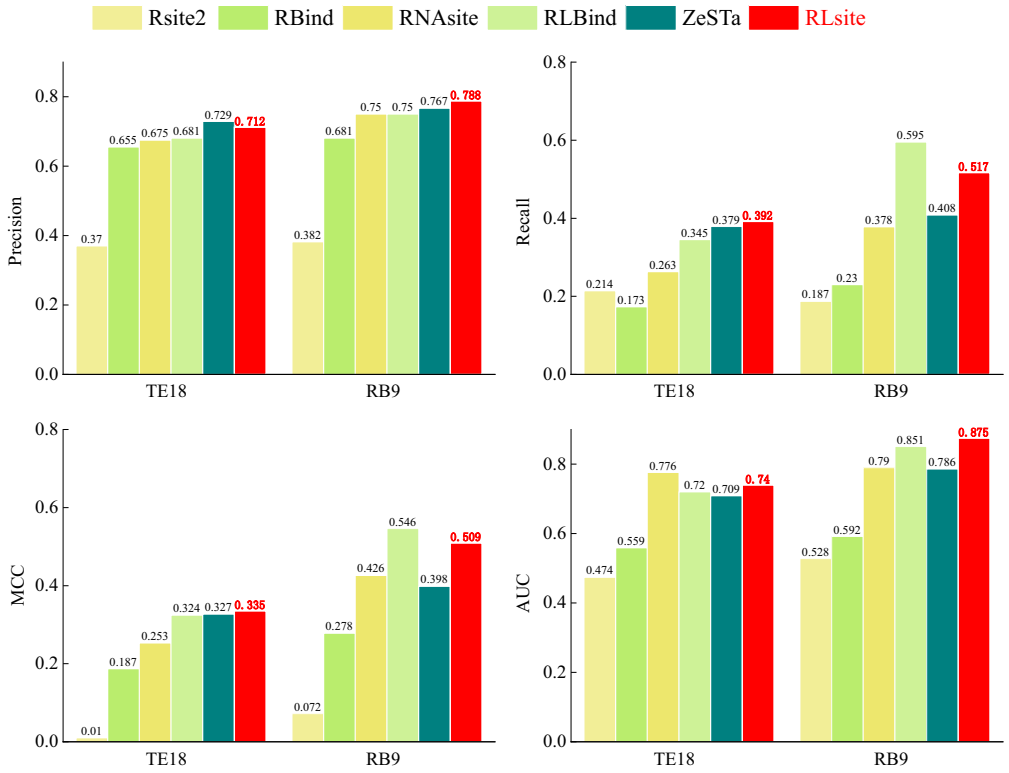


Fig. 2. Performance comparison between RLsite and competing methods in terms of precision, recall, MCC, and AUC on the TE18 and RB9 test sets.

Also, RLsite ranks second in precision on the TE18 dataset, slightly below ZeSTa. This minor difference in precision is primarily attributed to class imbalance, a well established factor affecting nucleotide level binding site prediction. In the TE18 set, the average ratio of positive to negative samples across RNA chains is approximately 0.374 to 0.626, indicating a notable skew toward negative samples. Since precision measures the proportion of true positives among predicted positives, while recall evaluates the proportion of true positives among actual positives, such imbalance tends to favor models that make more conservative predictions. These models often predict fewer positives, which may lead to higher precision but typically at the expense of lower recall. Although ZeSTa achieves slightly higher precision than RLsite (0.729 vs. 0.712), it does so with a lower recall (0.379 vs. 0.392). In contrast, RLsite provides a more balanced performance, as reflected in its higher F1-score (0.464 vs. 0.446), which harmonizes precision and recall, as well as its higher Matthews correlation coefficient (MCC) (0.335 vs. 0.327), despite the presence of class imbalance. To further assess the robustness of RLsite under varying class distributions, we analyzed the performance of RLsite on the RB9 dataset, which exhibits similar structural characteristics to TE18 but a slightly more imbalanced average positive-to-negative ratio (0.325 to

0.675). On RB9, RLsite achieves a substantially higher F1-score (0.596 vs. 0.451) and MCC (0.509 vs. 0.398) compared to ZeSTa, highlighting its consistent predictive performance across datasets with different levels of class imbalance. The detailed comparison is provided in Fig. 3

