PLOS Computational Biology

RESEARCH ARTICLE

Hierarchical analysis of RNA secondary structures with pseudoknots based on sections

Ryota Masuki[1,2], **Donn Liew**[2], **Ee Hou Yong**[2]*

1 Department of Physics, The University of Tokyo, Tokyo, Japan, **2** Division of Physics and Applied Physics, School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore, Singapore

* eehou@ntu.edu.sg

OPEN ACCESS

Citation: Masuki R, Liew D, Yong EH (2026) Hierarchical analysis of RNA secondary structures with pseudoknots based on sections. PLoS Comput Biol 22(1): e1013904. https://doi.org/10.1371/journal.pcbi.1013904

Editor: Arli Aditya Parikesit, Indonesia International Institute for Life Sciences, INDONESIA

Received: September 22, 2025

Accepted: January 9, 2026

Published: January 27, 2026

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: https://doi.org/10.1371/journal.pcbi.1013904

Abstract

Predicting RNA structures containing pseudoknots remains computationally challenging due to high processing costs and complexity. While standard methods for pseudoknot prediction require $O(N^6)$ time complexity, we present a hierarchical approach that significantly reduces computational cost while maintaining prediction accuracy. Our method analyzes RNA structures by dividing them into contiguous regions of unpaired bases ("sections") derived from known secondary structures. We examine pseudoknot interactions between sections using a nearest-neighbor energy model with dynamic programming. Our algorithm scales as $O(n^2\ell^4)$, offering substantial computational advantages over existing global prediction methods. Analysis of 726 transfer messenger RNA and 454 Ribonuclease P RNA sequences reveals that biologically relevant pseudoknots are highly concentrated among section pairs with large minimum free energy (MFE) gain. Over 90% of connected section pairs appear within just the top 3% of section pairs ranked by MFE gain. For 2-clusters, our method achieves high prediction accuracy with sensitivity exceeding 0.9 and positive predictive value above 0.8. For 3-clusters, we discovered asymmetric behavior where "former" section pairs (formed early in the sequence) are predicted accurately, while "latter" section pairs do not follow local energy predictions. This asymmetry suggests that complex pseudoknot formation follows sequential co-transcriptional folding rather than global energy minimization, providing insights into RNA folding dynamics.

Author summary

RNA molecules fold into structures to perform biological functions. However, predicting complex RNA structures known as "pseudoknots" is computationally expensive. Current methods often attempt to calculate the entire structure simultaneously, which requires significant computational resources. In this paper, we introduce a hierarchical approach that simplifies pseudoknot prediction. We break the RNA sequence into smaller "sections" of unpaired bases and calculate the

Data availability statement: All source code used to perform the analyses in this study, along with the complete datasets, is publicly available on Github. A permanent, citable version of the code and datasets used for this publication is archived on Zenodo.

Funding: R.M. acknowledges support by the GRI programme from Nanyang Technological University and the SVAP program from The University of Tokyo. D.L. and E.H.Y. acknowledges support from Nanyang Technological University, Singapore, under its Start Up Grant Scheme (04INS000175C230), Singapore Ministry of Education through the Academic Research Fund Tier 1 (RG140/22) and Academic Research Fund Tier 2 (MOE-T2EP50223-0014). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

energy required for these sections to bind locally, rather than solving for the global structure. Our analysis shows that strong local interactions are favored by biology; with over 90% of pseudoknots occurring within the top 3% of the most energetically favorable section pairs. This finding allows us to focus computational effort on the small subset of interactions that are most likely to form pseudoknots, rather than testing every possible combination. Our method achieves $> 90\%$ sensitivity for simple 2-section pseudoknots. However, for complex 3-section pseudoknots, only early-forming connections are predictable. This reveals that RNA does not simply fold into the most stable structure. Instead, folding is sequential, with earlier regions establishing interactions that constrain the final structure before synthesis of the later regions.

Introduction

Ribonucleic acid (RNA) plays fundamental roles across all life forms, carrying genetic information, regulating gene expression, and performing various catalytic functions. Non-coding RNA, which does not translate into proteins, plays important roles in many catalytic and regulatory cellular processes [1–4]. The biological functions of non-coding RNA have strong connections with its molecular structure, particularly pseudoknots [5–8]. Thus, predicting and understanding the structure of RNA for any given base sequence is important in biology, medicine, pharmacy, and other related fields. A fundamental challenge in RNA structure prediction is defining which interactions constitute the secondary structure versus pseudoknots. RNA secondary structure has multiple definitions in literature [9–11]. In this work, we define RNA secondary structure as nested base pairs that can be represented without crossing lines, while pseudoknots are non-planar interactions that create crossing patterns [12–16]. While no universal criterion exists for distinguishing secondary structure from pseudoknot base pairs [17,18], our definition enables computational tractability, aligns with standard structural notations (e.g. dot-bracket format), and provides a framework for our hiearchical method of identifying unpaired sections and predicting pseudoknot interactions between them. RNA secondary structure forms the basic frame of an RNA structure, typically exhibiting greater thermodynamic stability than pseudoknot interactions. Several reliable algorithms have been developed to predict RNA secondary structure for any given base sequence, such as dynamic programming algorithm (DPA) [19–24] and stochastic methods [25–27]. DPA is one of the most widely used methods and has been applied in many different areas [28,29]. DPA accurately predicts RNA secondary structures by minimizing the free energy contribution under a local energy model in $O(L^3)$ time for an RNA sequence of length L.

Predicting RNA structures that include pseudoknots remains challenging due to computational complexity and high processing costs. The most straightforward approach would be to DPA to find minimum free energy structures across all possible base pairings. However, this approach is NP-complete even with simple energy models that only consider individual base pair contributions [30–34]. In response,

algorithms have been developed that focuses on specific subsets of pseudoknotted structures [35–40]. These methods improve efficiency by eliminating physically impossible configurations early in the analysis. For example, a dynamic programming algorithm achieving $O(N^6)$ time complexity and $O(N^4)$ space complexity was introduced [10], providing the first method to fold optimal pseudoknotted RNAs using standard thermodynamic models. Though widely used, this approach becomes computationally intensive for longer RNA sequences. A later study showed that simpler dynamic programming techniques could achieve $O(N^5)$ time complexity [41], offering improved performance but lacking the ability to handle complex pseudoknot configurations like kissing-hairpins, which are commonly found in RNA structures. Alternative approaches have emerged, including stochastic methods [42] and novel computational strategies [43,44]. While these techniques show promise in specific scenarios, comparative analyses [45] reveal accuracy limitations, particularly when dealing with complex pseudoknot structures [46].

In order to overcome the trade-off between the accuracy of the model and the computational cost, there is now an increasing focus on hierarchical folding method [40,47,48]. The hierarchical folding method uses a two-step approach: first predicting the pseudoknot-free secondary structure, then calculating pseudoknot base pairs based on this foundation. A proposed DPA variant [49] finds pseudoknotted structures in $O(N^3)$ time and $O(N^2)$ space by applying this hierarchical principle, matching the time efficiency of algorithms that only predict secondary structures. This approach can handle complex configurations like kissing hairpins, which traditionally required $O(N^6)$ time to process. However, comparative studies show these methods face significant challenges with accuracy, particularly for complex pseudoknots in longer sequences [36,45,50]. These limitations appear largely due to the fact that pseudoknot structures are often long-range and thus affected by the many possible 3D RNA conformations, which are difficult to account for in current models.

In this paper, we analyze local pseudoknot interactions by examining how different "sections" of an RNA (derived from known secondary structures) are connected via pseudoknots. Using a nearest-neighbor energy model and dynamic programming, we calculate the minimum free energy of potential pseudoknot formations between unpaired regions. Our approach focuses strictly on local energy contributions between sections and excludes interactions between different pseudoknots.

Building on this section-wise interaction framework, we develop a computationally efficient method that scales as $O(n^2\ell^4)$, where n represents the number of sections and ℓ represents their typical length. This hierarchical approach provides further understanding into how local structural elements contribute to global RNA configurations.

Methods

Definitions of terms

We define the terminologies used in the research of RNA structure.

1. **Primary structure** of an RNA refer to its chemical sequence.
2. **Secondary structure** of an RNA is the set of all local nested pairing of complementary bases that can be represented in a planar graph without crossing lines. Even pairings between distant bases (such as initial and final bases) are considered part of the secondary structure if they maintain this planar property. Common secondary structure motif includes hairpin loops, hairpin stems, helical duplexes, bulges, internal loops, and multi-loops. Internal loops consist of unpaired nucleotides connecting exactly two helical segments ($h = 2$), whereas multibranch loops are junction points where three or more helices converge ($h \geq 3$).
3. **Pseudoknots** are base pairings between secondary motifs that cannot be represented without crossing lines in a planar path. These crossing interactions are tetiary structures. Pseudoknots are defined by these crossing patterns within the global RNA topology, not as isolated elements. The non-planarity emerges from the relationship between these interactions and the secondary structure, reflecting the topological complexity characteristic of pseudoknots.

