Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

Machine learning and language models for RNA structure prediction: Progress and perspectives

Lambert Moyon[1] and Annalisa Marsico[1,2]

RNA structure is central to the function of every RNA class yet the gap between annotated sequences and experimentally determined structures remains large. Computational methods to fill this gap have evolved from thermodynamic free energy minimization through supervised deep learning to self-supervised RNA language models trained on millions of sequences, progressively improving structure prediction. Here we review the state of the art in RNA structure prediction, covering key training datasets, community benchmarks, and the performance of current models. We further discuss perspectives on integrating other data modalities, such as chemical probing signals and RNA modifications, as well as the emerging role of generative models. Challenges in generalization, handling of noncanonical interactions, and contextual structure prediction remain open frontiers for the field.

Addresses
[1] Computational Health Center, Helmholtz Center Munich, Munich, Germany
[2] Cluster for Nucleic Acid Sciences and Technologies – NUCLEATE, Germany

Corresponding author: Moyon, Lambert (lambert.moyon@helmholtz-munich.de)

Current Opinion in Plant Biology 2026, **100**:1–11

This review comes from a themed issue on **Protein Nucleic Acid Interactions (2026)**

Edited by **Michael Sattler** and **Gian Gaetano Tartaglia**

For a complete overview see the Issue and the Editorial

Available online

https://doi.org/10.1016/j.sbi.2026.103339

Introduction

RNAs represent a central biomolecule in cells, whose functional complexity is reflected not only by the diversity of RNA families but also by the variety of transcriptional and post-transcriptional processes where RNA is involved either as a substrate or as an effector [1–3]. The bioactivity of these RNAs is determined through the folding of the single-stranded nucleic acid chain onto itself [4]. Helical structures arise from the stacking of nucleobases composing the primary sequence, which are further involved in hydrogen bonds between each other and the sugar-phosphate backbone, resulting in local secondary structures, and longer–range interactions ultimately leading to a 3D structure [5]. These structures are dynamic [6] and the kinetics of an RNA's structural landscape depends on the cellular context in which it is involved.

It is crucial to understand and characterize the structure(s) adopted by RNA to understand its bioactive state. Various experimental techniques [7] have enabled the measurement of these structures. Techniques such as X-ray crystallography, NMR, and cryo-EM provide information about the 3D conformations of sequences, at low throughput. Chemical probing methods like SHAPE and DMS, which reveal nucleotide accessibility and, when combined with sequencing, provide transcriptome-wide, single-nucleotide resolution in living cells. However, despite these advances, structural data have not kept pace with the rapid growth of known RNA sequences. The gap between known RNA sequences and determined 3D structures is enormous, likely even larger than the protein structure gap before AlphaFold. While databases such as RNAcentral [8] contain 10s of millions of RNA sequences and RFam [9] catalogs 1000s of families with millions of members, the Protein Data Bank (PDB) holds only ∼10,000–15,000 RNA-containing structures, mostly fragments, protein-bound RNAs, or redundant examples. After filtering for unique, high-resolution structures, fewer than ∼1000 distinct RNA folds remain [10]. As a result, most of the transcriptome lacks 3D structural characterization, limiting our understanding of RNA function and highlighting a major frontier for both experimental and computational structural biology.

Structural elements embedded within RNA sequences regulate the full lifecycle of a transcript, from transcription and processing, through translation (for coding RNAs), localization, and degradation, and can organize biomolecular condensates via structure-dependent interactions [4]. Their importance is exemplified by recent studies of the SARS-CoV-2 genome, where advances in experimental techniques have revealed complex regulatory mechanisms, including the hijacking of host RNA-binding proteins (RBPs) and the presence of intricate,

functionally relevant RNA structures [11]; the interplay between these structures and RBP binding has been further explored through computational and experimental approaches [12–14]. Beyond their fundamental roles, RNA structures also represent compelling, sequence-specific targets for therapeutic intervention in genetic diseases and viral infections [15].

Notably, there has been an explosion of methods applying large language-modeling architectures onto nucleic acid sequences [16,17,19], exploiting the conceptual analogy of nucleotides as words organized by context-specific rules into sentences, resulting in models designated as RNA language models (RNA-LM). Nonetheless, it remains difficult to identify across the methods whether one method truly outperforms the others, owing to the variety of datasets and preprocessing approaches adopted for each method, and a lack of standard evaluation approaches. Fortunately, recent works have proposed standardized datasets and benchmark evaluations, with the aim to properly quantify the relative performance of these models, identify the key architectural choices driving this performance, and assess the generalization potential of these models beyond their training data.

As illustrated in the graphical abstract, we present here a brief overview of state-of-the-art RNA structure prediction models, with a focus on recent RNA LM–based approaches, along with the key training data sources, widely used benchmarks, and their most relevant results. We also propose a perspective on potential avenues to improve computational predictions of RNA structures, from input datasets to modeling approaches. We note that each of the aspects discussed in the following is covered by specialized recent reviews, to which we refer the reader for more detailed treatment. Finally, we highlight novel avenues to predicting RNA structures, both from datasets and modeling approaches.

Datasets for training and benchmarking RNA structure prediction models

The proliferation of machine learning models in the past years has shown a large diversity of datasets, model architectures, and applications of these models on structure prediction and beyond. These models have been trained on a variety of datasets, which have become over time reference datasets for novel methods to be trained on or evaluated against. The content and scale of these datasets determines the type of models and training objective, as highlighted in detail in Table 1, either having been established by pioneer structure-prediction methods (such as the Archive II datasets or the bpRNA datasets [20]) or representing general databases (such as RNAcentral [8] or RFam [9]) that can be directly used by RNA LM as we will present in the next section. These pioneer datasets pair RNA sequences with their secondary structures in dot-bracket notation across curated RNA families, with ArchiveII offering a few thousand structures spanning ∼10 families and bpRNA-1m in particular aggregating over 100,000 structures from multiple sources, thereby establishing the scale required to train deep models.

One important question raised both for secondary structure and 3D structure prediction methods has been the overestimation of their capacity to generalize to unseen sequences, arising from high similarity in sequence and structures between training and testing datasets. This led to the definition of novel datasets, either relying on external sources and novel sequences and structures absent from the "pioneer" datasets (such as bpRNA-new [21] or RNAStralign [22]) or through careful data similarity reduction and minimization of similarity between data splits (such as RNA3DB [23] or RNAsolo [24]). RNAsolo provides a curated collection of experimentally determined RNA 3D structures extracted from the PDB and grouped into equivalence classes (isolated RNAs, RNA-protein complexes, and DNA-RNA hybrids) to limit redundancy. RNA3DB pushes this further by clustering PDB RNA chains into components that are nonredundant in both sequence and structure, so that any resulting train/validation/test split is guaranteed to avoid the sequence and structural overlap responsible for inflated generalization estimates.

