OXFORD

Briefings in Bioinformatics, 2026, **27(1)**, bbaf720
https://doi.org/10.1093/bib/bbaf720
Problem Solving Protocol

DeepRMSF: a deep learning-based automated approach for predicting atomic-level flexibility in RNA structure

Chenjie Feng[1,‡], Xiaowen Sun[2,‡], Xintao Song[2], Lei Bao[3], Weikang Gong[4,*], Renmin Han[2,5,*]

[1]College of Medical Information and Engineering, Ningxia Medical University, No. 1160 Shengli Road, Xingqing District, Yinchuan, Ningxia Province 750004, China
[2]Research Center for Mathematics and Interdisciplinary Sciences, Shandong University, No. 72, Binhai Road, Jimo District, Qingdao, Shandong Province 266237, China
[3]School of Public Health, Hubei University of Medicine, No. 16 Huangjiahuxi Road, Hongshan District, Shiyan, Hubei Province 442000, China
[4]Center for AI and Computational Biology, Suzhou Institute of Systems Medicine, Chinese Academy of Medical Sciences & Peking Union Medical College, 100 Chongwen Road, Suzhou Industrial Park, Suzhou, Jiangsu Province 215123, China
[5]Syneron Opal, George Town 10281, the Cayman Islands

*Corresponding authors. Weikang Gong, Center for AI and Computational Biology, Suzhou Institute of Systems Medicine, Chinese Academy of Medical Sciences & Peking Union Medical College, 100 Chongwen Road, Suzhou Industrial Park, Suzhou, Jiangsu Province 215123, China. E-mail: gwk@ism.cams.cn; Renmin Han, Research Center for Mathematics and Interdisciplinary Sciences, Shandong University, No. 72, Binhai Road, Jimo District, Qingdao, Shangdong Province 266237, China. E-mail: hanrenmin@sdu.edu.cn
‡Chenjie Feng and Xiaowen Sun joint-first authors.

Abstract

Understanding RNA conformational dynamics is essential to understand its roles in complex biological processes. While computational methods have revolutionized the prediction of static 3D RNA structures, predicting local flexibility directly from structure remains a significant challenge. We developed DeepRMSF, a deep learning-based method that leverages atomic-level descriptions of RNA to predict vibrational flexibility given a tertiary structure. Trained on MD-derived root-mean-square fluctuations(RMSF), DeepRMSF was benchmarked on 371 nonredundant RNAs, with 311 RNAs used for five-fold cross-validation (PCC = 0.7219–0.7464) and 60 RNAs as an independent test set (PCC = 0.734), ensuring minimal sequence/structural similarity between sets. DeepRMSF predicts the local flexibility of medium-sized RNAs (~75 nucleotides) in ~8.2 s, achieving >3000-fold speed-up over MD simulations while maintaining strong extrapolative accuracy. Rather than replacing MD, DeepRMSF offers a scalable and practical alternative for transcriptome-scale screening of RNA flexibility, facilitating studies on RNA structure-dynamics-function relationships and supporting computational modeling in RNA biology.

Keywords: 3D convolutional neural network; RNA local flexibility; molecular dynamics simulation; RNA dynamics prediction

Introduction

The local nucleotide flexibility of RNA plays a pivotal role in its biological functions, as seen in processes such as its intricate folding during transcription, structural transitions in riboswitches that regulate gene expression, and the formation of various specialized RNA structures with distinct functions, such as ribosomal RNA in protein synthesis or tRNA in amino acid transport [1]. In living systems, where molecular interactions are inherently dynamic, RNA molecules must adapt their conformations to engage with other biomolecules, ensuring the seamless progression of essential cellular processes. Even when folded into stable tertiary structures optimized for specific functions, RNA is not static. Driven by thermal fluctuations under physiological conditions (~300 K), RNA exhibits continuous local conformational motions around its lowest-energy state. These localized fluctuations not only characterize inherent flexibility but also modulate the likelihood of structural rearrangements that influence molecular recognition and regulatory mechanisms.

Experimental techniques used to probe local RNA flexibility include X-ray crystallography [2, 3], single-molecule fluorescence resonance energy transfer (smFRET) [4–8], and selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE) [9–15]. While these methods have greatly advanced our understanding, they have inherent limitations. For instance, B-factors from X-ray structures reflect atomic displacement at low temperature (~100 K) and may not represent flexibility under physiological conditions. smFRET primarily measures distance fluctuations between labeled sites and is constrained by the Förster distance range, limiting its utility for localized structural motions. These changes make it difficult to obtain atomic-level flexibility profiles for entire RNAs under physiological conditions.

Computational approaches have been developed to complement experimental measurements. Molecular dynamics (MDs) simulations [16, 17] provides detailed insights into RNA motions, with root mean square fluctuation (RMSF) widely used to quantify atomic mobility. However, MD simulations are computationally expensive, especially for large RNAs. Elastic network model (ENM)-based methods [18–28], such as the Gaussian network

Received: August 15, 2025. **Revised:** October 29, 2025. **Accepted:** December 16, 2025

model (GNM) [21], parameter-free GNM (pfGNM) [22], and multi-scale GNM (mGNM) [23], offer efficient approximations but may miss fine-scale dynamics. More recently, machine learning and deep learning approaches [29–34] have shown strong potential in predicting molecular properties, though their application to RNAs remains limited by the unique structural complexity of RNA molecules.

Recent experimental and integrative modeling studies, particularly those by Gab Varani and Rhiju Das, have provided valuable insights into RNA conformational flexibility and folding dynamics by combining structural biology with chemical probing and computational refinement [35–38]. These hybrids experimental-computational approaches have substantially deepened our understanding of RNA dynamic landscapes and ligand-induced conformational changes. However, such methods often require extensive experimental effort or long simulation times for each target RNA, which limits their scalability.

To address these limitations, we developed DeepRMSF, a fully automated deep learning approach for predicting atomic-level flexibility directly from a given RNA tertiary structure in PDB format. DeepRMSF integrated atomic coordinates with simulated density maps into a 3D convolutional neural network (U-Net++L3) trained to predict heavy-atom RMSF values consistent with MD simulations. Benchmarking on a nonredundant dataset of 371 RNAs with diverse lengths and secondary structures showed that DeepRMSF achieved high correlation with MD-derived RMSF while being thousands of times faster than MD simulations. This framework provides a rapid and accessible solution for RNA dynamics estimation, bridging the gap between static structure modeling and dynamic functional characterization.

Materials and methods

DeepRMSF pipeline

DeepRMSF is a fully automated pipeline for RNA dynamics modeling, consisting of two consecutive modules (Fig. 1). Starting with a given PDB-formatted RNA structure, the corresponding simulated density map was generated using UCSF Chimera [39], providing a volumetric representation that preserves both global architecture and local nucleotide environments. This map was subsequently segmented into a series of overlapping density boxes, enabling the model to focus on localized structural regions while retaining contextual information from surrounding nucleotides. These density boxes were then fed into the deep learning model, which utilizes U-Net++ L3 [40] as the backbone architecture to capture multiscale spatial features through nested skip connections. The regression head of the network processes each density box to output localized RMSF predictions, effectively mapping atomic-level flexibility within each region. Finally, the predicted RMSF sub-boxes were merged via weighted averaging into a continuous and comprehensive RMSF map, providing a global flexibility profile for the RNA target. By integrating static tertiary structure information with volumetric density features and deep learning–based representation learning, DeepRMSF bridges the gap between static RNA structures and their dynamic behaviors, enabling accurate and efficient predictions comparable to molecular dynamics simulations but orders of magnitude faster.

Root mean square fluctuation definition

RMSF is a widely used metric to assess the flexibility of molecular structures in MDs analysis and is defined as:

$$RMSF = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(x(t) - \bar{x}\right)^2}$$

Where $x(t)$ represents the position of the heavy atom at time t, and where $\bar{x}$ denotes the time-averaged position of that atom over the simulation time T. RMSF with larger values indicating greater atomic displacement and thus higher structural flexibility.