4. **Pseudoknot base pair** is a base pair that participates in a non-nested pairing arrangement, creating crossing interactions when represented in a planar diagram. These pairs are part of the global topological structure where bases i,j form a pair and bases k,l form another pair such that $i < k < j < l$, resulting in crossing lines in a 2D representation.
5. **Pseudoknotted secondary structure** is defined as the set of all base pairs, including pseudoknots. It is specifically used when the set of base pairs includes pseudoknots.

The definitions above are explained graphically in Fig 1A. Next, we will introduce terminologies to be used in our analysis.

1. A **section** is a contiguous subset of bases in the RNA sequence that are not involved in base pairing, as determined from the known structures in our dataset. Specifically, if the ith base and jth base ($i < j$) are not paired in the secondary structure, they belong to the same section if all bases between them (from position $i + 1$ to $j - 1$) are also not involved in base pairing. Each section represents a single-stranded region that may potentially form pseudoknot interactions with other sections. This definition relies on having accurate secondary structure information, which in our study comes from established RNA structure databases.
2. **Section pair** refers to a pair of sections. The ith section and jth section are **connected** if there exists a pseudoknot that connects a base in ith section and another in jth section. See Fig 1B for a graphical description of sections.
3. ***N*-cluster** is a cluster consisting of N sections connected by pseudoknots. The number of topologically distinct pseudoknot structures increases rapidly once $N \geq 3$. We will only consider 2-clusters and 3-clusters in this work. The different types of 3-clusters are shown in Fig 1C.
4. **Pseudoknot order** is a metric for RNA structural complexity defined by the minimum number of base pair decompositions to create a nested structure [9]. In our work, 2-clusters correspond to pseudoknot order 1 (1 decomposition; e.g. H/K-type) while 3 clusters correspond to pseudoknot order 2 (2 decompositions; L/M-type).

Energy model

We use the nearest-neighbor energy model which is based on the chemical and thermodynamic properties of the RNA chains. Nearest-neighbor energy model is widely used to predict secondary structure in conjunction with DPA [51–58]. In the nearest-neighbor energy model, values of free energy contribution are assigned to each closed loop (e.g. stackings, bulge loops, interior loops, multi-loops) in RNA secondary structure, and the total free energy of the structure is calculated as the sum of all the free energy contribution from the loops [59]. The parameters and detailed rules in the energy model used in our analysis are based on mfold version 3.6 [20,21]. While these parameters were derived from experiments in salt-free conditions, we apply them as an approximation for pseudoknot prediction in the absence of comprehensive energy tables calibrated for varying salt concentrations.

Formally, the energy model is defined as

$$E(i,j) = \sum E_{\text{stack}}(s) + \sum E_{\text{bulge}}(b) + \sum E_{\text{interior}}(n) + \sum E_{\text{multi}}(m) + \sum E_{\text{hairpin}}(h) \tag{1}$$

Using Eq 1, we calculate the free energy contribution of the pseudoknot between two sections in the example shown (in linear representation) in Fig 2. According to the conventional rule of the nearest-neighbor energy model, the free energy contribution of the pseudoknot in Fig 2A is given by:

$$\begin{aligned}&(\text{FE contribution of the pseudoknot in Fig 2A})\\&= -(\text{hairpin loop (1)}) - (\text{hairpin loop (2)}) + (\text{multi loop (3)}) + (\text{stacking (4)}) + (\text{bulge loop (5)})\\&+ (\text{stacking (6)}) + (\text{multi loop (7)}) + (\text{stacking (8)}).\end{aligned} \tag{2}$$

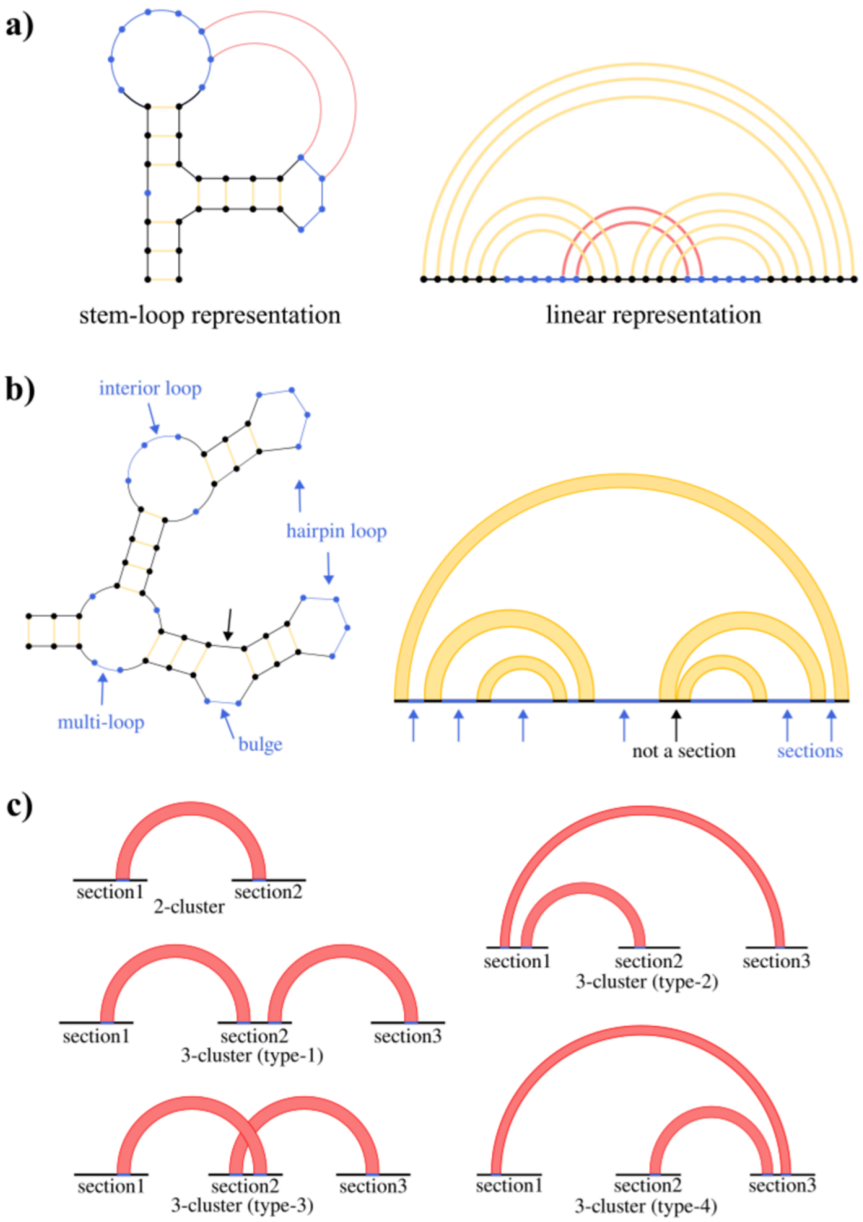

Fig 1. **Graphical representations of RNA structures and section-based pseudoknot classifications.** Yellow lines: secondary structure base pairs; red lines: pseudoknot base pairs; blue regions: sections (unpaired nucleotides); black regions: non-sections. a) RNA structure in stem-loop and linear representations. b) Stem-loop (left) and linear (right) representations of the same structure. Blue arrows indicate sections (contiguous unpaired regions). The black arrow indicates a non-section where nucleotides participate in secondary structure pairing. c) Topologically distinct pseudoknot structures of both 2-clusters and 3-clusters.

https://doi.org/10.1371/journal.pcbi.1013904.g001

However, in this work, the contribution of loops that include base pairs in secondary structures are neglected, and only loops that include pseudoknot base pairs are considered. Hence, the free energy contribution of the psuedoknot in Fig 2 is calculated as

$$\begin{aligned}&(\text{FE contribution of the pseudoknot in Fig 2A})\\&= (\text{stacking (4)}) + (\text{bulge loop (5)}) + (\text{stacking (6)}) + (\text{multi loop (7)}) + (\text{stacking (8)}).\end{aligned} \tag{3}$$

