

# Integrated experimental and AI innovations for RNA structure determination

Received: 9 August 2025

Accepted: 2 December 2025

Published online: 05 January 2026



Wenkai Wang<sup>1,2</sup>, Baoquan Su<sup>1,2</sup>, Zhenling Peng<sup>1</sup>✉ & Jianyi Yang<sup>1</sup>✉

RNAs act as crucial ‘social’ mediators within the cell, orchestrating a wide array of biological processes. Their functionality hinges on their complex three-dimensional structures, which dictate stability, binding specificity and molecular interactions. In recent years, a surge of research has focused on solving and/or predicting RNA structures to unlock their functional secrets. However, the dynamic nature and unique physicochemical properties of RNAs pose notable challenges to accurate structural determination. This Perspective reviews recent breakthroughs in RNA structure determination, driven by innovative experimental techniques, such as cryo-electron microscopy, alongside artificial intelligence-based tools inspired by advances in protein structure prediction. We explore how integrative approaches that combine experimental and computational methods are proving particularly powerful in illuminating the RNA world, offering enhanced resolution and scalability. We discuss remaining challenges and opportunities to overcome these hurdles. By integrating experiments with computation, the field is poised to deepen our understanding of RNA biology, paving the way for transformative applications in biotechnology and medicine.

RNA, historically regarded merely as a messenger for genetic information, is now recognized as a pivotal molecule in the origin of life, as proposed by the RNA World Hypothesis<sup>1</sup>. Like proteins, many RNAs, such as tRNA and rRNA, can fold into specific three-dimensional (3D) structures to perform diverse functions<sup>2</sup>. Through intricate base-pairing interactions, including both the canonical Watson–Crick base pairs and a rich vocabulary of noncanonical pairings, RNA forms complex secondary structures like stems, loops and bulges. These elements then twist and pack together in 3D space, forming tertiary structures that allow RNA to bind with other molecules, catalyze biochemical reactions and regulate cellular processes.

Since the first tRNA computational model (1969)<sup>3</sup> and crystal structure (1974)<sup>4</sup> were published, the number of experimental RNA structures has grown almost exponentially in the Protein Data Bank (PDB<sup>5</sup>; Fig. 1a). Despite this impressive growth, a substantial data gap persists between RNA-containing entries (~8,000 in total, with ~2,000

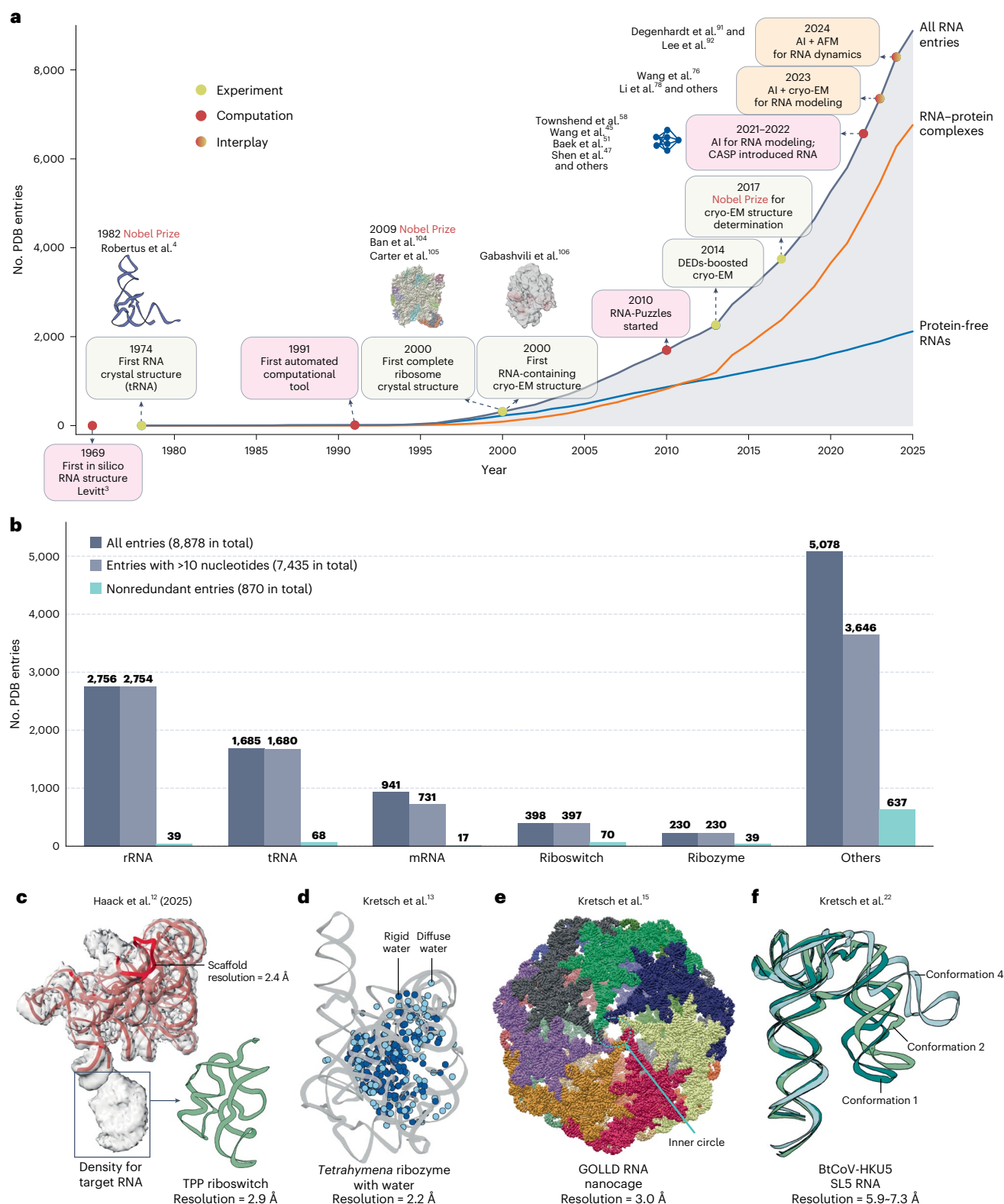
being protein free) and protein structures (>200,000). Furthermore, the existing RNA structure data are heavily biased toward certain types (for example, tRNAs and rRNAs). After removing redundancy and structures with fewer than ten nucleotides, the number of unique RNA structures reduces from ~8,000 to ~800 (Fig. 1b).

## Why RNA structure is hard to solve

RNA differs from protein in several properties that present considerable challenges to its structure determination<sup>6</sup>. First, RNA is highly flexible, adopting a vast ensemble of conformations to conduct various biological functions. This flexibility hinders the enhancement of resolution through averaging techniques in high-resolution experiments like X-ray crystallography and cryo-electron microscopy (cryo-EM). Second, its negatively charged backbone needs a precise ionic environment for stable folding, a condition that is difficult to replicate experimentally or model computationally. Third, beyond the

<sup>1</sup>MOE Frontiers Science Center for Nonlinear Expectations, State Key Laboratory of Cryptography and Digital Economy Security, Research Center for Mathematics and Interdisciplinary Sciences, Shandong University, Qingdao, China. <sup>2</sup>These authors contributed equally: Wenkai Wang, Baoquan Su.

