Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Structure-based discovery of small molecules that allosterically modulate RNA function

Zhaoming Su
zsulab2019@gmail.com

West China Hospital https://orcid.org/0000-0002-9279-1721
Zhiling Pan
Chengdu university
Hao Wu
Fudan University
Han Hu
Mingle Scope
Jian Zou
Sichuan University
Haiyun Ma
Sichuan University
Hongshan Ni
Sichuan university
Liu Wang
Sichuan University https://orcid.org/0000-0003-1243-9654
Jiaqi Zhao
Mingle Scope
Xiaoqi Sun
University of Chinese Academy of Sciences
Jiayu Dong
University of Chinese Academy of Sciences
Sheng Xu
Fudan University https://orcid.org/0000-0002-6507-9122
Yu Cheng
The Chinese University of Hong Kong https://orcid.org/0000-0002-7901-8662
Ling Liu
Institute of Microbiology, Chinese Academy of Sciences https://orcid.org/0000-0002-4882-0396
Ke Zheng
College of Chemistry, Sichuan University https://orcid.org/0000-0002-3345-7715
Siqi Sun

Fudan university https://orcid.org/0000-0001-7240-8724

Article

Keywords:

Posted Date: April 20th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9338412/v1

License: 🅭 🅯 This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: There is **NO** Competing Interest.

Structure-based discovery of small molecules that allosterically modulate RNA function

Zhiling Pan[1,2]†, Hao Wu[3,4]†, Han Hu[5,]†, Jian Zou[2,]†, Haiyun Ma[2,]†, Hongshan Ni[6,]†, Liu Wang[7], Jiaqi Zhao[5], Xiaoqi Sun,[8,9], Jiayu Dong[8], Sheng Xu[3,4], Yu Cheng[10], Ling Liu[8,]*, Ke Zheng[6,]*, Siqi Sun[3,4,]*, Zhaoming Su[2,]*

[1]School of Pharmacy, Chengdu University, Chengdu, 610051, China.
[2]State Key Laboratory of Biotherapy, West China Hospital, Sichuan University, Chengdu, 610044, China.
[3]Research Institute of Intelligent Complex Systems, Fudan University, Shanghai, 200437, China.
[4]Shanghai Artificial Intelligence Laboratory, Shanghai, 200030, China.
[5]Mingle Scope (Chengdu), 138 2nd Tianfu Street, Chengdu, 610096, China.
[6]Key Laboratory of Green Chemistry & Technology, Ministry of Education, College of Chemistry, Sichuan University, Chengdu 610064, China.
[7]State Key Laboratory of Oral Diseases, National Clinical Research Center for Oral Diseases, National Center for Stomatology, Department of Cariology and Endodontics, West China Hospital of Stomatology, Sichuan University, Chengdu, 610044, China.
[8]State Key Laboratory of Microbial Diversity and Innovative Utilization, Institute of Microbiology, Chinese Academy of Sciences, University of Chinese Academy of Sciences, Beijing 100101, China.
[9]School of Biomedical Sciences, Shandong First Medical University & Shandong Academy of Medical Sciences, Jinan 250117, China.
[10]Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong 999077, China.

†These authors contributed equally.
*Correspondence should be addressed to Z.S. (zsu@wchscu.cn), S.S. (siqisun@fudan.edu.cn), K.Z. (kzheng@scu.edu.cn), and L.L. (liul@im.ac.cn).

Abstract

RNAs play essential roles in biological processes and have emerged as promising therapeutic targets for both oligonucleotide- and small molecule-based interventions. However, targeting RNAs with small molecules remain challenging due to our limited understandings of the chemical space associated with RNA-binding small molecules, the intricate three-dimensional RNA structures and ligand-binding pockets, and the molecular basis of small-molecule modes of action (MOAs). In this study, we discovered Hit Small molecules TArgeting RNA Structures (HitSTARS) through the following three stages. We first curated a library of 5.2 million in-stock Commercial Small molecules TArgeting RNA (CSTAR) through training a molecular representation learning model based on 3D atom positions, and explored their physicochemical properties. Next, virtual screening of CSTAR against the paradigmatic *Tetrahymena* ScaI L-16 ribozyme, as a proof of concept, identified two novel allosteric pockets. Functional screening of the top-scoring 125 compounds, validated by multiple docking algorithms, resulted in nine hit RNA-binders capable of inhibiting catalysis. Finally, cryo-EM structures of the best hit of each allosteric site, ZPT-005 and ZPT-084, in complexes with L-16 were determined at 2.3 Å and 3.0 Å resolutions to elucidate their modes of action: ZPT-005 stabilizes J1/2 in the undocked conformation upon binding to site 1, whereas ZPT-084 occupies site 2 to displace the catalytic core away from the splice site. These results facilitated lead optimization of ZPT-084 that improved the binding affinity and activity by 104- and 15-fold. Collectively, HitSTARS provides a comprehensive framework for streamlining RNA structure-based chemical probe discovery while offering insights into their MOAs.

Introduction

RNAs have emerged as attractive drug targets due to their ability to encode disease-related proteins, especially those considered “undruggable”, and their capacity to form intricate structures, particularly in non-coding regions, which play critical roles in diverse disease development processes[1-3]. Current RNA-targeting strategies include sequence-specific approaches, such as antisense oligonucleotides (ASOs), and RNA structure-based small-molecule interventions. However, the clinical application of ASOs faces challenges owing to limited delivery strategies, off-target effects, and low efficacy on structured RNA targets[4,5]. Alternatively, small molecule is eminent for oral bioavailability, and leverages the extensive knowledge in structure-activity relationship (SAR) and experience in medicinal chemistry, enabling iterative optimizations of pharmacophores for improved potency, specificity and pharmacokinetics[6,7].

Recent phenotypic and RNA-binding screenings have resulted in drug-like small molecules that modulate transcription and translation[8,9], alternative splicing[10,11], protein-RNA interactions[12] and RNA degradation[13], some of which were already in clinical use or showed therapeutic promise. However, many RNA-binding compounds still suffer from poor drug-like properties, such as excessive positive charges, multivalency, or intercalation[1]. Moreover, routine discovery of small molecules targeting RNAs, particularly those with complex structures and high-quality binding pockets, remains challenging owing to the scarcity of RNA three-dimensional (3D) structures and limited understandings of RNA-binding propensity and modes of action (MOAs) of small molecules[14].

In the current study, we identified Hit Small molecules TArgeting RNA Structures (HitSTARS) and elucidated their MOAs through the following steps: 1) A deep-learning transformer architecture, RiboBIND, that integrates 3D spatial information from atomic coordinates with 2D topological data from molecular graphs[15]. Trained on thousands of RNA- and protein-binding small molecules[16,17], it enabled the curation of 5.2 million Commercial Small molecules TArgeting RNA (CSTAR)[18]; 2) Virtual screening of the CSTAR library against structured RNAs, where we selected the paradigmatic *Tetrahymena* ScaI L-16 ribozyme as a proof-of-concept target[19,20], to identify novel allosteric sites. The top-scoring 125 compounds, validated by multiple docking algorithms, were subjected to functional screening to yield nine hit L-16-binding compounds inhibiting catalysis; 3) MOA elucidation of the best hit from each site, ZPT-005 and ZPT-084, through single-particle cryo-electron microscopy (cryo-EM) that has been increasingly used for high-resolution structure determination of large RNAs complexed with small molecules[21-24].

The 2.3 Å resolution structure showed that ZPT-005 stacks with U106 and A306, forms anion-π and hydrogen bond interactions with the fully coordinated metal ion M_{16} through water molecules in site 1[19,23,25], and stabilizes the J1/2 (J denotes junction) in the P1 (P denotes paired stem) undocked conformation[23]. In the presence of ZPT-084, two conformations of the ribozyme, native (N) and inhibited (I), were resolved at 3.0 Å and 3.8 Å resolutions. In the N conformation, ZPT-084 binds to site 2 in the metal core (MC) via anion-π, hydrogen bond and outer-sphere metal coordination interactions. The I conformation showed disrupted P5/L9 (L denotes loop) tertiary interaction and translocations of P7, P9 and P9.2-P9.1-P13 away from the 5'-splice site (5'-SS) by 11º, 42º and 17º, suggesting that allosteric binding of ZPT-084 induces large conformational changes that displace the G-binding site more than 8 Å away from the 5'-SS. This MOA was further validated by lead optimization of ZPT-084 that enabled augmentation of ZPT-084.10 binding affinity and inhibition activity by 104- and 15-fold, yielding predominant I conformation in cryo-EM analysis.

Results

CSTAR library construction

Previous studies have established several libraries of RNA-binding small molecules and explored their preferred physicochemical properties and chemical space for RNA recognition [26]. Machine learning approaches have enabled identification of molecular features associated with RNA binding[16,27-29]. However, these methods have been primarily limited to two-dimensional (2D) molecular descriptors. Recent advances such as Uni-Mol have highlighted the importance of 3D molecular representations for accurate property predictions[15].

To leverage 3D structural information, we developed RiboBIND, a deep-learning model initialized with weights pre-trained on 200 million molecular conformations (Figure 1)[15]. Our curated training set comprises 994 RNA-binders as positives, sourced from various databases including R-BIND, R-SIM and the protein data bank (PDB), and rigorously filtered by drug-likeness and similarity [16,17,30]. Negatives consist of 3,767 protein-binders from PDB, excluding FDA-approved drugs (Figure 1A, Supplementary Data 1). RiboBIND comprises a molecular module and an application module, with SMILES strings, atomic coordinates, and molecular graphs as inputs (Figure 1B). The molecular module encodes 3D atomic coordinates using spatial positional embeddings and transformer-based self-attention, capturing atom-to-pair relationships while ensuring invariance to molecular orientation[31]. The application module employs a graph convolutional network (GCN) for downstream fine-tuning[32], The GCN is specifically trained to capture essential chemical bond features and atom-level relational patterns from the molecular graph. The resulting features are then processed by a multilayer perceptron (MLP) to compute the final RNA-binding preference over protein-binding.

RiboBIND demonstrated optimal model performance on the blinded benchmark set of 192 RNA- and 761 protein-binders (Figure 1C), achieving an area under the receiver operating characteristic (AUROC) of 0.96 and area under the precision-recall curve (AUPRC) of 0.90, outperforming Uni-Mol (AUPRC 0.87) (Figure S1A). We further evaluated RiboBIND performance on the independent, public accessible ROBIN data as benchmark set [27]. Using a five-fold cross-validation protocol, RiboBIND achieved a mean AUPRC of 0.92, outperforming both Uni-Mol (0.88) and ROBIN_MLP (0.77) using 2D molecular descriptors (Figure S1B). Based on this robust performance, a final model, trained on our entire dataset using the optimal hyperparameters, was applied to a collection of more than 13 million non-redundant in-stock small molecules (Figure 1D)[18], resulting in the CSTAR library of 5.2 million compounds predicted to preferentially bind RNAs over proteins (Figure 1E). We built a webserver (cstar.zsulab.com) for the CSTAR library, offering features including search, classification, download and prediction (Figure S1C). Feature analysis using the Shapley additive explanations (SHAP) identified properties associated to polarity, aromaticity, Van der Waals surface area and topology as major determinants of RNA-binding preference compared to protein-binders and FDA-approved drugs (Figure S1D-E)[33].

Virtual screening against RNA structure identified unforeseen ligand-binding pockets

Molecular docking has been well-established for high-throughput discovery of small molecules targeting given structures with defined ligand-binding pockets, allowing rapid screening of compound libraries[34]. Although recent advances in docking and scoring algorithms have facilitated accelerated virtual screening against RNA structures[35,36], successful identification of bioactive RNA-binding small molecules from virtual screening remains challenging owing to the paucity of protein-free RNA and RNA-small molecule complex structures[14,22].

We selected the *Tetrahymena* ScaI L-16 ribozyme as our model RNA target because of its extensively characterized structures and biochemical functions for subsequent virtual and functional screenings (Figure 2) [25,37,38]. Virtual screening was performed on the CSTAR library using DOCK6[39], which offers reliable performance with moderate computational costs [36]. Since RNA ligand-binding pockets and recognition patterns remain poorly understood, molecular docking was initially applied to the entire L-16 ribozyme structure with no defined pockets against 8,392 diversified CSTAR compounds of <0.25 similarity. Strikingly, the docking result revealed that 98% compounds localized to two high-quality binding sites according to fpocket scores (Figure 2B, Figure S2)[40,41]. Site 1 is formed by J3/4, P7 and J8/7 close to the catalytic site that accounts for 70% of the docked compounds, whereas site 2 resembles a “Y-shaped” pocket formed by P4, P5a-c and P14 in the MC that accounts for 28% (Figure 2B). The remaining 2% attributes to four sites, including the guanosine binding site, two helical-groove binding sites, and one site formed by P2, P2.1, P3 and P8 (Figure S2). Site 1 and 2 in L-16 were subsequently defined for virtual screening of the entire 5.2 million CSTAR library.

Functional screening discovered RNA-binding compounds that inhibit catalysis

The top-scoring 10,000 molecules were cross-validated by rDOCK[42] and RNAmigos2[36], and the top-ranking 125 compounds were purchased to assess L-16 binding and catalysis inhibition (Figure 2C-D, Supplementary Data 2). Biolayer interferometry determined dissociation constants (K_D) for 13 compounds, ranging from 20.7 to 441 μM (Figure 3A, Figure S3A). We then designed a Förster resonance energy transfer (FRET) assay monitoring the cleavage of RNA substrate 5'-CCCUCUAAACC-3' by L-16 ribozyme, in the presence of exogenous guanosine (exoG)[43]. The substrate was labeled with Cy3 and FAM at the 5' and 3' ends (Figure 2D), enabling detection of catalysis inhibition through the fluorescence signals of FAM at 1000 and 200 μM compound concentrations (Figure 3B). A total of nine hit compounds binding to L-16 showed inhibition of catalysis with IC_{50} ranging from 9.8 to 676 μM (Figure 3C and S3B), and ZPT-005 and ZPT-084 were the most potent inhibitors for site 1 and 2, with K_D and IC_{50} of 20.7 and 9.8 μM for ZPT-005, and 190 and 230 μM for ZPT-084, respectively (Figure 3D).