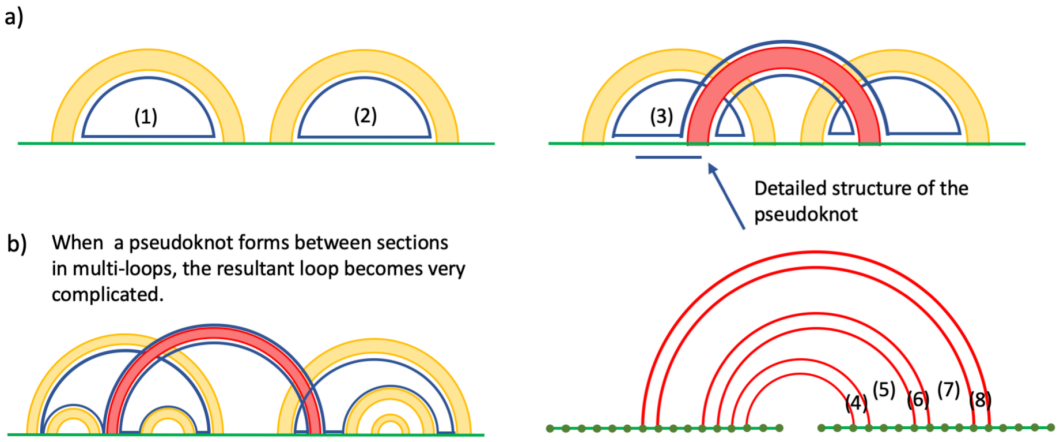

Fig 2. **Energy model calculation examples demonstrating local versus global structural considerations.** a) Pseudoknot base pair structure. The yellow bands represent sets of base pairs belonging to a secondary structure, while the red band represents a set of consecutive pseudoknots. Bottom right shows detailed base pair structure in the pseudoknot. b) Pseudoknot connecting sections across multi-loops. The newly formed loop that includes the pseudoknot contains all sections from the connected multi-loops, requiring global rather than local energy calculations.

https://doi.org/10.1371/journal.pcbi.1013904.g002

Our energy model uses two main nearest-neighbour thermodynamic parameters: stacking energies and loop energies. Stacking energies represent free energy contributions of stacked pairs formed by consecutive base pairs, with values ranging from –3.4 kcal/mol (most stable) to +1.5 kcal/mol (least stable). Loop energies compose of multiple parameter arrays such as destabilising values for interior loops (4.1 to 7.4 kcal/mol for lengths 1–30), bulge loops energies (3.9 to 6.7 kcal/mol), mismatch energy parameters (–2.7 to +0.2 kcal/mol) [21]. We exclude free energy contributions from loops containing secondary structure base pairs to focus on local interactions between sections under analysis. This exclusion is essential for our hierarchical approach. For example, when sections within our target pair reside in a multi-loop (Fig 2B), calculating their local pseudoknot structure would necessitate predicting pseudoknot formations across other section pairs simultaneously, forcing expansion from local to global structure prediction and defeating the purpose of our section-based approach. Moreover, global base pair structures predicted by DPA are known to be inaccurate even with additional energy parameters for pseudoknot motifs [49], likely due to inability to account for 3D conformation effects [60, 61]. Rather than incorporating 3D conformational effects into DPA, we focus on predicting local structure using only local information.

Our energy model further simplifies entropic contributions by treating loop contributions as purely additive. This additive approximation, independent analysis of section pairs, and loop exclusion achieve $O(n^2\ell^4)$ computational cost while maintaining focus on local interactions. The empirical validity of these approximations is demonstrated in our Results section. For exact thermodynamic parameter values, refer to our code. For exact thermodynamic parameter values used in this study, refer to our Github.

While our approach for energy calculations shares mathematical similarities with RNA hybridization algorithms through nearest-neighbor dynamic programming models, there are key differences in our implementation. Our method specifically addresses intramolecular interactions between sections within a single RNA molecule, unlike hybridization algorithms that focus on intermolecular binding [62,63]. Our approach also incorporates topological constraints specific to pseudoknot formation within an existing RNA secondary structure. For 3-clusters analysis, we apply weighted energy minimization (as shown in Fig 6, where $w = 0.8$ proved optimal), a consideration not typically present in standard hybridization algorithms [64–66]. Furthermore, our method distinguishes between different pseudoknot topologies (Fig 1C) and prevents crossing pseudoknot interactions within section pairs. We apply these techniques to analyse local pseudoknot interactions within known unpaired regions, which helps demonstrate how RNA structures can be examined in a hierarchical manner.

Algorithm

Our algorithm determines the MFE structure between section pairs by considering all possible structures where base pairs can form pseudoknots with the secondary structure, constrained so that pseudoknot base pairs within the same section pair cannot cross each other. This allows pseudoknots between different sections while preventing nested pseudoknots within the same section pair. Moreover, pairing between two sections from the same loop is prohibited because such base pairs would be nested rather than pseudoknotted. Sections are identified by analyzing the loop regions from the predicted secondary structure, where each continuous stretch of unpaired nucleotides within a loop forms a distinct section. The algorithm systematically identifies all sections by scanning the RNA sequence and grouping consecutive unpaired bases that belong to the same loop structure. Fig 3 shows the flowchart of our proposed algorithm.

We denote the two sections in the given section pair as "section1" and "section2", with length ℓ_1 and ℓ_2 respectively. We define a matrix C whose entry $C(i,j)$ $(1 \le i \le \ell_1,\ 1 \le j \le \ell_2)$ represents the minimum value of free energy contribution over candidate structures in which ith base in section1 and jth base in section2 are paired and i'th base in section1 or j'th base in section2 are not connected when either $i' < i$ or $j < j'$. We set $C(i,j) = \infty$ when the ith base in section1 and the jth base in section2 cannot form a base pair. By definition, the MFE between the section-pairs that we are calculating is the minimum entry of the C matrix. Thus, the problem of finding the MFE turns into the calculation of matrix C. The entries of matrix C can be calculated using the following recursion relation.

$$C(i,j) = \min \begin{cases} C(i+1,j-1) + (\text{stacking energy}) \text{ if } (i \neq \ell_1,\ j \neq 1) \\ C(i+d,j-1) + (\text{bulge loop energy}) \text{ if } (2 \le d \le \ell_1 - i,\ j \neq 1) \\ C(i+1,j-d) + (\text{bulge loop energy}) \text{ if } (i \neq \ell_1,\ 2 \le d \le j-1) \\ C(i+d_1,j-d_2) + (\text{interior loop energy}) \text{ if } (2 \le d_1 \le \ell_1 - i,\ 2 \le d_2 \le j-1) \\ 0 \ \ (\text{if the base pair is the innermost one}) \end{cases} \tag{4}$$

After calculating the C matrix (see Algorithm 1), we determine the MFE structure between section pairs by selecting the binding sites through dynamic programming that minimizes free energy, where the algorithm identifies the minimum entry in the C matrix, representing the optimal base pairing configuration. Base pairs are formed only between complementary bases (AU, UA, CG, GC, GU, UG). This restriction to canonical pairs (Watson-Crick and wobble) is due to the availability of thermodynamic parameters in mfold 3.6. Analysis of our datasets shows canonical pairs constitute over 99% of pseudoknot pairs in transfer messenger RNA (tmRNA) and over 90% in ribonuclease P RNA (RNase P RNA). For sections to interact, they must be from different loops, contain complementary bases, and have negative free energy contribution due to their interaction. For 2-clusters, we use the MFE structure between the section pair. For 3-clusters, we apply weighted energy minimization: (energy of former pair) + $w \times$ (energy of latter pair), where $0 < w \le 1$. Sections are ordered by their position in the RNA sequence (5' to 3' direction), with the former section pair connecting sections appearing earlier and the latter section pair connecting sections appearing later. This ordering reflects co-transcriptional folding, where 5' regions can form interactions before 3' regions are fully synthesized [67,68]. We ensure consistency by preventing bases already paired in former sections from forming pairs in latter sections. This hierarchical approach constructs the global structure by first identifying favorable local interactions, then combining them while maintaining structural constraints.

We prove the computational complexity of our algorithm is $O(n^2 l^4)$. If n is the number of sections and l the typical length of one section, we compute the MFE for all possible section pairs for n sections, yielding $n(n-1)/2 \approx O(n^2)$. For each section pair with lengths l_1 and l_2, we construct matrix C of size $l_1 \times l_2$. For each entry $C(i,j)$, there are three cases: stacking with $O(1)$ complexity, bulge loop with $O(l_1)$ and $O(l_2)$ complexity respectively, interior loop with $O(l_1 \times l_2)$ complexity due to two nested loops of length l_1 and l_2. Thus, the worst-case time complexity for each $C(i,j)$ entry is $O(l_1 \times l_2 \times l_1 \times l_2)$ or $O(l^4)$