Database of RNA structures

To build a high-quality dataset for model training and evaluation, we systematically curated monomeric RNA structures from Protein Data Bank (PDB, https://www.rcsb.org), and the RNA tertiary template library was constructed in three steps: (i) Chain and structure filtering: Nucleic acid structures containing protein fragments, small molecules (ligands, metal ions, and water molecules) were excluded, retaining only standalone RNA chains. Any modified or noncanonical nucleosides were converted to their corresponding standard nucleobases to ensure consistency. (ii) Length filtering: RNA chains shorter than 30 nucleotides were removed to eliminate structural fragments and ensure sufficient structural complexity. (iii) Redundancy removal: Based on the filtered set from steps (i) and (ii), a non redundant dataset was generated using cd-hit-est with a sequence identity cutoff of 90%. The resulting dataset contained high-quality, nonredundant, structurally complete RNA monomers, providing a reliable basis for downstream MD simulations and RMSF annotation.

All-atom molecular dynamics simulations

The initial dataset of 371 RNA structures was obtained from the Protein Data Bank (https://www.rcsb.org), with sequence lengths ranging from 30 to 414 nucleotides (nt). To focus exclusively on the intrinsic structural flexibility of RNA molecules, any ligands, DNA, or proteins present in the complexes were removed prior to simulation setup. Each RNA system was solvated in a truncated octahedron box of TIP3P [41] water molecules, ensuring at least a 12 Å buffer distance between the solute and the edges of the periodic box. Based on the total charge of each RNA, Na^+ counterions were added to neutralize the system [42]. In addition, to better match physiological conditions, 150 mM NaCl was included all systems according to the screening layer tally calculated using the SPLIT method [43]. All MD simulations were performed using the AMBER 20 software package [44] with the ff99bsc0 + χOL3 force field for RNA [45] on NVIDIA Tesla A100 GPUs. The parameters for Na^+ and Cl^- ions were adopted from the work of the Cheatham group. Energy minimization was conducted for 6000 steps using the conjugate gradient method. Systems were then gradually heated from 0 to 300 K over 400 ps under the NVT ensemble using the Langevin thermostat, with position restraints of 1000 kcal·mol^{-1}·Å^{-2} applied to RNA heavy atoms. Subsequently, restraints were progressively released in five consecutive 1 ns stages (1000, 100, 10, 1, and 0 kcal·mol^{-1}·Å^{-2}) under the NPT ensemble (P = 1 bar, T = 300 K). The equilibrated systems were subjected to 200 ns production MD simulations at constant temperature (300 K) and pressure (1 bar) using periodic boundary conditions and the particle mesh Ewald (PME) method. Further simulation details are provided in Section S2.

RMSD convergence and MD simulation equilibrium

To evaluate the equilibrium state of each RNA system, we assessed the convergence of RMSD values over the simulation time. Specifically, the initial 40 ns of each trajectory was discarded to remove the equilibration phase, and RMSD values were calculated separately for the 40–120 ns and 120–200 ns segments. The Pearson correlation coefficient between these two segments was computed for each RNA. A high correlation indicated that the RMSD profile had effectively converged, reflecting stable dynamic

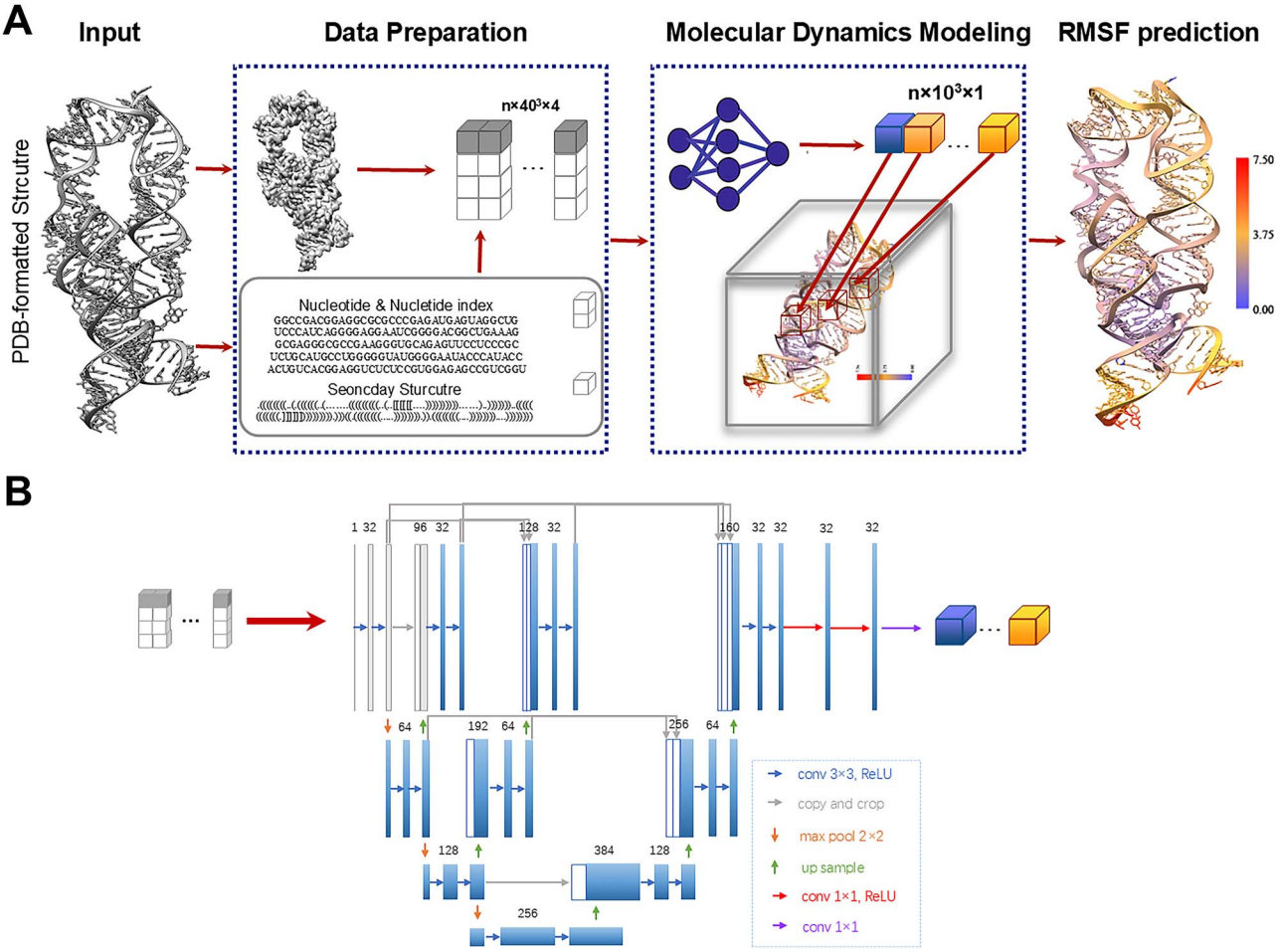

Figure 1. Overview of the DeepRMSF framework for RNA dynamics prediction. (a) The DeepRMSF workflow integrates sequence-derived secondary structure, 3D atomic coordinates, and simulated electron-density maps as input features. Each RNA structure is converted into voxelized density boxes that retain both global and local spatial information. (b) Architecture of the DeepRMSF model. The network employs a three-level nested 3D U-Net++ (L3) architecture consisting of encoder and decoder paths with multiscale skip connections. Input tensors (4 channels: density, nucleotide type, index, and pairing status) pass through convolutional blocks for spatial feature extraction, followed by a 3-layer convolutional regression head that outputs per-voxel RMSF predictions. The model is trained in a supervised manner against MD-derived RMSF ground truth to learn the relationship between local structural context and nucleotide flexibility.

behavior(see details in the RMSD definition section). In addition, to illustrate the robustness of the MD simulations across the dataset, we randomly selected 50 RNAs from the total of 371 and examined their RMSF variation over time. The results, shown in Supplementary Fig. S1, demonstrate stable RMSD convergence profiles and consistent equilibration behavior across the sampled RNAs, confirming that the trajectories had reached equilibrium prior to RMSF calculation.