ZPT-005 recognition mechanism and mode of action

Cryo-EM structure of ZPT-005 in complex with L-16 was resolved at 2.3 Å resolution in the N conformation (Figure S4, Table S1). Previous L-16 models with annotated metal ions and water molecules were docked and manually adjusted (Table S2)[19,23,25], enabling identification of ZPT-005 density in site 1 (Figure 4A). This compound was subsequently docked into the density to unveil its recognition pattern (Figure 4B), which is in close proximity to the molecular docking result (Figure S5). ZPT-005 contains two N-ethyl quinolines and a N-ethyl pyridine, connected through a conjugated branched site to fit in site 1 with overall positive charges (Figure 4C). The two quinoline rings stack with U106 and A306, and form anion-π interactions with non-bridging phosphate oxygens (OPs) of U259 and A265. The pyridine ring of ZPT-005 forms two outer-sphere contacts with the hydrated M_{16} that has full octahedral coordination through water molecules (Figure 4B). In the presence of ZPT-005, A29 and A30 form an A-platform while G27 stacks with A94, stabilizing J1/2 in an intermediate conformation with P1 in the undocked position (Figure 4D). This J1/2 conformation is drastically different from J1/2 in the previously reported *Tet*-S1 state, prohibiting P1 from docking (Figure 4E) [23].

This structure reveals novel RNA-small molecule interaction paradigms. Anion-π interactions have been widely observed in protein α-helix containing a glutamic acid and a tryptophan or tyrosine[44,45], or in a typical RNA GNRA tetraloop or U-turn motifs[46-48]. The adjacent metal ion contacts, as observed in ZPT-005 recognition pattern, have been suggested to increase the quadrupole moment of the aromatic rings, leading to enhanced anion-π interaction[44]. This combination is analogous to the previously reported recognition of anthranilic acid derivative inhibitor in the HCV NS5B polymerase thumb pocket[49], in which the acid group coordinate with the Mn^{2+} to enhance the anion-π interaction between the aromatic ring and E361 (Figure S5B)[44].

ZPT-084 and induces conformational change to translocate the catalytic center away from the splice site

Cryo-EM reconstruction of ZPT-084 bound L-16 ribozyme revealed two conformations: the N conformation resolved at 2.9 Å resolution and another extended I conformation with partially missing density in the peripheral P9 and P9.2 (Figure 5A, Figure S6A-B, Table S1). 3D classification focused on the P9 and P9.2 region resulted in a class with intact P9.2, resulting in a final cryo-EM reconstruction at 3.8 Å resolution (Figure 5B, Figure S6C). Focused classification of the N conformation yielded the final reconstruction at 3.0 Å resolution with the presence of ZPT-084 density in site 2 (Figure S7). The neutrally charged ZPT-084 adopts a bent conformation (Figure 5C), mediated by the central piperazine ring, to fit in the "Y-shaped" site 2, which is homologous to the docking result (Figure S5C). The recognition mechanism is elucidated by four hydrogen bonds formed by the piperazine nitrogens and the closeby hydroxyl groups with A183 and A184, two anion-π interactions formed by the aromatic ring on either end with the OPs of C165 and C170, and an outer-sphere contact between the hydroxyl group and M_5 with a distance of 3.5 Å (Figure 5D).

The I conformation revealed extensive structural rearrangements in the core regions P7 and P9, and the coaxially stacked peripheral domain P9.2-P9.1-P13 (Figure 5E). Such conformational changes originate from local shift of P5 and J5/5a that disrupts the critical P5/L9 interaction (Figure S5D)[19], which leads to translocations of P7, P9 and P9.2-P9.1-P13 away from P4-P6 by 11^{o}, 42^{o} and 17^{o}, respectively (Figure 5E). Previous studies have reported that disruption of the P5/L9 tertiary interaction primarily affects peripheral architecture and guanosine association to the G-binding site, leaving P1 docking largely unaffected [50,51]. This is validated by the unchanged spatial organization of A114, A207 and J8/7 that modulate P1 docking register compared to the N conformation, leading to homologous P1 docking position in the I conformation (Figure S5E). As a result, the conformational change in P7 extends the distance between the putative exoG and 5'-SS to over 8 Å, resulting in inhibition of catalysis (Figure 5F).

ZPT-084 binding mechanism enables SAR and lead optimization

Elucidation of ZPT-084 binding mechanism allows SAR investigation and lead optimization (Figure 6). Racemization of both chiral hydroxyl groups largely reduced binding affinity and activity (ZPT-084.14), whereas changing the phenolyl to anilinyl moieties further decreased potency (ZPT-084.1) (Figure 6A). Substitutions of the acetyl groups (R1) on both ends showed minimal effects, with the methoxyl substitution (ZPT-084.3) showing modest improvements of ~2-fold in both binding affinity and activity (Figure 6B). Benzoyl acylations of a single racemic hydroxyl group (R2) enhanced shape complementarity of small molecules occupying the "Y-shaped" site 2, leading to general improvements in binding affinity and inhibition activity across various R1 substitutions of acetyl, isopropyl, ethyl and trifluoromethyl (Figure 6C). In particular, acylating a benzyol moiety on one hydroxyl group of ZPT-084 yielded the most potent derivative with K_D of 1.82 μM and IC50 of 15.2 μM, increased by 104- and 15-fold compared to the parent compound (ZPT-084.10), respectively. As expected, phenylacetylations of both hydroxyl groups confirmed the importance of shape complementarity, as bulkier substitutions abolished the compound's binding and activity (ZPT-084.2). Cryo-EM analysis of L-16 in complex with ZPT-084.10 revealed predominant population of I conformation, validating that small-molecule binding to site 2 induces such conformational change (Figure S9). Both recognition patterns in site 1 and 2 were validated by introducing mutations to disrupt interactions with ZPT-005 and ZPT-084.1, which resulted in K_D values of 76.4 μM and 14.9 μM, reduced by 3.7- and 8.2-fold, respectively (Figure S9).

Discussion

Targeting RNAs has garnered increasing interest in the pharmaceutical industry, with several recent examples highlighting the potential of small molecules targeting non-ribosomal RNAs for clinical applications[7]. However, most RNA-binding compounds remain positively charged, multivalent, or intercalators, and primarily bind to less complexed pockets like loops and bulges[1]. This has led to perspectives that RNA-targeting drugs may not adapt the physicochemical properties of approved protein-targeting drugs following the classic guidelines for drug-likeness[6,52]. In this study, we developed a deep-learning model to retrieve molecular representations from thousands of carefully filtered RNA- and protein-binders, and curated the RNA-focused CSTAR library for analysis of RNA-binding propensities and identification of hit compounds (Figure 1-2). The scarcity of drug-like small molecules with validated RNA-binding capacity and biological activity has limited conventional deep-learning algorithms using only 2D small-molecule information [28,53]. Therefore, we leveraged 3D molecular representations of small molecules, generating multiple conformations per molecule, along with tens to hundreds of atom coordinates, atom pairs and bond information for each conformation (Figure 1B). Physicochemical property analysis revealed polarity, aromaticity, Van der Waals surface area and topology as critical determinants for RNA-binding propensities, consistent with previous findings (Figure S1)[26].

RNAs exist as dynamic conformational ensembles in solution[54]. Their inherent structural flexibility has mostly caused the paucity of experimentally determined RNA 3D structures that continues to hinder accurate RNA structure prediction and structure-based drug discovery[55]. Nuclear magnetic resonance can resolve RNA conformational ensembles and identify small molecules targeting dynamic short stem-loops, inducing RNA conformational changes, and stabilizing high-energy transient states[56-58]. However, these structural changes are typically localized and involve only a few nucleotides. Cryo-EM is now routinely used to capture global conformational changes and dynamics of large, intricately-structured RNAs[38].

Although the L-16 ribozyme is not pathogenic, previous studies have discovered small-molecule inhibitors of *Tetrahymena* group I intron, including orthosteric compounds competing with guanosine at the G-binding site[59-62], and allosteric inhibitors like pentamidine binding near P2/P2.1, adjacent to a characterized spermidine binding site in P2.1-P3-J8/7[23]. Therefore we used this RNA as a proof-of-concept target for the virtual and functional screening modules of HitSTARS. Virtual screening of the CSTAR library identified two primary allosteric binding pockets that have not been previously reported (Figure 2B), and four additional minor binding pockets, including the G-binding site (Figure S2). The top-scoring 125 compounds from virtual screening were subjected to functional screening, yielding 13 compounds with reliable K_D values (10% hit rate), and nine of which inhibited catalysis with IC_{50} values (7.2% hit rate).

ZPT-005 and ZPT-084 were identified as the most potent inhibitors for their corresponding allosteric binding pockets. Cryo-EM resolved L-16 ribozyme complexed with ZPT-005 at 2.3 Å resolution, and with ZPT-084 at 3.0 Å resolution. These structures revealed not only hydrogen bond and π-stacking commonly observed in RNA-small molecule interactions[14], but also new recognition patterns of anion-π and outer-sphere metal coordination in high-quality binding pockets. The 2.3 Å structure of L-16 in complex with ZPT-005 indicates that the hit compound locks J1/2 in the P1 undocked conformation (Figure 4), whereas ZPT-084 induces minor changes in P5 and J5/5a that eventually leads to disruption of the P5/L9 tertiary interaction, launching large conformational changes of P7, P9, and P9.2-P9.1-P13 domain that result in displacement of the G-binding site and inhibition of catalysis (Figure 5, Movie S1).

We further demonstrated its utility in facilitating downstream lead optimization and biological activity assessment. The structural information allowed optimization of the original hit ZPT-084 for site 2, with the modest K_D and IC_{50} enhanced by 104- and 15-fold. Although the L-16 ribozyme is not pathogenic, previous studies have discovered small-molecule inhibitors of *Tetrahymena* group I intron, including orthosteric compounds competing with guanosine at the G-binding site[59-62], and allosteric inhibitors like pentamidine binding near P2/P2.1, adjacent to a characterized spermidine binding site in P2.1-P3-J8/7[23]. The discovery of new allosteric inhibitors in this study expands the small molecules targeting this ribozyme in allostery.

Notably, the current HitSTARS framework has been designed to accommodate alternative methods and techniques in each module to fit customized needs. The virtual screening module can incorporate diverse compound libraries, structured RNA targets and molecular docking algorithms, while the functional screening module supports different biophysical methods and function-based screening assays. In the MOA elucidation state, complementary structural biology techniques can be employed: cryo-EM is especially useful for large structured and dynamic RNAs, X-ray crystallography fits smaller and more rigid RNAs, and NMR is useful for small and dynamic RNAs. This inherent adaptability ensures HitSTARS to remain cutting-edge as new methodologies for each module emerge.

In conclusion, HitSTARS establishes a streamlined strategy to discover bioactive hit small molecules targeting novel allosteric site in highly-structured RNAs, complemented by elucidation of their MOAs, allowing further lead optimization. This approach also provides unprecedented structural insights into high-quality RNA-binding pockets and RNA-small molecule interactions. As the repertoire of RNA structures expands and structure prediction algorithms improve, HitSTARS will become increaslingly scalable, advancing our fundamental understanding of RNA-small molecule recognition patterns while accelerating drug discovery against pathogenic RNAs.

Methods

CSTAR library construction

Data preparation. A total of 994 RNA-binding positives and 3,767 protein-binding negatives was partitioned into training and benchmark datasets. SMILES representations and 3D coordinates of these small molecules were prepared as inputs. To generate their 3D conformations, we employed the Experimental-Torsion Distance Geometry (ETKGD)[63] algorithm integrated with Merck Molecular Force Field (MMFF) optimization using the RDKit software, generating up to 11 conformations per molecule. We limited the optimizations to a maximum of 500 iterations for each molecule. For those molecules with infeasible 3D conformations, 2D structures were used instead. The molecular graph was constructed using RDKit to determine chemical bonds.

Model architecture. RiboBIND is a deep-learning model based on the transformer architecture, featuring 48 million parameters within its molecular and application modules. We use the UniMol weights pre-trained on 200 million small molecule conformations to initialize the molecular module. Specifically, the molecular module comprises 15 transformer blocks and takes atom types and their coordinates as inputs. Atom representations are initialized using a 512-dimensional embedding layer based on atom types, while pair representations utilize an invariant spatial positional encoding derived from atom coordinates. Specifically, we calculate the Euclidean distances between atom pairs and apply a pair-type aware Gaussian kernel[31]. For each atom pair $i\,j$, the calculation is expressed as follows:

$$\mathcal{G}\left(v_{ij}d_{ij} + u_{ij} - \mu\,, \sigma\right),$$

where $\mathcal{G}$ represents Gaussian basis functions with means μ and standard deviation σ, d_{ij} is the euclidean distances between the atom pairs i and j, and v_{ij} and u_{ij} are learnable parameters adjusting gain and bias, respectively. This positional encoding ensures that the pair representation remains invariant to global rotations and translations.