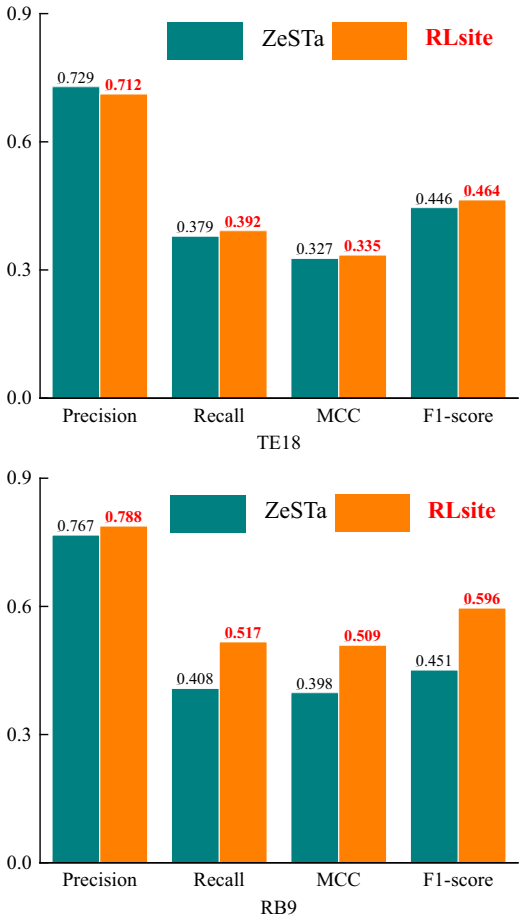


Fig. 3. Comparative performance of RLsite and ZeSTa on the TE18 and RB9 datasets across four metrics: precision, recall, MCC, and F1-score.

On the RB9 dataset, RLsite achieved the highest overall performance, with a precision of 0.788, a recall of 0.517, an MCC of 0.509, and an AUC of 0.875. Compared to ZeSTa, RLsite demonstrated absolute improvements of 2.1%, 10.9%, 11.1%, and 8.9% across precision, recall, MCC, and AUC, respectively. When compared against RNAsite (precision 0.750, recall 0.378, MCC 0.426, and AUC 0.790), RLsite showed corresponding improvements of 3.8%, 13.9%, 8.3%, and 8.5%. Although RLBind achieved a higher recall (0.595) and MCC (0.546) than RLsite on RB9, its overall performance across the remaining metrics on both TE18 and RB9 was consistently inferior.

To further investigate the observed differences, we merged the two test sets, which predominantly comprise low-complexity secondary structures. Figure 4 presents a performance comparison between RLsite and RLBind on the merged dataset, which includes 27 RNA chains derived from TE18 and RB9. RLsite achieved precision, recall, MCC, and AUC values of 0.737, 0.434, 0.393, and 0.768, respectively, whereas RLBind obtained corresponding values of 0.704, 0.428, 0.398, and 0.764. Although RLBind shows a slightly higher MCC than RLsite on the combined dataset (0.398 vs. 0.393), this observation can be interpreted in light of the variation in RNA chain composition and sample distribution. Specifically, the TE18 and RB9 datasets differ in both the number of RNA chains (18 and 9, respectively) and their positive sample ratios (0.374 for TE18, 0.325 for RB9), with chain lengths ranging from 20 to 94 nucleotides. Our analysis shows that RLBind's MCC is sensitive to such distributional shifts, increasing from 0.324 on TE18 to 0.546 on RB9 as the positive sample ratio decreases. A similar trend is observed in recall, which rises from 0.345 to 0.595 across the same datasets. After merging, the combined dataset contains 27 RNA chains with an average positive sample ratio of 0.358, slightly higher than RB9 but lower than TE18. Under these conditions, RLBind marginally outperforms RLsite in MCC (0.398 vs. 0.393). However, RLsite's performance remains stable and reliable. In particular, RLsite consistently maintains higher precision than RLBind across varying levels of class imbalance, suggesting that it adopts a more conservative strategy when positive classes are rare. This consistency indicates that RLsite exhibits greater robustness and generalizability, especially in datasets with imbalanced class distributions and varying RNA chain compositions.

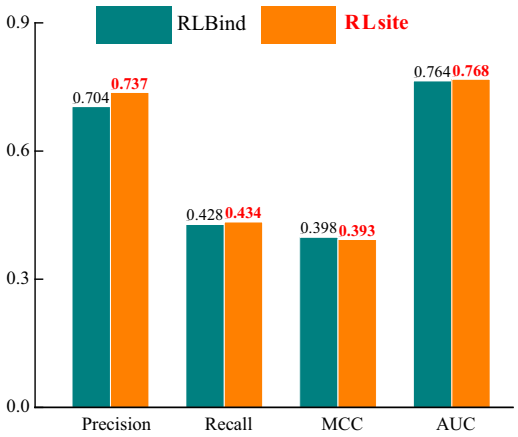


Fig. 4. Comparison of RLsite and RLBind in terms of precision, recall, MCC, and AUC on the combined test set TE18 and RB9 , which together include 27 RNA chains.

In addition, we assessed performance on TL12 sets which were generated from ZeSTa. TL12 test set consists of 12 RNA chains with low structural complexity and specifically excluding RNAs with junction loops. TL12 was used to assess overfitting and compare prediction performance with TE18 and RB9. Figure 5 show that, RLsite achieved a precision of 0.830, recall of 0.514, MCC of 0.521, and an AUC of 0.829. Compared with ZeSTa's 0.740, 0.514, 0.440, and 0.704, RLsite shows a clear advantage. These results indicate that RLsite maintains a balanced predictive profile on diverse datasets,

reflecting its strong generalizability and confirming that the model is not overfitting. These gains further affirm RLsite's superior performance across diverse RNA structural complexities.

