



# RNA-Targeted Natural Product Discovery: Computational Rules for Molecular Recognition, Dynamics, Selectivity, and Validation

Lina Hassan<sup>1\*</sup>, Omar Khalaf<sup>2</sup>, Reem Jaber<sup>1</sup>, Rania Mostafa<sup>3</sup>

<sup>1</sup>Department of RNA-Targeted Natural Product Discovery, Faculty of Pharmacy, University of Jordan, Amman, Jordan.

<sup>2</sup>Department of Molecular Recognition and Structural Dynamics for RNA, Faculty of Pharmacy, Jordan University of Science and Technology, Irbid, Jordan.

<sup>3</sup>Department of Selectivity and Functional Validation for RNA Binders, Faculty of Pharmacy, Al-Balqa Applied University, Salt, Jordan.

## ABSTRACT

RNA is increasingly tractable to small-molecule intervention, yet computational discovery still often inherits assumptions developed for protein targets: a comparatively stable receptor structure, transferable pocket definitions, simplified treatment of solvent and ions, and ranking by a single docking or affinity score. These assumptions are especially consequential for natural products, whose stereochemical density, ionizable functionality, conformational complexity, and frequent polypharmacology can amplify uncertainty in RNA recognition. This methodological article argues that RNA-targeted natural-product discovery requires a distinct computational grammar. The analysis integrates evidence on RNA-specific docking, electrostatics and solvation, conformational ensembles, motif recognition, RNA-binding chemical space, multi-conformation scoring, cellular selectivity, target engagement, functional modulation, and assay artefacts. The proposed framework treats receptor state, hydration and ion occupancy, motif context, ligand chemical-space priors, multi-conformation recognition, off-target RNA competition, and binding-to-function validation as coupled but non-substitutable decision layers. Natural-product origin is not treated as evidence of RNA selectivity; instead, natural products are considered chemically distinctive candidates whose behavior must be tested against alternative RNA states, competing nucleic acids, and assay liabilities. The framework does not establish universal scoring rules, prospective superiority over existing workflows, or translational readiness. Its purpose is to make uncertainty explicit and to define where computational prioritization should be challenged by orthogonal structural, biophysical, cellular, and functional evidence before an RNA-targeted natural product is advanced.

**Key Words:** RNA-targeted discovery, Natural products, Molecular docking, RNA ensembles, Selectivity, Target engagement

eIJPPR 2026; 16(2):164-173

**HOW TO CITE THIS ARTICLE:** Hassan L, Khalaf O, Jaber R, Mostafa R. RNA-Targeted Natural Product Discovery: Computational Rules for Molecular Recognition, Dynamics, Selectivity, and Validation. Int J Pharm Phytopharmacol Res. 2026;16(2):164-73. <https://doi.org/10.51847/Al0uS6OvXe>

## INTRODUCTION

Small-molecule RNA discovery has progressed from isolated examples to a field containing experimentally resolved binding sites, ligandable structural motifs, chemical probes, and increasingly specialized computational methods. That progress does not make RNA equivalent to a protein receptor. Structured RNAs combine a polyanionic backbone, extensive hydration, counterion

dependence, base stacking, local tertiary contacts, and conformational heterogeneity in ways that can alter both pocket definition and ligand pose. Contemporary syntheses of RNA-targeted small molecules therefore support an RNA-aware view of molecular recognition rather than automatic transfer of protein-derived assumptions [1–3]. The methodological problem is not that protein-oriented approaches are unusable. It is that their default assumptions about receptor rigidity, pocket persistence, scoring

**Corresponding author:** Lina Hassan

**Address:** Department of RNA-Targeted Natural Product Discovery, Faculty of Pharmacy, University of Jordan, Amman, Jordan.

**E-mail:** ✉ [lina.hassan@ju.edu.jo](mailto:lina.hassan@ju.edu.jo)

**Received:** 12 January 2025; **Revised:** 30 March 2026; **Accepted:** 06 April 2026

This is an **open access** journal, and articles are distributed under the terms of the Creative Commons Attribution-Non Commercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.



transferability, desolvation, and selectivity cannot be presumed valid for RNA without target-specific evidence. RNA sites may be shallow, solvent exposed, ion organized, transient, or generated by local rearrangement. A single static structure can therefore underrepresent the receptor states relevant to ligand binding.

Natural products sharpen this problem because their stereochemical richness, aromatic surfaces, constrained architectures, and ionizable groups may complement structured RNA while also increasing the possibility of promiscuous nucleic-acid association, aggregation, optical interference, or context-dependent activity. This article proposes a computational grammar that separates four questions commonly collapsed into one ranking exercise: whether a ligand can occupy a plausible RNA state; whether the interaction survives realistic solvent, ion, and conformational alternatives; whether the ligand remains selective across competing RNAs; and whether engagement changes the intended RNA function. These decisions are connected, but none substitutes for another. Accordingly, docking score, binding, target engagement, function, and translational relevance are treated as distinct evidential levels throughout the framework.

This separation matters operationally because the same numerical rank can arise from very different evidential states. A stable pose in an experimentally supported pocket and a favorable score obtained only after uncertain receptor preparation should not enter prioritization with equal confidence. The proposed grammar therefore makes the assumptions generating a computational result part of the result itself.