✉e-mail: [zhenling@email.sdu.edu.cn](mailto:zhenling@email.sdu.edu.cn); [yangjy@sdu.edu.cn](mailto:yangjy@sdu.edu.cn)



**Fig. 1 | Growth of RNA structural data and examples for recent cryo-EM advancements.** **a**, A timeline plotting the exponential growth of RNA-containing entries in PDB (data as of July 2025), annotated with key milestones in experimental (yellow), computational (red) and integrated approaches (hybrid)<sup>3,4,45,47,51,58,76,78,91,92,104–106</sup>. DEDs, direct electron detectors. **b**, Statistics for RNA types in PDB (data as of July 2025). The nonredundant entries are derived by removing sequences with over 80% identity using the CD-HIT-EST program<sup>103</sup>, which by default excludes short fragments (ten or fewer nucleotides) for improved cluster stability. **c–f**, Examples of recent advancements in RNA

structure determination by cryo-EM, including the structure of a small TPP riboswitch (PDB 9C6K; green) determined using an RNA scaffold (PDB 9C6J; red) strategy<sup>12</sup> (**c**; note that the different PDB IDs are a result of masked local refinement, a process where the scaffold and target RNA were computationally isolated and refined individually); the *Tetrahymena* ribozyme (gray) shown with its ordered (deep blue) and diffuse (light blue) water molecules (PDB 9CBU and 9CBX; **d**)<sup>13</sup>; the structure of the GOLLD RNA 14-mer nanocage (PDB 9MEE)<sup>15</sup>, with its inner circle highlighted (**e**) and three resolved conformations of a viral stem-loop 5 (SL5) RNA (PDB 8UYE, 8UYG and 8UYJ; **f**)<sup>22</sup>.

canonical Watson–Crick base pairs (A•U, G•C), RNA uses a variety of noncanonical interactions (for example, Hoogsteen pairs and wobble pairs) and tertiary motifs (for example, pseudoknots, G-quadruplexes and kissing loops) to build its functional structures, further stabilized by base stacking<sup>2</sup>. These interactions are highly sensitive to environmental fluctuations and are hard to capture *in silico* owing to poorly characterized energetic parameters.

These properties pose challenges to mainstream experimental techniques<sup>7</sup>. X-ray crystallography, the traditional gold standard, is hampered by the poor crystallization and phasing difficulties of RNA molecules, which primarily arise from their high surface charge, limited chemical diversity and susceptibility to degradation. Nuclear magnetic resonance (NMR) spectroscopy, although ideal for studying dynamics, is generally restricted to small RNAs (<50 nucleotides). For larger RNAs, the limited chemical shift dispersion among the four nucleotide types leads to intractable spectral overlap. Cryo-EM, the modern solution for large assemblies, struggles with smaller protein-free RNAs due to the low signal-to-noise ratios (SNRs) in their electron micrographs. The conformational heterogeneity of RNAs also complicates the particle alignment required for high-resolution reconstruction<sup>8</sup>. These difficulties result in a limited number of solved RNA 3D structures, especially for protein-free RNAs, hindering a comprehensive understanding of RNA structures and functions.

## Enhanced experimental approaches for RNA structure determination

The experimental landscape of RNA structural biology is evolving rapidly, with recent breakthroughs focused on two major aspects. First, the application of cryo-EM has broadened substantially to resolve structures for an expanding range of RNA molecules. Second, there is a growing emphasis on capturing conformational dynamics beyond static snapshots.

### Recent progress in cryo-EM for RNA structure determination

Cryo-EM, powered by single-particle analysis (SPA), has transformed structural biology, a breakthrough recognized with the 2017 Nobel Prize in Chemistry. However, cryo-EM was long considered most suitable for large assemblies such as ribosomes and viruses, while struggling to resolve small, protein-free RNAs (for example, <100 kDa). This limitation is reflected in PDB; before 2022, for instance, no protein-free RNA entry determined by cryo-EM had surpassed the resolution barrier of 3 Å (ref. 9). Fortunately, recent experimental efforts, along with advances in detector technology, have dramatically enhanced the capabilities of cryo-EM, increasing the number of cryo-EM entries for protein-free RNA from 50 in 2022 to 171 by July 2025. Crucially, several of these new structures have achieved sub-3-Å resolution (Supplementary Table 1).

One notable advancement is the use of RNA ‘scaffolds’ to obtain high-resolution structures of small RNAs. This strategy involves introducing a large scaffold RNA molecule (typically >100 kDa) with favorable biophysical properties and high resolution, such as group I or group II introns<sup>9–12</sup>. The scaffold increases the overall molecular weight of the sample, which in turn enhances the SNR and facilitates particle picking/alignment during 3D reconstruction. This approach has successfully resolved several RNA-only structures at the highest resolutions of ~3 Å. For example, the resolution of *Tetrahymena* group I intron was improved to sub-3 Å by scaffolding itself<sup>9</sup>, that is, assembling itself into closed rings by installing kissing-loop sequences onto its functionally nonessential stems, which can yield multiplied molecular weights and mitigated structural flexibility. A more typical example is the thiamine pyrophosphate (TPP) riboswitch with much smaller molecular weights (<30 kDa; Fig. 1c), which was scaffolded by a previously solved group IIC intron via covalent attachment through a rigid helix<sup>12</sup>. This strategy improved the resolution to 2.5 Å, allowing the observation of the ligand-binding pocket. Although the full potential of scaffold-based methods remains to be explored, this technique is a

crucial development for overcoming the previous inability of cryo-EM to determine the structures of small RNAs.

The improved resolution of cryo-EM also facilitates the direct observation of small ligands or water molecules. For example, through extensive data collection and a next-generation electron detector, researchers enhanced the resolution of the *Tetrahymena* ribozyme, a highly solvated RNA molecule, from 3.1 Å to ~2 Å (ref. 13). These high-resolution density maps enabled the visualization of distinct water networks surrounding the RNA. Specifically, ordered, rigid water molecules (Fig. 1d, deep blue) were initially identified by the automated segmentation-guided water and ion modeling (SWIM) algorithm, which identifies water molecules based on density map segmentation and physicochemical criteria. Surprisingly, the maps also revealed complex, diffuse water networks (Fig. 1d, light blue) that are typically absent from atomic models, an assignment supported by molecular dynamics simulations. This work highlights the growing capability of cryo-EM to provide a more complete picture of biomolecular structures, including their dynamic solvent environments.

Another recent advance is the determination of the RNA oligomeric structures without protein involvement. Using cryo-EM, researchers have successfully resolved the structures of several RNA oligomers in various states of assembly at resolutions of 2–4 Å (refs. 14–17; for example, the nanocage shown in Fig. 1e)<sup>15</sup>. These complex structures provide detailed insights into intermolecular interactions, the mechanisms of RNA multivalency and the principles governing the assembly of RNA-based molecular machines such as the RNA nanocage.

In addition to novel experimental strategies, sample preparation is also a critical procedure for successful cryo-EM reconstruction. To improve this process, researchers recently proposed a structured, feedback-driven protocol for optimizing RNA samples<sup>18</sup>. Unlike a standard linear workflow, this protocol is distinguished by its rigorous, iterative screening of preparation conditions, which is coupled with data processing pipelines designed to resolve conformational heterogeneity and modeling algorithms that build atomic coordinates from the resulting density maps. This protocol was subsequently applied to determine the structures of the aforementioned RNA oligomers<sup>14</sup>. Furthermore, while extending cryo-EM to *in situ* structure determination remains a substantial challenge, recent studies have begun to explore this frontier<sup>19–21</sup>. In conclusion, continued progress is expected to further solidify the key role of cryo-EM in advancing RNA structural biology.

### Determining alternative conformations and dynamic structures

Beyond static structure, understanding the inherent flexibility and dynamics of RNA molecules is crucial for a comprehensive elucidation of their functions, but is more challenging. Cryo-EM has emerged as a powerful technique to explore these dynamics, as it can directly visualize structural heterogeneity within numerous single particles from electron micrographs. This capability has been greatly enhanced by methodological advances designed to improve SNR and resolve conformational states, including sample preparation, sophisticated two-dimensional (2D)/3D classification and 3D variability analysis, which together characterize both discrete conformations and continuous motions<sup>22–25</sup>. For example, cryo-EM was used to determine the ‘cryoensemble’ of several coronavirus stem-loop 5 RNAs (such as the one shown in Fig. 1f)<sup>22</sup>, capturing both the discrete conformational states and the inherent flexibility of these molecules.

Meanwhile, time-resolved cryo-EM has been developed to capture transient states by flash-freezing samples at precise time points after reaction initiation<sup>26–28</sup>. This method enables the visualization of short-lived intermediates and has been applied to reveal the transient intermediates of bacterial RNA polymerase during transcription<sup>29</sup>. Although this technique remains challenging, as it requires precise



temporal control and numerous particles to observe low-abundance states, its application has shown potential for studying the transient dynamics of RNAs using cryo-EM, a field that is still nascent.