Each transformer block contains 64 attention heads and a feedforward network (FFN) with a hidden dimension of 2048. For the pair representation, the transformer block facilitates atom-to-pair communication using the multi-head query-key product within the self-attention mechanism. Each attention head employs separate weights $\boldsymbol{w}^{q} \in \boldsymbol{R}^{D}, \boldsymbol{w}^{k} \in \boldsymbol{R}^{D}$ to transform pair representation $\boldsymbol{x} \in \boldsymbol{R}^{L\times D}$ into queries $\boldsymbol{Q} = \boldsymbol{x}_i\boldsymbol{w}^{\boldsymbol{q}}$ and key $\boldsymbol{K} = \mathbf{x}_j\mathbf{w}^{\mathrm{k}}$, where D is the hidden dimension and L is the molecular length. The update for each attention head is defined as follows:

$$\boldsymbol{q}_{ij}^{l+1} = \boldsymbol{q}_{ij}^{l} + \frac{\boldsymbol{Q}_i^l\left(\boldsymbol{K}_j^l\right)^T}{\sqrt{d}},$$

where $\boldsymbol{q}_{ij}^{l}$ is the pair representation at layer l, d is the dimension of hidden representations, and $\left(\boldsymbol{K}_j^l\right)^T$ is the transpose of the key vector for atom j.

For atom representation, the transformer block incorporates the 3D representation as a bias in the attention function. Each attention head uses weights $\boldsymbol{w}^{q} \in \boldsymbol{R}^{D}, \boldsymbol{w}^{k} \in \boldsymbol{R}^{D}$ and $\boldsymbol{w}^{v} \in \boldsymbol{R}^{D}$ to transform atom representation into queries $\boldsymbol{Q}$ and keys $\boldsymbol{K}$ and values $\boldsymbol{V}$. The self-attention with pair-to-atom communication is formulated as follows:

$$Attention\left(\boldsymbol{Q}_i^l, \boldsymbol{K}_j^l, \boldsymbol{V}_j^l\right) = softmax\left(\frac{\boldsymbol{Q}_i^l\left(\boldsymbol{K}_j^l\right)^T}{\sqrt{d}} + \boldsymbol{q}_{ij}^{l-1}\right)\boldsymbol{V}_j^l,$$

where $\boldsymbol{V}_j^l$ is the value vector of atom j at layer l. The FFN module within each block comprises two linear layers with a hidden dimension of 2048, separated by a ReLU activation function. Layer normalization is applied both before and after the FFN to maintain stable training dynamics.

The molecular module effectively captures the spatial relationships between atoms based on their 3D coordinates. To add explicit information about the molecule's covalent bond structure to this spatial perspective, we subsequently employ a graph neural network (GNN), using the atom and pair representations generated by the molecular module as the initial node and edge embeddings for the molecular graph. These embeddings are then further refined by the GNN, which propagates information primarily along the chemical bonds.

We employ a Message Passing Neural Network (MPNN)[64] framework to generalize spatial convolution within the GNNs. For a molecular graph G with node embedding x_v and edge embedding e_{vm}, the MPNN operates in two phases: message passing and readout. During the message passing phase over T iterations, each node v updates its hidden representation (h_v^t) and message (m_v^t) as follows:

$$m_v^{t+1} = \sum_{w\in N(V)} M_t(h_v^t, h_w^t, e_{vw}),$$

$$h_v^{t+1} = U_t(h_v^t, m_v^{t+1}),$$

where $N(v)$ denotes the neighbors of node v. The readout phase computes a feature vector for the entire graph using a function R:

$$\hat{y} = R(\{h_v^T | v \in G\}).$$

After thoroughly updating the node and edge embeddings, we adopt a weighted Graph Convolutional Network (GCN)[32] as the application module to integrate them efficiently. Specifically, we transform the edge embeddings into scalar edge weights w_{vw} by applying an average pooling operation to each edge embedding e_{vw}. The message function is defined as $\mathrm{M}(\mathrm{h_v}, \mathrm{h_w}, \mathrm{e_{vw}}) = \mathrm{w_{vw}} \cdot \mathrm{h_w}$, where · denotes scalar multiplication. The vertex update function is $U_t(h_v^t, m_v^{t+1}) = \sigma\left(H_t^{\deg(v)} m_v^{t+1}\right)$, where σ is the sigmoid activate function, deg(v) is the degree of node v, and $\mathrm{H_t^N}$ is a learned matrix for each time step t and vertex degree N. The readout phase R incorporates skip connections to all previous hidden states and is defined as $R = f\left(\sum_{v,t} \mathrm{softmax}(W_t h_v^t)\right)$, where f is a neural network, and $\mathrm{W_t}$ are learned matrices for each timestep t.

After processing through the molecular and application modules, the final embedding is transformed by a multilayer perceptron (MLP), comprising two linear layers with a dropout rate of 0.1 and employing the hyperbolic tangent activation function. This MLP converts the final embedding into the probability of RNA binding.

Model training and evaluation. Our full dataset comprised 994 RNA-binders (positives) and 3,767 protein-binders (negatives). For model evaluation and hyperparameter selection, this dataset was randomly partitioned into a training set (80%) and a benchmark set (20%, consisting of 192 RNA-binders and 761 protein-binders). RiboBIND utilized weights initially pre-trained on a corpus of over 200 million small-molecule conformations via the Uni-Mol platform. This extensive pre-training was crucial for encoding the molecular knowledge essential for various predictive tasks. To ensure the robustness of our results, we performed five-fold cross-validation, where the training dataset was split into five parts, using four for training and one for validation in each fold. The performance metrics reported in the Results section (AUROC and AUPRC on the benchmark set) are the mean values obtained across these five folds. Fine-tuning was subsequently conducted on 4 A100 GPUs, each with 40 GB of memory, over the course of 50 epochs. We employed binary cross-entropy as the loss function, with a learning rate at 0.0001 and 5 epochs of warm-up. The batch size was set to 32. The Adam optimizer was employed, with default hyperparameter setting of $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 1 \times 10^{-6}$. Gradient clipping was applied with a maximum global norm of 1.0 to enhance training stability.

Application of the pre-trained model to construct CSTAR library. Upon identifying the optimal hyperparameters through cross-validation, we trained a single, final RiboBIND model on our entire dataset (all 994 positives and 3,767 negatives) using these parameters. This strategy ensures that the final model leverages information from all available data, maximizing its predictive power for the large-scale screening.This comprehensively trained model was subsequently applied to a collection of over 13 million in-stock small molecules to generate the CSTAR library comprising over 5 million. The inference process was executed on 4 A100

GPUs, each with 40 GB of memory, and was completed in 45 hours. The results of this large-scale prediction are accessible through our dedicated website.

SHAP analysis

10,000 compounds were randomly sampled from the CSTAR RNA-binding probability scores ranged from 0.5 to 1 (Supplementary Data). A total of 2,033 chemical descriptors of these compounds were generated using RDkit v2023.3.1 and Mordred v1.2.0[65]. Descriptors with more than 10% missing values, all 0s or overlapping between Mordred and RDKit were mamually excluded, resulting in 1467 final descriptors subjected to random forest regression model from Scikit-learn v1.5.1 to simulate the RiboBIND model, with all remaining descriptors serving as the feature data and predicted RNA-binding probability as the target variable. The TreeExplainer from SHAP v0.46.0 was used to assess the impact of different descriptors on the model output.

The 20 descriptors with highest SHAP impact values on the model output were divided into four major catagrories, including polarity (MinEStateIndex, BCUTd-1l, GATS2pe), aromaticity (n6HRing, piPC10, n6aHRing, fr_pyridine, GATS2are, SdsCH, n6Ring, SaaN), Van der Waals surface area (PEOE_VSA8, SlogP_VSA3, SMR_VSA5, Estate_VSA2, SlogP_VSA10, PEOE_VSA6) and topology (ATSC3s, Xc-5dv, AATSC0s). Five explainable representative descriptors, MinEstateIndex (Minimum E-state value), n6HRing (Number of 6-membered heterorings), PEOE_VSA8 (Van der Waals surface area with partial charges range from 0.20–0.25), Xc-5dv(5th-order valence connectivity index), and AATSC0s(Average E-state value for carbon atoms at zero topological distance) were selected for comparisons among four datasets. The datasets included 5,000 randomly sampled compounds from the CSTAR library (red), 994 compounds from the positive dataset (green), 3,767 compounds from the negative dataset (blue), and 1,615 FDA-approved drugs (yellow) (Supplementary Data 1). The whiskers in the boxplots represent 10-90% of data with outliers excluded, while the boxes encompass the middle 50% of the data. The black lines and × symbols denote the medians and means, respectively. Mann-Whitney U test were performed between the CSTAR representative set and the other three datasets.

Molecular docking

The CSTAR library was subjected to molecular docking against the *Tetrahymena* ScaI L-16 ribozyme using DOCK6 [66]. In brief, all CSTAR compounds were transformed into Mol2 format and processed in Open Babel 3.1.1 to retain only the largest contiguous fragment[67]. The *Tetrahymena* ScaI L-16 ribozyme model (PDB 7EZ0) was modified to remove the single-stranded 5' end (nts 22-27), P6 (nts 227-248) and P9.2 (nts 372-399) to allow grid generation. The hydrogen atoms were added and AM1-BCC charges were assigned for the modified RNA model, whose surface was used to generate sphere clusters. These sphere clusters were visually inspected to eliminate potential non-specific docking regions. The resulting sphere cluster created a grid with dimensions of 99 Å, 87 Å, 93 Å for molecular docking in DOCK6 with all other parameters set to default. A total of 8,392 diversified CSTAR library compounds with <0.25 similarity were docked to identify sphere clusters of high quality binding regions. The selected sphere clusters generasted a smaller grid with dimensions of 81 Å, 78 Å, 91 Å, which was then used for the entire CSTAR library. The structures of small molecules in complexes with L-16 were visually analyzed in Chimera to identify binding pockets [68].

The top 10,000 small molecules were rescored and ranked using rDock and validated by RNAmigos2. For rDock, two-sphere method were used to create cavity with center coordinate (84, 82, 86) and 49 Å radius. Then, the default parameter were used for the docking step. Each molecule were docked 50 times, and the pose of the lowest score were extracted and used for ranking. As for RNAmigos2, any residue within 3.5 Å range of the diversified CSTAR compounds were identified as binding sites, and binding score was predicted and rancked.

Finally, combining the result from DOCK6, rDock and RNAmigos2, the top 125 compounds were selected and purchased (from TargetMol), and their quantitative estimation of drug-likeness (QED) was evaluated[69] (Supplementary Data 2).

RNA preparation

The L-16 ribozyme and *Cal*-nuLSU intron RNA were prepared as previously described[19]. In brief, the DNA templates in pUC-19 plasmids were amplified by PCR with a forward primer (5'-TAATACGACTCACTATAGGTTTGGAGGGAAAGTTATCA-3') and a 2'-O-methylated reverse primer (5'-/i2OMeC/i2OMeT/CCAAAACTAATCAATATACTTT-3') for L-16 ribozyme, and a forward primer (5'-TAATACGACTCACTATAGGAAGGGAGGCAAAAG TAGGG-3') and a 2'-O-methylated reverse primer (5'-/i2OMeA/i2OMeC/TCCAAAACTAATC AATATACTTT-3') for *Cal*-nuLSU intron [70]. RNAs were then synthesized through *in vitro* transcription in reactions containing 0.2 µM DNA templates, 40 mM Tris-HCl, pH 8.1, 25 mM $MgCl_2$, 3.5 mM spermidine, 0.01% Triton X-100, 40 mM DTT, 4% PEG 8000, 3 mM NTPs, and 7.5 U/µl T7 RNA polymerase (New England Biolabs). The transcription reactions were incubated at 37 °C for 3 hours, after which the RNA products were isolated by ethanol precipitation, then mixed with the loading buffer containing 95% formamide, 10 mM EDTA, 0.1% xylene cyanol, and 0.1% bromophenol blue. The RNA samples were then purified on 8% 29:1 acrylamide:bis, 7 M urea polyacrylamide gels. The gels were run at 25 W for 2 hours, then visualized briefly with a 254-nm UV lamp. The target RNAs were excised from the gels, cut into small pieces and eluted overnight at 4 °C in an RNase-free elution buffer containing 30 mM sodium acetate pH 5.2 and 1 mM EDTA. The gel slurry was then filtered through 0.45 µm filters (Minisart Syringe Filter, Sartorius), concentrated using a 30 kDa cutoff filter (Ultrafiltration Centrifugal Tube, Millipore), and subjected to ethanol precipitation. The resulting pellets were washed with isopropyl alcohol, air-dried, dissolved in RNase-free water for quantification using a NanoDrop spectrophotometer (Thermo Scientific), and stored at -80 °C for future use.

RNA biotinylation

RNA biotinylation was performed following previously reported methods[71]. A total of 1 nmol of RNA was added to 10 µL of 0.5 M $NaIO_4$ and 10 µL of 0.5 M NaOAc (pH 5.2), and the mixture was diluted to 50 µL with RNase-free water. The reaction was incubated at room temperature, protected from light, for 1.5 hours. After incubation, the reaction was terminated, and the oxidized RNA at the 3' end was recovered through isopropanol precipitation. The resulting pellet was washed with isopropyl alcohol, air-dried, and dissolved in 40 µL of RNase-free water. Next, 10 µL of 50 mM biotin hydrazide was added, and the mixture was incubated at room temperature, protected from light, for 12 hours. Following incubation, phenol extraction was performed, and biotin-labeled RNA was recovered by isopropanol precipitation. The pellet was washed with isopropyl alcohol, air-dried, dissolved in RNase-free water, quantified using a NanoDrop spectrophotometer (Thermo Scientific), and stored at -80 °C for future use.