*Why protein-centric docking rules transfer poorly to rna*

Current RNA–ligand docking benchmarks show context-dependent performance across programs, RNA targets, binding modes, and ligand classes [4–6]. This matters because conventional workflows often presume a relatively fixed cavity, a receptor structure close to its binding-competent state, scoring terms trained largely on protein complexes, and limited receptor reorganization. RNA may violate each assumption. A ligand can recognize a shallow groove, a transient pocket, a base-displacement state, or an ion-structured microenvironment rather than a preformed hydrophobic cavity.

The proposed response is not to discard docking but to change what a credible docking decision means. For RNA-targeted natural products, docking should generate structural hypotheses that retain receptor-state provenance, ligand protonation and stereochemistry, and local solvent or ion assumptions. Agreement across docking programs or receptor states may be used as an uncertainty-management signal, but consensus cannot be treated as proof because correlated scoring errors can survive repeated calculations. The practical unit of interpretation should therefore be a pose family associated with explicit assumptions, not a single top-ranked pose.

For natural products, this requirement is particularly important when multiple stereocenters, macrocyclic conformations, protonation states, or intramolecular hydrogen bonds produce several credible ligand states. Ligand-state uncertainty should be explored alongside receptor-state uncertainty rather than fixed during preprocessing and forgotten during interpretation.

**Table 1** organizes the method components that should be made explicit before docking-based rank is allowed to influence lead priority.

**Table 1.** RNA-specific method components that condition docking-based prioritization

| Method component            | Evidence-supported concern                                        | Proposed decision use                                               | Boundary                                                    |
|-----------------------------|-------------------------------------------------------------------|---------------------------------------------------------------------|-------------------------------------------------------------|
| <b>Pose generation</b>      | Pose recovery varies across RNA targets and docking programs.     | Retain multiple plausible pose families when rankings are unstable. | Retrospective pose recovery does not establish binding.     |
| <b>Pose ranking</b>         | Score performance depends on receptor and ligand context.         | Interpret rank within a stated RNA state and scoring model.         | A high score is not target engagement.                      |
| <b>Receptor preparation</b> | RNA sites can depend on ion, solvent, and local structural state. | Record protonation, ion, solvent, and receptor-state assumptions.   | Preparation choices may dominate rank.                      |
| <b>Consensus</b>            | Different programs may agree or disagree across targets.          | Use agreement as an uncertainty signal.                             | Correlated errors can survive consensus.                    |
| <b>Escalation</b>           | Higher-fidelity calculations are costly and target dependent.     | Reserve them for decisions sensitive to physical uncertainty.       | More expensive computation is not automatically more valid. |

#### *Electrostatics and solvation in rna recognition*

RNA recognition occurs in an electrostatic environment that cannot be represented adequately by treating charge only as a ligand descriptor. The phosphate backbone creates a strongly negative field, while metal ions and other counterions can stabilize local folds, bridge interactions, and alter the energetic cost of ligand association. Atomistic work on the guanine-I riboswitch demonstrates that

metal ions can be integral to the interaction environment relevant to ligand binding [7]. The transferable lesson is not that one ion model should be used universally, but that ion occupancy, identity, and placement should be treated as uncertain receptor-state variables when structural or biochemical evidence indicates their relevance.

Hydration creates a related uncertainty. Water molecules can compete with ligand functional groups, mediate

contacts, stabilize RNA conformations, or define cavities that disappear when solvent is simplified. Comparative RNA–ligand docking indicates that explicit-solvent treatment can alter pose prediction and should be evaluated where hydration is structurally important [8]. A robust workflow should therefore compare plausible hydration states instead of treating either a water-free or fully hydrated receptor as a universal default.

Sensitivity analysis is useful here because ion and water treatment can be varied without pretending that one setup is definitively correct. If lead order reverses across physically defensible solvent or ionic conditions, that instability should be reported as a ranking uncertainty requiring experiment.

Higher-fidelity physical methods may become valuable after an initial recognition hypothesis survives screening. Advanced simulation and free-energy calculations have produced highly predictive affinity estimates in a defined riboswitch system [9]. That result supports selective escalation when structural definition and sampling are adequate; it does not establish equal accuracy across RNA classes. For natural products, the proposed rule is staged: inexpensive models explore possibilities, while ion-aware, explicit-solvent, or free-energy calculations are concentrated on cases where those variables could materially change prioritization.

#### *Structural ensembles and conformational plasticity*

A single RNA structure is often an incomplete receptor description. Helical regions may remain relatively stable while junctions, loops, bulges, and tertiary contacts redistribute among states that differ in pocket volume, base exposure, and electrostatic organization. Ensemble-based pocket detection with SHAMAN illustrates how ligandable regions can be evaluated across conformational states instead of one frozen structure [10]. A state-aware workflow should therefore ask whether a candidate pocket persists across plausible conformations, appears only in a minor state, or disappears after local relaxation.