Complementing SPA, cryo-electron tomography (cryo-ET) allows for the investigation of RNA dynamics in situ, providing a direct view of their structural states within the cell. A common strategy, resembling that used in SPA, involves aligning and averaging subtomograms to improve SNR for high-resolution 3D reconstruction, which has been used to observe the conformational changes of a ribosome during translation inside a bacterial cell<sup>30</sup>. By contrast, non-averaging methods such as individual-particle electron tomography<sup>31</sup> reconstruct each particle individually, thereby capturing transient and kinetically trapped states missed by averaging approaches. This capability has been demonstrated in the analysis of RNA origami self-folding. However, although individual-particle electron tomography provides a more comprehensive view of the conformational landscape in principle, its reliance on stringent conditions to overcome low SNR presents a barrier to its widespread adoption.

### Recent progress in other experimental approaches

Apart from cryo-EM and cryo-ET, other experimental approaches are also advancing RNA structural biology. For example, antibody-assisted RNA crystallography uses antigen-binding fragments (Fabs) as chaperones to facilitate the crystallization of otherwise intractable RNA molecules. This technique has become particularly useful with the discovery of Fab BL3-6, a versatile module that can be engineered into nearly any target RNA. The Fab BL3-6 system, including its engineered variants, has recently accelerated the structure determination of several challenging RNAs<sup>32–38</sup>, such as the *yjdF* riboswitch<sup>32</sup> and coxsackievirus B3 cloverleaf RNA<sup>33</sup>.

Other experimental techniques are also providing crucial insights into RNA dynamics. Atomic force microscopy (AFM), which resolves the topography of single molecules without averaging, is particularly suitable for studying structural heterogeneity. For example, low-resolution AFM snapshots visualizing the vast conformational ensemble of the cobalamin riboswitch<sup>39</sup> guided subsequent cryo-EM determination of the atomic models for this riboswitch in different states<sup>40</sup>. In parallel, NMR spectroscopy continues to yield new insights into RNA dynamics. For instance, a recent study used a new NMR strategy leveraging chemical shift perturbations to rapidly quantify the conformational propensity of HIV-1 transactivation response RNA, directly linking this biophysical property with the RNA's activity in cells<sup>41</sup>.

## Artificial intelligence for RNA structure prediction

Despite accelerating experimental efforts, the growth in available RNA structural data still lags far behind the demand from the research community. This gap highlights the urgent need for robust computational methods capable of rapid and accurate RNA structure prediction, a long-standing challenge since the 1960s<sup>3</sup>. Following the revolutionary success of artificial intelligence (AI) in protein structure prediction, especially the Nobel Prize-winning method AlphaFold2 (ref. 42), the field is now applying similar AI strategies to advance RNA 3D structure prediction.

### Recent AI methods for RNA structure prediction

AI methods follow a general workflow for RNA structure prediction (Fig. 2). These approaches typically rely on two critical input features: (1) multiple sequence alignment (MSA) and/or embedding from large language models, which provide coevolutionary signals for deducing long-range internucleotide contacts, and/or (2) predicted secondary structure, which offers base-pairing priors essential for 3D folding. Methodologically, these approaches can be divided into three groups, that is, two-step, end-to-end and complex prediction, which are introduced below.

The initial attempts of AI-based RNA 3D structure prediction adopted a two-step paradigm, largely inspired by AlphaFold2's transformer-based distogram prediction and trRosetta's energy minimization<sup>43</sup>. Representative approaches include DeepFoldRNA<sup>44</sup> and trRosettaRNA<sup>45</sup>. These methods first use transformer networks to predict a set of geometric restraints, which are then used to guide the folding and reconstruction of an all-atom 3D structure via energy minimization. The main benefit of this paradigm lies in its training efficiency: by focusing on one-dimensional (1D) and 2D geometric targets, it simplifies model training compared to fully end-to-end systems<sup>46</sup>. Furthermore, the rigorous energy minimization procedure improves the physicochemical plausibility of the generated structure models<sup>45</sup>.

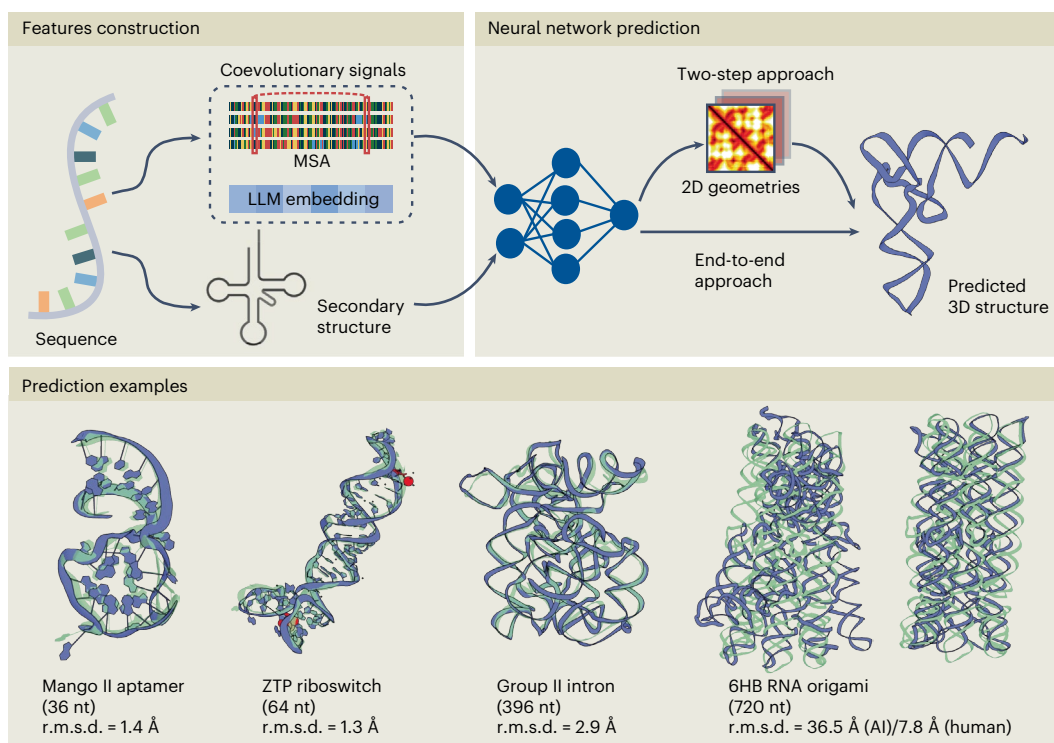
However, the efficiency of two-step approaches is constrained by two primary drawbacks: suboptimal optimization resulting from their simplified training objectives and slow inference speeds caused by the computationally intensive energy minimization step. This led the field to rapidly adapt the full end-to-end architecture of AlphaFold2 for RNA, resulting in methods such as RhoFold+<sup>47</sup> and NuFold<sup>48</sup> that directly predict all-atom coordinates using neural networks. These advances also involve integrating RNA language models to compensate for the often-sparse coevolutionary signals in MSAs, a feature leveraged by RhoFold+. Although pure AI prediction can sometimes result in severe structural violations, a third strategy, represented by DRfold<sup>49</sup>, provides a hybrid solution that combines coarse-grained end-to-end potentials with geometric restraints to improve both predictive accuracy and the physicochemical plausibility.

The scope of AI methods has rapidly expanded beyond RNA monomers, with a new generation of models such as AlphaFold3 (AF3)<sup>50</sup>, RoseTTAFoldNA (RFNA)<sup>51</sup> and RoseTTAFold All-Atom<sup>52</sup> aimed at predicting nucleic acid–protein assemblies and other vital biomolecular interactions. Specifically, RFNA pioneered the deep learning-based end-to-end prediction of protein–RNA complexes with its three-track network architecture. RoseTTAFold All-Atom then broadened this scope to include interactions with small molecules and ions. Currently, AF3 also aims to predict a wide range of complexes, but it uses a new diffusion model that enhances performance by enabling extensive sampling.