Biotinylation efficiency was analyzed using gel shift assays with excess streptavidin (Sigma-Aldrich). A total of 20 ng of biotinylated RNA was incubated with up to 1 µg of streptavidin for 15 minutes in a buffer containing 10 mM Tris (pH 7.4), 2.5 mM $MgCl_2$, and 100 mM NaCl. The mixture was then combined with loading buffer containing 250 mM Na-HEPES (pH 7.5), 5 mM EDTA pH 8.0, 50% glycerol, 0.1% xylene cyanol, and 0.1% bromophenol blue. Samples were loaded onto a 6% 29:1 acrylamide:bis polyacrylamide gel, which was run at 10 W for 1 hour. The gel was stained for 5 minutes with SYBR-Gold (Invitrogen) and visualized using a Bio-Rad ChemiDoc XRS+ ultraviolet transillumination molecular imager. Biotinylation efficiency was calculated by comparing the band intensity of the biotinylated RNA-streptavidin complex to that of the unmodified RNA.

Biolayer interferometry (BLI)

All BLI binding assays were performed using the Octet Red96e system. Before the experiments, biotinylated L-16 was refolded as described above and loaded onto the surface of super streptavidin biosensors (SSA, Sartorius) for 600 seconds. Optimal response levels were maintained between 0.5 and 2 nm, with variability within a row of sensors not exceeding 0.2 nm. A typical analysis cycle included baseline acquisition in 1× kinetics buffer (60 seconds), association in wells containing the compound (90 seconds), and dissociation in 1× kinetics buffer (90 seconds). After each cycle, the biosensors were washed in 1× kinetics buffer for 120 s before proceeding to the next cycle. A parallel set of biosensors without RNA loading were subjected to the same procedure. Compounds were initially screened at concentrations of 500 and 125 µM.

The resulting curves were corrected by first subtracting the response recorded from a sensor without RNA, followed by subtracting the response from a sensor loaded with RNA but incubated without compound. Compounds exhibiting a corrected response of >0.01 nm at both concentrations were further subjected to dose-response assays, with concentrations ranging from 500 to 7.8 μM. Data analysis and curve fitting were performed using Octet Data Analysis HT, version 12.0. The experimental data were fitted using a 2:1 heterogeneous ligand (HL) model. A global analysis of all data for different compound concentrations was performed under the assumption of reversible binding, using nonlinear least squares fitting. K_D values were calculated via steady-state kinetic analysis of the estimated equilibrium response (*Req*) according to the following equation, where [*Compound*] represents the compound concentration, *Req* is the estimated equilibrium response, k_{on} is the association constant, and k_{off} is the dissociation constant:

$$y = R_{max} \frac{[Compound]}{[Compound] + K_D}$$

$$R_{max} = R_{eq} \frac{k_{on} \times [Compound]}{k_{on} \times [Compound] + k_{off}}$$

FRET experiment

The Cy3 and FAM labeled RNA oligonucleotide substrate 5'-Cy3-CCCUCUAAACC-FAM-3' were purchased from Sangon Biotech. The FRET screening assay was performed using 384-well black polypropylene plates (Corning, part number 3677) in a 10 μL reaction volume containing 50 mM Tris-HCl pH 7.5, 100 mM $(NH_4)_2SO_4$, and 5 mM $MgCl_2$. Prior to screening, RNAs were denatured at 90°C for 3 minutes in 50 mM Na-HEPES pH 8.0, cooled to room temperature for 10 minutes, followed by the addition of $MgCl_2$ to a final concentration of 10 mM. The sample was incubated at 50°C for 30 minutes and cooled to 25°C for 10 minutes.

Reaction mixtures containing the buffer and refolded RNA were transferred to the assay plates, followed by the addition of 50 μM GTP. The reaction was initiated by adding the Cy3/FAM labeled substrate to a final concentration of 50/100 nM. Yeast tRNAs with final concentrations of 1 μM were added as needed. Plates were sealed with tape, protected from light, and incubated at 25°C for 15 minutes for the L-16 ribozyme reaction. After a brief centrifugation, fluorescence was measured. All fluorescence measurements were performed using a CLARIOstar Plus (BMG LABTECH) plate reader with an excitation filter of 482 nm (16 nm bandwidth) and an emission filter of 530 nm (20 nm bandwidth).

Small molecules were dissolved in DMSO and diluted to the desired concentrations (200 and 1000 μM) before being added to the reaction mixture, maintaining a final DMSO concentration of 5%. Fluorescence readings were taken prior to adding the substrate to assess compound fluorescence interference (I_{int}, representing interference intensity) and the plate fluorescence background (I_0, representing background intensity). After substrate addition, the mixture was incubated at room temperature for 15 minutes, and fluorescence was measured again following the same procedure (I_{cpd}, representing compound intensity). For the inhibition activity assay, the well without RNAs served as the positive control (I_{max}, representing 100% inhibition), while the well with RNAs served as the negative control (I_{min}, representing 0% inhibition). Both controls contained 5% DMSO. The normalized FAM intensity for each compound was calculated using the following formula:

$$Normalized\ FAM\ intensity = \frac{(I_{cpd} - I_{int}) - (I_{min} - I_0)}{(I_{max} - I_0) - (I_{min} - I_0)}$$

Compounds showing over 60% inhibitory activity against L-16 at 1000 μM concentration were further subjected to dose-response assays. Each compound was tested at six concentrations ranging from 1000 to 31 μM. The inhibition rate for each concentration was calculated using the following formula:

$$Inhibition(\%) = \frac{(I_{max} - I_0) - (I_{cpd} - I_{int})}{(I_{max} - I_0) - (I_{min} - I_0)} \times 100\%$$

IC_{50} values were calculated using dose-response inhibition analysis according to following equation, in which *Top* and *Bottom* are plateaus in the units of the Y axis, *HillSlope* describes the steepness of the family of curves.

$$y = Bottom + \frac{Top - Bottom}{1 + 10^{(log^{IC_{50}} - X) \times HillSlope}}$$

Cryo-EM sample preparation

For cryo-EM analysis, four samples were prepared: (i) 20 μM refolded L-16 with 500 μM ZPT-005, (ii) 1 μM refolded L-16 with 1 mM ZPT-084, and (iii) 2 μM refolded L-16 with 2 mM ZPT-084.10 each in 10 μL total volume. After incubating at room temperature for 15 min and cooling on ice, 3 μL aliquots were applied to glow-discharged (45 s) 200-mesh R2-1 Quantifoil Cu grids, blotted for 2.5 s at 4 °C (100% humidity), and vitrified in liquid ethane using a Vitrobot Mark IV (Thermo Fisher Scientific).

Cryo-EM data acquisition and processing

Frozen grids of L-16 in complex with ZPT-005 were loaded onto a Titan Krios G4 (Thermo Fisher) operated at 300kV, with a condenser lens aperture of 50 μm, spot size of 3, and a parallel beam with illuminated area of 0.63 μm in diameter. Microscope magnification was set to 165,000× (corresponding to a calibrated sampling of 0.74 Å per pixel). For L-16 in complex with ZPT-005, a total of 10,783 movie stacks were automatically collected using EPU software on a Falcon 4i direct electron camera equipped with a Seletrics X energy filter with an energy slit of

20 eV (Gatan), operating in counting mode at a recording rate of 10 raw frames per second and a total exposure time of 3 seconds, yielding 32 frames per stack, and a total dose of 48 $e^-/Å^2$. All movie stacks were collected with defocus values ranging from -0.7 to -1.1 µm. The movie stacks were motion-corrected using patch MotionCorr and CTF estimation via patch CTF in cryoSPARC[72], 10,751 micrographs were subjected to template picking with templates generated from the low-pass filtered (20 Å) holo L-16 map (EMD-31385), and 4,574,440 particles were extracted with a box size of 300 × 300 pixels and subsequently subjected to two rounds of 2D classification in cryoSPARC. Following 2D classification, 2,233,672 particles were selected for ab-initio 3D reconstruction and consequent heterogeneours refinement. The predominant class containing 1,118,181 particles (50.1% of selected particles) was subjected to iterative refinement, including reference-based motion correction (RBMC), homogeneous refinement and local resolution refinement, yielding a final map at 2.3 Å resolution. Visualization in Chimera showed extra density in site 1 for modeling of ZPT-005.

For L-16 in complexes with ZPT-084 and ZPT-084.10, frozen grids were loaded onto a Titan Krios G3i (Thermo Fisher) operated at 300kV, with a condenser lens aperture of 50 µm, spot size of 6, and a parallel beam with illuminated area of 1.03 µm in diameter. Microscope magnification was set to 165,000× (corresponding to a calibrated sampling of 0.85 Å per pixel). A total of 5,333 and 1,800 movie stacks were automatically collected for L-16-ZPT-084 and L-16-ZPT-084.10, using EPU software on a Gatan K2 direct electron camera equipped with a Bioquantum energy filter with an energy slit of 20 eV (Gatan), operating in counting mode at a recording rate of 5 raw frames per second and a total exposure time of 6 seconds, yielding 30 frames per stack, and a total dose of 62 $e^-/Å^2$. All movie stacks were collected with defocus values ranging from -1.2 to -2.0 µm. The movie stacks were motion-corrected using MotionCorr2[73]. After CTF correction via CTFFIND4[74], 5,283 micrographs of L-16-ZPT-084 of L-16-ZPT-084 were subjected to EMAN2 for neural network particle picking[75]. 3,686,007 particles were extracted with a box size of 256 × 256 pixels and were subjected to two rounds of 2D classification in Relion 4[76]. After 2D classification, 2,271,733 particles were selected for multi-reference 3D classification. The multi-reference 3D classification was initiated in Relion using low-pass filtered holo L-16 maps at 20, 30, 40, 60, 80, and 100 Å as references. Following 3D classification, two classes exhibited distinct RNA features, containing 989,224 and 683,172 particles, respectively, corresponding to the N and I conformations of L-16. For the N conformation, homogeneous and local refinement was performed in cryoSPARC to yield a density map at 2.9 Å resolution. For the I conformation, focused 3D classification was applied to the P9 and P9.2 regions in cryoSPARC, and the class of 154,740 particles with the most complete density map was selected for homogeneous and local refinement, resulting in a density map at 3.8 Å resolution. 1,781 micrographs of L-16-ZPT-084.1 were subjected to the same procedure, resulting in 565,713 particles of only I conformation at 4.3 Å resolution.

The N conformation was visualized in Chimera, with the apo N model (PDB 7EZ0) docked into the density[19]. Subtraction of the N density in the presence of ZPT-084 by the apo N density revealed additional densities in the core region and P9.2. Another round of 3D classification and subsequent refinement in cryoSPARC resulted in five reconstructions with additional densities in the P9.2 region, ranging from 3.0 to 3.3 Å resolution, accounting for 815,495 particles. Refinement of the remaining major class containing 173,729 particles yielded a cryo-EM density map at 3.0 Å resolution with additional densities in the core region. Another round of 3D classification focused on this core region resulted in four classes with individually separated additional densities. The major class containing 60,100 particles was subjected to local refinement, yielding a final reconstruction at 3.0 Å resolution. Visualization in Chimera showed extra bending density in site 2 for modeling of ZPT-084. The I conformation was processed under the same procedure, however, no apparent density corresponding to ZPT-084 was observed.

Cryo-EM model building, refinement and analysis

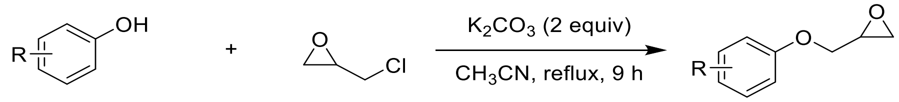
As the ZPT-005-bound N density map at 2.3 Å resolution reveals metal ion and water densities, the four recent models containing both concensus and non-concensus metal ion and water molecules (PDB 9CBU, 9CBW, 9CBX and 9CBY) were combined and docked in the ZPT-084-bound N density map[25], and manually adjustmented with removals of inappropriate metal ions and water molecules in Coot[77]. ZPT-005 and ZPT-084 were docked into the additional densities of their corresponding cryo-EM maps using EMERALD[78]. In brief, the protonation states of ZPT-005 and ZPT-084 were calculated using phenix.elbow, and partial charges were added via Chimera v1.16. Parameter files for ZPT-005 and ZPT-084 were generated with Rosetta[79], and used as input files along with the apo N model and small molecule-bound N density maps to run the EMERALD docking program.

The apo N model was docked in the cryo-EM map of I conformation and P7, P9, P9.2-P9.1-P13 were manually docked into the corresponding density in Chimera, followed by manual adjustments in Coot. All fragments were assembled by rigid-body docking in UCSF Chimera, and the full model was further refined manually in Coot to optimize geometry and correct local conformations.

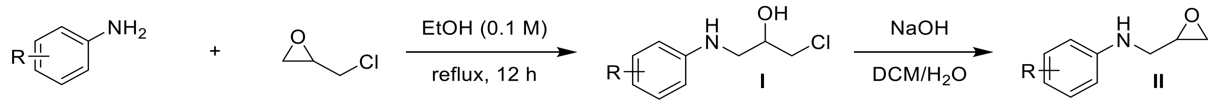
The complete models were refined with Phenix.real_space_refine[80], achieving average model-map correlation coefficiency (CC_{mask}) of 0.72 for ZPT-005-bound model, 0.73 for ZPT-084-bound model, and 0.67 for I model, respectively. The final models were validated with MolProbity and Q-score analysis[81,82]. ZPT-005 and ZPT-084 interactions were analyzed by Discovery Studio v21.1.0 assisted by visualizations in Chimera.