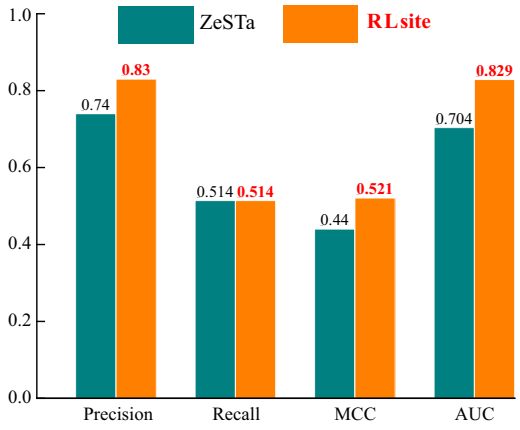


Fig. 5. Comparison between ZeSTa and RLsite in terms of precision, recall, MCC, and AUC on the TL12 test set.

3.2. Performance in challenging tasks

We further evaluated RLsite on the more challenging JL10 dataset, designed to assess prediction performance across varying levels of RNA structural complexity. JL10 comprises 10 RNA structures characterized by complex secondary architectures featuring junction loops with 3–5 connections each.

We benchmarked RLsite against state-of-the-art methods, including ZeSTa, RBind, and Rsite2, using precision, recall, Matthews correlation coefficient (MCC), and area under the ROC curve (AUC) as evaluation metrics. As shown in Fig. 6, RLsite achieved a precision of 0.622, recall of 0.320, MCC of 0.286, and AUC of 0.700, consistently outperforming all competing approaches. Compared to ZeSTa, RLsite improved precision by 7.3% (0.622 vs. 0.549), recall by 3.4% (0.320 vs. 0.296), MCC by 7.5% (0.286 vs. 0.211), and AUC by 10.8% (0.700 vs. 0.592). These results demonstrate that, despite the increased structural complexity, RLsite maintains robust predictive capability by effectively capturing critical spatial correlations within RNA architectures. Across all evaluated datasets, RLsite consistently outperforms existing methods across key metrics. On the simpler TE18 and RB9 datasets, RLsite provides marked improvements in precision, recall, MCC, and AUC. Its performance on the TL12 set further confirms its robustness in low-complexity scenarios. Notably, in the more demanding JL10 dataset, RLsite maintains a substantial advantage, demonstrating the generalizability of our framework integrating conformational energy and sequential descriptors. Together, these results establish RLsite as a powerful and reliable predictor of RNA binding sites across a broad range of structural complexities.

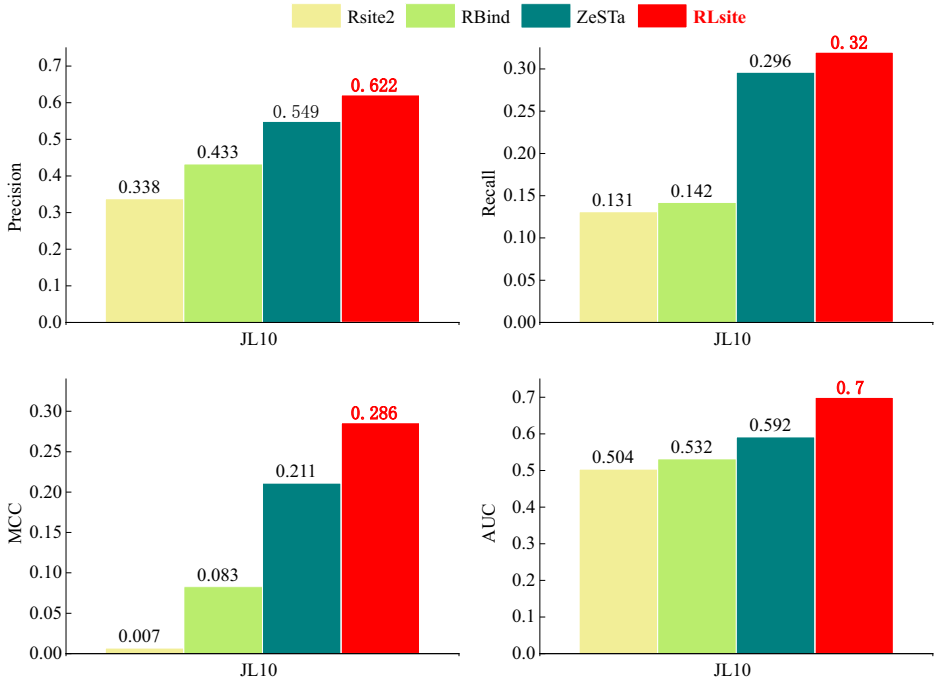


Fig. 6. Performance comparison between RLsite and competing methods in terms of precision, recall, MCC, and AUC on the JL10 test set.