Qualitative comparison between MD-derived RMSF and crystallographic B-factors

We qualitatively compared RMSF obtained from MD simulations with crystallographic B-factors reported in the Protein Data Bank. Only atoms with nonzero B-factor values were included in the analysis. The B-factors were converted to estimated RMSF values using the standard equation:

$$RMSF = \sqrt{\frac{3 \times B - factor}{8\pi^2}}$$

Because crystallographic B-factors are measured in the solid state at cryogenic temperatures, whereas MD simulations model solvated molecules at room temperature, the two datasets represent fundamentally different thermodynamic conditions. This comparison was therefore performed solely to assess whether regions of high and low flexibility exhibited similar spatial patterns, rather than to derive quantitative agreement.

Across the analyzed RNA samples (Supplementary Fig. S4), Pearson correlation coefficients between MD-derived and B-factor-derived RMSF values varied substantially (~−0.6–0.8), reflecting differences in structural stability among RNAs. Highly dynamic RNAs, such as hairpin loops (PDB ID: 1YXP), tended to show weaker correspondence, whereas more rigid RNAs, such as multihelix junctions (PDB ID: 3LA5), exhibited stronger agreement. All predictive performance evaluations in this study were benchmarked exclusively against MD-derived RMSF to maintain a consistent dynamic reference and to avoid systematic bias arising from differing experimental conditions.

Input embedding

In the data preparation module, we first used UCSF Chimera to convert each RNA PDB model into a simulated electron density map with a resolution of 4 Å and a grid spacing of 1.5 Å. Each density map was subsequently partitioned into a series of cubic boxes

of size 40 × 40 × 40 voxels. This box size was selected to ensure that the central 10 × 10 × 10 voxels, which corresponds to the region for which RMSF values were predicted, captured all possible local atomic and nucleotide interactions. The box size was also chosen to balance structural coverage with computational complexity in the neural network. Consequently, each density map was split into multiple 40 × 40 × 40 boxes with a stride of 10 voxels, generating overlapping regions that preserve structural continuity.

To enhance model performance, additional RNA structural information was incorporated. We tested several feature combinations and ultimately selected nucleotide type, nucleotide index, and secondary structure pairing status as auxiliary inputs. Secondary structure was assigned using x3DNA-DSSR [41]. For simplicity and to avoid overcomplicating the feature space, we distinguished only whether each nucleotide was base paired: atoms in paired nucleotides were labeled as 0, and atoms in unpaired nucleotides were labeled as 1. Nucleotide types (A, U, C, G) were extracted from the PDB files and encoded as integers (0 = A, 1 = U, 2 = C, 3 = G), then transformed into one-hot vectors. Nucleotide indices were similarly extracted and encoded to represent the sequential position of each nucleotide in the chain. Finally, the same box-splitting operation used for the density maps was applied to these structural feature maps. Boxes at corresponding positions were concatenated along the channel dimension, producing the multichannel input tensors for the deep learning model.

Network

In the dynamics modeling module, DeepRMSF employs a three-level nested U-Net++ (L3) architecture as the backbone, consisting of an encoder and a decoder structure designed for multi-scale feature extraction. Each convolution block in the U-Net++ contains two convolution layers, each followed by batch normalization and a Rectified Linear Unit (ReLU) activation function. The multichannel input tensors, formed by concatenating four feature maps(density, nucleotide type, nucleotide index, and secondary structure pairing status), are processed through three down-sampling blocks in the encoder to enhance robustness and enlarge the receptive field. The decoder then restores these features to the original spatial resolution through a series of up-sampling blocks, while extracting and fusing multiscale features from intermediate layers.

A regression head consisting of three convolutional layers is attached to the backbone to predict RMSF values. ReLU activation is applied in each layer, and a dropout layer is inserted after the first convolutional layer to prevent overfitting. For each RNA, model performance was evaluated by calculating the Pearson correlation coefficient between the MD-derived RMSF values and the predicted RMSF values for all atoms. Finally, the predicted RMSF sub-boxes were reassembled into a full RMSF map as the final output for each RNA, and visualization of the results were visualized using UCSF ChimeraX.

Training data and procedure

A 5-fold cross-validation strategy was employed to maximize dataset utilization and improve model generalization. The network was trained for a maximum of 100 epochs with a batch size of 16. The Adam optimizer was used for gradient descent with an initial learning rate of 4e-3. Model training was conducted on two NVIDIA A100 GPUs, requiring ∼7–10 h per fold, while inference on a single RNA structure took only a few seconds.

To ensure a robust and unbiased estimate of model performance, we adopted a repeated 5-fold cross-validation strategy. Specifically, the dataset of 311 RNAs was randomly partitioned into five folds, ensuring that RNAs sharing >80% sequence identity or a TM-score > 0.45 were assigned to the same fold. This procedure was repeated 10 times using different random seeds, resulting in 50 independent training and validation runs (10 × 5 CV). Such repetition captured both training stochasticity and variability due to data partitioning. The final performance was reported as the mean ± standard deviation of the Pearson correlation coefficient (PCC), mean absolute error (MAE), and root mean square error (RMSE) across all runs.

Results

Robust performance across cross-validation and external test set

We partitioned 311 RNAs into five cross-validation folds, ensuring that RNAs sharing >80% sequence identity and a TM-score > 0.45 [46] was assigned to the same fold. An independent test set of 60 RNAs was selected with <60% sequence identity and a TM-score < 0.45 relative to any RNA in the training folds (details are provided in Supplementary Section S2 and Fig. S2). Across the ten repetitions of 5-fold cross-validation, DeepRMSF achieved stable and consistent performance, with an average Pearson correlation coefficient (PCC) of 0.716 ± 0.011, MAE of 0.091 ± 0.004, and root mean square error (RMSE) of 0.107 ± 0.005 (Fig. 2A and Fig. S3). These results demonstrate that DeepRMSF maintains reliable predictive accuracy across different random initializations and data splits, highlighting its robustness against partitioning variability.

To further evaluate the model's generalization ability, we assessed its predictions on the independent test set comprising 60 RNAs. As shown in Fig. 2B, the best-performing model among the 10 runs (PCC = 0.736 on validation) achieved a strong correlation (PCC = 0.734) between predicted and MD-derived RMSF values. This fine-grained agreement indicates that DeepRMSF effectively captures both global and local patterns of RNA flexibility, closely approximating the dynamics observed from long-timescale MD simulations. Figure 2C presents the computational efficiency of DeepRMSF. While conventional molecular dynamics requires extensive computational resources and multiple hours to simulate a single RNA molecule, DeepRMSF predicts per-nucleotide RMSF values in only ∼8.2 s per RNA. The largest RNA in our dataset (414 nucleotides) was processed in 33 s, compared to ∼1 week of MD simulation. These results underscore DeepRMSF's potential as a scalable and practical alternative to MD simulations, enabling transcriptome-scale modeling of RNA dynamics in a fraction of the time.

To assess robustness with respect to input structure quality, we compared predictions generated from PDB-initial versus MD-equilibrated structures for 20 representative RNAs from the independent test set. The average PCC was 0.709 for PDB-initial inputs and 0.716 for MD-equilibrated inputs (Fig. S5), demonstrating that DeepRMSF maintains strong predictive performance even when relying on experimentally determined PDB structures. This highlights its practical utility for large-scale RMSF prediction across diverse RNAs.