General Synthetic Procedures of ZPT-084 analogues

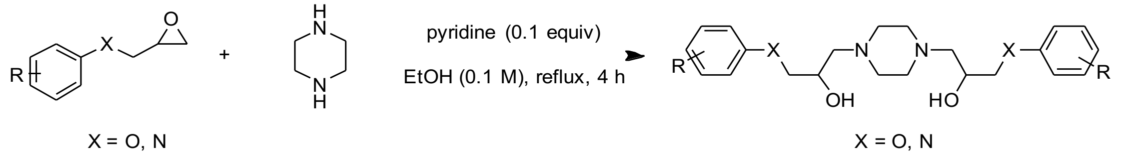
General procedure 1: synthesis of glycidyl ethers. To a magnetically stirred solution of K_2CO_3 (2 equiv) and aryl alcohol (1 equiv) in anhydrous CH_3CN (0.1 M) was added epichlorohydrin (1 equiv), and the mixture was stirred magnetically at 70 °C. After completion of the reaction (9 h, monitored by TLC), the reaction mixture was cooled to room temperature (RT), filtered, and the solvent was evaporated under vacuum. The crude residue was purified by flash column chromatography (hexane/EtOAc) to afford the glycidyl ethers[83].



General procedure 2: synthesis of epoxy amines. To a 100 mL round bottom flask was added aniline (1 equiv), epichlorohydrin (1.1 equiv) and ethanol (0.1 M). The mixture was heated to 80 °C and stirred for 12 h. After removing solvents, the intermediate **I** was isolated and purified by flash column chromatography. **I** was then dissolved in DCM and 0.3 M aqueous sodium hydroxide was added. The two-phase mixture was vigorously stirred overnight. The organic layer was washed twice with water and dried over $MgSO_4$. The organic layer was filtered through Celite and evaporated to give product **II**[84].



General procedure 3: synthesis of *β*-amino alcohol derivatives. A stirred solution of glycidyl ether (2.1-2.5 equiv), piperazine (1 equiv), and catalytic amount of pyridine (0.1 equiv) in ethanol (0.1 M) was heated at reflux (78 °C) for 4 h. The reaction mixture was concentrated under vacuum. The crude residue was purified by flash column chromatography (DCM/MeOH).



General procedure 4: synthesis of *β*-amino ester. In a round-bottom flask equipped with a PTFE-coated stirring bar, the appropriate carboxylic acid (1 equiv), EDCI·HCl (1.5 equiv), the β-amino alcohol (1.1equiv.), DMAP (1.2 equiv.) were dissolved in DCM (1.0 M. The reaction was stirred at RT overnight, then poured in water (approx. 5 mL/mmol). The layers were separated, then the aqueous layer was extracted once with CH_2Cl_2 (approx. 5 mL/mmol), then the combined organic layers were washed once with distilled water (approx. 5 mL/mmol), dried over $MgSO_4$ and the solvent was removed under vacuum. The crude residue was purified by flash column chromatography (DCM/MeOH) to afford the desired ester[85].

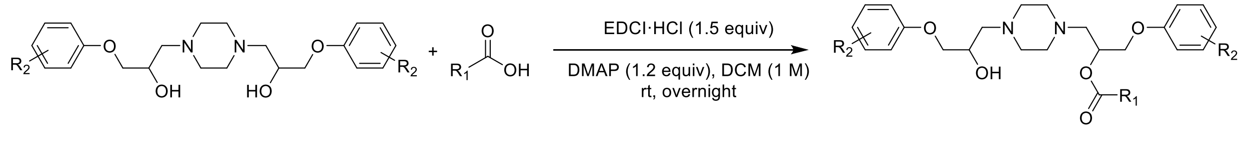


Figure preparation

Figures and illustrations were prepared using Chimera v1.16, ChimeraX v1.6, GraphPad Prism 10.0. and Adobe Illustrator.

Data availability

The cryo-EM maps and associated atomic coordinate models of *Tetrahymena* ScaI L-16 ribozyme in complexes with ZPT-005 and ZPT-084 have been deposited in the wwPDB OneDep System under EMD accession and PDB ID codes of EMD-64492 and 9UTI for the ZPT-005-bound N conformation, EMD-61283 and 9J9X for the ZPT-084-bound N conformation, and EMD-61359 and 9JC8 for the I conformation.

Code availability

The RiboBIND model is a Graph Isomorphism Network (GIN) implemented in PyTorch. Given a small molecule SMILES string, the model: (1) converts the molecule to a graph with atom and bond features using RDKit; (2) passes atom features through 3 GIN layers with edge MLP; (3) aggregates node embeddings via sum pooling; (4) passes the graph embedding through a 2-layer MLP classifier with ReLU and dropout (p=0.3); and (5) applies sigmoid to output a binding probability score (threshold 0.5 for binary prediction). The model was trained on a dataset comprising 994 RNA-binders (positives) and 3,767 protein-binders (negatives), provided in Supplementary Data 1. The code is provided as Supplementary Software with this submission and will be released at https://github.com/BEAM-Labs/RiboBIND under an MIT license upon acceptance.

Acknowledgements

Cryo-EM data were collected on Can Cong at SKLB West China Cryo-EM Center and the Cryo-EM Center of the Zhongshan Institute for Drug Discovery (ZIDD). Cryo-EM data and virtual screening were processed at SKLB Duyu High Performance Computing Center of West China Hospital and High Performance Computing Center of Chengdu University. This work was supported by National Key Research and Development Program of China (2022YFC2303700), Natural Science Foundation of China (NSFC 32222040, 32472320 and 32022002). This work was partially supported by Shanghai Artificial Intelligence Laboratory (to S.S.).

Author contributions

Z.S. conceived the project; Z.P., H.W., H.H., Y.C., S.S. constructed and trained the RiboBIND model to curate the CSTAR library; Z.P., H.H., H.N. and K.Z. performed virtual screening; Z.P., J.Z. and J.Zh. prepared RNAs and performed FRET and RNA binding experiments; Z.P., H.M., J.Z. and Z.S. collected and processed cryo-EM data, and generated RNA atomic coordinates; Z.P., H.W., H.H., H.M., L.W., S.X., S.S. and Z.S. prepared figure illustrations. All authors contributed to the preparation of the manuscript.

Competing interests

A patent that includes the work described in this manuscript has been filed.

Main figure

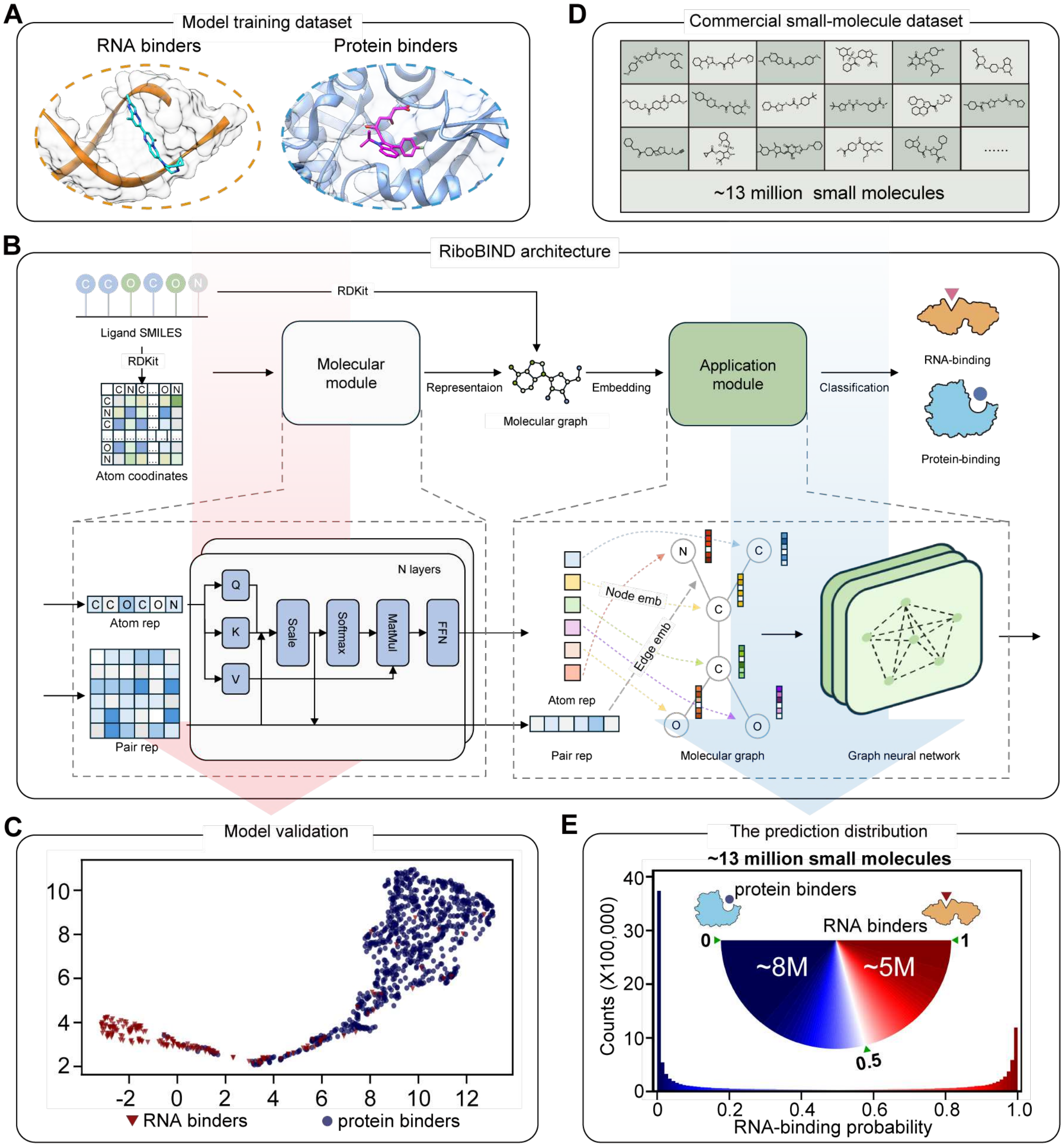


Fig. 1 | Curation of the CSTAR library using RiboBIND model. A, Model training dataset with RNA-binders as positives and protein-binders as negatives. **B,** The RiboBIND training scheme, in which the training dataset in SMILES format generates their 3D atomic coordinates and molecular graph, then proceed through molecular module and application module to train the corresponding atom and pair representations for RNA-binding preference. The final graph neural network generates output as RNA-binding probability for the input small molecules. **C,** A benchmark dataset kept blinded from the MRL training validates the model's capability of distinguishing RNA-binding from protein-binding compounds. **D,** A collection of about 13 million commercial small molecules was applied to the RiboBIND model. **E,** The model predicted more than 5.2 million commercial small molecules targeting RNA and this CSTAR library was deposited on a webserver for browse, search, prediction and download.