In RNA structure prediction, community challenges are used to benchmark methods, the most notable example being CASP experiments, where users apply models to predict protein and RNA structures before experimental data are released. RNA has only recently been included with dedicated targets in CASP15 (2022) and CASP16 (2024) [25,26], evaluated against 3D structures determined by X-ray crystallography, NMR, or cryo-EM. The recent Kaggle Ribonanza Challenge [27] has complemented this paradigm by benchmarking models on large-scale experimental datasets, using synthetic nucleotide-resolution chemical probing data (e.g., SHAPE/DMS) to predict RNA structure and reactivity from the sequence.

Additionally, datasets gathered in the context of online challenges have the property of maintaining a private, unreleased subset, known only to the authors of the challenge. While such a setting still requires proper assessment by the authors of sequence and structure similarities with the released datasets, this guarantees that the performance of trained models is not inflated by data leakage. Finally, while the majority of the datasets reported correspond to RNA sequences paired with RNA structures (either in the form of dot-brackets for secondary structures or commonly in 3D coordinates of mmCIF format or contact-maps for tertiary structures), we report about additional datasets used in the context of RNA structure prediction. Notably, chemical probing

Table 1

Datasets used for training and evaluating RNA structure prediction models.

Name	Date	Description	Example of methods exploiting the dataset	Type
ArchiveII [65]	2016	Sequence and structures from 3975 entities across 10 families.	UFold (test) MXfold (test) SincFold (interfamily cross-validation)	Structure
RNAStralign [22]	2017	Defined in the TurboFold II article. Total of 37,149 sequences and structures across 8 families, aggregated from different RNA structure repositories. This is considered complementary to the ArchiveII dataset.	E2Efold (train/validation/test) UFold (train/validation/test)	Structure
RNA mapping DB (RMDB) [66]	2018	Database storing 1055 entries, describing more than 1597 experiments of 4,576,919 sequences. A subset generated in the context of the Ribonanza Kaggle challenge contains probing signals for 66,351 unique sequences from 703 experiments, making use of 12 different experimental protocols, for a total of 142,300 profiles.	Specific to the Ribonanza Kaggle challenge models.	Chemical probing
bpRNA-1m [20]	2018	Sequences and structures from 102,318 entities across 10 families. Built notably from RFam 12.2. Usually post-processed to reduce sequence similarity, using tools such as CD-HIT.	SPOT-RNA (train) UFold (train) MXfold2 (train)	Structure
bpRNA-spot [67]	2019	Derived from bpRNA-1m, retaining sequences with less than 80% sequence similarity. The dataset is further split into train, validation, and test sets.	SPOT-RNA (test) MXfold2 (test) BPfold (test)	Structure
bpRNA-new [21]	2021	Defined in the MXfold2 article, built as a dataset independent from bpRNA-1m by accumulating 5430 low-similarity sequences from RFam v14.2 absent from bpRNA-1m to properly assess generalization performance on out-of-distribution sequences.	SincFold (test) MXfold2 (test)	Structure
RNAsolo [24]	2022	Derived from PDB, collection of RNA-structural data, grouped into classes to reduce redundancy. In March 2026: 3974 isolated RNAs, 33,951 RNA-protein complexes, and 161 DNA-RNA hybrids.	**RhoFold+** (test) **trRosettaRNA** (test)	Structure
RNA3DB [23]	2024	A dataset of nonredundant RNA structures from the PDB. RNA chains in the PDB were labeled with noncoding RNA families and then clustered to reduce redundancy. From 05/01/2026–full-release, 2199 sequences were clustered into 142 components that are nonredundant both in sequence and structure. This aims at preventing data leakage during data split for training models.	Benchmark of **AlphaFold3** by [68]	Structure
GARNET [30]	2024	Gtdb-acquired RNa with environmental temperatures: Database linking genomic RNA sequences (especially rRNAs from GTDB genomes) to optimal organism growth temperatures. Contains sequence and structural context (e.g. 16S/23S/5S RNAs with temperature metadata).	Specific to the GARNET paper.	Structure
RNAbench [69]	2024	Three benchmark datasets carefully curated and segmented to address homology issues and evaluated generalization performance of models.	Specific to the RNAbench benchmark paper.	Structure
Kaggle Ribonanza challenge [27]	2024	DMS-MapSeq and 2A3-MapSeq probing profiles for 806,573 synthetic sequences, processed with the Ubr pipeline from the Das lab.	Specific to the Ribonanza Kaggle challenge models.	Chemical probing
bpRNA–NF–15.0 [70]	2025	Defined in the DivideFold article, built from RFam v15.0 as a dataset independent from bpRNA-1m, resulting in 8788 low-similarity sequences.	DivideFold (defined it; independent test)	Structure

(*continued on next page*)

Table 1. (*continued*)

Name	Date	Description	Example of methods exploiting the dataset	Type
RNA family database (RFam) [9]	2025	Curated database of 10s of millions of RNA sequences grouped into RNA families (as of Jan 2026: 4227 families), each defined by multiple-sequence alignments, consensus secondary structures, and covariance models. Source for bpRNA-1m, bpRNA-new, RNASSTR.	RiNALMo (pretrain)	Structure
RASP DB [29]	2025	438 RNA structure probing datasets covering 85 articles, 24 species, and 39 experimental methods	Not used to our knowledge.	Chemical probing
RNASSTR [71]	2025	Datasets created by the authors of SincFold, curating sequences and structures from RFam, RefSeq, and GTDB. A total of 4,779,435 high-confidence RNA sequence–structure pairs, with increased depth across various RNA families.	SincFold (train/test)	Structure
Nucleotide affinity benchmark (NABench) [31]	2025	Aggregates 162 high-throughput assays (binding/activity) with 2.6 million mutated DNA/RNA sequences to assess fitness.	Not a structure prediction dataset, rather used to evaluate language models e.g. RNA-FM.	Sequence
CASP16 [26,72]	2026	While focused on protein folding, this edition and the previous one included RNA structures as part of the challenge. Here: 35 RNA monomer structures, covering a range of complexity (from homology to known structures to non-canonical interactions and tertiary motifs.)	AlphaFold3 (test) RhoFold (test)	Structure
RNAcentral [8]	2026	Comprehensive database of 45M RNA sequences from 54 databases.	**RNA-FM** (pretrain) **RiNALMo** (pretrain)	Sequence
RNAndria [52]	2026	Authors used DMS-MaPseq to probe and model the structures of 3′ ends of mRNAs and primary microRNAs (pri-miRNAs) of the human genome, totalling 1456 mRNA and 1098 pri-miRNA structures.	eFold (train)	Chemical probing