Structural rearrangement can also be part of recognition. Molecular analysis of de novo small-molecule recognition by an intron RNA shows that binding can be coupled to target-specific conformational organization [11]. A single pose can consequently represent different mechanisms: binding to a preorganized state, selection of a less populated conformation, or a geometry accessible only after ligand-induced rearrangement. Computational models should distinguish these possibilities where evidence permits rather than attributing all differences to scoring noise.

Ensemble generation itself is not ground truth. Molecular dynamics, enhanced sampling, experimental conformers, and structure-prediction methods can populate different

regions of conformational space. The workflow should therefore distinguish diversity generated computationally from states independently supported by structural or biochemical observations.

The energetic price of reaching a binding-competent conformation is equally important. Conformational penalties can offset apparently favorable intermolecular contacts and change interpretation of affinity [12]. No universal numerical correction follows. The proposed reporting rule is instead to carry receptor-state provenance and conformational uncertainty with every ranking. A natural product that scores well only in a strained or weakly supported state should not be treated as equivalent to one whose recognition is reproduced across independent, experimentally plausible conformations.

#### *Recurrent motifs versus sequence-specific binding*

RNA small-molecule recognition is neither purely sequence driven nor reducible to a few universal structural motifs. Local three-dimensional arrangements can create chemically discriminating environments even where primary sequence alone would not predict selectivity. Mechanistic work on ligands selective for RNA single-G bulges shows that motif geometry and immediate structural context can influence recognition [13]. Computational representations should therefore preserve local base orientation, accessibility, tertiary environment, and sequence neighborhood rather than encoding a motif as a context-free categorical label.

Motif recurrence simultaneously creates an off-target problem. Multiplexed RNA structure libraries reveal heterogeneous small-molecule binding patterns across many structural elements [14]. Such data can inform motif-level priors, but an in vitro preference within a library is not equivalent to engagement of the intended cellular RNA. Folding environment, competing structures, abundance, proteins, and compartmentalization can change accessibility.

This distinction also changes model validation: success on RNAs sharing a familiar motif may measure interpolation within a structural family rather than transfer to genuinely different RNA architectures. Motif-aware splits are therefore more informative than random train-test partitions when generalization is the question.

Cryptic recognition further complicates motif assignment. Structure-guided work has demonstrated that a ligandable RNA site can emerge through base displacement, so an apo structure or static motif description may miss a chemically accessible state [15]. The appropriate conclusion is conditional: cryptic-state hypotheses should be examined when local dynamics make them plausible, not presumed for every RNA. Recurrent motifs are therefore best treated as structured recognition environments whose sequence,

local geometry, and conformational state remain attached to the computational prediction.

#### *Natural product scaffolds in rna-relevant chemical space*

Natural products should not receive an RNA-targeting advantage simply because they occupy chemically distinctive space. Their potential lies in dense stereochemistry, extended aromatic surfaces, macrocyclic or constrained architectures, cationic functionality, and unusual heteroatom patterns that can engage RNA environments underrepresented in conventional screening collections. Yet binding chemistry and biological consequence remain separable. Programming otherwise inactive RNA-binding small molecules into bioactive degraders shows that detectable RNA recognition can become functionally consequential only when coupled to an effector mechanism [16]. Binding alone therefore should not confer lead status.

Direct natural-product studies establish that RNA can nevertheless be a relevant target class for chemically complex scaffolds. Integrated transcriptomic, physicochemical, and structural analyses have implicated RNA interactions in berberine activity [17]. Separately, an RNA-target-identification strategy applied to natural products identified nocathiacin as an inhibitor of an oncogenic RNA [18]. These examples do not show that natural products generally prefer RNA; they show that RNA-mediated activity can be experimentally investigated and assigned at the compound–target level.

Chemical-space analyses provide priors rather than admission rules. Bioactive RNA-targeted ligands share recurring physicochemical, structural, and spatial properties [19], and machine-learning studies indicate that RNA-binding chemical-space signals can be learned from curated datasets [20]. For natural-product libraries, such information can enrich screening or diversify selection, but dataset composition and label definitions limit transferability. The proposed rule is probabilistic: use RNA-relevant chemical features to enrich hypotheses, never as hard filters, proof of selectivity, or substitutes for structural and functional validation.

#### *Modeling rna–ligand recognition across multiple conformations*

Multi-conformation modeling should not mean docking a ligand into many structures and averaging scores. The central question is whether a recognition hypothesis is stable, state restricted, or model dependent. RNA-specific scoring functions such as SPRank provide one route for replacing unqualified protein-oriented ranking assumptions [21]. In the proposed framework, state-specific poses remain separate until receptor provenance, local interactions, and structural plausibility can be compared. A favorable score found only in one uncertain state should carry different evidential weight from a pose family recovered across independently supported states. Structure-aware learning can complement physical scoring. RNAmigos2 demonstrates that RNA binding-site information can be encoded in graph representations for structure-based virtual screening [22]. Such models can accelerate prioritization but cannot remove ensemble uncertainty, because their applicability remains constrained by the structures and chemotypes represented during training. Data perturbation and augmentation can improve RNA–ligand scoring under limited-data conditions [23], yet robustness to data perturbation should not be interpreted as recovery of unobserved tertiary states. This profile also allows contradictory models to remain visible. A physics-based pose and a learned score may disagree because they encode different approximations, not because one is automatically wrong. Reconciliation should focus on which assumptions are experimentally testable and which prediction changes when those assumptions are perturbed.