### Efforts for assessing RNA structure prediction

The scarcity of newly released RNA structures poses a challenge for the fair and rigorous assessment of prediction methods. To overcome this, community-wide blind assessment platforms like RNA-Puzzles and the Critical Assessment of Structure Prediction (CASP) provide crucial benchmarks. RNA-Puzzles, a classical platform, has provided over 60 targets for blind prediction since 2010 (ref. 53). More recently, CASP introduced an RNA category in its 15th experiment (CASP15, 2022)<sup>54</sup>, which involves 12 targets, including natural and synthetic RNAs, and attracted over 40 participating groups. CASP16, held in 2024, expanded this effort<sup>55</sup>, with the number of targets growing to over 40 and participation increasing to more than 60 groups. Notably, CASP16 also marked the first blind competition that introduced categories for nucleic acid–protein and nucleic acid–ligand complexes. These community-wide efforts are vital for objectively evaluating prediction methods and driving future innovation in the field.

Evaluation has also evolved. Classical RNA metrics like root mean square deviation (r.m.s.d.), interaction network fidelity and deformation index have known limitations, including dependency on RNA topology (r.m.s.d. and deformation index) and a narrow focus on local 2D interactions (interaction network fidelity). To address these issues, the field has adapted length-independent, 3D-aware metrics from protein structure assessment<sup>54,55</sup>, such as global distance test-total score (GDT-TS), template modeling score (TM-score) and local distance difference test (lDDT). The CASP community has used a scoring system that integrates this diverse set of metrics<sup>54</sup> and recently expanded to include motif and stoichiometry analyses, providing a



**Fig. 2 | Overview of the AI-driven RNA structure prediction pipeline and example predictions.** Top, a general pipeline of AI-driven RNA structure prediction. Starting from an input nucleotide sequence, two sets of features are constructed: the coevolutionary signals (MSA and/or large language model (LLM) embedding) and/or predicted secondary structure. The network then either predicts 2D geometries for subsequent 3D reconstruction (two-step approach) or directly outputs the full-atom 3D structure (end-to-end approach). Bottom, representative examples showing computational models (purple)

superimposed on their experimental structures (green). From left to right: a Mango II aptamer (PDB 8U5K), a ligand-binding ZTP riboswitch (PDB 8VQV; ligands and ions shown in red) and two challenging large RNAs, a group II intron (CASP16 R1241) and a synthetic RNA origami (CASP15 R1138, PDB 7PTL). The predictions for the first three examples and the left origami model were generated by trRosettaRNA2, whereas the right origami model was produced by the human expert group Alchemy\_RNA2; nt, nucleotides.

multidimensional view of predicted structure models<sup>55</sup>. Despite these efforts, a crucial challenge remains: the suitability of these metrics for RNA is not fully established. In contrast to proteins, no single metric has yet proven sufficient to reliably describe the accuracy of a predicted RNA structure.

In parallel with ground-truth-aware assessments, recent years have seen the emergence of scoring functions for evaluating predicted RNA structures. These span from knowledge-based approaches to advanced deep learning models<sup>56–60</sup>. A notable example, ARES<sup>58</sup>, uses an equivariant graph neural network to estimate model quality. These tools are valuable for guiding model selection and enhancing accuracy, yet their reliability is hampered by fundamental challenges: the difficulty of designing accurate potentials and the scarcity of training data for deep learning models. As a result, the development of highly reliable scoring functions for RNA remains a considerable challenge, leaving the field lagging behind the more established methods available for proteins.

### Advances and challenges of AI-based methods

Comprehensive benchmark tests demonstrate that AI-based methods outperform traditional automated approaches, such as those based on energy minimization or fragment assembly. For example, trRosettaRNA achieves an average r.m.s.d. of less than half that of the traditional automated methods<sup>45</sup>. In addition to improved accuracy, the prediction speed has also been increased by graphics processing unit (GPU)-powered neural network inference. For example, according to the estimation in the AF3 paper<sup>50</sup>, AF3 can predict the 3D structure of a 1,024-nucleotide RNA within 30 s starting from precomputed features. Undoubtedly, AI has boosted automated RNA 3D structure prediction.

The bottom of Fig. 2 presents three examples where AI methods have yielded accurate results, including an aptamer, a ligand-binding riboswitch and a larger RNA with ~400 nucleotides.

Despite this recent progress, AI in RNA structure prediction has not yet experienced its ‘AlphaFold moment’<sup>61</sup>, as automated methods still lag behind human experts. This performance gap is clearly demonstrated in the CASP and RNA-Puzzles blind assessments<sup>53–55</sup>, where no automated approach, whether traditional or AI driven, has matched the performance of top human groups, particularly on challenging targets with novel and complex topologies (such as the synthetic RNA origami shown in Fig. 2). This observation contrasts with the trend in protein structure prediction, where human expert interventions only have limited effects in improving automated models<sup>62</sup>. As noted by the leading human groups in CASP15 and CASP16 (refs. 63–67), human interventions remain essential, particularly for modeling challenging targets such as the synthetic RNA origami.

Several key challenges hinder AI-based modeling of RNA 3D structures. First, the aforementioned scarcity of 3D data is a primary constraint on the power of data-driven AI approaches. Second, due to algorithmic and database limitations, the MSAs for RNA are typically of low quality<sup>61</sup>, providing weak or even erroneous coevolutionary signals that hamper the prediction accuracy. Third, current AI predictors largely neglect the intrinsic dynamics and environmental sensitivity (for example, pH, ions) of RNA, which biases the learning process toward the static snapshots deposited in PDB. Moreover, although methods like AF3 and RFNA have been developed for modeling RNA–protein complexes, the sparsity of both intermolecule coevolutionary signals and available training data limits their ability to generalize to novel complexes.

Recent methods like trRosettaRNA2 (ref. 68) and DRfold2 (ref. 69) have been proposed to mitigate the 3D data limitations by leveraging 2D structure priors and language models, respectively. Still, the performance gap with leading human experts remains. This was evident in the recent CASP16 assessment, where our trRosettaRNA2-based server (Yang-Server)<sup>62</sup>, despite being the top-performing automated group, still ranked behind three human groups<sup>55</sup>.

## Interplay between experiment and computation

Despite their distinct challenges, experimental and computational approaches are highly complementary, and their integration is essential for future progress. Although computational techniques facilitate rapid and high-throughput analysis, experiments provide indispensable, target-specific restraints. Their interplay already shows potential in both fields. Computationally, the impact is clear in accelerating cryo-EM structure determination<sup>22,23,70,71</sup>. Versatile map-to-model pipelines like Phenix<sup>72</sup> provide an automated framework for diverse macromolecules, while specialized tools such as auto-DRRAFTER<sup>70</sup> optimize the process for RNA by incorporating experimentally derived secondary structure information within an iterative, consensus-driven modeling framework. Beyond atomic model building, computational methods also play a vital role in data interpretation: clustering and simulations help validate RNA dynamics observed experimentally<sup>22</sup>, and algorithms like SWIM can identify ordered water molecules within high-resolution density maps<sup>13</sup>. This relationship is reciprocal. Experimental data, such as chemical shifts and reactivities, are helpful not only for guiding and refining computational models<sup>73,74</sup> but also for training advanced foundation models like RibonanzaNet<sup>75</sup> by providing direct supervision.

The interplay between experiment and computation is entering a new phase, driven by the development of methods specifically designed for integration. These new approaches combine experimental data with computational engines, aiming to break new ground in RNA structure determination (Fig. 3).

## Predicting RNA all-atom structures from cryo-EM maps

Although traditional approaches like auto-DRRAFTER have been instrumental in cryo-EM-based RNA structure determination, they often rely on predefined secondary structures and time-consuming Monte Carlo searches, hindering their high-throughput application. To address these limitations, deep learning has recently been applied to improve both the quality and efficiency of this process, that is, predicting all-atom structures from density maps.

The general pipeline for these methods is illustrated in Fig. 3 (top). Initial efforts in this direction include CryoREAD<sup>76</sup> and DeepTracer-2.0 (ref. 77), which were developed for RNA and nucleic acid–protein complexes, respectively. These methods use a two-stage framework that first detects nucleotide components (phosphates, sugars and bases) using a 3D U-Net and then builds an all-atom model via backbone tracing and sequence assignment. This was subsequently improved by EMRNA<sup>78</sup>, which used an advanced Swin-Conv-UNet network and an optimized postprocessing procedure. Further extending this capability, EM2NA<sup>79</sup> enabled the automated identification and modeling of nucleic acids directly from raw complex maps. Moreover, the development of tools like ModelAngelo<sup>80</sup> demonstrates the power of AlphaFold2-like architectures for improving model-building accuracy across both proteins and RNAs.