To demonstrate that DeepRMSF outperforms coarse-grained traditional methods (GNM and pfGNM), we compare them. Since three of the 371 systems do not contain P atoms, we evaluated the PCC values of GNM and pfGNM based on the remaining 368 systems (see Table S1). According to the average PCC values, DeepRMSF (0.719 ± 0.02) significantly outperforms GNM (0.441 ± 0.3) and pfGNM (0.490 ± 0.326).

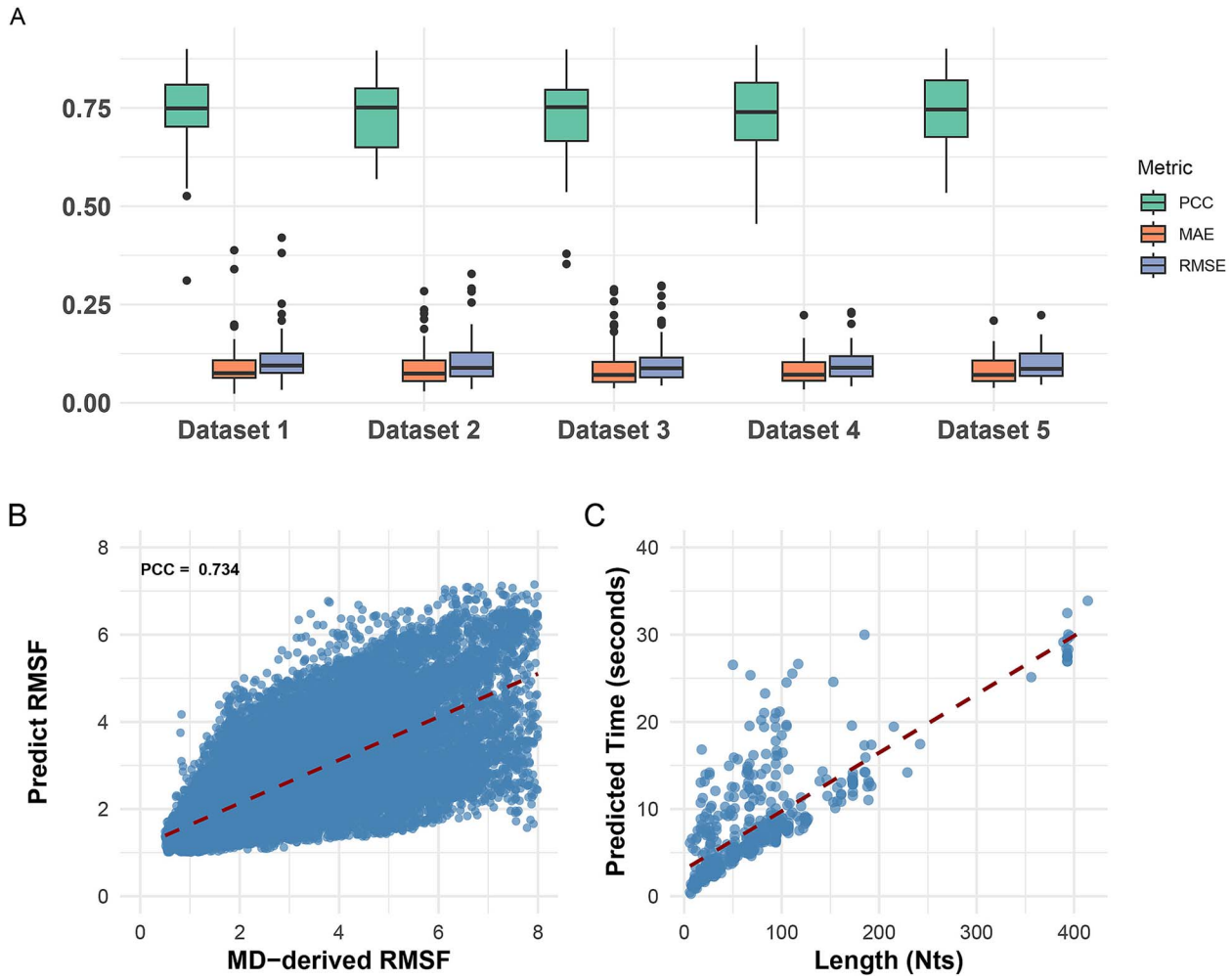

Figure 2. Evaluation of DeepRMSF performance, accuracy, and runtime scalability across RNA datasets. (a) Model performance in terms of Pearson correlation coefficient (PCC), mean absolute error (MAE), and root mean square error (RMSE) averaged over 10 repetitions of 5-fold cross-validation (10 × 5 CV). Error bars represent the standard deviation across 50 independent training and validation runs, capturing both model stochasticity and data-partition variability. (b) Correlation between predicted and MD-derived RMSF values across the independent test set (60 RNAs). A strong linear relationship (PCC = 0.734) indicates high predictive accuracy at the atomic level. (c) Runtime required by DeepRMSF to predict RMSF as a function of RNA length, showing near-linear scalability and a substantial computational advantage compared to MD simulations.

Impact of input feature combinations on root mean square fluctuation prediction

To assess the combinations of sequence and structural features, we benchmarked three input configurations under five-fold cross-validation: (i) nucleotide type together with density map, (ii) nucleotide type and nucleotide index combined with the simulated map, and (iii) nucleotide type, nucleotide, and secondary structure information combined with density map (Fig. 3A). The baseline model utilized only nucleotide type and simulated density map achieved an average PCC of 0.727(MAE = 0.0892, RMSE = 0.1066). When nucleotide index was added to the input feature set, slightly decreased to 0.721, suggesting that position specific information alone did not enhance prediction accuracy and may have introduced redundancy. Incorporating secondary structure annotations led to the best overall performance, with an average PCC of 0.7331, the lowest average MAE (0.086) and RMSE (0.1011), and reduced variance cross folds. These results indicate improved robustness and generalizability when secondary structure features are included.

We further examined a representative riboswitch RNA (PDB ID: 4FEP) to visualize the impact of feature combinations (Fig. 3B). RMSF values predicted by the three models were mapped onto the RNA 3D structure. The model using only nucleotide type and simulated map underestimated flexibility in dynamic regions, particularly within loops and junctions. In contrast, the full feature model, which included nucleotide type, nucleotide index, simulated map and secondary structure produced predictions that closely matched the MD-derived ground truth. In particular, a highly flexible loop region was accurately captured only when secondary structure information was incorporated. Collectively, these findings collectively demonstrate that enriching the input representation with position-specific and structural information significantly improves both the accuracy and reliability of RMSF prediction. Such improvements are especially critical for modeling localized flexibility in functionally important RNA motifs.

Case studies reveal DeepRMSF's ability to targets with complex structures

To assess the robustness and generalization ability of DeepRMSF across RNAs with varying lengths and structural complexities, we selected three representative RNAs: yeast tRNA (PDB ID: 3TRA, 64