3.3. Case studies

Identifying binding nucleotides in RNA sequences with low structural complexity is relatively straightforward, whereas the task becomes considerably more challenging for RNAs exhibiting high structural complexity. To further assess the performance of RLsite across these two categories, we conducted case-specific evaluations. Malachite green (MGR) is an ideal ligand for studying stacking and electrostatic effects, as it binds to RNA primarily through base stacking and electrostatic interactions (PDB ID: 1Q8N).[50] As illustrated

in Fig. 7(a), the secondary structure of 1Q8N is relatively simple, consisting of an internal loop, a hairpin loop, and a bulge, all of which are connected by stem regions. RLsite's prediction on the tertiary structure of 1Q8N successfully identified 6 out of 7 true binding nucleotides, achieving a precision of 0.75, recall of 0.857, and an MCC of 0.754 (Fig. 7(b)). Flavin mononucleotide (FMN), a redox-active cofactor, binds specifically to an RNA aptamer to form the FMN-RNA complex (PDB ID: 1FMN).[51] RLsite identified 4 of 6 true binding residues (Figs. 7(c) and 7(d)), yielding a precision of 1.0, recall of 0.667, and MCC of 0.79, demonstrating its effectiveness in capturing key nucleotide interactions. Neomycin B, an aminoglycoside antibiotic known to inhibit diverse RNA-mediated processes by binding tightly to structured RNAs (PDB ID: 1NEM)[52] was also assessed. RLsite correctly identified 6 of 12 true binding nucleotides (Figs. 7(e) and 7(f)), achieving a precision of 0.857, recall of 0.5, and MCC of 0.44.

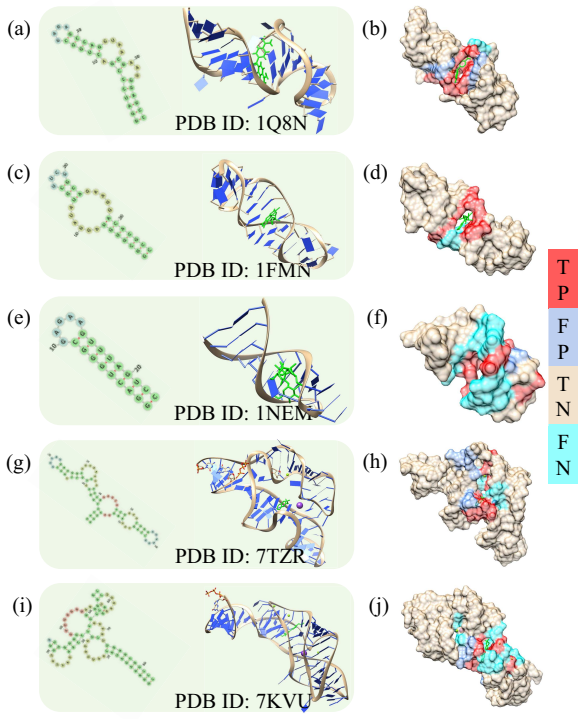


Fig. 7. Case studies of RLsite. (a) Secondary structure and ribbon representation of the MGR and RNA aptamer (PDB ID: 1Q8N). (b) RLsite prediction results for 1Q8N. (c) Secondary structure and ribbon representation of the FMN and RNA aptamer (PDB ID: 1FMN). (d) RLsite prediction results for 1FMN. (e) Secondary structure and ribbon representation of the neomycin B and RNA aptamer (PDB ID: 1NEM). (f) RLsite prediction results for 1NEM. (g) Secondary structure and ribbon representation of the N-methyl-1-(quinoxalin-6-yl)methanamine and thiamine pyrophosphate (TPP) riboswitch (PDB ID: 7TZR). (h) RLsite prediction results for 7TZR. (i) Secondary structure and ribbon representation of the DFHBI-1T and RNA aptamer with (PDB ID: 7KVU). (j) RLsite prediction results for 7KVU. In the prediction panels, true positive sites are shown in red, false positive sites in light blue, true negative sites in light brown, and false negative sites in light cyan.

The ligand N-methyl-1-(quinoxalin-6-yl) methanamine binds the structurally complex thiamine pyrophosphate (TPP) riboswitch RNA with high affinity and modulates its conformation during cotranscriptional folding (PDB ID: 7TZR).[53] The secondary structure of RNA 7TZR is markedly more complex, featuring two hairpin loops, three bulge loops, an internal loop, and a multi-branched loop (Fig. 7(g)). Among the 15 experimentally validated binding nucleotides in the 7TZR structure, RLsite accurately identified 8 (Fig. 7(h)), yielding a precision of 0.667, a recall of 0.533, and an MCC of 0.516. Despite the intricate architecture, RLsite demonstrates the capability to pinpoint small molecule-binding nucleotides with reasonable accuracy. Notably, even the false-positive predictions tend to cluster near the actual ligand-binding pocket, underscoring the strength of our conformational energy-based framework in capturing the geometric and physicochemical properties of RNA-ligand interfaces. Collectively, these case studies highlight RLsite as a robust and reliable tool for binding site prediction across RNAs of varying structural complexities. Lastly, we evaluated RLsite on the squash aptamer-DFHBI-1T complex, a fluorescence turn-on system (PDB ID: 7KVU).[5] RLsite identified 6 true positives of 8 predicted binding nucleotides (Figs. 7(i) and 7(j)), achieving a precision of 0.75, recall of 0.375, and MCC of 0.461.