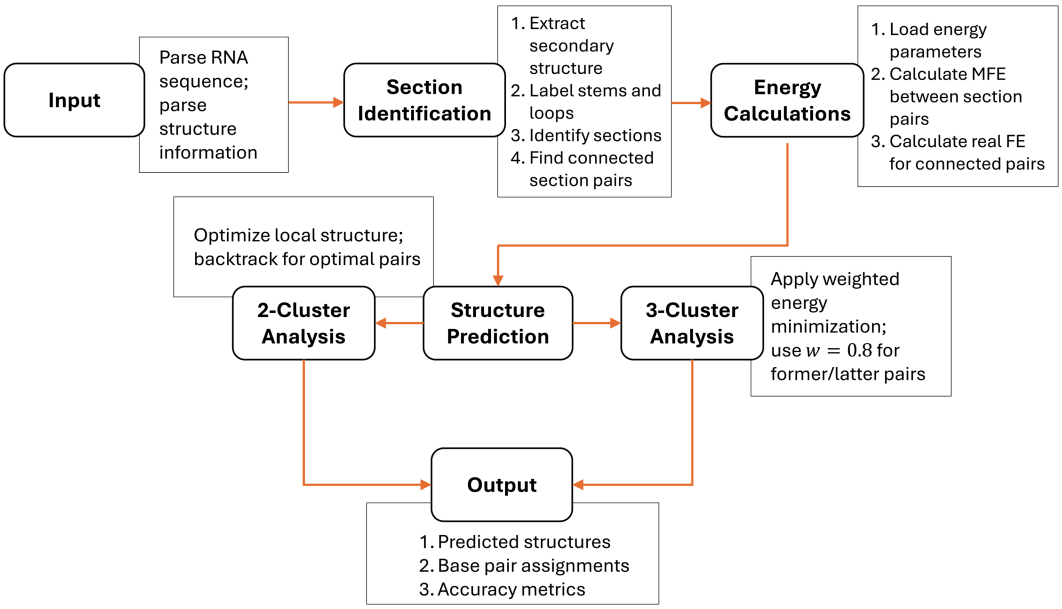

Fig 3. Implementation flowchart of the RNA pseudoknot structure prediction algorithm. The flowchart illustrates the hierarchical approach of our method, starting with input parsing, followed by section identification from the secondary structure. Energy calculations determine the minimum free energy (MFE) between section pairs. Structure prediction then branches into 2-cluster and 3-cluster analysis paths, with the latter using weighted energy minimization ($w = 0.8$) to balance former and latter section pair contributions. The output includes predicted structures, base pair assignments, and accuracy metrics.

https://doi.org/10.1371/journal.pcbi.1013904.g003

assuming $l_1 \approx l_2 \approx l$. Thus, for all $O(n^2)$ section pairs, the total computational cost is $O(n^2 l^4)$. For typical RNA molecules in our dataset, we observed $n \approx N/l$, where N is the total RNA length. This relationship suggests our algorithm's complexity in terms of total RNA length is approximately $O(N^2)$.

Our algorithm was implemented in C and compiled with gcc optimization flags. Runtime analysis on our complete dataset (1,180 sequences, 102–1,331 nt) showed a total execution time of 7.54 seconds, with mean runtime of 6.39 ms per sequence (median: 5.35 ms, IQR: 3.98–6.65 ms). Only 6.8% of sequences were classified as statistical outliers beyond 1.5×IQR, demonstrating consistent performance suitable for large-scale RNA database analysis.

Results and discussion

In this work, we analyzed 726 tmRNA sequences and 454 RNase P RNA sequences from the RNAstrand database [69]. These RNA families were selected because they represent a comprehensive and diverse dataset of biologically relevant structures with well-documented pseudoknots. Rather than selecting a small subset of examples, we analyzed the complete set of these RNA families available in the database to ensure a thorough and unbiased evaluation of our approach. Both families consist of relatively long RNA sequences with many pseudoknots, making them ideal test cases for our section-based prediction method.

Basic properties of investigated RNA sequences

Table 1 summarises key characteristics of tmRNA and RNase P RNA sequences from the RNAstrand database. Analysis of the RNAstrand database reveals structural similarities between the two RNA classes (Table 1). Mean sequence lengths are 368 nt (tmRNA) and 333 nt (RNase P RNA), with ∼30 sections per sequence and mean section lengths of 8 nt and 5 nt respectively. tmRNA sequences show greater variability in their measurements. The increase in standard deviation is

Algorithm 1 Dynamic programming algorithm for MFE calculation between section pairs.

```
ℓ1, ℓ2: section lengths;  s1, s2: starting positions;  C: DP matrix;  can_pair: possible base pairs;
mismatches: mismatch types between bases
function SEGMENT_MFE(segments, loops, bases, energy_params, ℓ, seg1, seg2)
   l1, l2 ← get_section_lengths(segments, ℓ, seg1, seg2)
   s1, s2 ← get_section_starting_points(segments, ℓ, seg1, seg2)
   if loops[s1] = loops[s2] then return ∞
   end if          //  Sections in same loop cannot form pseudoknots
   Initialize can_pair[l1][l2] and mismatches[l1][l2]
   for i ← 0 to l1 − 1 do
      for j ← 0 to l2 − 1 do
         can_pair[i][j] ← pair_type(bases[s1 + i], bases[s2 + j])
         mismatches[i][j] ← mismatch_type(bases[s1 + i], bases[s2 + j])
      end for
   end for
   Initialize C[l1][l2] and set MFE ← 0.0
   for i ← l1 − 1 downto 0 do
      for j ← 0 to l2 − 1 do
         if can_pair[i][j] = 0 then
            C[i][j] ← ∞          //  i,j cannot pair
         else
            current_energy ← 0.0          //  Case: i,j is the first bond
            if i < l1 − 1 and j > 0 then          //  Stacking
               current_energy ← min(current_energy, C[i + 1][j − 1] + stacking(can_pair[i + 1][j − 1], can_pair[i][j]))
            end if
            if i < l1 − 1 then          //  Bulge loop case 1
               for d ← 2 to j do
                  current_energy ← min(current_energy, C[i+1][j−d] + bulge_loop(d−1, can_pair[i+1][j−d], can_pair[i][j],
mismatches[i][j − d + 1], mismatches[i + 1][j − 1]))
               end for
            end if
            if j > 0 then          //  Bulge loop case 2
               for d ← 2 to l1 − i do
                  current_energy ← min(current_energy, C[i+d][j−1] + bulge_loop(d−1, can_pair[i+d][j−1], can_pair[i][j],
mismatches[i + d − 1][j], mismatches[i + 1][j − 1]))
               end for
            end if
            for all d1 ∈ [2, l1 − i], d2 ∈ [2, j] do          //  Interior loop
               current_energy ← min(current_energy, C[i + d1][j − d2] + interior_loop(d1 − 1, d2 − 1, can_pair[i + d1][j −
d2], can_pair[i][j], mismatches[i + d1 − 1][j − d2 + 1], mismatches[i + 1][j − 1]))
            end for
            C[i][j] ← current_energy
            MFE ← min(MFE, current_energy)
         end if
      end for
   end for
   return MFE
end function
```

due to sequences exceeding 1000 bases, with each sequence containing a single extended section that forms no pseudoknot connections with other sections. These outlier sequences do not impact our analysis of pseudoknot interactions between section pairs. The substantial length of these RNA sequences presents a challenge, as conventional prediction methods struggle to accurately determine the structure of such extended sequences [36].

Table 1. Structural characteristics of RNA sequences analyzed from RNAstrand database. Section length refers to the number of unpaired nucleotides within each contiguous unpaired region.

Parameter	tmRNA	RNase P RNA
Sequences analyzed (n)	726	454
Mean sequence length ± SD (nt)	367.7 ± 86.3	332.6 ± 49.6
Median sequence length (nt)	363	330
Sequence length Q1–Q3 (nt)	353–370	301–365
Mean sections per sequence	29.8	28.1
Median sections per sequence	30	28
Mean section length ± SD (nt)	7.7 ± 23.1	5.2 ± 6.1
Median section length (nt)	4	4

https://doi.org/10.1371/journal.pcbi.1013904.t001

Predicting the connected section pairs based on MFE

Fig 4A shows the distribution of MFE gain (absolute value of MFE) for all section pairs. Frequency decays exponentially with increasing MFE gain. The exponential decay follows from two assumptions. Firstly, the probability of forming each additional base pair between two sections is independent of existing structure ("probability of base pairing"). Secondly, each base pair contributes approximately 1.0 kcal/mol to the free energy, representing the average stacking energy in our model. Under these assumptions, the histogram in Fig 4 fits an exponential function with parameter a:

$$\text{frequency} \propto \exp(-a \times (\text{MFE gain})). \tag{5}$$

The probability of base pairing can be calculated as

$$\begin{aligned}\text{probability of base pairing} &= \exp(-a \times (\text{free energy contribution per one base pair})) \\ &= \exp(-a \times 1.0\,(\text{kcal/mol})).\end{aligned} \tag{6}$$