The proposed decision unit is a state-conditioned recognition profile containing ligand identity, RNA state, solvent/ion assumptions, pose family, model provenance, and model disagreement. If experimental state populations are unavailable, alternative state assumptions should be reported rather than replaced with invented probabilities. Sensitivity to removing one plausible conformation is itself useful evidence of ranking fragility.

**Figure 1** shows how state-conditioned recognition is embedded within a broader RNA-specific prioritization architecture and separated from downstream validation.

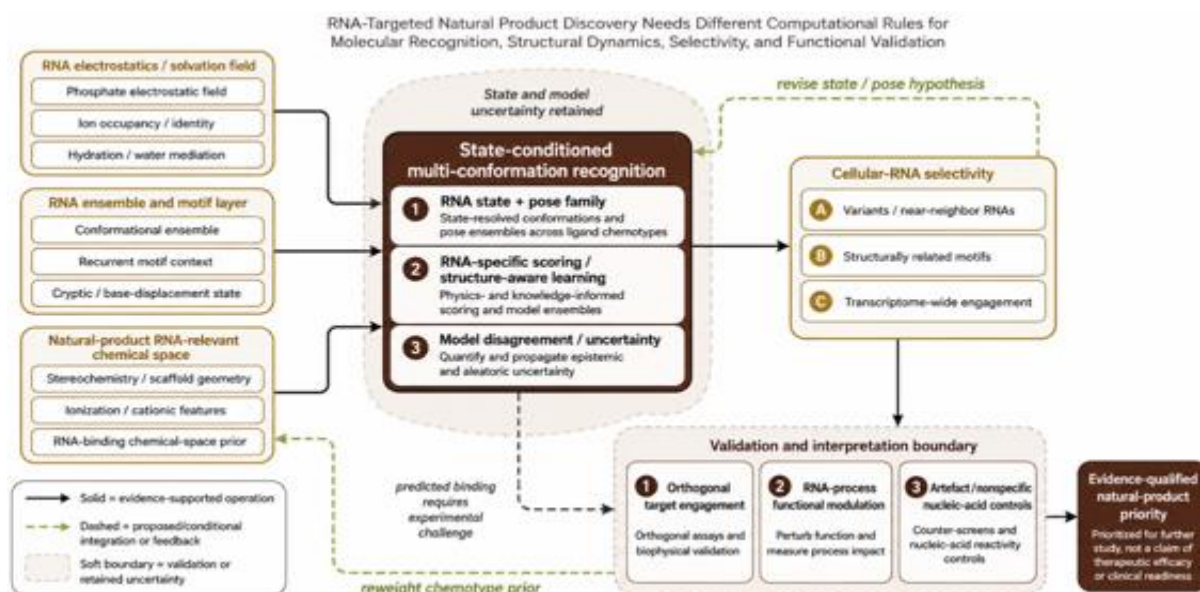


Figure 1. From RNA state to experimentally constrained natural-product priority

Table 2 summarizes what each stage of the state-conditioned workflow is allowed to decide.

Table 2. State-conditioned workflow for RNA–ligand recognition and prioritization

| Workflow step            | Capability represented                                                | Proposed decision use                                  | Interpretation limit                              |
|--------------------------|-----------------------------------------------------------------------|--------------------------------------------------------|---------------------------------------------------|
| RNA-specific scoring     | Pose and virtual-screening assessment tailored to RNA complexes.      | Compare pose families within explicit receptor states. | Does not establish universal scoring superiority. |
| Structure-aware learning | Graph or structure representations for accelerated screening.         | Prioritize candidates for higher-fidelity analysis.    | Does not establish binding or selectivity.        |
| Data perturbation        | Robustness testing under limited or biased data.                      | Expose sensitivity to training composition.            | Does not reconstruct missing tertiary states.     |
| State integration        | Comparison across alternative conformations and physical assumptions. | Carry model disagreement forward as uncertainty.       | No universal state weighting is established.      |
| Decision escalation      | Selective use of more expensive physical modeling.                    | Focus computation on rank-sensitive uncertainties.     | Computational cost is not evidence quality.       |

### Selectivity against the cellular rna landscape

For RNA targets, selectivity is a structural-landscape problem as much as a target-versus-off-target affinity problem. Variant-specific mapping shows that sequence changes can alter RNA structure and small-molecule interaction patterns [24]. A useful counterscreen should therefore include near-neighbor variants and structurally related RNAs whenever the intended recognition determinant could recur elsewhere. This does not imply that every sequence change alters pharmacology; it means that sequence identity alone is insufficient evidence of preserved or selective recognition.

Cellular mapping broadens the question from an isolated construct toward competition within the transcriptome. Transcriptome-wide identification of small-molecule RNA-binding sites has informed isoform-specific design in cells [25]. Such approaches reveal engagement landscapes inaccessible to purified-target assays, but capture chemistry and detection limits prevent interpretation as an exhaustive census of native interactions. Photoaffinity

sequencing provides another route to map small-molecule–RNA interactions [26], while derivatization and crosslinking remain potential sources of perturbation.