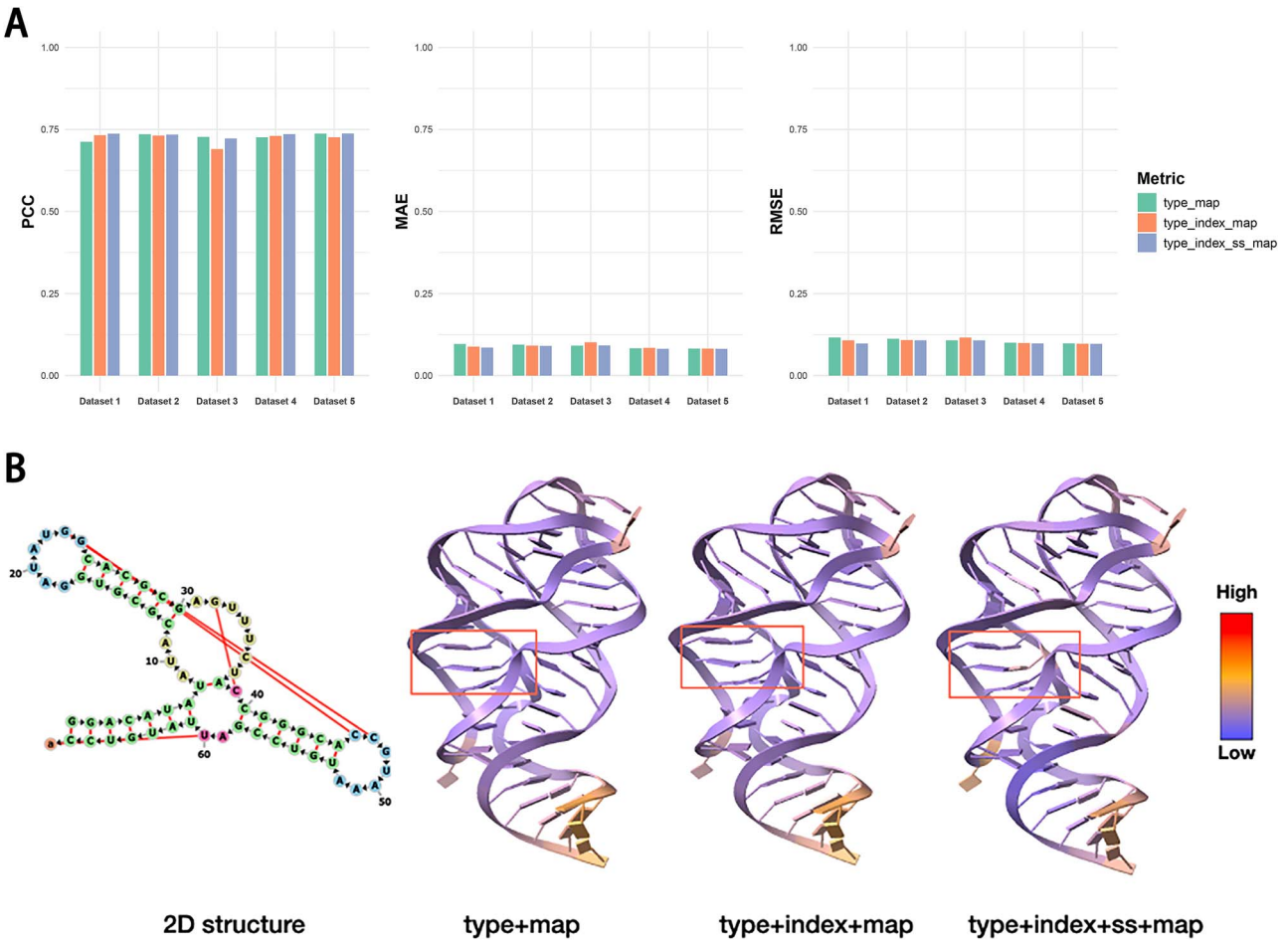

Figure 3. Impact of input feature combinations on RNA RMSF prediction performance. (a) Benchmarking results of DeepRMSF using three different input feature combinations: type+map, type+index+map, and type+index+ss + map, evaluated via 5-fold cross-validation. (b) Structural visualization of RMSF predictions for the riboswitch RNA (PDB ID: 4FEP). The RMSF values predicted by the three models are mapped onto the RNA tertiary structure.

nts), the SAM-I riboswitch (PDB ID: 2GIS, 145 nts), and the Tetrahymena group I intron ribozyme (PDB ID: 1X8W, 248 nts). These RNAs represent distinct structural motifs and topological organizations, ranging from compact tRNA scaffolds to large catalytic ribozymes. As shown in Fig. 4, DeepRMSF predicted nucleotide-level flexibility profiles that closely matched MD-derived RMSF values, with PCCs of 0.7925, 0.74009, and 0.8007 for 3TRA, 2GIS, and 1X8W, respectively. These results indicate that DeepRMSF reliably captures both rigid and flexible regions across diverse RNA architectures. The results indicate that DeepRMSF reliably captures both rigid and flexible across diverse RNA architectures. First, a 64-nucleotide yeast tRNA (PDB ID: 3TRA) was analyzed to evaluate the prediction accuracy of DeepRMSF on a compact, functionally conserved RNA. This tRNA(Asp) forms a dimer in the crystal, featuring four canonical helices, several highly structured loops and junctions. DeepRMSF accurately predicted flexible regions in the D-loop and T-loop, which are known to undergo dynamic shifts depending on base stacking and twist angle propagation from the anticodon loop. In particular, the tertiary contact between G19 in the D-loop and C56 in the T-loop is known to be more labile in the dimeric Asp tRNA than in Phe tRNA, which is consistent with the observed fluctuation patterns. Our predictions reflect this behavior, showing increased flexibility in the D and T loops, and reduced motion in the anticodon region, as also observed in MD simulations. The high PCC (0.7925) confirms DeepRMSF's ability to reproduce such dynamic behavior with atomic-level resolution.

Second, the S-adenosylmethionine (SAM) riboswitch (PDB ID: 2GIS, 145 nts) contains a compact architecture organized by four paired stems (P1–P4), interlinked by joining regions and a pseudoknot. The L2 loop and junctions J3/4 and J4/1 form ligand-independent tertiary contacts, establishing a prefolded architecture even in the absence of SAM. Upon SAM binding, the P1 helix is stabilized through hydrogen bonding and van der Waals interactions with the P3 and J1/2 region. DeepRMSF successfully predicted low flexibility around this ASM-binding pocket, consistent with its structural compactness and in-line probing data. The 5′ region of P4, which remains single-stranded after folding, showed the highest RMSF peaks, reflecting its inherent flexibility. The correlation between prediction and MD results reached a PCC of 0.74009, demonstrating strong predictive accuracy for complex ligand-binding RNAs.

Last, the Tetrahymena thermophila group I intron ribozyme (PDB ID: 1X8W, 248 nts) adopts a well-organized tertiary structure critical for RNA self-splicing. It folds into two domains: P4–P6 and P3–P9, which pack against each other to organize the active site, including the G-binding site (G-site) in P7. Multiple base triples and a catalytic magnesium ion stabilize the active site, even in the absence of substrate RNA. DeepRMSF accurately identified regions of flexibility, particularly the external loops and the