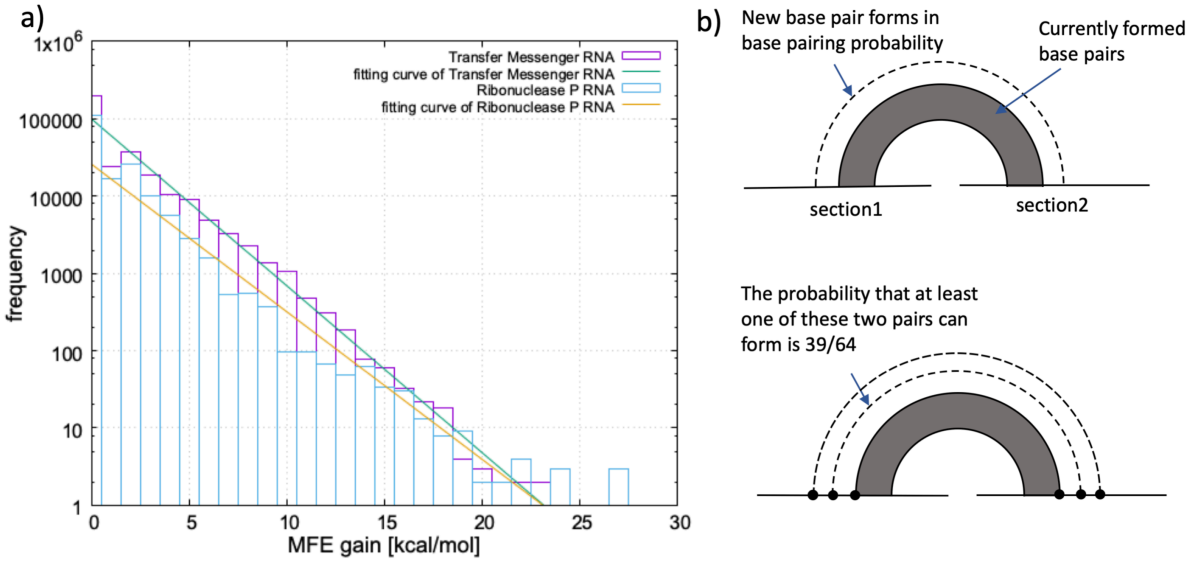

Fig 4. Distribution of MFE gain for section pairs in RNA databases. a) Histogram showing exponential decay of MFE gain frequency for all section pairs in tmRNA and RNase P RNA sequences from RNAstrand database. b) Simple probabilistic model explaining the observed exponential decay, assuming independent base pairing probability and ~1.0 kcal/mol energy contribution per base pair.

https://doi.org/10.1371/journal.pcbi.1013904.g004

Curve fitting analysis reveals base pairing probabilities of 0.608 for tmRNA and 0.632 for RNase P RNA. These values reflect fundamental RNA base pairing constraints. With six possible pairing types (AU, UA, CG, GC, UG, GU), the probability of two adjacent bases forming a new base pair is 3/8 (0.375). The probability of at least one of two adjacent base pairs forming next to an existing pair is $1 - (5/8)^2 = 39/64$ (0.609), matching our observed pairing probabilities.

Our results show that MFE structures between section pairs primarily consist of stacking interactions and small internal loops of length 2. The exponential decay in the distribution reflects the rarity of section pairs with large MFE gain. Specifically, only 5.5% of pseudoknots in tmRNA and 2.8% in RNase P RNA have MFE gain exceeding 5.0 kcal/mol. Even fewer exceed 10.0 kcal/mol: just 0.47% in tmRNA and 0.23% in RNase P RNA.

We investigated the connecting probability for both tmRNA and RNase P RNA in our database (Fig 5). Connecting probability is defined as the ratio between connected section pairs and all section pairs with a given MFE:

$$\text{Connecting probability} = \frac{\text{Total number of connected section pairs with MFE}}{\text{Total number of all section pairs with MFE}}, \tag{7}$$

Connecting probability correlates with high MFE gain as expected, since larger energy gains stabilize structures. However, we observed non-connecting section pairs with unexpectedly large MFE gains. These sections have large total MFE gain but small MFE gain per base pair compared to genuinely connected pairs. The loss of conformational entropy, which our nearest-neighbor energy model ignores, outweighs the stabilizing effect of each individual pair.

Connected section pairs concentrate among those with large MFE gains, which represent a small fraction of all section pairs. We sorted all section pairs by MFE gain and examined the cumulative distribution of connected pairs. We find that more than 90% of connected section pairs in tmRNA are included in the top 10^4 section pairs (3%) with largest maximum free energy gains (out of all the 3.5×10^5 possible section pairs). Over 90% of connected section pairs fall within the top

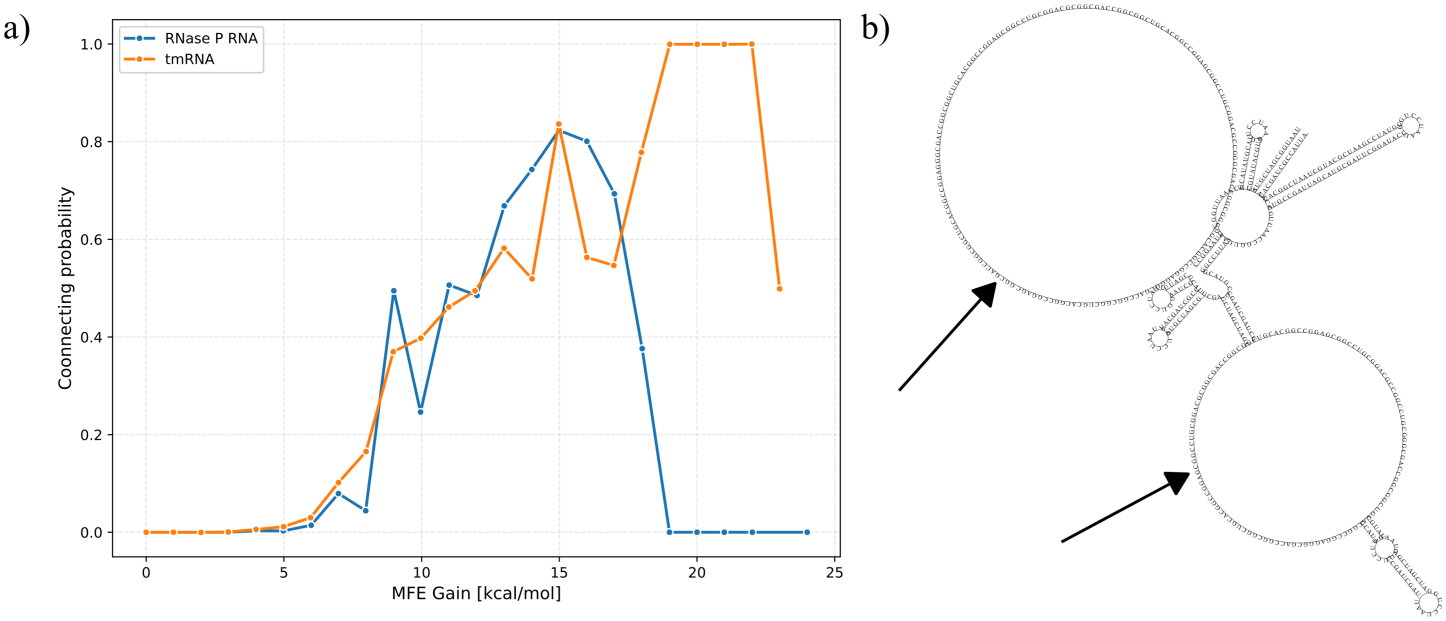

Fig 5. Analysis of pseudoknot formation likelihood based on energetic predictions. a) Connecting probability as a function of MFE gain for tmRNA and RNase P RNA sequences, defined as the ratio of actually connected section pairs to all section pairs with a given MFE value. b) Representative example of a non-connecting section pair with substantial MFE gain (80 kcal/mol), illustrating how conformational entropy losses can outweigh local energy gains in long sections (~104 and ~48 nucleotides).

https://doi.org/10.1371/journal.pcbi.1013904.g005

3% by MFE gain for tmRNA (10^4 of 3.5×10^5 pairs) and top 1% for RNase P RNA (1,500 of 1.8×10^5 pairs). Therefore, it is sufficient to only consider a small fraction of section pairs with large MFE gain in order to predict RNA structures with pseudoknots. On average, we only need to consider ∼14 section pairs per tmRNA sequence (10,000/726) or ∼3 section pairs per RNase P RNA sequence (1,500/454) to account for 90% of pseudoknots.

Comparison between MFE structures and real structures for connected section pairs

Connected section pairs exclusively form either 2-clusters or 3-clusters, with no observed instances of larger cluster formations (Table 2). Structures with pseudoknot order 1(2-clusters: 2337 in tmRNA, 543 in RNase P RNA) substantially outnumber structures with pseudoknot order 2 (3-clusters: 110 in tmRNA, 1 in RNase P RNA). This distribution is consistent with hierarchical folding, where lower-order pseudoknots form before higher-order structures. We compared MFE predictions with actual free energies (real FE) for connected 2-cluster section pairs. While the proportion of 2-clusters where real FE = MFE varies considerably between RNA classes (68.7% for tmRNA versus 39.9% for RNase P RNA), >80% of all 2-clusters demonstrate free energy gain >0.8 × MFE gain.