Selectivity analysis should also distinguish absence of detected engagement from evidence of nonbinding. Limited abundance, inaccessible structure, insufficient probe capture, or assay sensitivity may suppress an off-target signal. Negative cellular maps therefore require the same qualification as positive ones.

The proposed selectivity rule is hierarchical: test closely related structures and variants, broaden to cellular RNA engagement where feasible, and interpret off-target signals together with exposure and function. The comparison set should be mechanistically selected rather than convenient. RNAs sharing a recognized motif, local electrostatic environment, or conformational state may be more informative than unrelated abundant transcripts. A natural product should not be called RNA selective because it binds one purified RNA better than a small convenience panel, nor should transcriptome-wide engagement alone be

called pharmacological selectivity without preferential functional consequences.

#### *A computational workflow designed specifically for rna-targeted natural products*

The proposed workflow begins by matching representation to RNA hierarchy. RNA language-model and graph-attention approaches illustrate how sequence-aware and structure-aware information can be combined for binding-site prediction [27]. For natural products, these receptor representations should be coupled to explicit ligand stereochemistry, ionization, conformers, and experimentally plausible RNA states. Site prediction remains hypothesis generation, not evidence that a compound binds.

Evaluation must also resist shortcut learning. RNA–small-molecule models can exploit dataset biases rather than intended structural determinants; bias-matched decoys and structure-aware network design have been developed to expose and reduce this behavior [28]. The workflow therefore favors chemically meaningful negatives, family-aware or chemotype-aware splits where possible, and explicit reporting of performance under harder evaluation regimes. These safeguards reduce known shortcut routes but cannot prove that hidden leakage is absent.

RNA-specific pocket-druggability prediction provides another useful prior [29]. In the proposed architecture, it gates computational effort rather than conferring lead status. Pocket plausibility, ensemble support, conditional docking, RNA-relevant chemical-space assessment, selectivity testing, and functional validation remain distinct. Rejection or reprioritization should also retain cause: an unstable pocket, state-specific pose, extrapolative model, competing RNA, absent cellular engagement, or discordant function are different failures requiring different corrective experiments. Later evidence can therefore return a candidate to an earlier state or representation hypothesis instead of merely reducing one composite score.

The resulting workflow is deliberately modular. A project can strengthen one weak layer without claiming that stronger docking compensates for weak selectivity evidence or that a cellular signal rescues an ill-defined binding hypothesis. Each layer contributes a different reason to continue, revise, or stop. Iteration is therefore treated as part of scientific falsification rather than a defect in workflow linearity.

#### *From binding prediction to functional rna modulation*

Binding becomes pharmacologically informative only when the molecular interaction is connected to an RNA process. Structural and functional work on small-molecule modulators of conserved splicing machinery demonstrates

that selective recognition can be linked to measurable modulation in a defined RNA-centered system [30]. Functional readouts should therefore enter prioritization as a separate layer: splicing, translation, processing, stability, localization, or another target-appropriate process should be measured rather than inferred from docking or affinity. Disease-relevant repeat RNA provides a complementary example. Precise recognition of a toxic CUG repeat expansion has been linked to functional effects in its specific disease context [31]. Repeat RNAs, however, can create unusual multivalent recognition environments, so this evidence should not be generalized across structured RNAs.

The proposed evidence progression is structural plausibility, binding, target engagement, RNA-process modulation, and then phenotype. Each transition poses a different falsifiable question. A compound may bind a purified construct yet fail to encounter the same structure in cells; engage the RNA without changing the relevant process; or change phenotype through another mechanism. Binding–function discordance is therefore informative. It can indicate neutral occupancy, insufficient engagement, an incorrect cellular state, or a mistaken mechanistic hypothesis.

This separation prevents phenotype-first reasoning from being projected backward onto the molecular model. A desirable cellular response cannot validate a predicted pose unless the causal connection between that pose, RNA engagement, and the downstream process is independently interrogated. Functional assays should consequently be selected around the RNA's biological role and, where feasible, paired with perturbations that test whether the proposed RNA interaction is required for the response.

#### *Orthogonal validation of target engagement*

RNA target engagement should be triangulated with methods that interrogate association through different physical principles rather than inferred from a single optical assay. Affinity-selection mass spectrometry, covalent capture in living cells, and native mass spectrometry provide complementary ways to detect RNA–small-molecule association or engagement [32–34]. None is a universal gold standard. Affinity selection depends on assay conditions; covalent methods require compatible probe chemistry and can modify native kinetics; and native mass spectrometry measures complexes under instrument-specific conditions.

Evidence becomes more persuasive when engagement is demonstrated under increasingly relevant conditions and connected to function, but this ordering is proposed rather than a validated universal hierarchy. Living-cell covalent RNA engagement shows that direct cellular interrogation is experimentally possible for suitable chemistry [33].

Orthogonality should therefore mean independence of measurement principle, not simply repetition with another reporter. Two fluorescence assays sharing the same optical liability provide weaker triangulation than fluorescence paired with mass spectrometry, structural measurement, or cellular capture.

The same principle applies to assay scale: adding more measurements does not create orthogonality if they share the same interference mechanism. Method choice should maximize independence of failure modes, not merely the number of readouts.