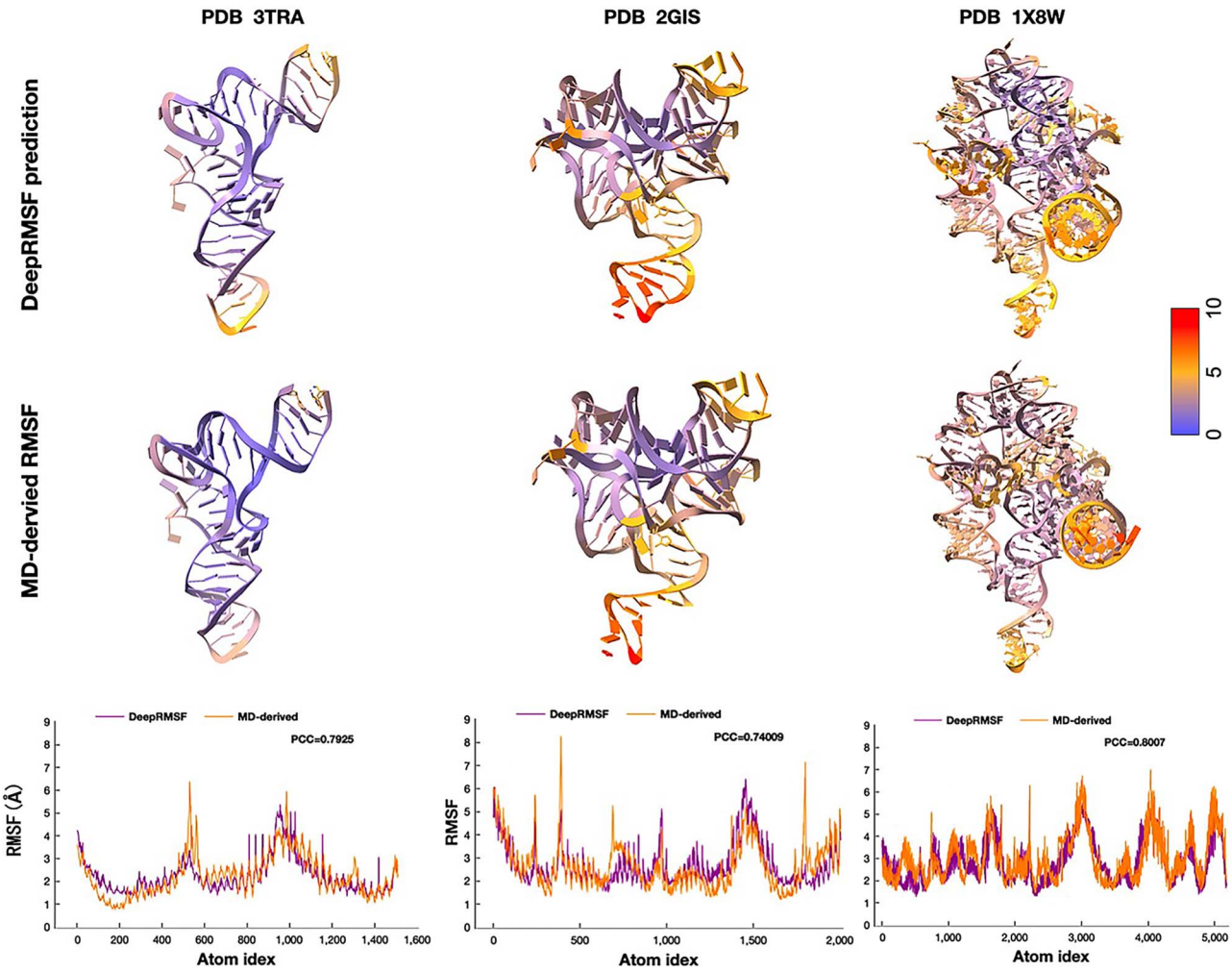

Figure 4. DeepRMSF performance for three RNAs. RMSF predictions for yeast tRNA (PDB ID: 3TRA, 64 nts), SAM-I riboswitch (PDB ID: 2GIS, 145 nts) and Tetrahymena group I intron ribozyme (PDB ID: 1X8W, 248 nts). Colors represent normalized RMSF values, as indicated by the color bar on the right. The first row is colored based on DeepRMSF predictions, while the second row shows the ground truth RMSF derived from molecular dynamics simulations. The third row plots the normalized RMSF values along the nucleotide sequence for each RNA, comparing DeepRMSF predictions (purple lines) with MD-based ground truth (orange lines).

G-site entry, while capturing stable core structure. The overall correlation with MD simulations reached a PCC of 0.8007, highlighting its capability to model dynamics in large and catalytically active RNAs.

We loaded the ligand positions from the PDB into the structure mapped by DeepRMSF and found that most of the flexible regions are the binding sites of the ligands (see Fig. S6), especially the magnesium ion. DeepRMSF predictions align with functional motifs.

Collectively, these case studies demonstrate that DeepRMSF consistently reproduces atomic-level flexibility patterns across RNAs of varying sizes and architectures. The method maintains high accuracy even for large RNAs with intricate tertiary folds, offering a scalable and computationally efficient alternative to long time scale MD simulations for RNA dynamics analysis.

DeepRMSF standalone

Tool input

DeepRMSF is available as a standalone offline tool that allows users to run the program locally on their own computers(Fig. 5). To the best of our knowledge, DeepRMSF is the first standalone tool dedicated to fast and accurate prediction of local nucleotide flexibility for a given fixed RNA tertiary structure, rather than simulating long time scale global conformational changes. The required input is an RNA tertiary structure in PDB format.

Tool output

For RNA structures shorter than 1000 nucleotides, the program typically finishes within minutes. The DeepRMSF results consist of the following three sections: (i) sequence in FASTA format obtained from the input structure, which provides the sequence information clearly; (ii) secondary structure in dot-bracket notation, which summarizes base-pairing and overall secondary structure of the input model; (iii) predicted RMSF: first, the user can inspect a line plot of the predicted RMSF, from which the positional fluctuation level of each atom can be examined. Then, two structure files named {pdbid}.pdb and {pdbid}_pre_nor.pdb are provided. {pdbid}.pdb is the original input structure. {pdbid}_pre_nor.pdb is the same structure annotated with predicted RMSF. These files can be visualized with UCSF Chimera or UCSF ChimeraX. For quantitative evaluation and comparison across systems, RMSF values in Å should be used.

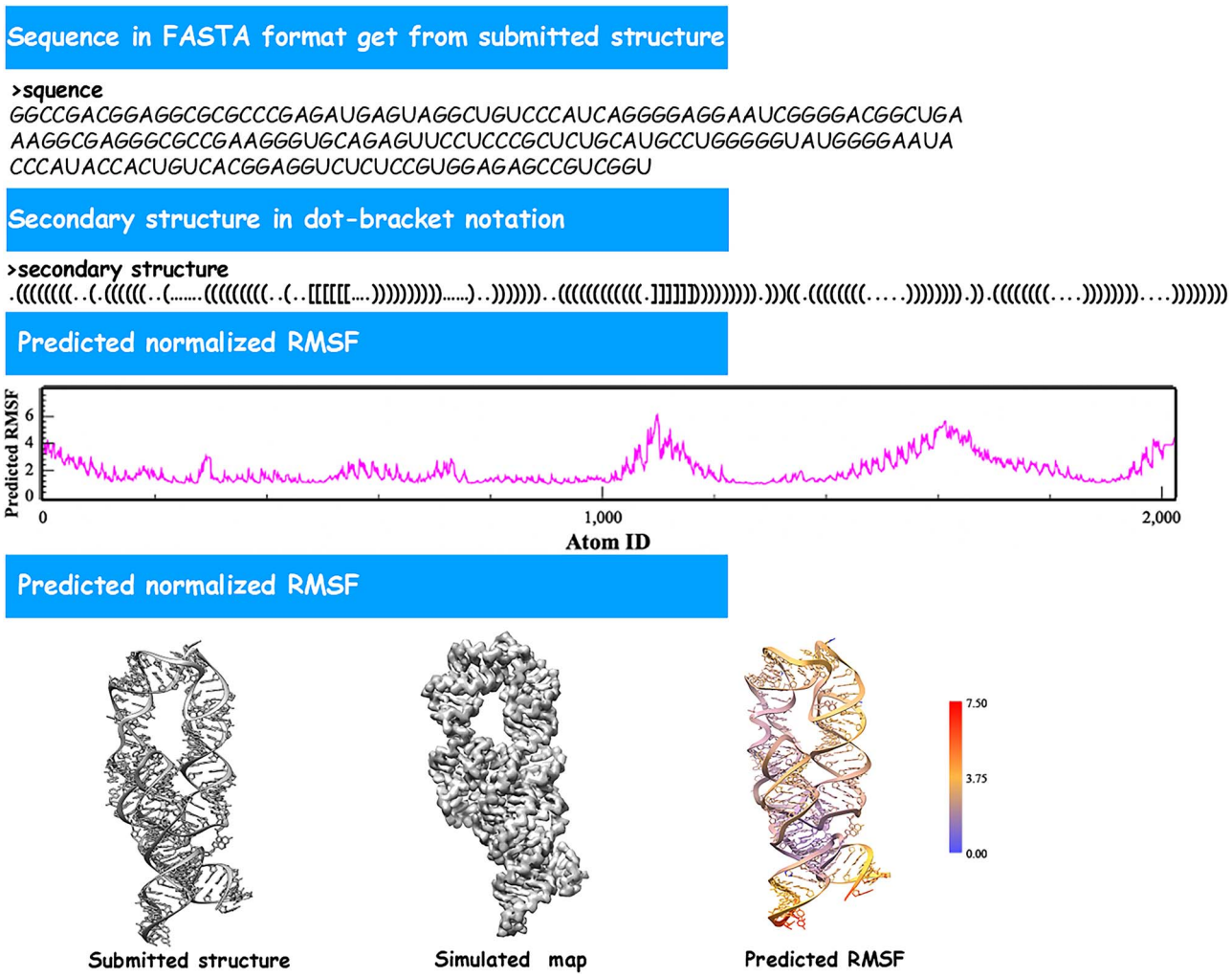

Figure 5. Illustration of the DeepRMSF output, including sequence information and secondary structure extracted from input structure in PDB format. And the dynamics predicted results for all atom among this structure. Finally, the input 3D model, simulated maps translated by input structure and predicted dynamics structure showed by UCSF Chimera and UCSF ChimeraX.