PDB, Protein Data Bank.

datasets provide a readout of RNA reactivities within experimental conditions, which can be interpreted as a proxy of the flexibility or accessibility of the nucleobases in the sequence. Many methods can use these readouts to guide folding algorithms [28] but they can also be used as a signal to predict, closer to the readouts from the experiments (as compared to dot-brackets, which involve an algorithmic transformation, that introduce specific biases from priors). The Kaggle Ribonanza Challenge provided a large-scale coherent set of probing signals for close to a million sequences. Besides, the RASPdb [29] compiled a variety of published datasets into a unified interface to query. Two other datasets worth mentioning are the GARNET dataset [30] and the NAbench dataset [31], enabling the exploration of the relationships between RNA and fitness. Beyond the training objectives and annotations emphasized so far, the datasets surveyed here also differ in the RNA classes they represent. Most structure-oriented datasets are dominated by noncoding RNAs and are skewed toward a handful of strongly structured families, such as ribosomal RNAs, tRNAs, riboswitches, ribozymes, SRP, and RNase P. Hence, long RNA sequences are represented almost exclusively by rRNAs, while long noncoding RNAs, messenger RNAs, and microRNAs remain scarce unless explicitly targeted (e.g. RNAndria for mRNA 3' ends and pri-miRNAs). Large-scale probing experiments such as the Kaggle Ribonanza dataset and NAbench instead rely largely on synthetic or mutated sequences, for which natural RNA class labels are not meaningful, so that the choice of dataset implicitly determines which RNA types a model encounters during training and evaluation. These families, moreover, are not merely well structured but evolutionarily conserved in structure. This property underpins notably the use of large sequence-only corpora in RNA language models, where structure-preserving covariation is assumed to dominate the evolution of structured ncRNAs, enabling masked-language pretraining to infer base-pairing dependencies from the sequence alone.

State-of-the-art prediction methods and benchmarks

From the 1970s to 2006, RNA structure prediction relied on thermodynamic free energy minimization methods (e.g., RNAfold [32]) using Zuker's dynamic programming and Turner's parameters, alongside comparative approaches like SCFG-based Pfold [33] that exploited evolutionary covariation from sequence alignments. Early machine learning for RNA structure began with CONTRAfold (2006) [34], using conditional log-linear models with ∼300 parameters to match thermodynamic methods, while ContextFold (2011) [35] exposed overparametrization by overfitting unseen RNA families. A further advance came from experiment-trained neural networks: CROSS (2017) [36] was among the first methods trained directly on high-throughput probing data (SHAPE, PARS, and icSHAPE) and NMR/X-ray structures to predict single- versus double-stranded structural profiles from sequence alone, at single-nucleotide resolution and without length restrictions, and its outputs could be fed back into thermodynamic methods to improve their accuracy. Its successor CROSSalive (2020) [37], built on the same architecture but incorporated predicted protein interactions (via catRAPID) and m6A-methylation—dependent models, captures how RNA-binding proteins and modifications reshape folding inside the cell. The foundation model era for RNA structure prediction began with large-scale RNA LMs trained via self-supervised masked sequence modeling on millions to billions of RNA or DNA sequences, learning task-agnostic representations useful for downstream tasks including structure prediction. An early example is RNA-FM (2022) [38], a 12-layer bidirectional transformer pretrained on 23.7 million noncoding RNAs from RNAcentral. RNA-MSM (2024) [39] incorporated MSA-based coevolution signals, with attention maps revealing base-pairing contacts, and RhoFold+ (2024) [40] combined RNA-FM embeddings with geometric deep learning to achieve state-of-the-art accuracy on RNA-Puzzles and CASP15. StructRFM (2025) [41] further advanced the field with structure-aware pretraining on 21 million sequence—structure pairs, boosting tertiary structure prediction over AlphaFold3. Recently, RiNALMo [42] demonstrated that such language model embeddings can also be integrated into hybrid architectures for RNA structure prediction, offering competitive performance while remaining MSA-free. These models harness large unlabeled sequence datasets to capture evolutionary and structural patterns, enable single-sequence predictions for orphan RNAs, and provide versatile embeddings for structure and functional analyses.

The aforementioned datasets have been used in various combinations across publications, comparing the model introduced by the respective authors against a selection of relevant state-of-the-art models and algorithms. This prevents a straight-forward comparison of all against all methods. Recent benchmarks have tried to address this issue by exploring greater numbers of methods on specific tasks. We compiled a list of such benchmarks published in the past years (Table 2) and report hereafter the notable points of interest.

From the EternaBench benchmark (2022) [43] provided with a comparison of thermodynamics-based models and machine learning—based models, showing that machine learning models such as CONTRAfold2 perform better than the classical models such as ViennaRNA or RNAstructure. Overall, RNA-LMs appear competitive against other model types across different tasks, with RNA-FM and RiNALMo consistently leading in noncoding RNA classification and secondary structure

Table 2

Benchmark frameworks and analyses.