3.4. Discussion

3.4.1. The impact of the dynamic feature of RNA structure

To evaluate the robustness of RLsite in the context of RNA structural flexibility, we systematically assessed its performance on RNA structures subjected to perturbations. Perturbed models for the TE18 test set were generated using three distinct modeling and simulation tools: RNAComposer,[54] simRNA,[55] and OpenMM.[56] Structural deviations were quantified by calculating the root mean square deviation (RMSD) between the native and perturbed structures. RMSD calculations involved aligning the native and perturbed structures, followed by coarse-graining each model using the P, C4′, and C1′ atoms. The RMSD was computed using the MDAnalysis package. For each perturbed model, RLsite's performance was evaluated across varying RMSD ranges, with metrics including precision (Fig. 8(a)), recall (Fig. 8(b)), and Matthews correlation coefficient (MCC) (Fig. 8(c)). As the RMSD increased, all three performance metrics-precision, recall, and MCC-exhibited a gradual decline, indicating a consistent reduction in predictive accuracy under increasing structural perturbation. Remarkably, when the structural perturbation was minimal (RMSD < 9.0 Å), RLsite maintained high prediction accuracy Fig. 8(a), achieving a precision of 0.689. Precision, defined as the proportion of correct predictions among all predictions, is crucial for reliably identifying the key binding sites where small-molecule ligands interact with RNA.

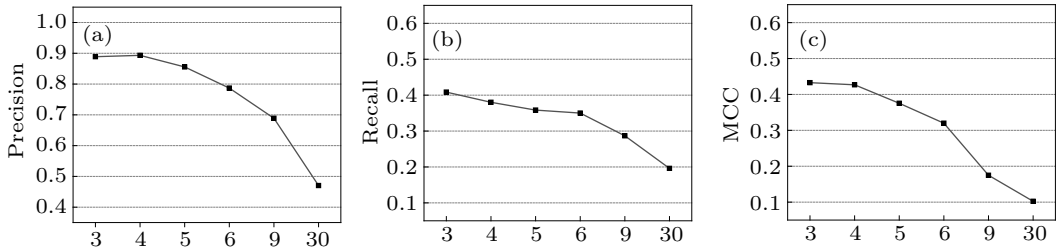


Fig. 8. (a) Precision, (b) recall, and (c) MCC performance of RLsite on the TE18 test set using perturbed RNA structures.

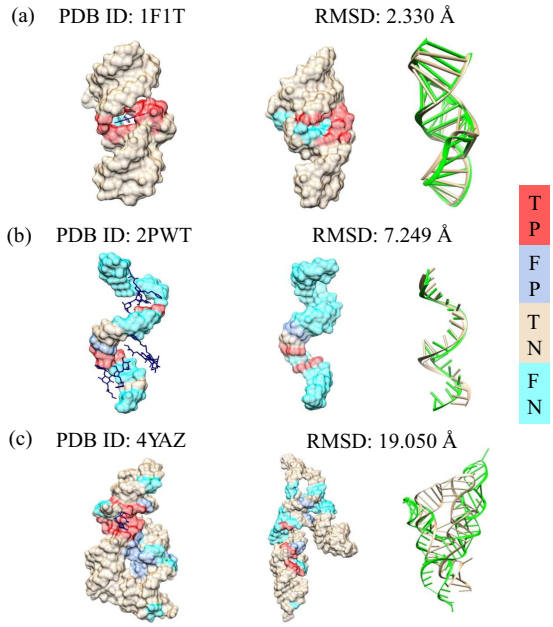


Fig. 9. Case studies on data set TE18 after perturbation generated by RNAComposer. (a) Tetramethylrosamine and RNA aptamers (PDB ID: 1F1T). (b) Aminoglycoside containing the L-HABA group and its bacterial ribosomal aptamer (PDB ID: 2PWT). (c) 3′, 3′-cGAMP and 3′, 3′-cGAMP riboswitch (PDB ID: 4YAZ). Left: The predicted result of RLsite on RNAs with a native structure. Small molecules are represented by navy blue. Middle: The predicted result of RLsite on RNAs with a perturbated structure. In the prediction panels, true positive sites are shown in red, false positive sites in light blue, true negative sites in light brown, and false negative sites in light cyan. Right: The ribbon representation of the native RNA structures and perturbated RNA structures. Native RNA structures are colored tan, and perturbated RNA structures are shown in green.

To further demonstrate RLsite's robustness against RNA structural perturbations, we showcase representative examples from the TE18 test set. The structural deviation from the native conformation is quantified by RMSD, and corresponding performance metrics are reported. In Fig. 9(a), among the 10 experimentally validated binding nucleotides in the RNA structure (PDB ID: 1F1T, RMSD = 2.330 Å), RLsite identified 5, achieving a precision of 1.000 and a recall of 0.500, reflecting accurate predictions with moderate coverage. In Fig. 9(b), the more distorted structure (PDB ID: 2PWT, RMSD = 7.249 Å) yielded a precision of 0.667 and a recall of 0.105. Specifically, RLsite predicted 3 binding nucleotides, of which 2 were correct. Figure 9(c) shows a highly perturbed case (PDB ID: 4YAZ, RMSD = 19.095 Å). Despite the significant deviation, RLsite predicted 8 binding nucleotides with 3 correct, resulting in a precision of 0.375 and a recall of 0.130, further illustrating the RLsite's consistent high precision but declining recall as structural perturbation increases. These examples confirm that RLsite maintains high precision even under substantial structural perturbations, though recall gradually diminishes as the RNA conformation deviates from its native state.