We compared MFE-predicted base pair arrangements with actual structures for 2-clusters. For each cluster, we calculated sensitivity (TPR) and positive predictive value (PPV) using true positives (TP), false positives (FP), and false negatives (FN). When multiple MFE structures existed, we averaged these metrics:

$$\mathrm{TPR(sensitivity)} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}, \quad \mathrm{PPV} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}. \tag{8}$$

With sensitivity >0.9 and PPV >0.8 for both RNA classes, our algorithm outperforms global predictions methods, which typically achieves a sensitivity score of 0.5–0.7 and PPV <0.7 for complex pseudoknotted structures [45].

It is important to acknowledge that our analysis relies on knowledge of the true secondary structure from which we derive the unpaired sections. This differs from the more challenging problem of *de novo* pseudoknot prediction without prior structural information. While our results demonstrate that local energetic considerations can effectively predict pseudoknot formations between known unpaired regions, additional research would be needed to determine whether similar accuracy could be achieved when starting with predicted rather than true secondary structures.

3-clusters exhibit several topologically distinct configurations (Fig 1C), with 110 type-1 3-clusters in tmRNA and only one type-2 3-cluster in RNase P RNA. Given limited type-2 samples, we analyze only type-1 3-clusters, designating section1-section2 (see Fig 1C) connections as “former pairs” and section2-section3 as “latter pairs”.

We analysed the 110 type-1 3-clusters in tmRNA by comparing MFE calculations with actual structures for both pair types, treating them independently (similar to 2-cluster analysis). For former section pairs, approximately 50% have real structures identical to predicted MFE structures, while over 80% have free energy contributions lower than 0.8×MFE (Table 3). These results closely resemble our 2-cluster prediction accuracy. In contrast, latter section pairs show markedly different behavior, with only a small fraction having free energy contributions approximating their MFE. Former section pairs demonstrate prediction accuracy comparable to 2-clusters, while latter section pairs exhibit particularly low PPV

Table 2. The result of analyses based on section pairs.

Parameter	tmRNA	RNase P RNA
The number of all 2-clusters	2,337	534
The number of all 3-clusters	110	1
The number of 2-clusters for which MFE = real FE	1,605 (68.7%)	213 (39.9%)
The number of 2-clusters for which real FE < 0.8×MFE	1,972 (84.4%)	438 (82.0%)
Sensitivity of base pair prediction for 2-clusters	0.9159	0.9087
PPV of base pair prediction for 2-clusters	0.8400	0.8056

https://doi.org/10.1371/journal.pcbi.1013904.t002

Table 3. The comparison of MFE and free energy contribution of real structure for connected section pairs in 3-clusters.

Parameter	Former section pair	Latter section pair
Total number	110	110
The number of section pair for which real FE = MFE	55 (50.0%)	10 (9.1%)
The number of section pair for which real FE<0.8×MFE	91 (82.7%)	18 (16.3%)
Sensitivity of base pair prediction	0.8343	0.5529
PPV of base pair prediction	0.7164	0.2568

https://doi.org/10.1371/journal.pcbi.1013904.t003

values. Across both pair types, sensitivity consistently exceeds PPV, indicating our method tends to predict excess base pairs. This over-prediction is expected, as our approach treats these connections as independent 2-clusters despite their presence in more complex 3-cluster structures.

The failure suggests that former section pairs form independently, whereas latter section pair formations are strongly influenced by pre-existing base-pair connections. This asymmetry appears to reflect the fundamental biology of RNA synthesis, which proceeds directionally from 5' to 3' along the sequence. During transcription, former section pairs likely establish their interactions before latter sections are even fully synthesised, as supported by previous research on co-transcriptional folding dynamics [70,71]. The consistency of this pattern across our dataset suggests it represents a general feature of RNA folding rather than a family-specific characteristic. These findings challenge the assumption that RNA structures in cells form solely through global free energy minimization [72–75]. Instead, our results suggest that the dynamic and sequential nature of RNA synthesis/folding plays a crucial role in determining the final structural configuration, particularly for complex pseudoknot arrangements. This sequential dependency within pseudoknot order 2 structures demonstrates that even within a single higher-order pseudoknot, simpler interactions must establish before complex interactions can form.

To evaluate whether our method's accuracy depends on sequence length, we analyzed all RNA sequences from our dataset. For each sequence, we compared predicted connections (negative MFE values) against actual connections in known structures. We calculated the true positive rate (TPR) and positive predictive value (PPV) for each sequence and examined their correlation with sequence length using Pearson correlation analysis.

For TPR versus sequence length, $r = -0.026$ ($p = 0.392$), while for PPV versus sequence length, $r = 0.046$ ($p = 0.129$). Both metrics show negligible correlation with sequence length ($|r| < 0.05$) with high p-values ($p > 0.05$), confirming no significant dependency. Under the null hypothesis of no correlation, these results support the conclusion that prediction accuracy remains consistent across the tested sequence length range.

The prediction accuracy independence is due to our hierarchical approach. Our algorithm calculates minimum free energy locally between section pairs rather than globally across the entire RNA structure. Consequently, prediction accuracy is determined by local section-pair interactions and section length ℓ rather than total RNA length L, consistent with our $O(n^2\ell^4)$ computational complexity. For a given section length, computational cost scales with the number of section pairs, not with total sequence length, ensuring that prediction accuracy remains stable regardless of overall RNA size.

The comparison between MFE structures and real structures for 3-clusters

Our previous analysis treated section pairs independently, ignoring that bases in section2 cannot simultaneously pair with both section1 and section3. We now predict complete 3-cluster structures while enforcing this section pair dependency with two prediction strategies:

1. Equal weighting: Minimize the combined free energy contribution from both former and latter section pairs, treating them with equal importance.
2. Sequential determination: First identify the minimum free energy structure of the former section pair, then determine the latter section pair structure while ensuring no bases in section2 are reused.

These approaches represent special cases of a weighted method that minimizes:

$$(\text{free energy contribution of former pair}) + w \times (\text{free energy contribution of latter pair}), \quad (9)$$

where $0 < w < 1$. We implemented an energy minimisation algorithm that evaluates all possible base pairings between sections (Fig 6), calculating MFE contributions for each former-latter pair combination using nearest-neighbor thermodynamic parameters from mfold 3.6. The algorithm considers stacking interactions, bulge loops, and interior loops while preventing bases already paired in the former pair from participating in latter pair interactions. We determined optimal w by testing values from 0 to 1, calculating weighted energies, predicted base pairs, and accuracy metrics (sensitivity and PPV) for each. Optimal performance occurs at $w = 0.8$, indicating preferential weighting of former pairs while maintaining latter pair contributions. However, 3-cluster latter pair prediction remains poor across all w values, suggesting their formation is influenced by global structure or 3D conformation beyond our local energy model's scope.

A limitation of our approach is the use of energy parameters derived from salt-free experimental conditions (mfold 3.6) to predict structures that typically form in the presence of salt. This approximation was necessary due to the lack of comprehensive energy tables for RNA in various salt concent rations. Despite this limitation, our method demonstrates strong predictive power for 2-clusters, suggesting that local pseudoknot formation may be reasonably approximated by these parameters even without explicitly modeling ionic effects. Future work could significantly improve prediction accuracy by incorporating salt-dependent energy parameters as they become available.

Our method is successful despite using additive entropy approximation for pseudoknots. Although previous work has shown that entropy contributions in pseudoknotted structures are not strictly additive [76,77], our results for 2-clusters (sensitivity >0.90, PPV >0.80) suggest this simplification works well within our hierarchical section-based framework.

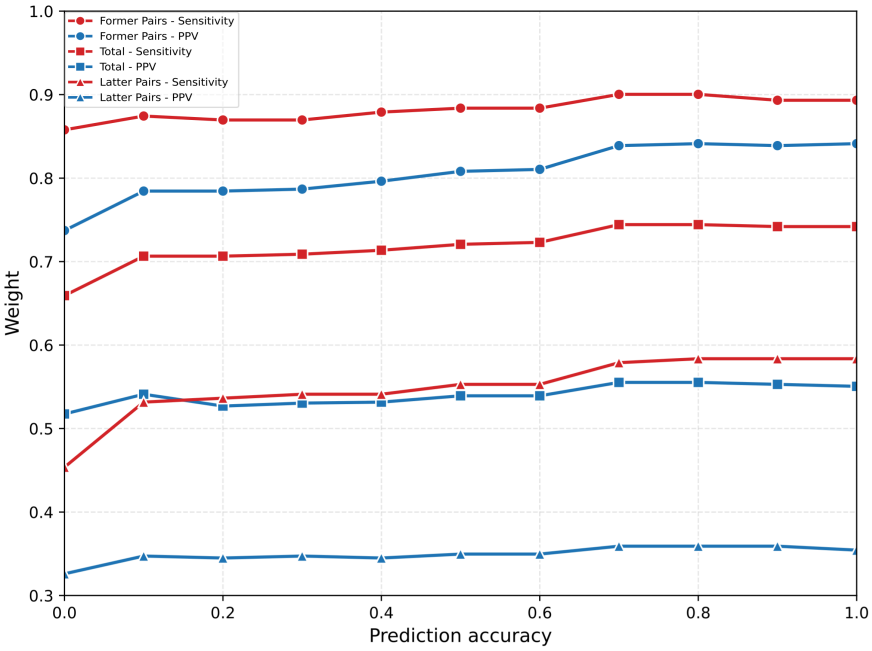

Fig 6. **Optimization of weighting parameter for 3-cluster pseudoknot structure prediction.** Prediction accuracy metrics (sensitivity and positive predictive value) as a function of weight parameter w applied to the free energy contribution of latter section pairs in the weighted energy minimization approach. Optimal performance occurs at $w = 0.8$, indicating preferential weighting of former section pair interactions while maintaining contribution from latter pairs in type-1 3-cluster configurations.