Name	Date	Description	Structure focus	Best models
EternaBench [43]	2022	Evaluates secondary structure packages using high-throughput chemical mapping and riboswitch affinity data.	2D	CONTRAfold 2 EternaFold RNAsoft
BEnchmArk for COmprehensive RNA tasks and language models (BEACON) [64]	2024	A benchmark for 13 tasks covering RNA structural analysis, functional studies, and engineering applications	Both 2D and 3D	RNA-FM SpliceBERT-MS1024 BEACON-B
State-of-the-RNArt [73]	2024	State-of-the-RNArt: Compares RNA 3D structure prediction tools (*ab initio*, template-based, and deep learning) on common test sets.	3D	**trRosettaRNA** RNAComposer IsRNA1 RNAJP
Sinc-lab RNA-LLM-Folding [44]	2025	Focus on RNA-LM, evaluated on secondary structure prediction with increasing difficulty and low homology.	Both 2D and 3D	ERNIE-RNA **RiNALMo** **RNA-FM**
RNAGym [74]	2025	Evaluates RNA fitness (variant effects), secondary structure, and tertiary structure prediction across 70 assays.	Both 2D and 3D	Evo 2 **RNA-FM** RNAErnie Evo 1.5 EternaFold CONTRAfold **AlphaFold3** NuFold
Wang et al., 2026 zero-shot [45]	2026	Zero-shot evaluation of 21 language models, for secondary structure, classification, and fitness prediction. Strong focus on scaling, data leakage, and architecture choices.	2D	RNA-MSM AIDO.RNA MP-RNA ProtRNA RNA-FM
Wyss et al., 2025 [75]	2025	Evaluate RNA 3D structure–function modeling, attempting to reduce structural similarity across splits to test model generalization.	3D	gRNAde RNA-frameflow RNAflow R3design
NCBench [76]	2025	Evaluation on predicting noncanonical base-pair edges and orientations.	3D	NCfold UTR-language models **RiNALMo** **RNA-FM**
RNAScope [77]	2025	RNAscope: Benchmarks RNA LM across 15 subtasks gauging structure, interaction, and function.	Both 2D and 3D	**RiNALMo** **RNA-FM** SpliceBERT 3UTRBERT
RNATx-bench [63]	2025	Focused on RNA therapeutics (ASOs, siRNAs, etc.); evaluation of RNAGenesis on 3D structure.	3D	RNAGenesis **RhoFold+** Evo 2 **RNA-FM**
AlphaFold3 benchmark [68]	2025	A comprehensive evaluation of AlphaFold3 for RNA 3D structure prediction against 10 state-of-the-art methods.	3D	**AlphaFold3**
CASP16 [25,72,78]	2026	CASP16: Large-scale assessment of nucleic acid (RNA/DNA) 3D structure prediction for monomers and multimers	3D	Vfold GuangzhouRNA-human KiharaLab Yang-server GeneSilico
RNA-puzzles round 5 [79]	2026	RNA-puzzles round V: A blind community challenge for predicting RNA 3D structures from sequences.	3D	RNAComposer SimRNA Das group Chen group Szachniuk group Bujnicki group

LM, language models.

prediction [44]. The benchmark from Wang et al. [45] represents a particularly interesting approach to the evaluation of RNA-LMs: the authors focus on zero-shot evaluations, i.e. directly using the numerical representations (embeddings) produced by the models from the input sequences and minimally transforming those to adapt to the considered downstream tasks. This addresses the idea that through their training RNA-LM learn complex dependency rules related to the sequence content that should in part reflect structure folding. Across benchmarks, the question of generalization still remains as many models do show noticeable prediction quality drop-offs when applied interfamily. In addition, deep learning for RNA faces limited structural data (25x fewer than proteins in PDB) and high computational demands for long sequences, while even top foundation models must account for RNA conformational ensembles, chemical modifications affecting stability, environmental context, and scaling to full-length transcripts. Finally, the CASP16 challenge results highlight that overall machine learning approaches frequently struggle with noncanonical interactions, pseudoknots, and the long-range dependencies necessary for stable tertiary folding. Best performing teams showed that thermodynamically informed predictions serve as a critical bridge between accurate 2D scaffolds and complex 3D architectures, by integrating biophysical constraints that purely data-driven models often overlook. Such hybrid approaches enforce thermodynamic stability and stereochemical correctness, leading to 3D models that respect the underlying physical laws of RNA folding.

Perspectives: chemical probing signals, multimodal data, and modeling improvements

On chemical probing signals

Chemical probing signals, indicating local nucleotide flexibility or accessibility at each position in the RNA, where high reactivity indicates an unpaired, conformationally flexible nucleotide and low reactivity indicates a base-paired or otherwise constrained one, represent an interesting data source, providing researchers with contextualized information on the structure(s) adopted by a given sequence [46]. State-of-the-art RNA-folding algorithms such as ViennaRNA [47] and RNAStructure [48] handle these probing signals as additional constraints to consider when folding input sequences. The pioneering machine learning methods ShaKer [49] and StructureImpute [50] demonstrated that it was possible to model the probing signal from input sequences, using graph-kernel modeling (ShaKer) or combining multilayer convolutional neural network and recurrent neural network modeling. More recently, the Atom-1 [51] was presented as a well-performing (but closed-source) foundation model trained on such data. The Kaggle Ribonanza Challenge [27] has then been a notable milestone, providing large-scale chemical probing datasets through the Kaggle platform, which lead to a large number of deep learning models from participants. The final model developed by the authors of the challenge, RibonanzaNet, is based on a "dual-tower" architecture where sequence representations (processed through layers of convolutions and attention) and 2D pairwise representations (built from the individual nucleotide representations from the first "tower") are mutually updating during training of the model, leveraging motif detections and long-range dependencies between any pairs of nucleotides. They demonstrated great performance as a foundation model when applying it notably to 2D and 3D structure predictions. We note that the datasets correspond to synthetic sequences and thus does not address the idea of *in vivo* contextualized RNA structures. On the other hand, the very recent eFold model [52], trained on authors' newly defined RNAndria dataset of *in vivo* chemical probing data, is the latest model trained on such task, with promising performance, considering great generalization on the Ribonanza datasets. While *in vitro* RNA folding primarily reflects the intrinsic thermodynamic and biophysical constraints encoded by the RNA sequence, *in vivo* RNA structure probing captures additional layers of cellular regulation, including protein interactions, molecular crowding, active remodeling, and cotranscriptional effects [53]. Thus, *in vivo* chemical probing provides better contextualized snapshots of RNA structures that may significantly differ from *in vitro* folded structures. Beyond contextualization of RNA folding, recent chemical probing studies have enabled the identification of alternative structures adopted by RNAs depending on context [54,55]. This is a particularly interesting avenue for computational approaches to go beyond the static prediction of a single structure given an input sequence. This concept is exploited by the FoldARE method [56], although the probing signal there is entirely simulated rather than predicted after training on experimental data.