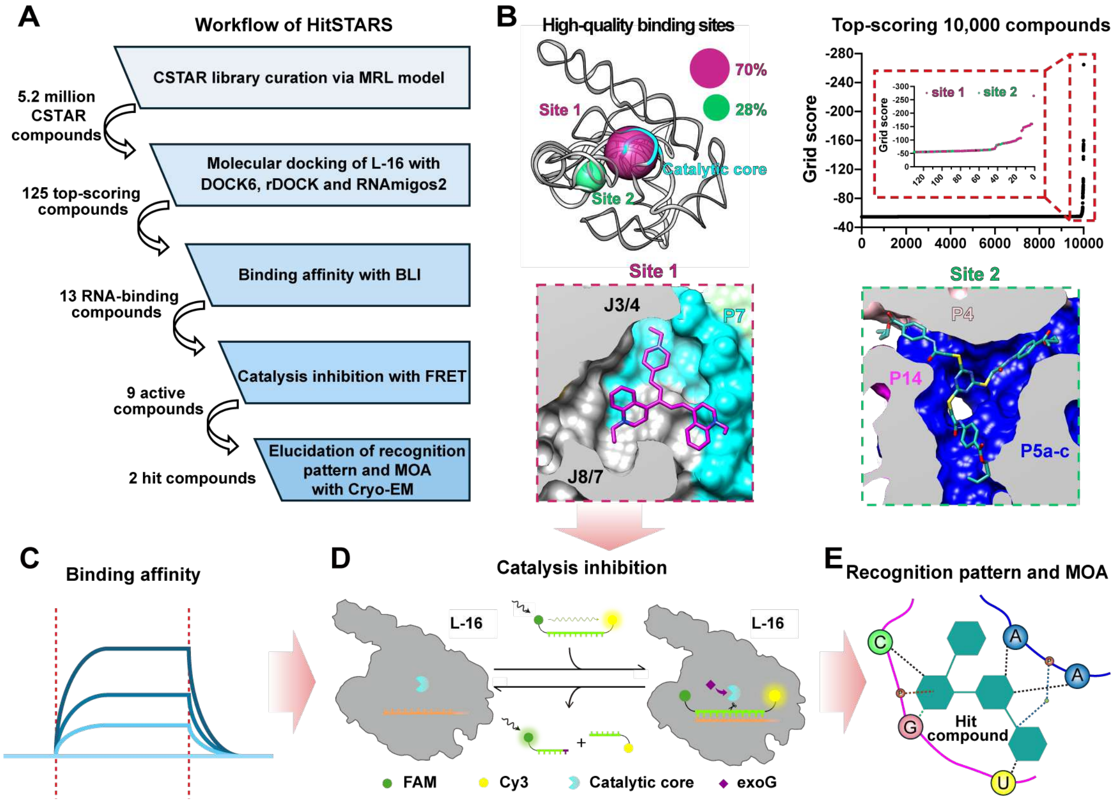
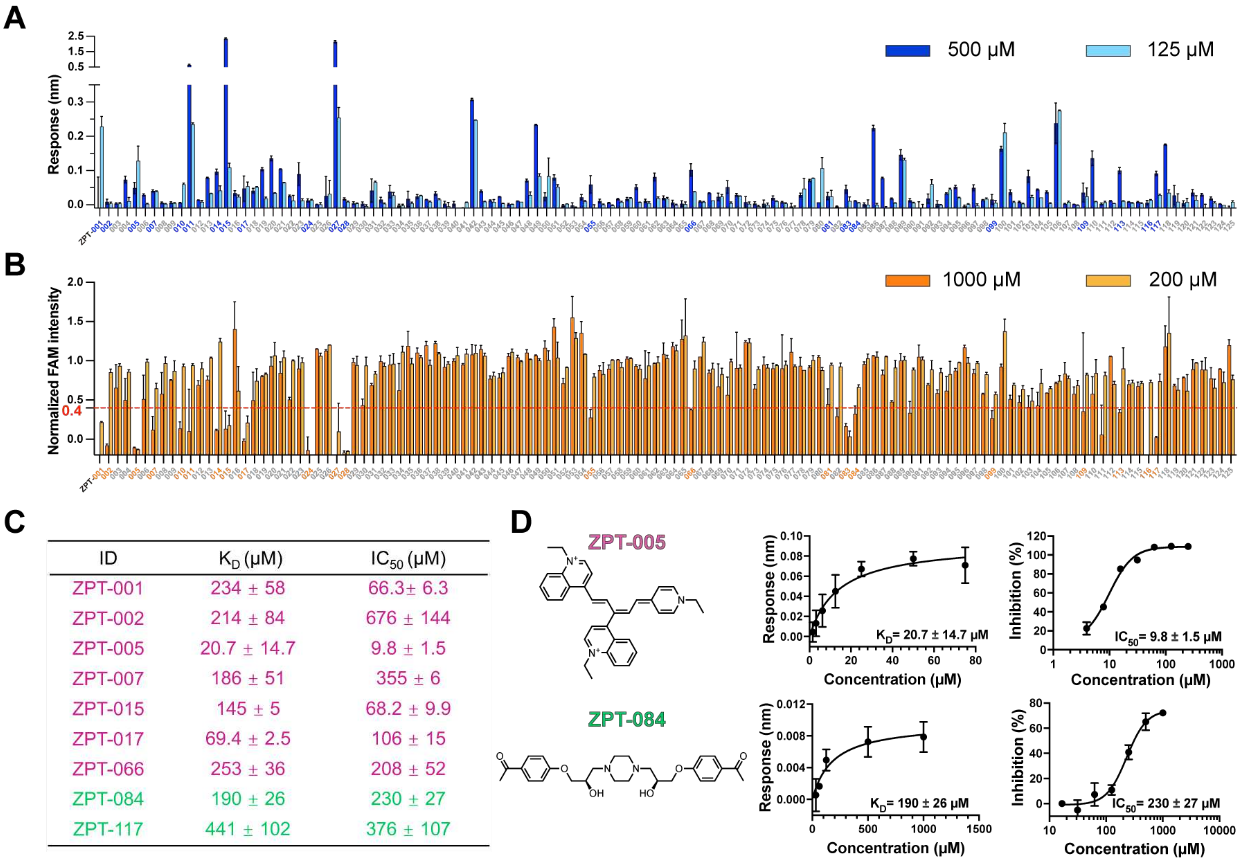


Fig. 2 | Schematic workflow of HitSTARS. A, HitSTARS workflow. **B,** Virtual screening revealed primary allosteric binding site 1 and site 2. **C-D,** Schematic drawing of functional screening to validate RNA-binding and catalysis inhibition. **E,** Cartoon illustration of the small-molecule recognition and MOA.



ID	K_D (μM)	IC_{50} (μM)
ZPT-001	234 ± 58	66.3± 6.3
ZPT-002	214 ± 84	676 ± 144
ZPT-005	20.7 ± 14.7	9.8 ± 1.5
ZPT-007	186 ± 51	355 ± 6
ZPT-015	145 ± 5	68.2 ± 9.9
ZPT-017	69.4 ± 2.5	106 ± 15
ZPT-066	253 ± 36	208 ± 52
ZPT-084	190 ± 26	230 ± 27
ZPT-117	441 ± 102	376 ± 107

Fig. 3 | Functional screening revealed hit compounds bound to L-16 and inhibited catalysis. **A,** All purchased compounds were screened for binding capacity using BLI at 500 and 125 μM concentrations, 23 compounds were subjected to K_D determination (labeled in blue). **B,** FRET screening of all compounds at 1,000 and 200 μM concentrations, 22 compounds were subjected to dose-response analysis to determine IC_{50} values (labeled in orange). **C,** K_D and IC_{50} values of nine hit compounds bound to site 1 (magenta) and site 2 (green). **D,** Structures, binding affinity and inhibition curves for ZPT-005 and ZPT-084. All data were presented as mean values ± standard deviation (SD) from two independent experiments.

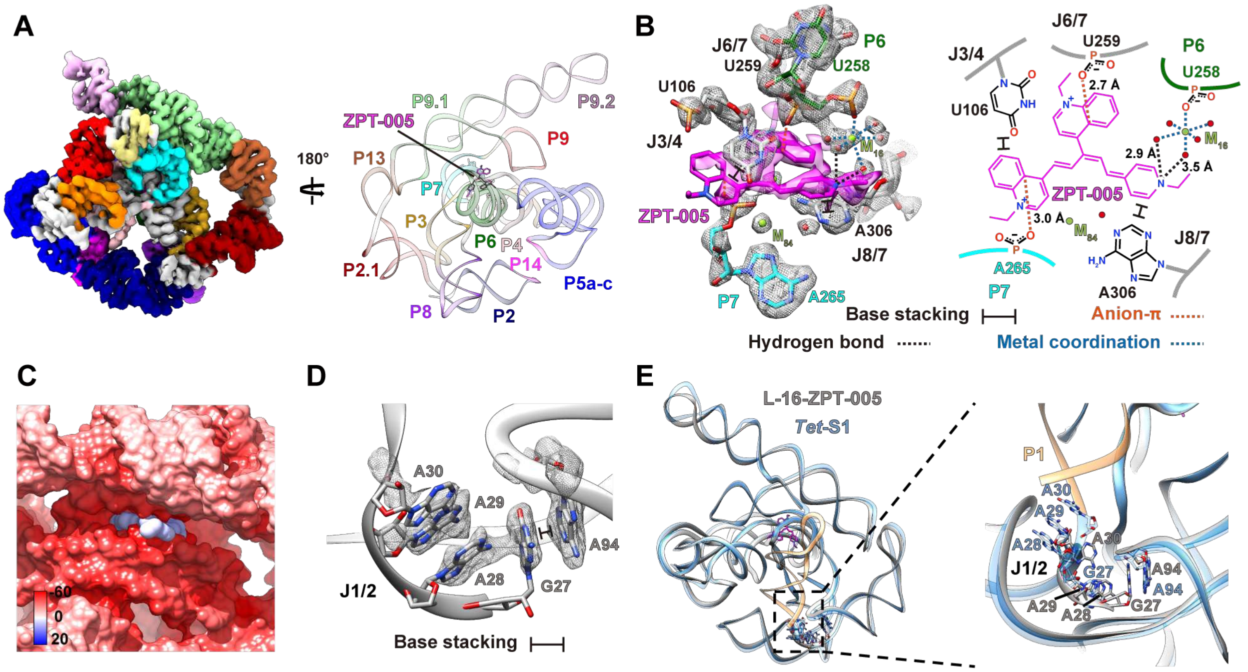


Fig. 4 | Cryo-EM structure of L-16 complexed with ZPT-005 revealed its recognition pattern and MOA. A, Cryo-EM map (left) and model (right) of L-16 complexed with ZPT-005 (magenta). **B,** Zoom-in view of the cryo-EM density and model (left) of ZPT-005 (magenta surface at 1 σ threshold) binding at site 1 (grey mesh at 3 σ threshold), and cartoon illustration of ZPT-005 binding mechanism (right). **C,** Electrostatic potential surface of ZPT-005 binding at site 1. **D,** Cryo-EM density and model of J1/2 in the presence of ZPT-005. **E,** Comparison of overall structures (left) of L-16-ZPT-005 (grey) and *Tet*-S1 (light blue, 8I7N) and zoom-in view of P1 (sandy brown) and J1/2 (right). Black labels indicate base stackings, red dashed lines indicate anion-π interactions, black dashed lines indicate hydrogen bonds, blue dashed lines indicate metal coordinations, green spheres indicate metal ions, and red spheres indicate water molecules.

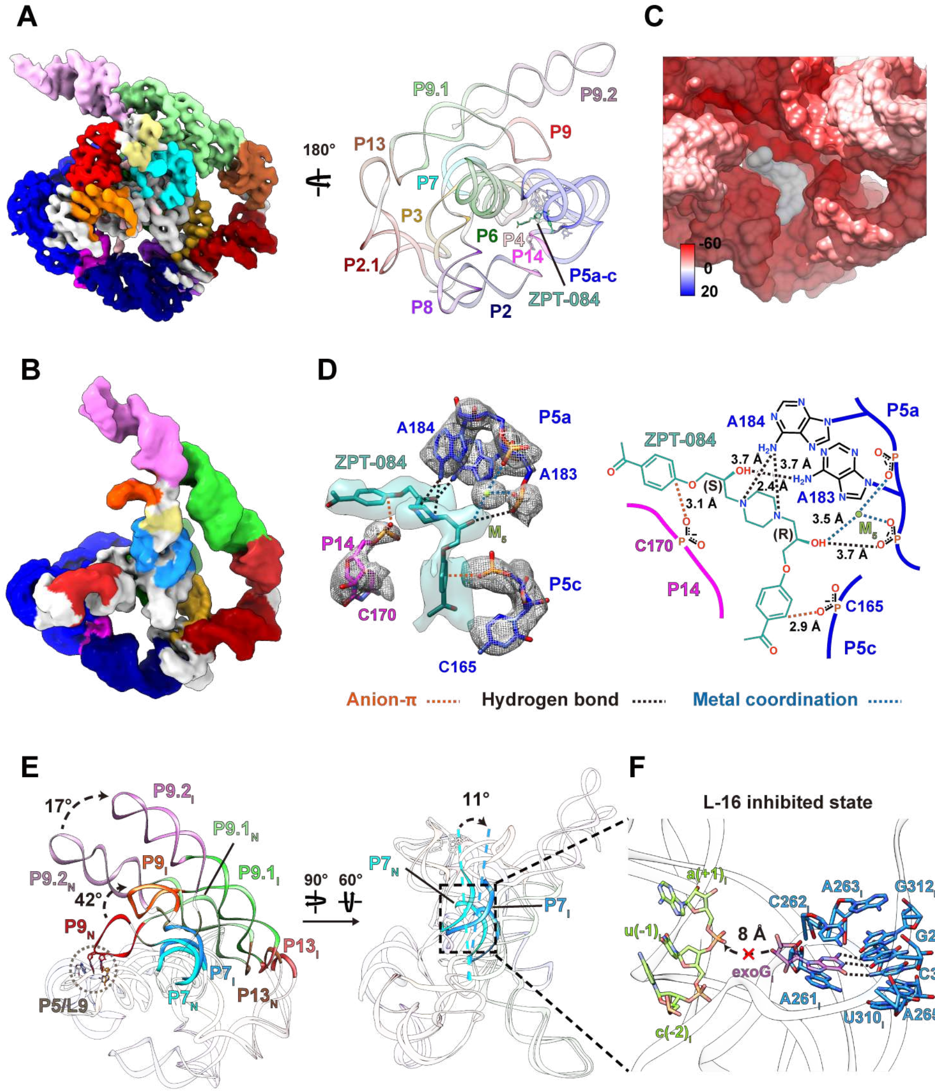


Fig. 5 | Cryo-EM structure of L-16 complexed with ZPT-084 revealed its recognition pattern. A, Cryo-EM map (left) and model (right) of L-16 complexed with ZPT-084 (light sea green) in N conformation. The map is colored according to the model. **B,** Cryo-EM map of the L-16 I conformation, colored according to the model. **C,** Electrostatic potential surface of ZPT-084 binding at site 2. **D,** Zoom-in view of the cryo-EM density and model (left) of ZPT-084 (light sea green surface at 1 σ threshold) binding at site 2 (grey mesh at 3 σ threshold), and cartoon illustration of ZPT-084 binding mechanism (right). **E,** Superposition of L-16 N and I conformations revealed large conformational changes of P7, P9 and P9.2-P9.1-P13. **F,** Zoom-in view of L-16 catalytic site in I conformation, in which the distance between putative exoG (transparent dark magenta) in P7 G-binding site (dodger blue) and 5'-SS (transparent chartreuse) was extended to 8 Å to inhibit L-16 catalysis. The putative P1 and exoG were modeled according to *Tet*-S1 (PDB 8I7N). Red dashed lines indicate anion-π interactions, black dashed lines indicate hydrogen bonds, blue dashed lines indicate metal coordinations, and green sphere indicates metal ion.