3.4.2. Interpretation of feature contribution

To assess the relative importance of each feature group in our predictive framework, we performed a feature ablation study by introducing random perturbations to individual feature subsets and evaluating the resulting changes in model performance. In this setting, a greater decline in predictive accuracy upon perturbation reflects a higher contribution of the perturbed feature to the model. As shown in Figs. 10(a) and 10(b), we report the area under the ROC curve (AUC) and Matthews correlation coefficient (MCC) on the TE18 dataset under various perturbation scenarios.

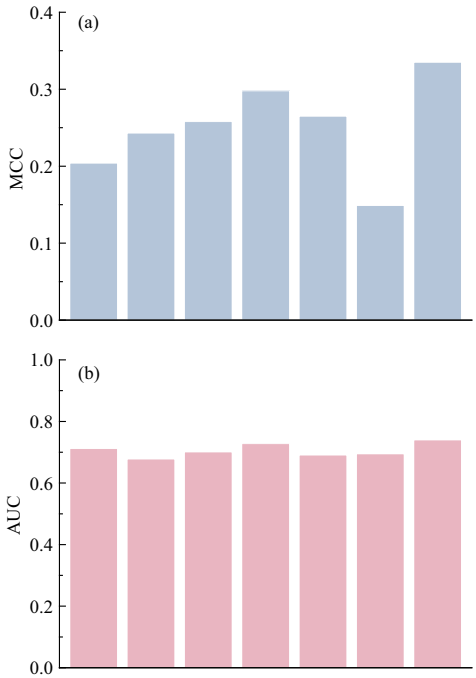


Fig. 10. (a) AUC performance on TE18 using feature groups with selective information perturbation within the sample set, alongside results from the complete (unshuffled) feature group. (b) MCC performance on TE18 using feature groups with selective information perturbation within the sample set, alongside results from the complete (unshuffled) feature group.

The Full setting, in which all features were utilized without perturbation, served as the baseline, yielding an AUC of

approximately 0.74 and an MCC of around 0.335. Notably, perturbing the 3D energy-based features (Peng) resulted in the most pronounced performance drop (AUC: 0.695; MCC: 0.149), followed closely by the perturbation of descriptors encoded in sequential format (PSeq) (AUC: 0.691; MCC: 0.265). These findings underscore the indispensable roles of both conformational energy descriptors and sequential descriptors in capturing the intricate principles governing RNA-ligand interactions. Importantly, the comparable magnitude of performance degradation suggests that neither modality alone is sufficient to comprehensively characterize binding specificity and affinity.

To further dissect the contributions of the sequential descriptors, we independently perturbed their constituent components. Randomization of the network-based (PNW), Laplacian norm (PLN), and evolutionary conservation score (PECS) features each led to moderate reductions in performance, suggesting that these features provide complementary, albeit less dominant, predictive signals. In contrast, perturbation of the solvent-accessible surface area (PSASA) feature resulted in only a marginal decrease in model performance, indicating a relatively limited individual contribution within the current framework.

Together, these results highlight the dominant influence of 3D conformational energy descriptors (Peng) and reinforce the complementary role of integrating diverse descriptors represented in a sequential format. The consistent performance degradation observed under targeted perturbations validates the robustness of our multimodal feature design and demonstrates the synergistic advantage of jointly leveraging spatial and sequential representations to enhance the prediction of RNA-small molecule interactions.

To rigorously assess the individual contributions of different feature modalities, we separately trained models using only the 3D spatial energy descriptors via 3D-CNN and only the sequential descriptors via BiLSTM. All experiments were performed under identical hyperparameter configurations and consistent data splits to ensure a fair comparison. As summarized in Table 3, the standalone 3D-CNN achieves a substantial performance advantage over the BiLSTM across all metrics, with improvements in precision (0.657 vs. 0.221), recall (0.294 vs. 0.160), MCC (0.287 vs. 0.125), and AUC (0.720 vs. 0.689), underscoring the critical role of conformational energy-based features in capturing RNA-ligand binding patterns. Notably, the integrated 3D-CNN and BiLSTM architecture further enhances predictive performance, achieving the highest scores in precision (0.712), recall (0.392), MCC (0.335), and AUC (0.740). These results demonstrate that the fusion of conformational energy and sequential representations yields a more comprehensive feature encoding, thereby enhancing predictive accuracy and robustness.

Table 3. Evaluation of 3D-CNN, BiLSTM, and their combination on the TE18 dataset of 18 RNA structures, with performance assessed by precision, recall, MCC, and AUC.