These deep learning-based methods achieve notable improvement in both accuracy and speed. For example, for input maps at resolutions of 2–6 Å, EMRNA achieved a median r.m.s.d. of ~2 Å (for example, the group I intron shown in Fig. 3), a substantial improvement over auto-DRRAFTER (>6 Å). Moreover, EMRNA can build a 100-nucleotide structure within 3 min. The development of such rapid and accurate tools is crucial for accelerating the pace of RNA structural biology. However, a key limitation is that current deep learning-based methods are generally applicable only to high-resolution cryo-EM maps (for

example, <4 Å). Accurate and automated structure determination from medium- and low-resolution maps remains a substantial challenge that requires future collaborative efforts to address<sup>81</sup>.

## Exploring RNA dynamics from experimental data

Although recent experimental efforts have demonstrated the potential for exploring RNA structural dynamics, their success typically relies on manual case-by-case analysis, such as elaborate sample preparation, particle selection and clustering or expert interpretation of 3D map variability. Such reliance on manual intervention hinders the automated and universal applications, highlighting the urgent need for robust computational solutions that can systematically characterize RNA dynamics.

In principle, 2D particle images offer a direct window into RNA structural dynamics, as they reflect the full range of conformational heterogeneity among samples. The primary challenge, however, is to overcome the low SNR of these images to distinguish true structural variations from experimental noise and different particle orientations. Deep learning can offer a solution<sup>82–88</sup>. A prominent approach, exemplified by methods like CryoDRGN<sup>82</sup>, uses deep generative models (for example, variational autoencoder) to reconstruct heterogeneous 3D maps from 2D particle images (Fig. 3, middle). These models encode the structural information from each particle into a low-dimensional ‘latent space’. By traversing this latent space, one can generate a smooth trajectory of 3D maps that visualizes a continuous conformational change. This methodology was subsequently extended to *in situ* reconstruction from cryo-ET data<sup>89,90</sup>. However, while promising for large complexes like the ribosome, their applicability to protein-free RNAs and the subsequent automated building of all-atom models from their output heterogeneous maps has yet to be validated. Moreover, the requirement to retrain the latent representation for each new dataset is computationally demanding. Nevertheless, these generative approaches represent an important advance in the automated exploration of structural dynamics directly from raw particle images and can serve as a blueprint for developing similar methods specifically for RNA.

Another key example of integrating computation and experiments is the AFM-based determination of RNA conformers. Recognizing the low-resolution nature of AFM, a computational framework was developed to improve the usage of these data (Fig. 3, bottom)<sup>91</sup>. In this approach, initial models are generated either computationally or from other low-resolution experimental data such as small-angle X-ray scattering. These models are then fitted to AFM topographic images using coarse-grained molecular dynamics simulations. A machine learning pipeline named HORNET is then used to select the high-confidence structures for final atomic structure construction. Specifically, HORNET first applies unsupervised clustering to filter the simulated models and then uses a supervised deep neural network to score the models. This integrated framework has characterized the heterogeneous conformers of RNase P RNA (Fig. 3) and the HIV-1 Rev response element RNA at atomic resolution<sup>91,92</sup>. This work demonstrates that even low-resolution topographic data can yield high-resolution insights when coupled with computational techniques, further confirming the potential of such integrative approaches.

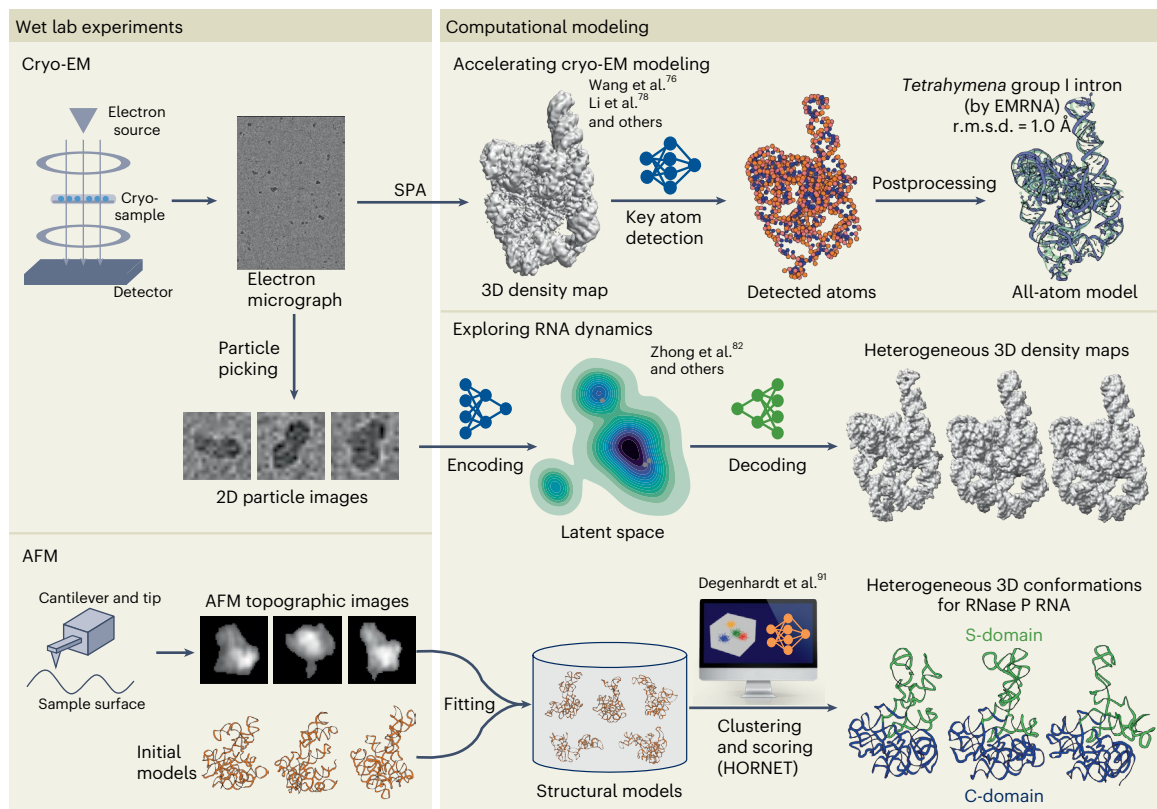
## Discussion

Building on the proven potential of integrating experimental and computational approaches, the next wave of innovation in RNA structural biology will likely emerge from their deeper interplay. We highlight three critical directions that will be discussed below: the discovery of new structured RNAs, the acceleration of structure determination pipelines and the continued advancement of predictive AI-based algorithms.

## Identification of a diverse set of new and structured RNAs

The currently solved RNAs in PDB exhibit considerable redundancy: only <10% of structures (870 of 8,878) are nonredundant after removing





**Fig. 3 | Representative efforts for integrating experimental data with computational modeling to elucidate RNA structure and dynamics.** Cryo-EM and AFM-based efforts are shown. Based on specific strategies, this figure can be divided into three parts from top to bottom. Top, cryo-EM for static structure prediction. A 3D density map can be obtained from electron micrographs via SPA. Deep learning methods like EMRNA then use a neural network to detect the key nucleotide components from the map. Subsequent postprocessing steps, including backbone tracing and sequence assignment, are then used to build a final all-atom model. Middle, cryo-EM for dynamics. Alternatively, 2D particle images picked from raw micrographs can be leveraged by AI methods like CryoDRGN to reconstruct a series of heterogeneous 3D density maps with a

variational autoencoder encoder-decoder architecture. Such methods enable the exploration of dynamics for large biomolecules such as the ribosome. Bottom, AFM for dynamics. The HORNET framework demonstrates the power of AFM for studying RNA dynamics. In this framework, initial models (from low-resolution experiments or structure prediction) are iteratively fitted into the AFM topographic images to simulate numerous structural models. Unsupervised clustering and a supervised scoring neural network are then used to filter and select high-confidence models for final atomic structure construction. This framework has successfully characterized the structural heterogeneity of targets like RNase P RNA.

redundancy at 80% sequence identity (Fig. 1b). For example, the ROOL RNA nanocage structure has been independently solved by four different research groups<sup>14–17</sup>. Although these repeated structures reinforce the validity of this structure, it would be beneficial to allocate research efforts toward a broader range of diverse RNAs.