Disagreement among methods should trigger investigation rather than automatic rejection, because buffer composition, RNA construct, ligand concentration, probe derivatization, and conformational state can legitimately alter observations. The proposed standard is concordance with explained discordance: confidence rises when independent methods converge, while unresolved divergence remains attached to the candidate as uncertainty.

**Table 3** distinguishes the evidential contribution of complementary engagement approaches without treating any method as sufficient on its own.

**Table 3.** Complementary approaches for RNA target-engagement validation

| Approach                                    | Primary evidential contribution                                                               | Principal boundary                                                          |
|---------------------------------------------|-----------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------|
| <b>Affinity-selection mass spectrometry</b> | Detects ligand association with selected RNA targets without relying on one optical reporter. | Selection conditions can reshape observed association.                      |
| <b>Cellular covalent capture</b>            | Provides direct evidence of engagement in living cells for probe-compatible ligands.          | Probe derivatization and covalency can alter native behavior.               |
| <b>Native mass spectrometry</b>             | Detects noncovalent RNA–ligand complexes through a distinct physical readout.                 | Instrument conditions are not equivalent to cellular context.               |
| <b>Cross-method triangulation</b>           | Tests whether independent measurement principles converge on the same target hypothesis.      | Convergence strengthens attribution but does not by itself prove mechanism. |

#### *Avoiding fluorescent, aggregation, and nonspecific binding artefacts*

Natural products can generate convincing assay signals for reasons unrelated to the intended RNA mechanism. Optical interference, aggregation, reactivity, reporter perturbation, and other small-molecule assay liabilities are established sources of false or misleading activity [35]. RNA-focused workflows should therefore challenge apparent activity with orthogonal readouts, concentration dependence, aggregation controls where relevant, and counterscreens against unrelated nucleic acids. A genuine binder can still trigger an interference alert, so such tests should diagnose possible mechanisms rather than operate as automatic exclusion filters.

Computational frequent-hitter tools can flag compounds associated with aggregation, spectroscopic interference, reactivity, or promiscuity [36]. These alerts support triage but cannot establish that a compound is artefactual. For RNA-targeted natural products, general interference assessment should be paired with tests for nonspecific electrostatic, intercalative, or broad nucleic-acid binding. Competition with unrelated RNA or DNA, salt sensitivity, reporter-free assays, and orthogonal physical detection can help distinguish a structure-dependent interaction from nonspecific association, although appropriate controls remain assay specific. Lead ranking should therefore respond to the diagnosed source of signal rather than to a simple pass/fail nuisance label.

#### *Boundaries of current rna modeling*

RNA-targeted machine learning remains constrained by data volume and diversity, benchmark construction, representation mismatch, and uncertain generalization beyond familiar targets and chemotypes [37]. Structural datasets are enriched for experimentally tractable RNAs and successful complexes, while apparent negative examples may represent untested pairs, assay-dependent inactivity, or artificial decoys rather than demonstrated nonbinding. Strong retrospective metrics may therefore fail on new RNA families, new natural-product scaffolds, or different cellular contexts.

The proposed framework inherits these limitations. It does not define universal ensemble sizes, ion placements, score thresholds, affinity cutoffs, or evidence weights. Nor can it resolve uncertainties that precede computation, including incorrect RNA constructs, missing post-transcriptional modifications, context-dependent protein partners, or cellular structures that differ from the modeled receptor. Natural-product uncertainty adds stereochemistry, protonation, tautomerism, aggregation, and metabolite formation, any of which can alter the species presented to RNA.

Prospective evaluation should therefore separate RNA families and chemotypes, use near-neighbor RNAs and credible negatives, and connect prediction to orthogonal engagement, cellular selectivity, and functional assays. Bias-aware evaluation can challenge shortcut performance [28], but no approved study tests the entire proposed package end to end. The framework should consequently be judged first by whether it improves falsifiability,

uncertainty tracking, and decision traceability, not by untested claims of predictive superiority.

## CONCLUSION

RNA-targeted natural-product discovery should not be organized around the assumption that an RNA receptor is simply a protein pocket built from different atoms. RNA recognition couples electrostatics, hydration, ions, conformational states, recurrent motifs, and chemically distinctive ligands within a cellular landscape containing many competing RNAs. Pose plausibility, state dependence, selectivity, target engagement, function, and assay credibility therefore answer different questions. The proposed computational grammar makes these distinctions operational. It carries receptor-state provenance and model disagreement forward, treats natural-product chemical features as priors rather than guarantees, and requires computational ranking to face progressively stronger structural, cellular, and functional challenges. Its value is not a claim of universal predictive accuracy but a more falsifiable route from molecular hypothesis to evidence-qualified priority. Prospective benchmarking across new RNA families, natural-product chemotypes, and experimental contexts will be required to determine whether this organization improves discovery decisions. Until then, RNA-specific computation is best treated as a disciplined generator of testable recognition hypotheses rather than a replacement for selectivity, engagement, artefact control, and functional evidence.