Model	TE18 dataset			
	Precision	Recall	MCC	AUC
3D-CNN	0.657	0.294	0.287	0.720
BiLSTM	0.221	0.160	0.125	0.689
3D-CNN + BiLSTM	0.712	0.392	0.335	0.740

4. Conclusions

We present RLsite, a deep learning-based model for predicting RNA-ligand binding sites. To thoroughly assess its generalizability, RLsite was tested across multiple datasets with varying structural complexities and compared to state-of-the-art RNA binding site prediction methods. Overall, RLsite consistently demonstrated superior performance, providing robust results across diverse datasets. On the TE18 dataset, comprising 18 RNA structures, RLsite achieved a precision of 71.2%, recall of 39.2%, Matthews correlation coefficient (MCC) of 33.5%, and area under the curve (AUC) of 74.0%. Similarly, on the RB9 dataset, containing 9 RNA structures, RLsite attained a precision of 78.8%, recall of 51.7%, MCC of 50.9%, and AUC of 87.5%, significantly outperforming existing methods and demonstrating its robustness. Moreover, on the JL10 test set, consisting of RNA structures with junction secondary structures, RLsite achieved a precision of 62.2%, recall of 32.0%, MCC of 28.6%, and AUC of 70.0%, surpassing ZeSTa, which yielded 54.9%, 29.6%, 21.1%, and 59.2%, respectively. These results underscore RLsite's enhanced capability to handle complex structural features. Notably, RLsite maintained high accuracy when tested on predicted RNA structures. When the root-mean-square deviation (RMSD) of the perturbed structure was small, the prediction accuracy remained high. For larger RMSD values, RLsite reliably identified binding nucleotides within structural pockets, regions more likely to accommodate small-molecule binding. Further analysis revealed that conformational energy-based features were crucial for accurate binding site prediction, emphasizing the importance of integrating such features into the model. In summary, RLsite offers a promising and robust solution for RNA-ligand binding site prediction, with significant potential to accelerate RNA-targeted drug discovery and inhibitor design.

Program availability

The source code and examples are available at GitHub (https://github.com/fine1231/RLsite) and also in Science Data Bank at https://doi.org/10.57760/sciencedb.j00113.00264.

Acknowledgements

Project supported by the National Natural Science Foundation of China (Grant No. 12204118), the Guizhou Uni-

versity Talent Fund (Grant No. [2022]30), and Guizhou Provincial Major Scientific and Technological Program (Grant No. XKBF (2025)010).

References

[1] Cao Y, Liu H C, Shannon S Lu, Krysten A Jones, Anitha P Govind, Okunola Jeyifous, Christine Q Simmons, Negar Tabatabaei, William N Green, Jimmy. L Holder Jr., Soroush Tahmasebi, Alfred L George Jr. and Bryan C Dickinson 2023 *Nat. Commun.* **14** 6827

[2] Singh S, Sinha T and Panda A C 2024 *Wiley Interdisciplinary Reviews: RNA* **15** e1820

[3] Delaunay S, Helm M and Frye M 2024 *Nat. Rev. Genet.* **25** 104

[4] Peng X, Liao W, Lin X, Lilley D M J and Huang L 2023 *Nucleic Acids Res.* **51** 2904

[5] Truong L, Kooshapur H, Dey S K, Li X, Tjandra N, Jaffrey S R and Ferré-D'Amaré A R 2022 *Nat. Chem. Biol.* **18** 191

[6] Li T, He J H, Cao H, Zhang Y, Chen J, Xiao Y and Huang S Y 2025 *Nat. Biotechnol.* **43** 97

[7] Yang C, Song X, Feng Y, Zhao G and Liu Y 2023 *J. Phys.: Condens. Matter* **35** 265101

[8] Song L, Yu S X, Wang X X, Tan Y L and Tan Z J 2022 *Commun. Theor. Phys.* **74** 075602

[9] He X L, Wang J, Wang J and Xiao Y 2020 *Chin. Phys. B* **29** 078702

[10] Zhu Y R, Zhu L Y, Wang X and Jin H C 2022 *Cell Death Dis.* **13** 644

[11] Toden S and Goel A 2022 *Br. J. Cancer* **126** 351

[12] Pardi N, Hogan M J, Porter F W and Weissman D 2018 *Nat. Rev. Drug Discov.* **17** 261

[13] Disney M D 2019 *J. Am. Chem. Soc.* **141** 6776

[14] Yu A M, Choi Y H and Tu M J 2020 *Pharmacol Rev. 27* 862

[15] Zhao R Y, Fu J H, Zhu L J, Chen Y and Liu B 2020 *J. Hematol Oncol.* **15** 14

[16] de Souza Neto L R, Moreira-Filho J T, Neves B J, Maidana R L B R, Guimarães A C R, Furnham N, Andrade C H and Silva F P Jr 2020 *Front. Chem. 8* 93

[17] Xu W and Kang C 2025 *J. Med. Chem.* **68** 5000

[18] Ursu A, Childs-Disney J L, Andrews R J, O'Leary C A, Meyer S M, Angelbello A J, Moss W N and Disney M D 2020 *Chem. Soc. Rev.* **49** 7252

[19] Velagapudi S P, Cameron M D, Haga C L, Rosenberg L H, Lafitte M, Duckett D R, Phinney D G and Disney M D 2016 *Proc. Natl. Acad. Sci. USA* **113** 5898

[20] Feng Y Y, Yan Y M, He J H, Tao H Y, Wu Q L and Huang S Y 2022 *Drug Discov Today* **27** 838

[21] Feng Y, Zhang K, Wu Q and Huang S Y 2021 *J. Chem. Inf. Model.* **61** 4771

[22] Sun L Z, Jiang Y W, Zhou Y Z and Chen S J 2020 *J. Chem. Theory Comput.* **16** 7173