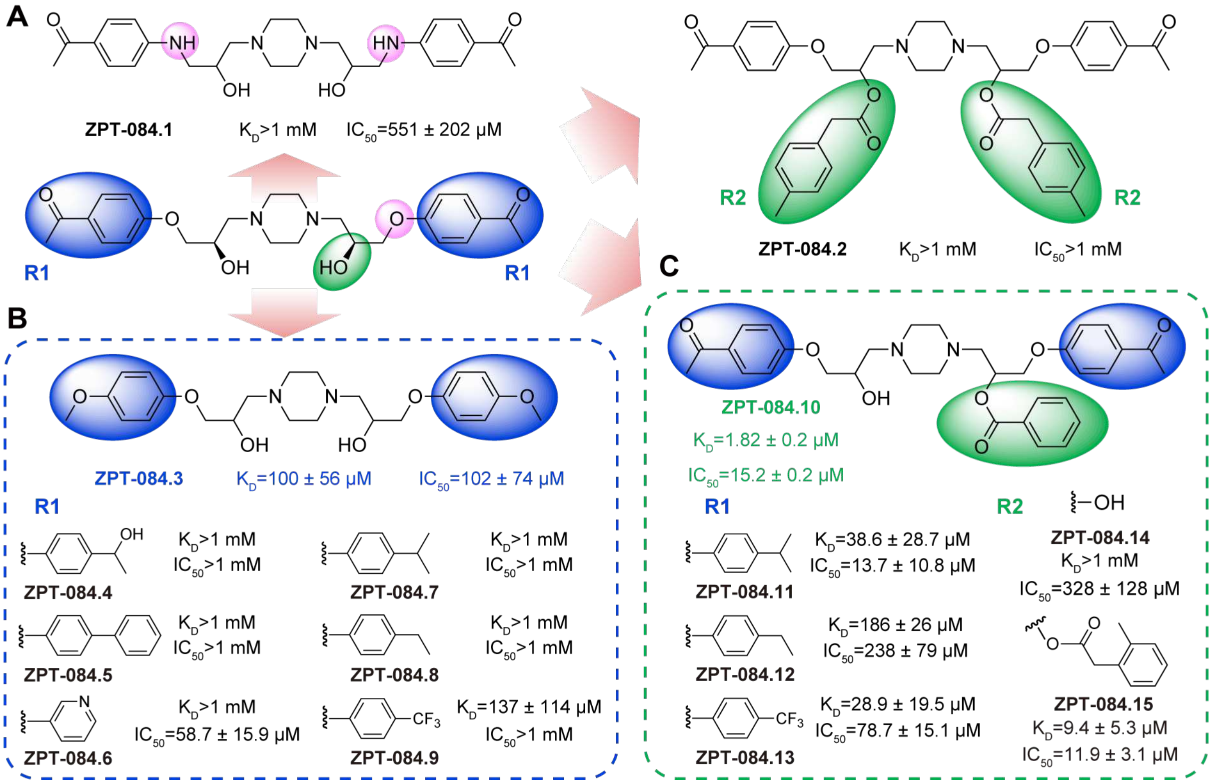
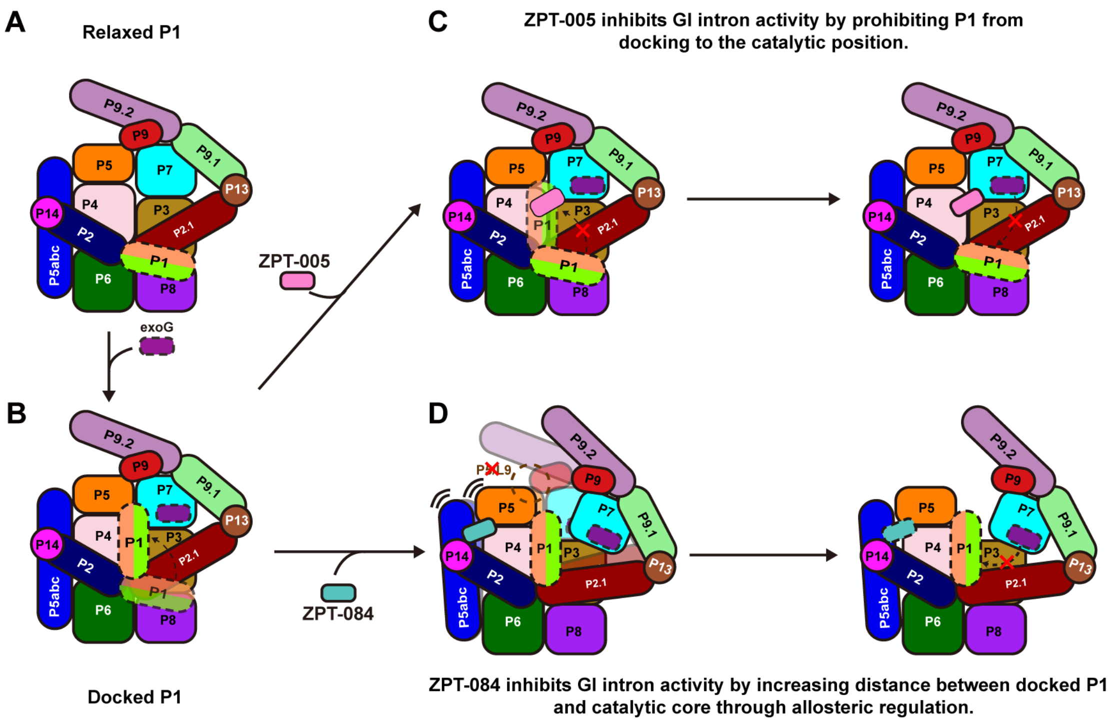


Fig. 6 | MOA of ZPT-70 and lead optimization to improve binding affinity and inhibition activity. A, Replacement of phenolyl to anilinyl moieties (pink circle) and acetylations of both R2 hydroxyl groups (green oval-shapes) deteriorated compound potency. **B,** Substitution of both R1 groups (blue oval-shapes) resulted in moderate potency enhancement by ~2-fold (ZPT-70.3). **C,** Benzoyl acylations and phenylacetylation of one R2 hydroxyl group (green oval-shape) facilitated maximal enhancement of the potency by ~20 fold (ZPT-70.15). All data were presented as mean values ± standard deviation (SD) from two independent experiments.


Relaxed P1

ZPT-005 inhibits GI intron activity by prohibiting P1 from docking to the catalytic position.

Docked P1

ZPT-084 inhibits GI intron activity by increasing distance between docked P1 and catalytic core through allosteric regulation.

Fig. 7 | Schematic illustration of ZPT-005 and ZPT-084 MOA. A-B, L-16 catalysis is initiated upon exoG binding and P1 translocation from the relaxed (A) to docked conformation (B). **C,** ZPT-005 binds to L-16 site 1 and stabilizes the J1/2 in P1 undocked conformation to inhibit catalysis. **D,** ZPT-084 binds to L-16 site 2 and disrupts P5/L9 tertiary interaction to induce large conformational changes to translocate the catalytic site away from the 5'-SS to inhibit catalysis.

References

1 Warner, K. D., Hajdin, C. E. & Weeks, K. M. Principles for targeting RNA with drug-like small molecules. *Nature Reviews Drug Discovery* **17**, 547-558, doi:10.1038/nrd.2018.93 (2018).

2 Cooper, T. A., Wan, L. & Dreyfuss, G. RNA and Disease. *Cell* **136**, 777-793, doi:10.1016/j.cell.2009.02.011 (2009).

3 Nemeth, K., Bayraktar, R., Ferracin, M. & Calin, G. A. Non-coding RNAs in disease: from mechanisms to therapeutics. *Nature Reviews Genetics* **25**, 211-232, doi:10.1038/s41576-023-00662-1 (2024).

4 Crooke, S. T., Baker, B. F., Crooke, R. M. & Liang, X.-h. Antisense technology: an overview and prospectus. *Nature Reviews Drug Discovery* **20**, 427-453, doi:10.1038/s41573-021-00162-z (2021).

5 Bennett, C. F. Therapeutic Antisense Oligonucleotides Are Coming of Age. *Annual Review of Medicine* **70**, 307-321, doi:https://doi.org/10.1146/annurev-med-041217-010829 (2019).

6 Childs-Disney, J. L. *et al.* Targeting RNA structures with small molecules. *Nature Reviews Drug Discovery* **21**, 736-762, doi:10.1038/s41573-022-00521-4 (2022).

7 Kovachka, S. *et al.* Small molecule approaches to targeting RNA. *Nature Reviews Chemistry* **8**, 120-135, doi:10.1038/s41570-023-00569-9 (2024).

8 Howe, J. A. *et al.* Selective small-molecule inhibition of an RNA structural element. *Nature* **526**, 672-677, doi:10.1038/nature15542 (2015).

9 Su, Z. *et al.* Discovery of a Biomarker and Lead Small Molecules to Target r(GGGGCC)-Associated Defects in c9FTD/ALS. *Neuron* **83**, 1043-1050, doi:10.1016/j.neuron.2014.07.041 (2014).

10 Naryshkin, N. A. *et al.* *SMN2* splicing modifiers improve motor function and longevity in mice with spinal muscular atrophy. *Science* **345**, 688-693, doi:doi:10.1126/science.1250127 (2014).

11 Palacino, J. *et al.* SMN2 splice modulators enhance U1–pre-mRNA association and rescue SMA mice. *Nature Chemical Biology* **11**, 511-517, doi:10.1038/nchembio.1837 (2015).

12 Aguilar, R. *et al.* Targeting Xist with compounds that disrupt RNA structure and X inactivation. *Nature* **604**, 160-166, doi:10.1038/s41586-022-04537-z (2022).

13 Tong, Y. *et al.* Programming inactive RNA-binding small molecules into bioactive degraders. *Nature* **618**, 169-179, doi:10.1038/s41586-023-06091-8 (2023).

14 Padroni, G., Patwardhan, N. N., Schapira, M. & Hargrove, A. E. Systematic analysis of the interactions driving small molecule–RNA recognition. *RSC Medicinal Chemistry* **11**, 802-813, doi:10.1039/D0MD00167H (2020).

15 Zhou, G. *et al.* in *The Eleventh International Conference on Learning Representations.*

16 Donlic, A. *et al.* R-BIND 2.0: An Updated Database of Bioactive RNA-Targeting Small Molecules and Associated RNA Secondary Structures. *ACS Chemical Biology* **17**, 1556-1566, doi:10.1021/acschembio.2c00224 (2022).

17 Ramaswamy Krishnan, S., Roy, A. & Michael Gromiha, M. R-SIM: A Database of Binding Affinities for RNA-small Molecule Interactions. *Journal of Molecular Biology* **435**, 167914, doi:https://doi.org/10.1016/j.jmb.2022.167914 (2023).

18 Tingle, B. I. *et al.* ZINC-22—A Free Multi-Billion-Scale Database of Tangible Compounds for Ligand Discovery. *Journal of Chemical Information and Modeling* **63**, 1166-1176, doi:10.1021/acs.jcim.2c01253 (2023).

19 Su, Z. *et al.* Cryo-EM structures of full-length *Tetrahymena* ribozyme at 3.1 Å resolution. *Nature* **596**, 603-607, doi:10.1038/s41586-021-03803-w (2021).

20 Cech, T. R. The RNA Worlds in Context. *Cold Spring Harbor Perspectives in Biology* **4**, doi:10.1101/cshperspect.a006742 (2012).

21 Zhang, K. *et al.* Cryo-EM structure of a 40 kDa SAM-IV riboswitch RNA at 3.7 Å resolution. *Nature Communications* **10**, 5511, doi:10.1038/s41467-019-13494-7 (2019).

22 Ma, H., Jia, X., Zhang, K. & Su, Z. Cryo-EM advances in RNA structure determination. *Signal Transduction and Targeted Therapy* **7**, 58, doi:10.1038/s41392-022-00916-0 (2022).

23 Luo, B. *et al.* Cryo-EM reveals dynamics of *Tetrahymena* group I intron self-splicing. *Nature Catalysis* **6**, 298-309, doi:10.1038/s41929-023-00934-3 (2023).

24 Haack, D. B. *et al.* Scaffold-enabled high-resolution cryo-EM structure determination of RNA. *Nature Communications* **16**, 880, doi:10.1038/s41467-024-55699-5 (2025).

25 Kretsch, R. C. *et al.* Complex water networks visualized by cryogenic electron microscopy of RNA. *Nature*, doi:10.1038/s41586-025-08855-w (2025).

26 Taghavi, A. *et al.* The evolution and application of RNA-focused small molecule libraries. *RSC Chemical Biology* **6**, 510-527, doi:10.1039/D4CB00272E (2025).

27 Yazdani, K. *et al.* Machine Learning Informs RNA-Binding Chemical Space. *Angewandte Chemie International Edition* **62**, e202211358, doi:https://doi.org/10.1002/anie.202211358 (2023).

28 Rizvi, N. F. *et al.* Targeting RNA with Small Molecules: Identification of Selective, RNA-Binding Small Molecules Occupying Drug-Like Chemical Space. *SLAS Discovery* **25**, 384-396, doi:https://doi.org/10.1177/2472555219885373 (2020).

29 Wicks, S. L., Morgan, B. S., Wilson, A. W. & Hargrove, A. E. Probing Bioactive Chemical Space to Discover RNA-Targeted Small Molecules. *bioRxiv*, 2023.2007.2031.551350, doi:10.1101/2023.07.31.551350 (2023).

30 consortium, w. Protein Data Bank: the single global archive for 3D macromolecular structure data. *Nucleic Acids Research* **47**, D520-D528, doi:10.1093/nar/gky949 (2018).

31 Shuaibi, M. *et al.* Rotation invariant graph neural networks using spin convolutions. *arXiv preprint arXiv:2106.09575* (2021).