Discussion

We developed and validated DeepRMSF, a fully automated and computationally efficient method for predicting atomic-level flexibility from RNA tertiary structures in PDB format. The method converts input PDB models into simulated maps, segments them into local density boxes, and processes them using a 3D voxel-based convolutional neural network (U-Net++ L3) with structural encoding. A regression head then outputs atom-level RMSF values.

DeepRMSF was trained and evaluated on 371 nonhomologous RNAs, each simulated for 200 ns of all atom molecular dynamics under explicit solvent conditions. For these targets, the method achieved an average PCC of 0.733 with MD-derived RMSF, while requiring only seconds for inference. This balance between accuracy and computational efficiency makes DeepRMSF particularly suitable for large-scale local flexibility screening, where long time scale MD simulations would be computationally prohibitive. Furthermore, the neural network framework can be extended to model RNA–protein interactions and RNA complexes, offering new opportunities to elucidate the molecular and cellular functions of noncoding RNAs. Studies along these lines are currently underway.

To further enhance the biological relevance of flexibility prediction, future work will incorporate RNA-ligand molecular dynamics simulations into future model training. While DeepRMSF demonstrates robust performance for intrinsic RNA flexibility, it does not yet account for ligand-induced conformational changes. Accurate modeling of RNA–ligand systems poses specific challenges, including: (i) the need for extended simulation timescales to capture binding-induced motions; (ii) the requirement for accurate parameterization of diverse ligand chemistries; and (iii) the increased computational cost associated with RNA–ligand complexes simulations. To address these limitations, future development work will explore enhanced sampling strategies and refined force fields to better capture ligand-dependent flexibility. Collectively, these planned extensions will enable DeepRMSF to model functional RNA motifs such as riboswitches and aptamers, thereby broadening its application to RNA structural biology and RNA-targeted drug discovery.

Key Points

- DeepRMSF predicts atom-level root mean square fluctuation(RMSF) directly from RNA tertiary structure, achieving high accuracy with markedly reduced computational cost compared to molecular dynamics (MD) simulations.
- The method integrates simulated electron density maps with nucleotide type, index, and secondary structure features into a 3D U-Net++(L3) architecture for multi-scale feature extraction and robust RMSF prediction.
- Benchmarking on 371 nonredundant RNAs shows consistent performance in five-fold cross-validation and independent test sets, with Pearson correlation coefficients above 0.73 and over 3000-fold speed-up relative to MD.
- Case studies on tRNA, riboswitch, and ribozyme highlight the framework's ability to capture rigid and flexible regions across diverse RNA architectures, enabling scalable transcriptome-wide RNA dynamics analysis.

Acknowledgements

We would like to thank Guojun Li and Xin Gao for fruitful discussions.

Supplementary data

Supplementary data is available at *Briefings in Bioinformatics* online.

Competing interests: None declared.

Funding

This research was supported by the National Key Research and Development Program of China [2021YFF0704300], This work was supported by the National Natural Science Foundation of China Projects Grant [62072280, 32241027, 61932018, 62072441], Chinese Academy of Medical Sciences & Peking Union Medical College, Union Medical College Young Scholar Support Program, No. 2023086; China Postdoctoral Science Foundation, the Postdoctoral Fellowship Program (Crade B), Grant Number GZB20230084, the Natural Science Foundation of Shandong Province ZR2023YQ057, Cultivating Project for Young Scholar at Hubei University of Medicine 2020QDJZR015, the Natural Science Foundation of Ningxia Province [2023AAC05036, 2024YCZX0041, 2024AAC03247] and the University-level scientific research project of Ningxia Medical University [XT2022024, XJKF230109], the Shanghai Municipal Science and Technology Major Project [No.2018SHZDZX01], Key Laboratory of Computational Neuroscience and Brain-Inspired Intelligence (LCNBI) and ZJLab.

Data and code availability

The source code and datasets underlying this article are publicly available online at https://figshare.com/articles/dataset/data_zip/24533122.

References

1. Rambo RP, Tainer JA. Bridging the solution divide: Comprehensive structural analyses of dynamic RNA, DNA, and protein assemblies by small-angle X-ray scattering. *Curr Opin Struct Biol* 2010;**20**:128–37. https://doi.org/10.1016/j.sbi.2009.12.015
2. Hinsen K. Structural flexibility in proteins: Impact of the crystal environment. *Bioinformatics* 2008;**24**:521. https://doi.org/10.1093/bioinformatics/btm625
3. Choy JS, Lee T-H. Structural dynamics of nucleosomes at single-molecule resolution. *Trends Biochem Sci* 2012;**37**:425–35. https://doi.org/10.1016/j.tibs.2012.06.006
4. Lerner E, Cordes T, Ingargiola A. *et al.* Toward dynamic structural biology: Two decades of single-molecule Förster resonance energy transfer. *Science* 2018;**359**:1133. https://doi.org/10.1126/science.aan1133
5. Roy R, Hohng S, Ha T. A practical guide to single-molecule FRET. *Nat Methods* 2008;**5**:507–16. https://doi.org/10.1038/nmeth.1208
6. Jeng SCY, Trachman RJ, Weissenboeck F. *et al.* Fluorogenic aptamers resolve the flexibility of RNA junctions using orientation-dependent FRET. *RNA* 2021;**27**:433–44. https://doi.org/10.1261/rna.078220.120
7. Willkomm S, Jakob L, Kramm K. *et al.* Single-molecule FRET uncovers hidden conformations and dynamics of human Argonaute 2. *Nat Commun* 2022;**13**:3825. https://doi.org/10.1038/s41467-022-31480-4
8. Mlýnský V, Bussi G. Molecular dynamics simulations reveal an interplay between SHAPE reagent binding and RNA flexibility. *The Journal of Physical Chemistry Letters* 2018;**9**:313–8. https://doi.org/10.1021/acs.jpclett.7b02921
9. HD, Cate Jamie. The Mechanisms of RNA SHAPE Chemistry. 2012.
10. Mortimer SA, Weeks KM. Time-resolved RNA SHAPE chemistry. *J Am Chem Soc* 2008;**130**:16178–80. https://doi.org/10.1021/ja8061216
11. Siegfried NA, Busan S, Rice GM. *et al.* RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). *Nat Methods* 2014;**11**:959–65. https://doi.org/10.1038/nmeth.3029
12. Deigan KE, Li TW, Mathews DH. *et al.* Accurate SHAPE-directed RNA structure determination. *Proc Natl Acad Sci* 2009;**106**:97–102. https://doi.org/10.1073/pnas.0806929106
13. Spitale RC, Flynn RA, Torre EA. *et al.* RNA structural analysis by evolving SHAPE chemistry. *Wiley Interdisciplinary Reviews: RNA* 2014;**5**:867–81.
14. Wilkinson KA, Merino EJ, Weeks KM. Selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE): Quantitative RNA structure analysis at single nucleotide resolution. *Nat Protoc* 2006;**1**:1610–6. https://doi.org/10.1038/nprot.2006.249
15. Dethoff EA, Chugh J, Mustoe AM. *et al.* Functional complexity and regulation through RNA dynamics. *Nature* 2012;**482**:322–30. https://doi.org/10.1038/nature10885
16. Karplus M, McCammon JA. Molecular dynamics simulations of biomolecules. *Nat Struct Biol* 2002;**9**:646–52. https://doi.org/10.1038/nsb0902-646
17. Dror RO, Dirks RM, Grossman JP. *et al.* Biomolecular simulation: A computational microscope for molecular biology. *Annu Rev Biophys* 2012;**41**:429–52. https://doi.org/10.1146/annurev-biophys-042910-155245
18. Kmiecik S, Kouza M, Badaczewska-Dawid AE. *et al.* Modeling of protein structural flexibility and large-scale dynamics: Coarse-grained simulations and elastic network models. *Int J Mol Sci* 2018;**19**:3496. https://doi.org/10.3390/ijms19113496