**Acknowledgments:** None

**Conflict of interest:** None

**Financial support:** None

**Ethics statement:** None

## REFERENCES

- [1] Kovachka S, Panosetti M, Grimaldi B, Azoulay S, Di Giorgio A, Duca M. Small molecule approaches to targeting RNA. *Nat Rev Chem*. 2024;8(2):120-35. doi:10.1038/s41570-023-00569-9
- [2] Childs-Disney JL, Yang X, Gibaut QMR, Tong Y, Batey RT, Disney MD. Targeting RNA structures with small molecules. *Nat Rev Drug Discov*. 2022;21(10):736-62. doi:10.1038/s41573-022-00521-4
- [3] Chen S, Mao Q, Cheng H, Tai W. RNA-binding small molecules in drug discovery and delivery: An overview from fundamentals. *J Med Chem*. 2024;67(18):16002-17. doi:10.1021/acs.jmedchem.4c01330
- [4] Gunasinghe KKJ, Ginjom IRH, San HS, Rahman T, Wezen XC. Can current molecular docking methods accurately predict RNA inhibitors? *J Chem Inf Model*. 2024;64(15):5954-63. doi:10.1021/acs.jcim.4c00235
- [5] Agarwal R, Tatikonda RR, Smith JC. Comparative assessment of pose prediction accuracy in RNA–ligand docking. *J Chem Inf Model*. 2023;63(23):7444-52. doi:10.1021/acs.jcim.3c01533
- [6] Wang K, DeFalco L, Jaladanki CK, Huber RG, Chai CLL, Fan H. Evaluating molecular docking programs for RNA-targeted ligand screening: Influence of binding modes and ligand types. *J Chem Inf Model*. 2026;66(10):6120-30. doi:10.1021/acs.jcim.6c00152
- [7] Franke L, Globisch C, Karakurt MC, Stephan T, Peter C. Atomistic simulations reveal crucial role of metal ions for ligand binding in guanine-I riboswitch. *Macromol Rapid Commun*. 2024;45(24):e2400606. doi:10.1002/marc.202400606
- [8] Almendra Rodriguez L, Kersten C. Comparative evaluation of explicit solvent models for RNA-ligand docking. *J Chem Inf Model*. 2026;66(11):6719-32. doi:10.1021/acs.jcim.6c00498
- [9] Ansari N, Liu C, Hédin F, Hénin J, Ponder JW, Ren P, et al. Targeting RNA with small molecules using state-of-the-art methods provides highly predictive affinities of riboswitch inhibitors. *Commun Biol*. 2025;8(1):1405. doi:10.1038/s42003-025-08809-y
- [10] Panci FP, Gkeka P, Bonomi M. Identifying small-molecules binding sites in RNA conformational ensembles with SHAMAN. *Nat Commun*. 2024;15(1):5725. doi:10.1038/s41467-024-49638-7
- [11] Liu T, Xu L, Chung K, Sisto LJ, Hwang J, Zhang C, et al. Molecular insights into de novo small-molecule recognition by an intron RNA structure. *Proc Natl Acad Sci U S A*. 2025;122(19):e2502425122. doi:10.1073/pnas.2502425122
- [12] Geng A, Roy R, Al-Hashimi HM. Conformational penalties: New insights into nucleic acid recognition. *Curr Opin Struct Biol*. 2024;89:102949. doi:10.1016/j.sbi.2024.102949
- [13] Hegde S, Akhter S, Tang Z, Qi C, Yu C, Lewicka A, et al. Mechanistic studies of small molecule ligands selective to RNA single G bulges. *Nucleic Acids Res*. 2025;53(12):gkaf559. doi:10.1093/nar/gkaf559
- [14] Nagasawa R, Onizuka K, Komatsu KR, Miyashita E, Murase H, Ojima K, et al. Large-scale analysis of small molecule-RNA interactions using multiplexed RNA structure libraries. *Commun Chem*. 2024;7(1):98. doi:10.1038/s42004-024-01181-8
- [15] Olenginski LT, Wierzbka AJ, Laursen SP, Batey RT. Designing small molecules targeting a cryptic RNA