Perspective on multimodal input data

In the review [18], the authors provide a list of databases compiling experimental results measuring RNA modifications (such as m6A) across sequences. As discussed by authors, these modifications are of particular importance in the folding of RNA sequences as they can greatly influence and redefine the landscape of interactions that are possible within an unmodified sequence. While challenges remain such as the sparsity of these data, their integration was successful for tools such as the ViennaRNA suite [57] and RNAStructure suite [48]. The review additionally reports a list of methods aimed at predicting RNA modifications from input sequences. But to date, it seems that no deep learning method has attempted to integrate both RNA

sequence and RNA modifications to address the task of RNA folding, which would be an interesting future perspective. Beyond RNA modifications, the structure ensemble accessible to an RNA can be redefined by interactions with other biomolecules, either endogenous (such as RBPs, in particular helicases unwinding the mRNA during translation) or exogenous (such as RNA therapeutics or small molecules [58]). Thus, models trained on such additional signals, both experimental or predicted [59], should provide richer, contextually accurate structure predictions. Causal relationships will require additional experimental data and modeling approaches though, which goes beyond the task of predicting structure from input modalities.

From predictive models to generative AI approaches

Besides prediction models, we note that some benchmarks included generative models. Whereas, the predictive models discussed so far learn to map an input sequence to a single most-likely structure, generative models instead learn the overall distribution of structures present in the training data, and can then draw new samples from it. This distinction is biologically meaningful for RNA: a given sequence does not adopt one fixed shape but rather samples an ensemble of alternative conformations. By learning this data distribution and its intrinsic variability, generative models can propose several plausible structures for the same sequence and thus better reflect this conformational diversity. For instance, diffusion models are trained by progressively adding noise to real data and learning to reverse that process; once trained, they start from pure noise and denoise it step by step to generate new samples resembling the training data [60]. It remains to be clarified whether the variability of generated signals reflect a biologically relevant variability learned by the model, such as structure dynamics, flexibility, energy states, or merely the uncertainty of the model in reconstructing the signal. Despite these open questions, recent models such as RNADiffFold [61] or DynaRNA [62] demonstrate successful application of such architectures to the prediction of RNA structures, while RNAGenesis [63] was presented as a foundation model with great performance on the BEACON benchmark [64]. Nonetheless, a more exhaustive evaluation of these models is required to understand their advantages and limitations compared to alternatives.

Conclusion

We reported here some of the latest results in applying machine learning approaches to RNA structure prediction, notably progress from RNA LM–based approaches. Challenges remain in properly evaluating these models and in identifying clear architecture choices or data modeling approaches that consistently set them apart. Current models may be complementary to one another and new methods should provide critical looks at where they fail and interrogate the nature of the ground truths they compare against. Finally, we suggest that novel generative deep learning architectures, as well as multimodal or multitask approaches would enable both contextualized and uncertainty-aware folding of RNA sequences.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Annalisa Marsico reports financial support was provided by the Helmholtz Center Munich German Research Center for Environmental Health. Annalisa Marsico reports a relationship with the Helmholtz Center Munich German Research Center for Environmental Health that includes employment. Lambert Moyon reports a relationship with the Helmholtz Center Munich German Research Center for Environmental Health that includes employment. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Work in the laboratory of A.M. was supported by the Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung, BMBF) Clusters4Future program (Cluster for Nucleic Acid Therapeutics Munich, CNATM, ID: 03ZU1201AA), SFB TRR267 (Project-ID 403584255) and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) — Project-ID 533767322 — EXC 3113/1, Cluster for Nucleic Acid Sciences and Technologies — NUCLEATE.

Data availability

No data was used for the research described in the article.

References

Papers of particular interest, published within the period of review, have been highlighted as:

* of special interest
** of outstanding interest

1. Nemeth K, Bayraktar R, Ferracin M, Calin GA: **Non-coding RNAs in disease: from mechanisms to therapeutics**. *Nat Rev Genet* 2024, **25**:211–232.

2. Haseltine WA, Patarca R: **The RNA Revolution in the Central Molecular Biology Dogma Evolution**. *Int J Mol Sci* 2024, **25**, 12695.

3. Delaunay S, Helm M, Frye M: **RNA modifications in physiology and disease: towards clinical applications**. *Nat Rev Genet* 2024, **25**:104–122.

4. ** Cao X, Zhang Y, Ding Y, Wan Y: **Identification of RNA structures and their roles in RNA functions**. *Nat Rev Mol Cell Biol* 2024, **25**:784–801.
This review provides one of the most comprehensive treatments of the relationship between experimentally derived structural data,

particularly from chemical probing, and the functional consequences of RNA structure across the full range of RNA biogenesis and regulatory processes, making it a key reference for framing the biological motivation behind RNA structure prediction

5. Vicens Q, Kieft JS: **Thoughts on how to think (and talk) about RNA structure**. *Proc Natl Acad Sci* 2022, **119**, e2112677119.

6. Ganser LR, Kelly ML, Herschlag D, Al-Hashimi HM: **The roles of structural dynamics in the cellular functions of RNAs**. *Nat Rev Mol Cell Biol* 2019, **20**:474–489.

7. Deng J, Fang X, Huang L, Li S, Xu L, Ye K, Zhang J, Zhang K, Zhang QC: **RNA structure determination: From 2D to 3D**. *Fundam Res* 2023, **3**:727–737.

8. The RNAcentral Consortium: **RNAcentral in 2026: genes and literature integration**. *Nucleic Acids Res* 2026, **54**:D303–D313.

9. Ontiveros-Palacios N, Cooke E, Nawrocki EP, Triebel S, Marz M, Rivas E, Griffiths-Jones S, Petrov AI, Bateman A, Sweeney B: **RNA families database in 2025**. *Nucleic Acids Res* 2025, **53**: D258–D267. Rfam 15.

10. Nithin C, Kmiecik S, Błaszczyk R, Nowicka J, Tuszyńska I: **Comparative analysis of RNA 3D structure prediction methods: towards enhanced modeling of RNA-ligand interactions**. *Nucleic Acids Res* 2024, **52**:7465–7486.

11. Manfredonia I, Incarnato D: **Structure and regulation of coronavirus genomes: state-of-the-art and novel insights from SARS-CoV-2 studies**. *Biochem Soc Trans* 2020, **49**:341–352.

12. Schmidt N, Ganskih S, Wei Y, Gabel A, Zielinski S, Keshishian H, Lareau CA, Zimmermann L, Makroczyova J, Pearce C, *et al.*: **SND1 binds SARS-CoV-2 negative-sense RNA and promotes viral RNA synthesis through NSP9**. *Cell* 2023, **0**.