32 Chen, M., Wei, Z., Huang, Z., Ding, B. & Li, Y. in *Proceedings of the 37th International Conference on Machine Learning* Vol. 119 (eds Daumé Hal, III & Singh Aarti) 1725--1735 (PMLR, Proceedings of Machine Learning Research, 2020).

33 Lundberg, S. M. & Lee, S.-I. in *Proceedings of the 31st International Conference on Neural Information Processing Systems* 4768–4777 (Curran Associates Inc., Long Beach, California, USA, 2017).

34 Irwin, J. J. & Shoichet, B. K. Docking Screens for Novel Ligands Conferring New Biology. *Journal of Medicinal Chemistry* **59**, 4103-4120, doi:10.1021/acs.jmedchem.5b02008 (2016).

35 Manigrasso, J., Marcia, M. & De Vivo, M. Computer-aided design of RNA-targeted small molecules: A growing need in drug discovery. *Chem* **7**, 2965-2988, doi:https://doi.org/10.1016/j.chempr.2021.05.021 (2021).

36 Carvajal-Patiño, J. G. *et al.* RNAmigos2: accelerated structure-based RNA virtual screening with deep graph learning. *Nature Communications* **16**, 2799, doi:10.1038/s41467-025-57852-0 (2025).

37 Hougland, J. L., Piccirilli, J. A., Forconi, M., Lee, J. & Herschlag, D. How the group I intron works: A case study of RNA structure and function. *Cold Spring Harbor Monograph Series* **43**, 133 (2006).

38 Chen, X. *et al.* RNA sample optimization for cryo-EM analysis. *Nature Protocols*, doi:10.1038/s41596-024-01072-1 (2024).

39 Allen, W. J. *et al.* DOCK 6: Impact of new features and current docking performance. *Journal of Computational Chemistry* **36**, 1132-1156, doi:https://doi.org/10.1002/jcc.23905 (2015).

40 Le Guilloux, V., Schmidtke, P. & Tuffery, P. Fpocket: An open source platform for ligand pocket detection. *BMC Bioinformatics* **10**, 168, doi:10.1186/1471-2105-10-168 (2009).

41 Veenbaas, S. D., Felder, S. & Weeks, K. M. fpocketR: A platform for identification and analysis of ligand-binding pockets in RNA. *bioRxiv*, 2025.2003.2025.645323, doi:10.1101/2025.03.25.645323 (2025).

42 Ruiz-Carmona, S. *et al.* rDock: A Fast, Versatile and Open Source Program for Docking Ligands to Proteins and Nucleic Acids. *PLOS Computational Biology* **10**, e1003571, doi:10.1371/journal.pcbi.1003571 (2014).

43 Blount, K., Puskarz, I., Penchovsky, R. & Breaker, R. Development and Application of a High-Throughput Assay for glmS Riboswitch Activators. *RNA Biology* **3**, 77-81, doi:10.4161/rna.3.2.3102 (2006).

44 Kuzniak-Glanowska, E., Glanowski, M., Kurczab, R., Bojarski, A. J. & Podgajny, R. Mining anion–aromatic interactions in the Protein Data Bank. *Chemical Science* **13**, 3984-3998, doi:10.1039/D2SC00763K (2022).

45 Cavalier, M. C., Kim, S.-G., Neau, D. & Lee, Y.-H. Molecular basis of the fructose-2,6-bisphosphatase reaction of PFKFB3: Transition state and the C-terminal function. *Proteins: Structure, Function, and Bioinformatics* **80**, 1143-1153, doi:https://doi.org/10.1002/prot.24015 (2012).

46 Esmaeeli, R., Piña, M. d. l. N., Frontera, A., Pérez, A. & Bauzá, A. Importance of Anion−π Interactions in RNA GAAA and GGAG Tetraloops: A Combined MD and QM Study. *Journal of Chemical Theory and Computation* **17**, 6624-6633, doi:10.1021/acs.jctc.1c00756 (2021).

47 D'Ascenzo, L., Leonarski, F., Vicens, Q. & Auffinger, P. Revisiting GNRA and UNCG folds: U-turns versus Z-turns in RNA hairpin loops. *RNA* **23**, 259-269, doi:10.1261/rna.059097.116 (2017).

48 Chawla, M., Kalra, K., Cao, Z., Cavallo, L. & Oliva, R. Occurrence and stability of anion–π interactions between phosphate and nucleobases in functional RNA molecules. *Nucleic Acids Research* **50**, 11455-11469, doi:10.1093/nar/gkac1081 (2022).

49 Stammers, T. A. *et al.* Discovery of a novel series of non-nucleoside thumb pocket 2 HCV NS5B polymerase inhibitors. *Bioorganic & Medicinal Chemistry Letters* **23**, 2585-2589, doi:https://doi.org/10.1016/j.bmcl.2013.02.110 (2013).

50 Benz-Moy, T. L. & Herschlag, D. Structure–Function Analysis from the Outside In: Long-Range Tertiary Contacts in RNA Exhibit Distinct Catalytic Roles. *Biochemistry* **50**, 8733-8755, doi:10.1021/bi2008245 (2011).

51 Shi, X. *et al.* Roles of Long-Range Tertiary Interactions in Limiting Dynamics of the *Tetrahymena* Group I Ribozyme. *Journal of the American Chemical Society* **136**, 6643-6648, doi:10.1021/ja413033d (2014).

52 Lipinski, C. A., Lombardo, F., Dominy, B. W. & Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Reviews* **23**, 3-25, doi:https://doi.org/10.1016/S0169-409X(96)00423-1 (1997).

53 Morgan, B. S., Forte, J. E., Culver, R. N., Zhang, Y. & Hargrove, A. E. Discovery of Key Physicochemical, Structural, and Spatial Properties of RNA-Targeted Bioactive Ligands. *Angewandte Chemie International Edition* **56**, 13498-13502, doi:https://doi.org/10.1002/anie.201707641 (2017).

54 Ganser, L. R., Kelly, M. L., Herschlag, D. & Al-Hashimi, H. M. The roles of structural dynamics in the cellular functions of RNAs. *Nature Reviews Molecular Cell Biology* **20**, 474-489, doi:10.1038/s41580-019-0136-0 (2019).

55 Kretsch, R. C. *et al.* Functional relevance of CASP16 nucleic acid predictions as evaluated by structure providers. *bioRxiv*, 2025.2004.2015.649049, doi:10.1101/2025.04.15.649049 (2025).

56 Ganser, L. R. *et al.* High-performance virtual screening by targeting a high-resolution RNA dynamic ensemble. *Nature Structural & Molecular Biology* **25**, 425-434, doi:10.1038/s41594-018-0062-4 (2018).

57 Ganser, L. R., Kelly, M. L., Patwardhan, N. N., Hargrove, A. E. & Al-Hashimi, H. M. Demonstration that Small Molecules can Bind and Stabilize Low-abundance Short-lived RNA Excited Conformational States. *Journal of Molecular Biology* **432**, 1297-1304, doi:https://doi.org/10.1016/j.jmb.2019.12.009 (2020).

58 Shibata, T. *et al.* Small molecule targeting r(UGGAA)n disrupts RNA foci and alleviates disease phenotype in Drosophila model. *Nature Communications* **12**, 236, doi:10.1038/s41467-020-20487-4 (2021).

59 Bass, B. L. & Cech, T. R. Ribozyme inhibitors. Deoxyguanosine and dideoxyguanosine are competitive inhibitors of self-splicing of the *Tetrahymena* ribosomal ribonucleic acid precursor. *Biochemistry* **25**, 4473-4478, doi:10.1021/bi00364a001 (1986).

60 Yarus, M., Illangesekare, M. & Christian, E. Selection of small molecules by the *Tetrahymena* catalytic center. *Nucleic Acids Research* **19**, 1297-1304, doi:10.1093/nar/19.6.1297 (1991).

61 Hanna, R. L., Gryaznov, S. M. & Doudna, J. A. A phosphoramidate substrate analog is a competitive inhibitor of the *Tetrahymena* group I ribozyme. *Chemistry & Biology* **7**, 845-854, doi:https://doi.org/10.1016/S1074-5521(00)00033-8 (2000).

62 Liu, T. *et al.* Molecular insights into de novo small-molecule recognition by an intron RNA structure. *Proceedings of the National Academy of Sciences* **122**, e2502425122, doi:doi:10.1073/pnas.2502425122 (2025).

63 Riniker, S. & Landrum, G. A. Better Informed Distance Geometry: Using What We Know To Improve Conformation Generation. *J Chem Inf Model* **55**, 2562-2574, doi:10.1021/acs.jcim.5b00654 (2015).

64 Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O. & Dahl, G. E. in *International Conference on Machine Learning.*

65 Moriwaki, H., Tian, Y.-S., Kawashita, N. & Takagi, T. Mordred: a molecular descriptor calculator. *Journal of Cheminformatics* **10**, 4, doi:10.1186/s13321-018-0258-y (2018).

66 Balius, T. E., Tan, Y. S. & Chakrabarti, M. DOCK 6: Incorporating hierarchical traversal through precomputed ligand conformations to enable large-scale docking. *Journal of Computational Chemistry* **45**, 47-63, doi:https://doi.org/10.1002/jcc.27218 (2024).

67 O'Boyle, N. M. *et al.* Open Babel: An open chemical toolbox. *Journal of Cheminformatics* **3**, 33, doi:10.1186/1758-2946-3-33 (2011).

68 Pettersen, E. F. *et al.* UCSF Chimera—A visualization system for exploratory research and analysis. *Journal of Computational Chemistry* **25**, 1605-1612, doi:https://doi.org/10.1002/jcc.20084 (2004).

69 Bickerton, G. R., Paolini, G. V., Besnard, J., Muresan, S. & Hopkins, A. L. Quantifying the chemical beauty of drugs. *Nature Chemistry* **4**, 90-98, doi:10.1038/nchem.1243 (2012).

70 Kao, C., Zheng, M. & Rüdisser, S. A simple and efficient method to reduce nontemplated nucleotide addition at the 3 terminus of RNAs transcribed by T7 RNA polymerase. *RNA* **5**, 1268-1272 (1999).

71 Willkomm, D. K. & Hartmann, R. K. in *Handbook of RNA Biochemistry* 86-94 (2005).

72 Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods* **14**, 290-296, doi:10.1038/nmeth.4169 (2017).

73 Zheng, S. Q. *et al.* MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nature Methods* **14**, 331-332, doi:10.1038/nmeth.4193 (2017).

74 Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *Journal of Structural Biology* **192**, 216-221, doi:https://doi.org/10.1016/j.jsb.2015.08.008 (2015).

75 Chen, M. *et al.* Convolutional neural networks for automated annotation of cellular cryo-electron tomograms. *Nature Methods* **14**, 983-985, doi:10.1038/nmeth.4405 (2017).

76 Kimanius, D., Dong, L., Sharov, G., Nakane, T. & Scheres, S. H. W. New tools for automated cryo-EM single-particle analysis in RELION-4.0. *Biochemical Journal* **478**, 4169-4185, doi:10.1042/bcj20210708 (2021).

77 Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallographica Section D* **66**, 486-501, doi:doi:10.1107/S0907444910007493 (2010).

78 Muenks, A., Zepeda, S., Zhou, G., Veesler, D. & DiMaio, F. Automatic and accurate ligand structure determination guided by cryo-electron microscopy maps. *Nature Communications* **14**, 1164, doi:10.1038/s41467-023-36732-5 (2023).

79 Combs, S. A. *et al.* Small-molecule ligand docking into comparative models with Rosetta. *Nature Protocols* **8**, 1277-1298, doi:10.1038/nprot.2013.074 (2013).

80 Afonine, P. V. *et al.* Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallographica Section D* **74**, 531-544, doi:doi:10.1107/S2059798318006551 (2018).

81 Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D* **66**, 12-21, doi:doi:10.1107/S0907444909042073 (2010).

82 Pintilie, G. *et al.* Measurement of atom resolvability in cryo-EM maps with Q-scores. *Nature Methods* **17**, 328-334, doi:10.1038/s41592-020-0731-1 (2020).

83 Seth, K., Roy, S. R., Pipaliya, B. V. & Chakraborti, A. K. Synergistic dual activation catalysis by palladium nanoparticles for epoxide ring opening with phenols. *Chemical Communications* **49**, 5886-5888, doi:10.1039/C3CC42507J (2013).

84 Lee, Y., Choi, J. & Kim, H. Stereocontrolled, Divergent, Al(lll)-Catalyzed Coupling of Chiral N-Aryl Epoxy Amines and CO(2). *Org Lett* **20**, 5036-5039, doi:10.1021/acs.orglett.8b02186 (2018).

85 Wang, H., Bellotti, P., Zhang, X., Paulisch, T. O. & Glorius, F. A base-controlled switch of SO2 reincorporation in photocatalyzed radical difunctionalization of alkenes. *Chem* **7**, 3412-3424, doi:https://doi.org/10.1016/j.chempr.2021.10.007 (2021).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- RiboBindv2.zip
- NstateZPT0059UTIvalreport.pdf
- NstateZPT0849J9Xvalreport.pdf
- Istate9JC8valreportfull.pdf
- SupplementaryData1.xlsx
- SIV3.pdf
- movieS1.mp4
- SupplementaryData2.xlsx
- machinelearningchecklist.pdf
- NstateZPT0059UTI2.txt
- nrsoftwarepolicy.pdf
- NstateZPT0849J9X.txt
- Istate9JC8.txt
- nrreportingsummary.pdf
- NstateZPT084EMD61283.mrc
- IstateEMD61359.mrc
- NstateZPT005EMD64492.mrc