- binding site through base displacement. *Nat Chem Biol.* 2026;22(6):1004-13. doi:10.1038/s41589-025-02018-8
- [16] Tong Y, Lee Y, Liu X, Childs-Disney JL, Suresh BM, Benhamou RI, et al. Programming inactive RNA-binding small molecules into bioactive degraders. *Nature.* 2023;618(7963):169-79. doi:10.1038/s41586-023-06091-8
- [17] Satpathi S, Endoh T, Podbevšek P, Plavec J, Sugimoto N. Transcriptome screening followed by integrated physicochemical and structural analyses for investigating RNA-mediated berberine activity. *Nucleic Acids Res.* 2021;49(15):8449-61. doi:10.1093/nar/gkab189
- [18] Ye F, Haniff HS, Suresh BM, Yang D, Zhang P, Crynen G, et al. Rational approach to identify RNA targets of natural products enables identification of nocathiacin as an inhibitor of an oncogenic RNA. *ACS Chem Biol.* 2022;17(2):474-82. doi:10.1021/acscchembio.1c00952
- [19] Morgan BS, Forte JE, Culver RN, Zhang Y, Hargrove AE. Discovery of key physicochemical, structural, and spatial properties of RNA-targeted bioactive ligands. *Angew Chem Int Ed Engl.* 2017;56(43):13498-502. doi:10.1002/anie.201707641
- [20] Yazdani K, Jordan D, Yang M, Fullenkamp CR, Calabrese DR, Boer R, et al. Machine learning informs RNA-binding chemical space. *Angew Chem Int Ed Engl.* 2023;62(11):e202211358. doi:10.1002/anie.202211358
- [21] Zhou Y, Jiang Y, Chen SJ. SPRank—a knowledge-based scoring function for RNA-ligand pose prediction and virtual screening. *J Chem Theory Comput.* 2024;20(16):7358-69. doi:10.1021/acs.jctc.4c00681
- [22] Carvajal-Patiño JG, Mallet V, Becerra D, Niño Vasquez LF, Oliver C, Waldispühl J. RNAmigos2: Accelerated structure-based RNA virtual screening with deep graph learning. *Nat Commun.* 2025;16(1):2799. doi:10.1038/s41467-025-57852-0
- [23] Ma H, Gao L, Lu ZJ. RNA–ligand interaction scoring via data perturbation and augmentation modeling. *Nat Comput Sci.* 2025;5(8):648-60. doi:10.1038/s43588-025-00820-x
- [24] Miyashita E, Onizuka K, Chen Y, Yoshida H, Hatayama H, Ishikawa S, et al. Systematic identification of variant-specific RNA structure-small molecule interactions exemplified by RNA G-quadruplexes. *Nat Commun.* 2026;17(1):2243. doi:10.1038/s41467-026-70097-9
- [25] Tong Y, Gibaut QMR, Rouse W, Childs-Disney JL, Suresh BM, Abegg D, et al. Transcriptome-wide mapping of small-molecule RNA-binding sites in cells informs an isoform-specific degrader of QSOX1 mRNA. *J Am Chem Soc.* 2022;144(26):11620-5. doi:10.1021/jacs.2c01929
- [26] Mukherjee H, Blain JC, Vandivier LE, Chin DN, Friedman JE, Liu F, et al. PEARL-seq: A photoaffinity platform for the analysis of small molecule-RNA interactions. *ACS Chem Biol.* 2020;15(9):2374-81. doi:10.1021/acscchembio.0c00357
- [27] Sun S, Yang J, Gao L, Li P, Liu Y. RNA language model and graph attention network for RNA and small molecule binding sites prediction. *Bioinformatics.* 2025;41(9):btaf447. doi:10.1093/bioinformatics/btaf447
- [28] Wen Y, Han Y, Wang D. Overcoming shortcut learning in RNA-small molecule modeling via bias-matched decoys and structure-aware network design. *J Chem Inf Model.* 2026;66(11):6403-17. doi:10.1021/acs.jcim.6c00360
- [29] Krishnan SR, Roy A, Wong L, Gromiha MM. DRLiPS: A novel method for prediction of druggable RNA-small molecule binding pockets using machine learning. *Nucleic Acids Res.* 2025;53(6):gkaf239. doi:10.1093/nar/gkaf239
- [30] Silvestri I, Manigrasso J, Andreani A, Brindani N, Mas C, Reiser JB, et al. Targeting the conserved active site of splicing machines with specific and selective small molecule modulators. *Nat Commun.* 2024;15(1):4980. doi:10.1038/s41467-024-48697-0
- [31] Rzuczek SG, Colgan LA, Nakai Y, Cameron MD, Furling D, Yasuda R, et al. Precise small-molecule recognition of a toxic CUG RNA repeat expansion. *Nat Chem Biol.* 2017;13(2):188-93. doi:10.1038/nchembio.2251
- [32] Rizvi NF, Howe JA, Nahvi A, Klein DJ, Fischmann TO, Kim HY, et al. Discovery of selective RNA-binding small molecules by affinity-selection mass spectrometry. *ACS Chem Biol.* 2018;13(3):820-31. doi:10.1021/acscchembio.7b01013
- [33] Bereiter R, Flemmich L, Nykiel K, Heel S, Geley S, Hanisch M, et al. Engineering covalent small molecule–RNA complexes in living cells. *Nat Chem Biol.* 2025;21(6):843-54. doi:10.1038/s41589-024-01801-3
- [34] Sternicki LM, Klose JW, Poulsen SA. Small molecule screening of RNA aptamers using native mass spectrometry. *Bioorg Chem.* 2025;165:109015. doi:10.1016/j.bioorg.2025.109015
- [35] Tan L, Hirte S, Palmacci V, Stork C, Kirchmair J. Tackling assay interference associated with small molecules. *Nat Rev Chem.* 2024;8(5):319-39. doi:10.1038/s41570-024-00593-3
- [36] Shi S, Fu L, Yi J, Yang Z, Zhang X, Deng Y, et al. ChemFH: An integrated tool for screening frequent

false positives in chemical biology and drug discovery. Nucleic Acids Res. 2024;52(W1):W439-W49. doi:10.1093/nar/gkae424

- [37] Karroucha W, Oliver C, Stoven V, Mallet V. Machine learning for RNA-targeting drug design. J Chem Inf Model. 2026;66(14):7840-61. doi:10.1021/acs.jcim.6c01094