19. Pinamonti G, Bottaro S, Micheletti C. *et al.* Elastic network models for RNA: A comparative assessment with molecular dynamics and SHAPE experiments. *Nucleic Acids Res* 2015;**43**:7260–9. https://doi.org/10.1093/nar/gkv708
20. Tirion MM. Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Phys Rev Lett* 1996;**77**:1905–8. https://doi.org/10.1103/PhysRevLett.77.1905
21. Yang LW, Rader AJ, Liu X. *et al.* oGNM: Online computation of structural dynamics using the gaussian network model. *Nucleic Acids Res* 2006;**34**:W24–31. https://doi.org/10.1093/nar/gkl084
22. Bahar I, Jernigan RL. Vibrational dynamics of transfer RNAs: Comparison of the free and synthetase-bound forms. *J Mol Biol* 1998;**281**:871–84. https://doi.org/10.1006/jmbi.1998.1978
23. Hyeon C, Thirumalai D. Ripping RNA by force using gaussian network models. *J Phys Chem B* 2017;**121**:3515–22. https://doi.org/10.1021/acs.jpcb.6b09402
24. Romo D, Tod AG. Validating and improving elastic network models with molecular dynamics simulations. *Proteins: Structure, Function, and Bioinformatics* 2010;**79**:23–34. https://doi.org/10.1002/prot.22855
25. Wang Y, Rader AJ, Bahar I. *et al.* Global ribosome motions revealed with elastic network model. *J Mol Biol* 2004;**147**:302–14. https://doi.org/10.1016/j.jsb.2004.01.005
26. Delarue M, Sanejouand YH. Simplified normal mode analysis of conformational transitions in DNA-dependent polymerases: The elastic network model. *J Mol Biol* 2022;**320**:1011–24. https://doi.org/10.1016/S0022-2836(02)00562-4
27. Li C, Lv D, Zhang L. *et al.* Approach to the unfolding and folding dynamics of add A-riboswitch upon adenine dissociation using a coarse- grained elastic network model. *J Chem Phys* 2016;**145**. https://doi.org/10.1063/1.4954992
28. Han Z, Shao Q, Gong W. *et al.* Interpreting the dynamics of binding interactions of snRNA and U1A using a coarse-grained model. *Biophys J* 2019;**116**:1625–36. https://doi.org/10.1016/j.bpj.2019.03.008
29. Yang J, Wang Y, Zhang Y. ResQ: An approach to unified estimation of B-factor and residue-specific error in protein structure prediction. *J Mol Biol* 2016;**428**:693–701. https://doi.org/10.1016/j.jmb.2015.09.024
30. Pandey A, Liu E, Graham J. *et al.* B-factor prediction in proteins using a sequence-based deep learning model. *Patterns* 2023;**4**:100805. https://doi.org/10.1016/j.patter.2023.100805
31. Wang Q, Xiao X, Miao Z. *et al.* Prediction of protein B-factor profiles based on bidirectional long short-term memory network. *IEEE Transactions on Computational Biology and Bioinformatics* 2025;**22**:1674–81. https://ieeexplore.ieee.org/abstract/document/10976581
32. Bramer D, Wei G-W. Blind prediction of protein B-factor and flexibility. *J Chem Phys* 2018;**149**:4107.
33. Bandyopadhyay S, Mondal J. A deep autoencoder framework for discovery of metastable ensembles in biomacromolecules. *J Chem Phys* 2021;**155**:114106. https://doi.org/10.1063/5.0059965
34. Guruge I, Taherzadeh G, Zhan J. *et al.* B-factor profile prediction for RNA flexibility using support vector machines. *J Comput Chem* 2018;**39**:407–11. https://doi.org/10.1002/jcc.25124
35. Barnwal RP, Loh E, Katherine SG. *et al.* Structure and mechanism of a molecular rheostat, an RNA thermometer that modulates immune evasion by Neisseria meningitidis. *Nucleic Acids Res* 2016;**44**:9426–37. https://doi.org/10.1093/nar/gkw584
36. Barnwal RP, Yang F, Varani G. Applications of NMR to structure determination of RNAs large and small. *Arch Biochem Biophys* 2017;**4**:2–56.
37. Sharma A, Gopi P, Trivedi R. *et al.* Small molecule based targeting of the CssA RNA thermometer: Insights from computational and biophysical approaches. *Biochimie* 2025;**237**:110–24. https://doi.org/10.1016/j.biochi.2025.07.017
38. Mundlia P, Singh SP, Kaushal A. *et al.* RNA-targeting therapeutics: Drugging the undruggable. *Future. Med Chem* 2025;**17**:1–3.
39. Pettersen EF, Goddard TD, Huang CC. *et al.* UCSF chimera-a visualization system for exploratory research and analysis. *Journal of Computational Chemistry: Organic, Inorganic, Physical, Biological* 2004;**25**:1605–12. https://doi.org/10.1002/jcc.20084
40. Zhou Z, Siddiquee MMR, Tajbakhsh N. *et al.* UNet++: Redesigning skip connections to exploit multiscale features in image segmentation. *IEEE Trans Med Imaging* 2019;**39**:1856–67. https://doi.org/10.1109/TMI.2019.2959609
41. Jorgensen WL, Chandrasekhar J, Madura JD. *et al.* Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 1983;**79**:926–35. https://doi.org/10.1063/1.445869
42. Machado MR, Pantano S. Split the charge difference in two! A rule of thumb for adding proper amounts of ions in MD simulations. *J Chem Theory Comput* 2020;**16**:1367–72. https://doi.org/10.1021/acs.jctc.9b00953
43. Case DA, Cheatham TE III, Darden T. *et al.* The Amber biomolecular simulation programs. *J Comput Chem* 2005;**26**:1668–88. https://doi.org/10.1002/jcc.20290
44. Salomon-Ferrer R, Case DA, Walker RC. An overview of the Amber biomolecular simulation package. *Wiley Interdisciplinary Reviews: Computational Molecular Science* 2013;**3**:198–210. https://doi.org/10.1002/wcms.1121
45. Xiang-Jun L, Bussemaker HJ, Olson WK. DSSR: An integrated software tool for dissecting the spatial structure of RNA. *Nucleic Acids Res* 2015;**21**:e142–2.
46. Gong S, Zhang C, Zhang Y. RNA-align: Quick and accurate alignment of RNA 3D structures based on size-independent TM-score$_{RNA}$. *Bioinformatics* 2019;**35**:4459–61. https://doi.org/10.1093/bioinformatics/btz282

Briefings in Bioinformatics, 2026, **27(1)**, bbaf720
https://doi.org/10.1093/bib/bbaf720
Problem Solving Protocol