However, not all RNAs are capable of folding into stable 3D structures. Therefore, the initial step should be to identify a diverse set of new and structured RNAs, which can help minimize the costs associated with experimental determination. A comparative analysis of specific subsets of intergenic regions identified 224 promising candidates<sup>93</sup>, including the recently solved ROOL RNA nanocage<sup>14–17</sup>. New AI-driven approaches could be developed to further expand this list. Nevertheless, it is essential to carefully benchmark these methods to minimize the false-positive rate. Ultimately, solving the structures of a diverse array of new and structured RNAs will help illuminate the RNA world.

### Accelerating RNA structure determination

The integrative efforts provide a path to accelerating experimental structure determination. In cryo-EM, AI-powered tools are streamlining complex workflows. For instance, 3D reconstruction methods such as CryoDRGN simplify the classification of heterogeneous conformational states, whereas new map-to-model tools like EMRNA enable the rapid and accurate interpretation of density maps. Furthermore, AI-based structure prediction provides valuable molecular hypotheses

for experimental analysis. As evidenced by the CASP15 assessment<sup>94,95</sup>, even predicted models of moderate accuracy can be effectively refined against density maps to generate plausible initial structures, bypassing the laborious process of traditional de novo model building. High-quality models can also be used to perform molecular replacement to resolve the persistent phase problem in X-ray crystallography<sup>94</sup>.

Another promising direction is leveraging AI's predictive power to lower the resolution requirements for structure determination, thereby saving the considerable time and resources spent on high-resolution experiments. Computational frameworks like HORNET exemplify how atomic models can be yielded from low-resolution restraints. Our prior work has shown that guiding trRosettaRNA2 with diverse secondary structure inputs can recapitulate the 3D conformational heterogeneity that mirrors AFM observations<sup>68</sup>. The clear path forward, therefore, is to develop conditional generative models that systematically incorporate a wider spectrum of experimental data by translating them into appropriate conditions, from chemical probing profiles as 1D conditions and cross-linking distances as 2D conditions to abstract AFM data as an energy-based sampling guidance.

### Improving AI-based RNA structure prediction algorithms

The promising interplay between experiment and computation hinges on robust AI-based predictive models. Current methods, which rely on MSAs and/or unsupervised language models, are limited by the simple

four-letter nucleotide alphabet and the lack of reliable base-pairing knowledge in standard homolog searches or masked language modeling. Even supervised language models can be skewed by biased experimental data for training. A promising path forward is to develop hybrid RNA language models that merge broad, unsupervised pretraining with targeted, experiment-guided fine-tuning, incorporating smarter, structure-aware tokenization and masking strategies. As seen in the protein field, the improved language model powerfully enhances structure prediction in two ways: by generating rich sequence embeddings as extra inputs<sup>96</sup> or by improving homolog searches to produce higher-quality MSAs<sup>97–100</sup>.

In addition to sequence priors, structural templates offer more powerful restraints if they exist, as shown in recent CASP challenges<sup>55,62,66,67,101</sup>. Nevertheless, their application is frequently limited by immature search algorithms and sparse template databases relative to proteins. Future work should focus on better template detection, perhaps using language models to boost sensitivity or by integrating known local motifs like kink-turns and T-loops directly into deep learning pipelines<sup>102</sup>.

## Conclusions

For decades, the intrinsic complexity of RNA molecules has hindered both experimental determination and computational prediction of RNA 3D structures. However, this landscape is rapidly evolving, propelled by the ongoing integration of experimental techniques and AI innovations. As highlighted in this Perspective, this interplay is now the principal force driving the field forward. Although a complete understanding of the RNA world remains a distant goal, the path toward it is clearer than ever. Thanks to the strengthening collaboration between structural biologists and computational scientists, we have every reason to be optimistic about our ability to illuminate the RNA world.

## Data availability

All data used in this work were obtained from publicly available sources. The 3D structure statistics and examples were sourced from PDB (<https://www.rcsb.org/>) or CASP16 (<https://predictioncenter.org/casp16/index.cgi>). Cryo-EM particle images and 3D density maps were obtained from EMPIAR (<https://www.ebi.ac.uk/empair/>) and EMDB (<https://www.ebi.ac.uk/emdb/>), respectively. The AFM images and 3D conformational ensembles for RNase P RNA were sourced from the data repository provided by the original authors<sup>91</sup> at [https://home.ccr.cancer.gov/csb/pnai/data/conformational\\_space/Conf\\_space\\_RNasePRNA/](https://home.ccr.cancer.gov/csb/pnai/data/conformational_space/Conf_space_RNasePRNA/).

## References

- Gilbert, W. Origin of life: the RNA world. *Nature* **319**, 618 (1986).
- Vicens, Q. & Kieft, J. S. Thoughts on how to think (and talk) about RNA structure. *Proc. Natl Acad. Sci. USA* **119**, e2112677119 (2022).
- Levitt, M. Detailed molecular model for transfer ribonucleic acid. *Nature* **224**, 759–763 (1969).
- Robertus, J. D. et al. Structure of yeast phenylalanine tRNA at 3 Å resolution. *Nature* **250**, 546–551 (1974).
- Burley, S. K. et al. Updated resources for exploring experimentally-determined PDB structures and Computed Structure Models at the RCSB Protein Data Bank. *Nucleic Acids Res.* **53**, D564–D574 (2025).
- Kwon, D. RNA function follows form—why is it so hard to predict? *Nature* **639**, 1106–1108 (2025).
- Deng, J. et al. RNA structure determination: from 2D to 3D. *Fundam. Res.* **3**, 727–737 (2023).
- Wu, M. & Lander, G. C. How low can we go? Structure determination of small biological complexes using single-particle cryo-EM. *Curr. Opin. Struct. Biol.* **64**, 9–16 (2020).
- Liu, D., Th  lot, F. A., Piccirilli, J. A., Liao, M. & Yin, P. Sub-3-   cryo-EM structure of RNA enabled by engineered homomeric self-assembly. *Nat. Methods* **19**, 576–585 (2022).
- Langeberg, C. J. & Kieft, J. S. A generalizable scaffold-based approach for structure determination of RNAs by cryo-EM. *Nucleic Acids Res.* **51**, e100 (2023).
- Sampedro Vallina, N., McRae, E. K. S., Hansen, B. K., Boussebayle, A. & Andersen, E. S. RNA origami scaffolds facilitate cryo-EM characterization of a Broccoli–Pepper aptamer FRET pair. *Nucleic Acids Res.* **51**, 4613–4624 (2023).
- Haack, D. B. et al. Scaffold-enabled high-resolution cryo-EM structure determination of RNA. *Nat. Commun.* **16**, 880 (2025).
- Kretsch, R. C. et al. Complex water networks visualized by cryogenic electron microscopy of RNA. *Nature* **642**, 250–259 (2025).
- Wang, L. et al. Cryo-EM reveals mechanisms of natural RNA multivalency. *Science* **388**, 545–550 (2025).
- Kretsch, R. C. et al. Naturally ornate RNA-only complexes revealed by cryo-EM. *Nature* **643**, 1135–1142 (2025).
- Ling, X. et al. Cryo-EM structure of a natural RNA nanocage. *Nature* **664**, 1107–1115 (2025).
- Zhang, S. et al. Structural insights into higher-order natural RNA-only multimers. *Nat. Struct. Mol. Biol.* **32**, 2012–2021 (2025).
- Chen, X. et al. RNA sample optimization for cryo-EM analysis. *Nat. Protoc.* **20**, 1114–1157 (2025).
- Ni, T. et al. High-resolution in situ structure determination by cryo-electron tomography and subtomogram averaging using emClarity. *Nat. Protoc.* **17**, 421–444 (2022).
- Hoffmann, P. C. et al. Structures of the eukaryotic ribosome and its translational states in situ. *Nat. Commun.* **13**, 7435 (2022).
- Cheng, J. et al. Capturing eukaryotic ribosome dynamics in situ at high resolution. *Nat. Struct. Mol. Biol.* **32**, 698–708 (2025).
- Kretsch, R. C. et al. Tertiary folds of the SL5 RNA from the 5' proximal region of SARS-CoV-2 and related coronaviruses. *Proc. Natl Acad. Sci. USA* **121**, e2320493121 (2024).
- Bonilla, S. L., Sherlock, M. E., MacFadden, A. & Kieft, J. S. A viral RNA hijacks host machinery using dynamic conformational changes of a tRNA-like structure. *Science* **374**, 955–960 (2021).
- Yang, W. et al. Structural insights into dynamics of the BMV TLS aminoacylation. *Nat. Commun.* **16**, 1276 (2025).
- Chen, J. et al. Ensemble cryo-EM reveals conformational states of the nsp13 helicase in the SARS-CoV-2 helicase replication–transcription complex. *Nat. Struct. Mol. Biol.* **29**, 250–260 (2022).
- Dandey, V. P. et al. Time-resolved cryo-EM using Spotiton. *Nat. Methods* **17**, 897–900 (2020).
- Fischer, N., Konevega, A. L., Wintermeyer, W., Rodnina, M. V. & Stark, H. Ribosome dynamics and tRNA movement by time-resolved electron cryomicroscopy. *Nature* **466**, 329–333 (2010).
- Torino, S., Dhurandhar, M., Stroobants, A., Claessens, R. & Efremov, R. G. Time-resolved cryo-EM using a combination of droplet microfluidics with on-demand jetting. *Nat. Methods* **20**, 1400–1408 (2023).
- Saecker, R. M. et al. Early intermediates in bacterial RNA polymerase promoter melting visualized by time-resolved cryo-electron microscopy. *Nat. Struct. Mol. Biol.* **31**, 1778–1788 (2024).
- Xue, L. et al. Visualizing translation dynamics at atomic detail inside a bacterial cell. *Nature* **610**, 205–211 (2022).
- Liu, J. et al. Non-averaged single-molecule tertiary structures reveal RNA self-folding through individual-particle cryo-electron tomography. *Nat. Commun.* **15**, 9084 (2024).
- Krochmal, D., Roman, C., Lewicka, A., Shao, Y. & Piccirilli, J. A. Structural basis for promiscuity in ligand recognition by yjdf riboswitch. *Cell Discov.* **10**, 37 (2024).
- Das, N. K. et al. Crystal structure of a highly conserved enteroviral 5' cloverleaf RNA replication element. *Nat. Commun.* **14**, 1955 (2023).