13. Horlacher M, Oleshko S, Hu Y, Ghanbari M, Cantini G, Schinke P, Vergara EE, Bittner F, Mueller NS, Ohler U, *et al.*: **A computational map of the human-SARS-CoV-2 protein–RNA interactome predicted at single-nucleotide resolution**. *NAR Genom Bioinf* 2023, **5**. lqad010.

14. Sun L, Li P, Ju X, Rao J, Huang W, Ren L, Zhang S, Xiong T, Xu K, Zhou X, *et al.*: ***In vivo* structural characterization of the SARS-CoV-2 RNA genome identifies host proteins vulnerable to repurposed drugs**. *Cell* 2021, **184**:1865–1883.e20.

15. ** Tong Y, Childs-Disney JL, Disney MD: **Targeting RNA with small molecules, from RNA structures to precision medicines: IUPHAR review: 40**. *Br J Pharmacol* 2024, **181**: 4152–4173.
A review exploring RNA-as-target for the design of RNA therapeutics, with focus on RNA structures. Through the presentation of different experimentally validated examples, the authors highlight how important the RNA structures are to the correct regulatory processing of RNAs

16. Sacco G, Bussi G, Sanguinetti G: **Machine learning for RNA secondary structure prediction: a review of current methods and challenges**. *RNA* 2026, **32**:443–456.

17. Kagaya Y, Liu B, Kihara D: **Computational approaches for RNA structure prediction and design**. *Cell Rep Phys Sci* 2026, **7**, 103097.

18. Yang S, Pham NT, Li Z, Baik JY, Lee J, Zhai T, Yu W, Hou B, Shang T, He W, *et al.*: *Advances in RNA secondary structure prediction and RNA modifications: methods, data, and applications*. 2025, https://doi.org/10.48550/arXiv.2501.04056.

19. Shu L, Tang J, Guan X, Zhang D: **A comprehensive survey of genome language models in bioinformatics**. *Briefings Bioinf* 2026, **27**. bbaf724.

20. Danaee P, Rouches M, Wiley M, Deng D, Huang L, Hendrix D: **bpRNA: large-scale automated annotation and analysis of RNA secondary structure**. *Nucleic Acids Res* 2018, **46**: 5381–5394.

21. Sato K, Akiyama M, Sakakibara Y: **RNA secondary structure prediction using deep learning with thermodynamic integration**. *Nat Commun* 2021, **12**:941.

22. Tan Z, Fu Y, Sharma G, Mathews DH: **TurboFold II: RNA structural alignment and secondary structure prediction informed by multiple homologs**. *Nucleic Acids Res* 2017, **45**: 11570–11581.

23. Szikszai M, Magnus M, Sanghi S, Kadyan S, Bouatta N, Rivas E: **RNA3DB: A structurally-dissimilar dataset split for training and benchmarking deep learning models for RNA structure prediction**. *J Mol Biol* 2024, **436**, 168552.

24. Adamczyk B, Antczak M, Szachniuk M: **RNAsolo: a repository of cleaned PDB-derived RNA 3D structures**. *Bioinformatics* 2022, **38**:3668–3670.

25. * Kretsch RC, Albrecht R, Andersen ES, Chen H-A, Chiu W, Das R, Gezelle JG, Hartmann MD, Höbartner C, Hu Y, *et al.*: **Functional Relevance of CASP16 Nucleic Acid Predictions as Evaluated by Structure Providers**. *Proteins: Struct, Funct, Bioinf* 2026, **94**: 51–78.
The CASP challenges are a well-known benchmark that provides an overview of the current state of research in protein folding, and since CASP15, RNA as well. While the number of RNA sequences evaluated in the challenge is limited, this article underlines the limitations and difficulties that remain in predicting RNA structures with machine/deep learning and language models approaches

26. Kretsch RC, Hummer AM, He S, Yuan R, Zhang J, Karagianes T, Cong Q, Kryshtafovych A, Das R: **Assessment of Nucleic Acid Structure Prediction in CASP16**. *Proteins: Struct, Funct, Bioinf* 2026, **94**:192–217.

27. ** He S, Huang R, Townley J, Kretsch RC, Karagianes TG, Cox DBT, Blair H, Penzar D, Vyaltsev V, Aristova E, *et al.*: *Ribonanza: deep learning of RNA structure through dual crowdsourcing*. 2024, https://doi.org/10.1101/2024.02.24.581671.
This paper is of major importance because of the scale of the Ribonanza Kaggle Challenge: by the scale of chemical probing data generated, and the scale of participants to the challenge, it resulted in a large amount of computational modeling approaches, enabling the exploration of many alternative model architectures and data processing. Besides, the architecture of RibonanzaNet, which is the model finally adopted by the authors, presents many interesting concepts, notably the "two-tower" design where both nucleotide-embeddings and pair-wise 2D embeddings (from the outer product of the first "tower") update one-another during training.

28. Saaidi A, Allouche D, Regnier M, Sargueil B, Ponty Y: **IPANEMAP: integrative probing analysis of nucleic acids empowered by multiple accessibility profiles**. *Nucleic Acids Res* 2020, **48**:8276–8289.

29. * Mu K, Fei Y, Xu Y, Zhang QC: **RASP v2.0: an updated atlas for RNA structure probing data**. *Nucleic Acids Res* 2025, **53**: D211–D219.
This resource represents an important effort in collecting and harmonizing chemical probing datasets from various sources. As we discussed in the perspective section, we believe that chemical probing experimental data is currently under-exploited in the modeling of RNA structures. This is thus an appealing resource to be used for future models.

30. Shulgina Y, Trinidad MI, Langeberg CJ, Nisonoff H, Chithrananda S, Skopintsev P, Nissley AJ, Patel J, Boger RS, Shi H, *et al.*: **RNA language models predict mutations that improve RNA function**. *Nat Commun* 2024, **15**, 10627.

31. Li Z, Ma R, Tan J, Tan C, Zheng S: NABench: large-scale benchmarks of nucleotide foundation models for fitness prediction. *arXiv.org* 2025.

32. Gruber AR, Lorenz R, Bernhart SH, Neuböck R, Hofacker IL: **The Vienna RNA Websuite**. *Nucleic Acids Res* 2008, **36**:W70–W74.

33. Knudsen B, Hein J: **Pfold: RNA secondary structure prediction using stochastic context-free grammars**. *Nucleic Acids Res* 2003, **31**:3423–3428.

34. Do CB, Woods DA, Batzoglou S: **CONTRAfold: RNA secondary structure prediction without physics-based models**. *Bioinformatics* 2006, **22**:e90–e98.