https://doi.org/10.1371/journal.pcbi.1013904.g006

The high accuracy indicates that for common pseudoknot configurations, local energy minimization with simplified entropy terms captures the essential physics governing structure formation. The method's limitations with latter section pairs in 3-clusters likely indicate where more complex entropic considerations become necessary. Nonetheless, this simplification allows for structural prediction of longer RNA sequences while maintaining good accuracy for the most common pseudoknot configurations.

Conclusion

We have presented a section-based approach to RNA pseudoknot analysis that employs mfold nearest-neighbor energy model with DPA. This method offers both computational efficiency and prediction accuracy advantages over traditional approaches. Our algorithm scales as $O(n^2\ell^4)$, where n represents the number of sections and ℓ represents the average section length. Assuming section length remains independent of total RNA sequence length, the computational complexity reduces to $O(L^2)$ for an RNA of length L. A key finding of our research is the highly concentrated distribution of biologically relevant pseudoknots among section pairs with large MFE gains. By focusing on just the top 3% of section pairs ranked by MFE gain, we can capture approximately 90% of all connected section pairs. This observation enables substantial reductions in the search space required for structure prediction.

Our approach predicts pseudoknot order 1 structures (2-clusters) with sensitivity above 0.90 and positive predictive value exceeding 0.80. These results significantly outperform conventional methods that attempt to predict global pseudoknotted structures in a single step. This indicates that once unpaired regions are identified, 2-cluster pseudoknot formations between them appear to be primarily determined by local information. This observation suggests potential value in hierarchical approaches to RNA structure prediction, though such approaches would need to address the challenge of accurately identifying relevant unpaired regions and resolving potential competition between pseudoknot and regular secondary structure formation. Our work provides evidence for the effectiveness of local energy calculations specifically for the subproblem of pseudoknot prediction between known unpaired regions. However, significant limitations exist for pseudoknot order 2 structures (3-clusters). Nearly all 3-clusters conform to the type-1 configuration (Fig 1C). While our approach can predict former section pairs in these 3-clusters with accuracy comparable to 2-clusters, it performs poorly when predicting latter section pairs. The inadequacy of our local energy model for latter section pairs suggests that these pseudoknots are significantly influenced by factors beyond local interactions, such as the dynamic process of RNA folding and the global 3D RNA conformation. These findings highlight the limitations of purely thermodynamic approaches to complex RNA structure prediction and suggest that accurate modeling of 3-clusters may require incorporating kinetic folding pathways and long-range structural constraints.

Acknowledgments

R.M., D.L., and E.H.Y thank Dr. Michael Zuker for his permission to use the mfold energy model that he developed.

Author contributions

Conceptualization: Ryota Masuki, Ee Hou Yong.

Data curation: Ryota Masuki.

Formal analysis: Ryota Masuki, Ee Hou Yong.

Funding acquisition: Ee Hou Yong.

Investigation: Ryota Masuki, Ee Hou Yong.

Methodology: Ryota Masuki, Ee Hou Yong.

Project administration: Ee Hou Yong.

Supervision: Ee Hou Yong.

Validation: Ee Hou Yong.

Visualization: Ryota Masuki, Donn Liew, Ee Hou Yong.

Writing – original draft: Ryota Masuki, Ee Hou Yong.

Writing – review & editing: Donn Liew, Ee Hou Yong.

References

1. Gupta RA, Shah N, Wang KC, Kim J, Horlings HM, Wong DJ, et al. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. Nature. 2010;464(7291):1071–6. https://doi.org/10.1038/nature08975 PMID: 20393566
2. Mattick JS, Makunin IV. Non-coding RNA. Hum Mol Genet. 2006;15 Spec No 1:R17-29. https://doi.org/10.1093/hmg/ddl046 PMID: 16651366
3. Esteller M. Non-coding RNAs in human disease. Nat Rev Genet. 2011;12(12):861–74. https://doi.org/10.1038/nrg3074 PMID: 22094949
4. Guttman M, Rinn JL. Modular regulatory principles of large non-coding RNAs. Nature. 2012;482(7385):339–46. https://doi.org/10.1038/nature10887 PMID: 22337053
5. Staple DW, Butcher SE. Pseudoknots: RNA structures with diverse functions. PLoS Biol. 2005;3(6):e213. https://doi.org/10.1371/journal.pbio.0030213 PMID: 15941360
6. Tholstrup J, Oddershede LB, Sørensen MA. mRNA pseudoknot structures can act as ribosomal roadblocks. Nucleic Acids Res. 2012;40(1):303–13. https://doi.org/10.1093/nar/gkr686 PMID: 21908395
7. Wang C, Le SY, Ali N, Siddiqui A. An RNA pseudoknot is an essential structural element of the internal ribosome entry site located within the hepatitis C virus 5' noncoding region. RNA. 1995;1(5):526–37. PMID: 7489514
8. Powers T, Noller HF. A functional pseudoknot in 16S ribosomal RNA. EMBO J. 1991;10(8):2203–14. https://doi.org/10.1002/j.1460-2075.1991.tb07756.x PMID: 1712293
9. Antczak M, Popenda M, Zok T, Zurkowski M, Adamiak RW, Szachniuk M. New algorithms to represent complex pseudoknotted RNA structures in dot-bracket notation. Bioinformatics. 2018;34(8):1304–12. https://doi.org/10.1093/bioinformatics/btx783 PMID: 29236971
10. Rivas E, Eddy SR. A dynamic programming algorithm for RNA structure prediction including pseudoknots. J Mol Biol. 1999;285(5):2053–68. https://doi.org/10.1006/jmbi.1998.2436 PMID: 9925784
11. Brion P, Westhof E. Hierarchy and dynamics of RNA folding. Annu Rev Biophys Biomol Struct. 1997;26:113–37. https://doi.org/10.1146/annurev.biophys.26.1.113 PMID: 9241415
12. Gray M, Will S, Jabbari H. SparseRNAfolD: optimized sparse RNA pseudoknot-free folding with dangle consideration. Algorithms Mol Biol. 2024;19(1):9. https://doi.org/10.1186/s13015-024-00256-4 PMID: 38433200
13. Benedetti G. Rna secondary structure prediction using a genetic algorithm with a selection method based on free energy value and topological index. bioRxiv. 2024.
14. Bon M, Vernizzi G, Orland H, Zee A. Topological classification of RNA structures. J Mol Biol. 2008;379(4):900–11. https://doi.org/10.1016/j.jmb.2008.04.033 PMID: 18485361
15. Danaee P, Rouches M, Wiley M, Deng D, Huang L, Hendrix D. bpRNA: large-scale automated annotation and analysis of RNA secondary structure. Nucleic Acids Res. 2018;46(11):5381–94. https://doi.org/10.1093/nar/gky285 PMID: 29746666
16. Stadler PF. Alignments of biomolecular contact maps. Interface Focus. 2021;11(4):20200066. https://doi.org/10.1098/rsfs.2020.0066 PMID: 34123355
17. Smit S, Rother K, Heringa J, Knight R. From knotted to nested RNA structures: a variety of computational methods for pseudoknot removal. RNA. 2008;14(3):410–6. https://doi.org/10.1261/rna.881308 PMID: 18230758
18. Zok T, Badura J, Swat S, Figurski K, Popenda M, Antczak M. New models and algorithms for RNA pseudoknot order assignment. International Journal of Applied Mathematics and Computer Science. 2020;30(2). https://doi.org/10.34768/amcs-2020-0024
19. Nussinov R, Pieczenik G, Griggs JR, Kleitman DJ. Algorithms for Loop Matchings. SIAM J Appl Math. 1978;35(1):68–82. https://doi.org/10.1137/0135006
20. Zuker M, Mathews DH, Turner DH. Algorithms and thermodynamics for RNA secondary structure prediction: a practical guide. RNA biochemistry and biotechnology. Springer Netherlands. 1999. p. 11–43. https://doi.org/10.1007/978-94-011-4485-8_2
21. Mathews DH, Sabina J, Zuker M, Turner DH. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. J Mol Biol. 1999;288(5):911–40. https://doi.org/10.1006/jmbi.1999.2700 PMID: 10329189
22. Hofacker IL. Vienna RNA secondary structure server. Nucleic Acids Res. 2003;31(13):3429–31. https://doi.org/10.1093/nar/gkg599 PMID: 12824340