34. Ojha, M. et al. Structure of saguaro cactus virus 3' translational enhancer mimics 5' cap for eIF4E binding. *Proc. Natl Acad. Sci. USA* **121**, e2313677121 (2024).
35. Das, N. K., Vogt, J., Patel, A., Banna, H. A. & Koirala, D. Structural basis for a highly conserved RNA-mediated enteroviral genome replication. *Nucleic Acids Res.* **52**, 11218–11233 (2024).
36. Radakovic, A. et al. A potential role for RNA aminoacylation prior to its role in peptide synthesis. *Proc. Natl Acad. Sci. USA* **121**, e2410206121 (2024).
37. Hegde, S. et al. Mechanistic studies of small molecule ligands selective to RNA single G bulges. *Nucleic Acids Res.* **53**, gkaf559 (2025).
38. Lewicka, A. et al. Crystal structure of a cap-independent translation enhancer RNA. *Nucleic Acids Res.* **51**, 8891–8907 (2023).
39. Ding, J. et al. Visualizing RNA conformational and architectural heterogeneity in solution. *Nat. Commun.* **14**, 714 (2023).
40. Ding, J. et al. Capturing heterogeneous conformers of cobalamin riboswitch by cryo-EM. *Nucleic Acids Res.* **51**, 9952–9960 (2023).
41. Ken, M. L. et al. RNA conformational propensities determine cellular activity. *Nature* **617**, 835–841 (2023).
42. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
43. Yang, J. et al. Improved protein structure prediction using predicted interresidue orientations. *Proc. Natl Acad. Sci. USA* **117**, 1496–1503 (2020).
44. Pearce, R., Omenn, G. S. & Zhang, Y. De novo RNA tertiary structure prediction at atomic resolution using geometric potentials from deep learning. Preprint at *bioRxiv* <https://doi.org/10.1101/2022.05.15.491755> (2022).
45. Wang, W. et al. trRosettaRNA: automated prediction of RNA 3D structure with transformer network. *Nat. Commun.* **14**, 7266 (2023).
46. Peng, Z., Wang, W., Han, R., Zhang, F. & Yang, J. Protein structure prediction in the deep learning era. *Curr. Opin. Struct. Biol.* **77**, 102495 (2022).
47. Shen, T. et al. Accurate RNA 3D structure prediction using a language model-based deep learning approach. *Nat. Methods* **21**, 2287–2298 (2024).
48. Kagaya, Y. et al. NuFold: end-to-end approach for RNA tertiary structure prediction with flexible nucleobase center representation. *Nat. Commun.* **16**, 881 (2025).
49. Li, Y. et al. Integrating end-to-end learning with deep geometrical potentials for ab initio RNA structure prediction. *Nat. Commun.* **14**, 5745 (2023).
50. Abramson, J. et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500 (2024).
51. Baek, M. et al. Accurate prediction of protein–nucleic acid complexes using RoseTTAFoldNA. *Nat. Methods* **21**, 117–121 (2024).
52. Krishna, R. et al. Generalized biomolecular modeling and design with RoseTTAFold All-Atom. *Science* **384**, ead12528 (2024).
53. Bu, F. et al. RNA-Puzzles Round V: blind predictions of 23 RNA structures. *Nat. Methods* **22**, 399–411 (2025).
54. Das, R. et al. Assessment of three-dimensional RNA structure prediction in CASP15. *Proteins* **91**, 1747–1770 (2023).
55. Kretsch, R. C. et al. Assessment of nucleic acid structure prediction in CASP16. *Proteins* <https://doi.org/10.1002/prot.70072> (2025).
56. Zhang, T., Hu, G., Yang, Y., Wang, J. & Zhou, Y. All-atom knowledge-based potential for RNA structure discrimination based on the distance-scaled finite ideal-gas reference state. *J. Comput. Biol.* **27**, 856–867 (2019).
57. Tan, Y.-L., Wang, X., Shi, Y.-Z., Zhang, W. & Tan, Z.-J. rsRNASP: a residue-separation-based statistical potential for RNA 3D structure evaluation. *Biophys. J.* **121**, 142–156 (2022).
58. Townshend, R. J. L. et al. Geometric deep learning of RNA structure. *Science* **373**, 1047–1051 (2021).
59. Tarafder, S. & Bhattacharya, D. lociPARSE: a locality-aware invariant point attention model for scoring RNA 3D structures. *J. Chem. Inf. Model.* **64**, 8655–8664 (2024).
60. Liu, X. et al. Quality assessment of RNA 3D structure models using deep learning and intermediate 2D maps. Preprint at *bioRxiv* <https://doi.org/10.1101/2025.07.25.666746> (2025).
61. Schneider, B. et al. When will RNA get its AlphaFold moment?. *Nucleic Acids Res.* **51**, 9522–9532 (2023).
62. Wang, W., Luo, Y., Peng, Z. & Yang, J. Accurate biomolecular structure prediction in CASP16 with optimized inputs to state-of-the-art predictors. *Proteins* <https://doi.org/10.1002/prot.70030> (2025).
63. Chen, K., Zhou, Y., Wang, S. & Xiong, P. RNA tertiary structure modeling with BRiQ potential in CASP15. *Proteins* **91**, 1771–1778 (2023).
64. Li, J., Zhang, S. & Chen, S.-J. Advancing RNA 3D structure prediction: exploring hierarchical and hybrid approaches in CASP15. *Proteins* **91**, 1779–1789 (2023).
65. Sarzynska, J., Popena, M., Antczak, M. & Szachniuk, M. RNA tertiary structure prediction using RNAComposer in CASP15. *Proteins* **91**, 1790–1799 (2023).
66. Kagaya, Y. et al. Structure modeling protocols for protein multimer and RNA in CASP16 with enhanced MSAs, model ranking, and deep learning. *Proteins* <https://doi.org/10.1002/prot.70033> (2025).
67. Zhang, S., Li, J., Zhou, Y. & Chen, S.-J. Enhancing RNA 3D structure prediction in CASP16: integrating physics-based modeling with machine learning for improved predictions. *Proteins* <https://doi.org/10.1002/prot.26856> (2025).
68. Wang, W., Peng, Z. & Yang, J. Predicting RNA 3D structure and conformers using a pre-trained secondary structure model and structure-aware attention. Preprint at *bioRxiv* <https://doi.org/10.1101/2025.04.09.647915> (2025).
69. Li, Y., Feng, C., Zhang, X. & Zhang, Y. Ab initio RNA structure prediction with composite language model and denoised end-to-end learning. Preprint at *bioRxiv* <https://doi.org/10.1101/2025.03.05.641632> (2025).
70. Kappel, K. et al. Accelerated cryo-EM-guided determination of three-dimensional RNA-only structures. *Nat. Methods* **17**, 699–707 (2020).
71. Zhang, K. et al. Cryo-EM and antisense targeting of the 28-kDa frameshift stimulation element from the SARS-CoV-2 RNA genome. *Nat. Struct. Mol. Biol.* **28**, 747–754 (2021).
72. Terwilliger, T. C., Adams, P. D., Afonine, P. V. & Sobolev, O. V. A fully automatic method yielding initial models from high-resolution cryo-electron microscopy maps. *Nat. Methods* **15**, 905–908 (2018).
73. Watkins, A. M., Rangan, R. & Das, R. FARFAR2: improved de novo Rosetta prediction of complex global RNA folds. *Structure* **28**, 963–976 (2020).
74. Li, J., Zhang, S., Zhang, D. & Chen, S.-J. Vfold-Pipeline: a web server for RNA 3D structure prediction from sequences. *Bioinformatics* **38**, 4042–4043 (2022).
75. He, S. et al. Ribonanza: deep learning of RNA structure through dual crowdsourcing. Preprint at *bioRxiv* <https://doi.org/10.1101/2024.02.24.581671> (2024).
76. Wang, X., Terashi, G. & Kihara, D. CryoREAD: de novo structure modeling for nucleic acids in cryo-EM maps using deep learning. *Nat. Methods* **20**, 1739–1747 (2023).
77. Nakamura, A. et al. Fast and automated protein–DNA/RNA macromolecular complex modeling from cryo-EM maps. *Brief. Bioinform.* **24**, bbac632 (2023).
78. Li, T. et al. All-atom RNA structure determination from cryo-EM maps. *Nat. Biotechnol.* **43**, 97–105 (2025).