35. Zakov S, Goldberg Y, Elhadad M, Ziv-ukelson M: **Rich Parameterization Improves RNA Structure Prediction**. *J Comput Biol* 2011, **18**:1525–1542.

36. Delli Ponti R, Marti S, Armaos A, Tartaglia GG: **A high-throughput approach to profile RNA structure**. *Nucleic Acids Res* 2017, **45**:e35.

37. Ponti RD, Armaos A, Vandelli A, Tartaglia GG: **CROSSalive: a web server for predicting the in vivo structure of RNA molecules**. *Bioinformatics* 2020, **36**:940–941.

38. Chen J, Hu Z, Sun S, Tan Q, Wang Y, Yu Q, Zong L, Hong L, Xiao J, Shen T, *et al*.: *Interpretable RNA foundation model from unannotated data for highly accurate RNA structure and function predictions*. 2022, https://doi.org/10.48550/arXiv.2204.00300.

39. ** Zhang Y, Lang M, Jiang J, Gao Z, Xu F, Litfin T, Chen K, Singh J, Huang X, Song G, *et al*.: **Multiple sequence alignment-based RNA language model and its application to structural inference**. *Nucleic Acids Res* 2024, **52**, e3.
The RNA-MSM model has been a well performing deep learning model. While Multiple-Sequence-Alignment methods tend to underperform on poorly represented sequences, evolutionary conservation remains an important feature to consider when evaluating the relevance and functional potential of structures. This model and its architecture represent a high-quality standard approach to this kind of data.

40. Shen T, Hu Z, Sun S, Liu D, Wong F, Wang J, Chen J, Wang Y, Hong L, Xiao J, *et al*.: **Accurate RNA 3D structure prediction using a language model-based deep learning approach**. *Nat Methods* 2024, **21**:2287–2298.

41. Zhu H, Li R, Zhang F, Tang F, Ye T, Li X, Gu Y, Xiong P, Zhou SK: *A fully-open structure-guided RNA foundation model for robust structural and functional inference*. 2025, https://doi.org/10.1101/2025.08.06.668731.

42. ** Penić RJ, Vlašić T, Huber RG, Wan Y, Šikić M: **RiNALMo: general-purpose RNA language models can generalize well on structure prediction tasks**. *Nat Commun* 2025, **16**:5671.
RiNALMo appears across the different benchmarks as one of the best performing models for predicting RNA structures from sequence. Adopting a Language-Modeling based architecture, its unsupervised training has proven successful in learning highly informative sequence representations. The performance reported on various downstream tasks beyond RNA structure prediction further emphasize it can serve as a robust foundation model for RNA sequence representation

43. Wayment-Steele HK, Kladwang W, Strom AI, Lee J, Treuille A, Becka A, Das R: **RNA secondary structure packages evaluated and improved by high-throughput experiments**. *Nat Methods* 2022, **19**:1234–1242.

44. * Zablocki LI, Bugnon LA, Gerard M, Di Persia L, Stegmayer G, Milone DH: **Comprehensive benchmarking of large language models for RNA secondary structure prediction**. *Briefings Bioinf* 2025, **26**:bbaf137.
One of the important benchmarks produced in the last years, formally exploring the limitations in generalization of RNA-Language-models for RNA structure prediction

45. ** Wang H, Zhang Y, Chen J, Zhan J, Zhou Y: **Zero-shot benchmarking of RNA language models in structural, functional, and evolutionary learning**. *Briefings Bioinf* 2026, **27**. bbag098.
Another important benchmark which we report about in the article. It is interesting both by its scale (21 models evaluated) and its focus on "zero-shot" predictions. This is notable as this properly evaluates how good the numerical representations of sequence are "as is." Their benchmark is exhaustive in terms of tasks. This is a fundamental benchmark that should inspire future models' evaluation

46. Spitale RC, Incarnato D: **Probing the dynamic RNA structurome and its functions**. *Nat Rev Genet* 2023, **24**:178–196.

47. Lorenz R, Bernhart SH, Höner zu Siederdissen C, Tafer H, Flamm C, Stadler PF, Hofacker IL: **ViennaRNA Package 2.0**. *Algorithm Mol Biol* 2011, **6**:26.

48. Bellaousov S, Reuter JS, Seetin MG, Mathews DH: **RNAstructure: web servers for RNA secondary structure prediction and analysis**. *Nucleic Acids Res* 2013, **41**:W471–W474.

49. Mautner S, Montaseri S, Miladi M, Raden M, Costa F, Backofen R: **ShaKer: RNA SHAPE prediction using graph kernel**. *Bioinformatics* 2019, **35**:i354–i359.

50. Gong J, Xu K, Ma Z, Lu ZJ, Zhang QC: **A deep learning method for recovering missing signals in transcriptome-wide RNA structure profiles from probing experiments**. *Nat Mach Intell* 2021, **3**:995–1006.

51. Boyd N, Anderson BM, Townshend B, Chow R, Stephens CJ, Rangan R, Kaplan M, Corley M, Tambe A, Ido Y, *et al*.: *ATOM-1: a foundation model for RNA structure and function built on chemical mapping data*. 2023, https://doi.org/10.1101/2023.12.13.571579.

52. * de Lajarte AA, Taillades YJM des, Aruda J, Bongrand P, Wightman FF, Salazar D, Allan MF, Kalicki C, L'Esperance-Kerckhoff C, Kashi A, *et al*.: **Diverse database and machine learning model to narrow the generalization gap in RNA structure prediction**. *Sci Adv* 2026, **12**. eadz4967.
A very recent article proposing both a novel database for Chemical probing data and an associated model eFold. The authors report very good performance of the model. It is overall an appealing model, considering our perspective on chemical probing as a data modality to be more systematically integrated by future models

53. Leamy KA, Assmann SM, Mathews DH, Bevilacqua PC: **Bridging the Gap Between In Vitro and In Vivo RNA Folding**. *Q Rev Biophys* 2016, **49**:e10.