[23] Detering C and Varani G 2004 *J. Med. Chem.* **47** 4188

[24] Ruiz-Carmona S, Alvarez-Garcia D, Foloppe N, Beatriz Garmendia-Doval A, Juhos S, Schmidtke P, Barril X, Hubbard R E and David Morley S 2014 *PLoS Comput. Biol.* **10** e1003571

[25] Therese Lang P, Brozell S R, Mukherjee S, Pettersen E F, Meng E C, Thomas V, Rizzo R C, Case, James T L and Kuntz I D 2009 *RNA* **15** 1219

[26] Guilbert C and James T L 2008 *J. Chem. Inf. Model.* **48** 1257

[27] Zeng P, Li J, Ma W and Cui Q 2015 *Sci. Rep.* **5** 9179

[28] Zeng P and Cui Q 2016 *Sci. Rep.* **6** 19016

[29] Wang K, Jian Y, Wang H, Zeng C and Zhao Y 2018 *Bioinformatics* **34** 3131

[30] Su H, Peng Z and Yang J 2021 *Bioinformatics* **37** 36

[31] Wang K, Zhou R, Wu Y and Li M 2023 *Brief Bioinform* **24** bbac486

[32] Gao J, Liu H, Zhuo C, Zeng C and Zhao Y 2024 *J. Chem. Inf. Model.* **64** 6979

[33] Berman H M, Westbrook J, Feng Z, Gilliland G, Bhat T N, Weissig H, Shindyalov I N and Bourne P E 2000 *Nucleic Acids Res.* **28** 235

[34] Gong S, Zhang C and Zhang Y 2019 *Bioinformatics* **35** 4459

[35] Li W and Adam G 2006 *Bioinformatics* **22** 1658

[36] Altschul S F, Gish W, Miller W, Myers E W and Lipman D J 1990 *J. Mol. Biol.* **215** 403

[37] Fu L, Niu B, Zhu Z, Wu S and Li W 2012 *Bioinformatics* **28** 3150

[38] Huang Y, Niu B, Gao Y, Fu L and Li W 2010 *Bioinformatics* **26** 680

[39] Feng Y and Huang S Y 2020 *J. Chem. Inf. Model.* **60** 6698

[40] Amitai G, Shemesh A, Sitbon E, *et al.* 2004 *J. Mol. Biol.* **344** 1135

[41] Bonnel N and Marteau P F 2012 *IEEEACM Transactions on Computational Biology and Bioinformatics* **9** 1451

[42] Sun S, Wu Q, Peng Z and Yang J 2019 *Bioinformatics* **35** 1686

[43] Zhang Z, Scott Schwartz, Lukas Wagner and Webb Miller 2000 *J. Comput. Biol.* **7** 203

[44] Chen K, Thomas Litfin, Jaswinder Singh, Zhan J and Zhou Y Q 2024 *Genomics, Proteomics & Bioinformatics* **22** qzae018

[45] Marks D S, Hopf T A and Sander C 2012 *Nat. Biotechnol.* **30** 1072

[46] Jian Y, Wang X, Qiu J, Wang H, Liu Z, Zhao Y and Zeng C 2019 *BMC Bioinformatics* **20** 497

[47] Ekeberg M, Lövkvist C, Lan Y, Weigt M and Aurell E 2013 *Phys. Rev. E Stat Nonlin Soft Matter Phys.* **87** 012707

[48] Zhou P, Xie X, Lin Z and Yan S. 2024 *IEEE Trans Pattern Anal Mach Intell.* **46** 6486

[49] He K M, Zhang X Y, Ren S Q and Sun J 2015 *IEEE International Conference on Computer Vision (ICCV)* 1026

[50] Flinders J, DeFina S C, Brackett D M, Baugh C, Wilson C and Dieckmann T 2004 *Chembiochem.* **5** 62

[51] Fan P, Suri A K, Fiala R, Live D and Patel D J 1996 *J. Mol. Biol.* **258** 480

[52] Jiang L, Majumdar A, Hu W, Jaishree TJ, Xu W and Patel DJ 1999 *Structure* **7** 817

[53] Zeller M J, Favorov O, Li K, Nuthanakanti A, Hussein D, Michaud A, Lafontaine D A, Busan S, Serganov A, Aubé J and Weeks K M 2022 *Proc. Natl. Acad. Sci. USA* **119** e2122660119

[54] Biesiada M, Pachulska-Wieczorek K, Adamiak R W and Purzycka K J 2016 *Methods* **103** 120

[55] Boniecki M J, Lach G, Dawson W K, Tomala K, Lukasz P, Soltysinski T, Rother K M and Bujnicki J M 2016 *Nucleic Acids Res.* **44** e63

[56] Eastman P, Galvelis R, Peláez R P, Abreu C R A, Farr S E, Gallicchio E, Gorenko A, Henry M M, Hu F, Huang J, Krämer A, Michel J, Mitchell J A, Pande V S, Rodrigues J P, Rodriguez-Guerra J, Simmonett A C, Singh S, Swails J, Turner P, Wang Y, Zhang I, Chodera J D, De Fabritiis G and Markland T E 2024 *J. Phys. Chem. B* **128** 109