79. Li, T., Cao, H., He, J. & Huang, S.-Y. Automated detection and de novo structure modeling of nucleic acids from cryo-EM maps. *Nat. Commun.* **15**, 9367 (2024).
80. Jamali, K. et al. Automated model building and protein identification in cryo-EM maps. *Nature* **628**, 450–457 (2024).
81. Su, B., Huang, K., Peng, Z., Amunts, A. & Yang, J. CryoAtom improves model building for cryo-EM. *Nat. Struct. Mol. Biol.* <https://doi.org/10.1038/s41594-025-01713-3> (2025).
82. Zhong, E. D., Bepler, T., Berger, B. & Davis, J. H. CryoDRGN: reconstruction of heterogeneous cryo-EM structures using neural networks. *Nat. Methods* **18**, 176–185 (2021).
83. Luo, Z., Ni, F., Wang, Q. & Ma, J. OPUS-DSD: deep structural disentanglement for cryo-EM single-particle analysis. *Nat. Methods* **20**, 1729–1738 (2023).
84. Li, Y., Zhou, Y., Yuan, J., Ye, F. & Gu, Q. CryoSTAR: leveraging structural priors and constraints for cryo-EM heterogeneous reconstruction. *Nat. Methods* **21**, 2318–2326 (2024).
85. Punjani, A. & Fleet, D. J. 3DFlex: determining structure and motion of flexible proteins from cryo-EM. *Nat. Methods* **20**, 860–870 (2023).
86. Chen, M. & Ludtke, S. J. Deep learning-based mixed-dimensional Gaussian mixture model for characterizing variability in cryo-EM. *Nat. Methods* **18**, 930–936 (2021).
87. Huang, Y., Zhu, C., Yang, X. & Liu, M. High-resolution real-space reconstruction of cryo-EM structures using a neural field network. *Nat. Mach. Intell.* **6**, 892–903 (2024).
88. Schwab, J., Kimanius, D., Burt, A., Dendooven, T. & Scheres, S. H. W. DynaMight: estimating molecular motions with improved reconstruction from cryo-EM images. *Nat. Methods* **21**, 1855–1862 (2024).
89. Powell, B. M. & Davis, J. H. Learning structural heterogeneity from cryo-electron sub-tomograms with tomoDRGN. *Nat. Methods* **21**, 1525–1536 (2024).
90. Rangan, R. et al. CryoDRGN-ET: deep reconstructing generative networks for visualizing dynamic biomolecules inside cells. *Nat. Methods* **21**, 1537–1545 (2024).
91. Degenhardt, M. F. S. et al. Determining structures of RNA conformers using AFM and deep neural networks. *Nature* **637**, 1234–1243 (2025).
92. Lee, Y.-T. et al. The conformational space of RNase P RNA in solution. *Nature* **637**, 1244–1251 (2025).
93. Weinberg, Z. et al. Detection of 224 candidate structured RNAs by comparative analysis of specific subsets of intergenic regions. *Nucleic Acids Res.* **45**, 10811–10823 (2017).
94. Kretsch, R. C. et al. RNA target highlights in CASP15: evaluation of predicted models by structure providers. *Proteins* **91**, 1600–1615 (2023).
95. Mulvaney, T. et al. CASP15 cryo-EM protein and RNA targets: refinement and analysis using experimental maps. *Proteins* **91**, 1935–1951 (2023).
96. Wang, W., Peng, Z. & Yang, J. Single-sequence protein structure prediction using supervised transformer protein language models. *Nat. Comput. Sci.* **2**, 804–814 (2022).
97. Kaminski, K., Ludwiczak, J., Pawlicki, K., Alva, V. & Dunin-Horkawicz, S. pLM-BLAST: distant homology detection based on direct comparison of sequence representations from protein language models. *Bioinformatics* **39**, btad579 (2023).
98. Hamamsy, T. et al. Protein remote homology detection and structural alignment using deep learning. *Nat. Biotechnol.* **42**, 975–985 (2024).
99. Liu, W. et al. PLMSearch: protein language model powers accurate and fast sequence search for remote homology. *Nat. Commun.* **15**, 2775 (2024).
100. Hong, L. et al. Fast, sensitive detection of protein homologs using deep dense retrieval. *Nat. Biotechnol.* **43**, 983–995 (2025).
101. Xiao, B., Shi, Y. & Huang, L. Enhancing RNA 3D structure prediction: a hybrid approach combining expert knowledge and computational tools in CASP16. *Proteins* <https://doi.org/10.1002/prot.70034> (2025).
102. Karan, A. & Rivas, E. All-at-once RNA folding with 3D motif prediction framed by evolutionary information. *Nat. Methods* **22**, 2094–2106 (2025).
103. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659 (2006).
104. Ban, N., Nissen, P., Hansen, J., Moore, P. B. & Steitz, T. A. The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* **289**, 905–920 (2000).
105. Carter, A. et al. Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics. *Nature* **407**, 340–348 (2000).
106. Gabashvili, I. S. et al. Solution structure of the E. coli 70S ribosome at 11.5 Å resolution. *Cell* **100**, 537–549 (2000).

## Acknowledgements

This work is supported by the National Natural Science Foundation of China (NSFC T2225007, T2222012, 32430063, 62501364, T25B2009), Postdoctoral Fellowship Program and China Postdoctoral Science Foundation (BX20240212) and Fundamental Research Funds for the Central Universities.

## Author contributions

J.Y. conceived and supervised the project. W.W. and B.S. collected data and performed the analysis. Z.P. cosupervised the project. All authors wrote and revised the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41587-025-02974-5>.

**Correspondence** should be addressed to Zhenling Peng or Jianyi Yang.

**Peer review information** *Nature Biotechnology* thanks Rhiju Das and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© Springer Nature America, Inc. 2026