54. * Borovská I, Zhang C, Dülk S-LJ, Morandi E, Cardoso MFS, Bourkia BM, van den Homberg DAL, Wolfinger MT, Velema WA, Incarnato D: **Identification of conserved RNA regulatory switches in living cells using RNA secondary structure ensemble mapping and covariation analysis**. *Nat Biotechnol* 2025, https://doi.org/10.1038/s41587-025-02739-0.
The experiments and results presented in this paper show how interesting chemical probing data is, to explore not only the RNA structures themselves, but also their dynamics across different conditions

55. Allan MF, Aruda J, Plung JS, Grote SL, Taillades YJM des, Lajarte AA de, Bathe M, Rouskin S: *Discovery and quantification of long-range RNA base pairs in coronavirus genomes with SEARCH-MaP and SEISMIC-RNA*. 2024, https://doi.org/10.1101/2024.04.29.591762.

56. Marino SM, Husak V, Tebaldi T: *FoldARE, an RNA secondary structure analysis and prediction tool via generative pseudo-SHAPE modeling*. 2026, https://doi.org/10.64898/2026.03.04.709501.

57. Varenyk Y, Spicher T, Hofacker IL, Lorenz R: **Modified RNAs and predictions with the ViennaRNA Package**. *Bioinformatics* 2023, **39**. btad696.

58. Zhang C, Borovská I, Iobashvili T, Morandi E, Lionnez M, Olayinka OS, de Boer R, Clamer M, Witte MD, Pors K, *et al*.: **RNA functional modulation by Mitoxantrone via RNA structural ensemble repartitioning**. *Nat Commun* 2026, https://doi.org/10.1038/s41467-026-70801-9.

59. Horlacher M, Wagner N, Moyon L, Kuret K, Goedert N, Salvatore M, Ule J, Gagneur J, Winther O, Marsico A: **Towards in silico CLIP-seq: predicting protein-RNA interaction via sequence-to-signal learning**. *Genome Biol* 2023, **24**:180.

60. Ho J, Jain A, Abbeel P: *Denoising diffusion probabilistic models*. 2020. *arXiv.org*.

61. Wang Z, Feng Y, Tian Q, Liu Z, Yan P, Li X: **RNADiffFold: generative RNA secondary structure prediction using discrete diffusion models**. *Briefings Bioinf* 2025, **26**. bbae618.

62. ** Li Z, Zhu J, Hong X, Mu J, Zheng Z, Cui T, Sun Y, Wei T, Chen H-F: **DynaRNA: accurate dynamic RNA conformation ensemble generation with diffusion model**. *Commun Biol* 2025, **8**:1472.
A particularly interesting model based on the diffusion-based architecture. As we suggested in our perspective section, such architectures are particularly appealing in the context of RNA sequences populating a landscape of alternative structures with varying proportions and kinetics, rather than a single static structure. This diffusion-based model is successfully exploiting this concept.

63. Zhang Z, Jin R, Chao L, Xu G, Zhang Y, Zhou G, Yin D, Guo Y, Fu Y, Yang Y, *et al*.: *RNAGenesis: a generalist foundation model for functional RNA therapeutics*. 2025, https://doi.org/10.1101/2024.12.30.630826.

64. Ren Y, Chen Z, Qiao L, Jing H, Cai Y, Xu S, Ye P, Ma X, Sun S, Yan H, *et al*.: **BEACON: benchmark for comprehensive RNA tasks and language models**. In *Proceedings of the 38th*

international conference on neural information processing systems. Curran Associates Inc.; 2024:92891–92921.

65. Sloma MF, Mathews DH: **Exact calculation of loop formation probability identifies folding motifs in RNA secondary structures**. *RNA* 2016, **22**:1808–1818.

66. Cordero P, Lucks JB, Das R: **An RNA Mapping DataBase for curating RNA structure mapping experiments**. *Bioinformatics* 2012, **28**:3006–3008.

67. Singh J, Hanson J, Paliwal K, Zhou Y: **RNA secondary structure prediction using an ensemble of two-dimensional deep neural networks and transfer learning**. *Nat Commun* 2019, **10**:5407.

68. Bernard C, Postic G, Ghannay S, Tahi F: **Has AlphaFold3 achieved success for RNA?** *Acta Crystallogr Sect Struct Biol* 2025, **81**:49–62.

69. Runge F, Farid K, Franke JKH, Hutter F: *RnaBench: a comprehensive library for in silico RNA modelling*. 2024, https://doi.org/10.1101/2024.01.09.574794.

70. Omnes L, Angel E, Bartet P, Radvanyi F, Tahi F: **A divide-and-conquer approach based on deep learning for long RNA secondary structure prediction: Focus on pseudoknots identification**. *PLoS One* 2025, **20**, e0314837.

71. Langeberg CJ, Kim T, Nagle R, Meredith C, Garuadapuri DA, Doudna JA, Cate JHD: *Improving RNA secondary structure prediction through expanded training data*. 2025, https://doi.org/10.1101/2025.05.03.652028.

72. Westhof E, Sun H, Bu F, Miao Z: **The RNA-Puzzles Assessments of RNA-Only Targets in CASP16**. *Proteins: Struct, Funct, Bioinf* 2026, **94**:218–229.

73. Bernard C, Postic G, Ghannay S, Tahi F: **State-of-the-RNArt: benchmarking current methods for RNA 3D structure prediction**. *NAR Genom Bioinf* 2024, **6**. lqae048.

74. Arora R, Angelo M, Choe CA, Kollasch AW, Qu F, Shearer CA, Weitzman R, Gazizov A, Gurev S, Xie E, *et al*.: *RNAGym: benchmarks for RNA fitness and structure prediction*. 2025.

75. Wyss L, Mallet V, Karroucha W, Borgwardt K, Oliver C: A comprehensive benchmark for RNA 3D structure-function modeling. *arXiv.org* 2025.

76. Zhu H, Li R, Chang A, Li M, Chen H, Xiong P, Zhou SK: *NCBench and NCfold: a benchmark and closed-loop framework for RNA non-canonical base-pair prediction*. 2025.

77. Wang H, Lin W, Xiao H, Ye Q, Zhang Y: *RNAscope: benchmarking RNA language models for RNA sequence understanding*. 2025.

78. Kryshtafovych A, Schwede T, Topf M, Fidelis K, Moult J: **Progress and Bottlenecks for Deep Learning in Computational Structure Biology: CASP Round XVI**. *Proteins: Struct, Funct, Bioinf* 2026, **94**:5–14.

79. Bu F, Adam Y, Adamiak RW, Antczak M, de Aquino BRH, Badepally NG, Batey RT, Baulin EF, Boinski P, Boniecki MJ, *et al*.: **RNA-Puzzles Round V: blind predictions of 23 RNA structures**. *Nat Methods* 2025, **22**:399–411.