23. Huang L, Zhang H, Deng D, Zhao K, Liu K, Hendrix DA, et al. LinearFold: linear-time approximate RNA folding by 5'-to-3' dynamic programming and beam search. Bioinformatics. 2019;35(14):i295–304. https://doi.org/10.1093/bioinformatics/btz375 PMID: 31510672

24. Sato K, Akiyama M, Sakakibara Y. RNA secondary structure prediction using deep learning with thermodynamic integration. Nat Commun. 2021;12(1):941. https://doi.org/10.1038/s41467-021-21194-4 PMID: 33574226

25. Flamm C, Fontana W, Hofacker IL, Schuster P. RNA folding at elementary step resolution. RNA. 2000;6(3):325–38. https://doi.org/10.1017/s1355838200992161 PMID: 10744018

26. Zhang H, Li S, Zhang L, Mathews DH, Huang L. LazySampling and LinearSampling: fast stochastic sampling of RNA secondary structure with applications to SARS-CoV-2. Nucleic Acids Res. 2023;51(2):e7. https://doi.org/10.1093/nar/gkac1029 PMID: 36401871

27. Fiscon G. A perspective on the algorithms predicting and evaluating the RNA secondary structure. J Genet Genome Res. 2016;3(1). https://doi.org/10.23937/2378-3648/1410023

28. Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, et al. RNA-guided human genome engineering via Cas9. Science. 2013;339(6121):823–6. https://doi.org/10.1126/science.1232033 PMID: 23287722

29. Lagos-Quintana M, Rauhut R, Lendeckel W, Tuschl T. Identification of novel genes coding for small expressed RNAs. Science. 2001;294(5543):853–8. https://doi.org/10.1126/science.1064921 PMID: 11679670

30. Lyngsø RB, Pedersen CN. RNA pseudoknot prediction in energy-based models. J Comput Biol. 2000;7(3–4):409–27. https://doi.org/10.1089/106652700750050862 PMID: 11108471

31. Yuan Y, Yang E, Zhang R. Wfold: a new method for predicting RNA secondary structure with deep learning. Comput Biol Med. 2024;182:109207. https://doi.org/10.1016/j.compbiomed.2024.109207 PMID: 39341115

32. Fox DM, MacDermaid CM, Schreij AMA, Zwierzyna M, Walker RC. RNA folding using quantum computers. PLoS Comput Biol. 2022;18(4):e1010032. https://doi.org/10.1371/journal.pcbi.1010032 PMID: 35404931

33. Saman Booy M, Ilin A, Orponen P. RNA secondary structure prediction with convolutional neural networks. BMC Bioinformatics. 2022;23(1):58. https://doi.org/10.1186/s12859-021-04540-7 PMID: 35109787

34. Fu L, Cao Y, Wu J, Peng Q, Nie Q, Xie X. UFold: fast and accurate RNA secondary structure prediction with deep learning. Nucleic Acids Res. 2022;50(3):e14. https://doi.org/10.1093/nar/gkab1074 PMID: 34792173

35. Marchand B, Will S, Berkemer SJ, Ponty Y, Bulteau L. Automated design of dynamic programming schemes for RNA folding with pseudoknots. Algorithms Mol Biol. 2023;18(1):18. https://doi.org/10.1186/s13015-023-00229-z PMID: 38041153

36. Gong T, Ju F, Bu D. Accurate prediction of RNA secondary structure including pseudoknots through solving minimum-cost flow with learned potentials. Commun Biol. 2024;7(1):297. https://doi.org/10.1038/s42003-024-05952-w PMID: 38461362

37. Dirks RM, Pierce NA. A partition function algorithm for nucleic acid secondary structure including pseudoknots. J Comput Chem. 2003;24(13):1664–77. https://doi.org/10.1002/jcc.10296 PMID: 12926009

38. Reeder J, Giegerich R. Design, implementation and evaluation of a practical pseudoknot folding algorithm based on thermodynamics. BMC Bioinformatics. 2004;5:104. https://doi.org/10.1186/1471-2105-5-104 PMID: 15294028

39. Jabbari H, Wark I, Montemagno C, Will S. Knotty: efficient and accurate prediction of complex RNA pseudoknot structures. Bioinformatics. 2018;34(22):3849–56. https://doi.org/10.1093/bioinformatics/bty420 PMID: 29868872

40. Gray M, Trinity L, Stege U, Ponty Y, Will S, Jabbari H. CParty: hierarchically constrained partition function of RNA pseudoknots. Bioinformatics. 2024;41(1):btae748. https://doi.org/10.1093/bioinformatics/btae748 PMID: 39700413

41. Akutsu T. Dynamic programming algorithms for RNA secondary structure prediction with pseudoknots. Discrete Applied Mathematics. 2000;104(1–3):45–62. https://doi.org/10.1016/s0166-218x(00)00186-4

42. Bon M, Micheletti C, Orland H. McGenus: a Monte Carlo algorithm to predict RNA secondary structures with pseudoknots. Nucleic Acids Res. 2013;41(3):1895–900. https://doi.org/10.1093/nar/gks1204 PMID: 23248008

43. Knudsen B, Hein J. Pfold: RNA secondary structure prediction using stochastic context-free grammars. Nucleic Acids Res. 2003;31(13):3423–8. https://doi.org/10.1093/nar/gkg614 PMID: 12824339

44. Do CB, Woods DA, Batzoglou S. CONTRAfold: RNA secondary structure prediction without physics-based models. Bioinformatics. 2006;22(14):e90-8. https://doi.org/10.1093/bioinformatics/btl246 PMID: 16873527

45. Sperschneider J, Datta A, Wise MJ. Heuristic RNA pseudoknot prediction including intramolecular kissing hairpins. RNA. 2011;17(1):27–38. https://doi.org/10.1261/rna.2394511 PMID: 21098139

46. Justyna M, Antczak M, Szachniuk M. Machine learning for RNA 2D structure prediction benchmarked on experimental data. Brief Bioinform. 2023;24(3):bbad153. https://doi.org/10.1093/bib/bbad153 PMID: 37096592

47. Jabbari H, Wark I, Montemagno C. RNA secondary structure prediction with pseudoknots: contribution of algorithm versus energy model. PLoS One. 2018;13(4):e0194583. https://doi.org/10.1371/journal.pone.0194583 PMID: 29621250

48. Tinoco I Jr, Bustamante C. How RNA folds. J Mol Biol. 1999;293(2):271–81. https://doi.org/10.1006/jmbi.1999.3001 PMID: 10550208

49. Jabbari H, Condon A, Zhao S. Novel and efficient RNA secondary structure prediction using hierarchical folding. J Comput Biol. 2008;15(2):139–63. https://doi.org/10.1089/cmb.2007.0198 PMID: 18312147

50. Sato K, Kato Y. Prediction of RNA secondary structure including pseudoknots for long sequences. Brief Bioinform. 2022;23(1):bbab395. https://doi.org/10.1093/bib/bbab395 PMID: 34601552

51. Opuu V, Merleau NSC, Messow V, Smerlak M. RAFFT: efficient prediction of RNA folding pathways using the fast Fourier transform. PLoS Comput Biol. 2022;18(8):e1010448. https://doi.org/10.1371/journal.pcbi.1010448 PMID: 36026505

52. Zuber J, Cabral BJ, McFadyen I, Mauger DM, Mathews DH. Analysis of RNA nearest neighbor parameters reveals interdependencies and quantifies the uncertainty in RNA secondary structure prediction. RNA. 2018;24(11):1568–82. https://doi.org/10.1261/rna.065102.117 PMID: 30104207

53. Sloma MF, Mathews DH. Exact calculation of loop formation probability identifies folding motifs in rna secondary structures. RNA 22, 1808–18 (2016).

54. Kierzek E, Zhang X, Watson RM, Kennedy SD, Szabat M, Kierzek R, et al. Secondary structure prediction for RNA sequences including N6-methyladenosine. Nat Commun. 2022;13(1):1271. https://doi.org/10.1038/s41467-022-28817-4 PMID: 35277476

55. Zuber J, Sun H, Zhang X, McFadyen I, Mathews DH. A sensitivity analysis of RNA folding nearest neighbor parameters identifies a subset of free energy parameters with the greatest impact on RNA secondary structure prediction. Nucleic Acids Res. 2017;45(10):6168–76. https://doi.org/10.1093/nar/gkx170 PMID: 28334976

56. Zuber J, Schroeder SJ, Sun H, Turner DH, Mathews DH. Nearest neighbor rules for RNA helix folding thermodynamics: improved end effects. Nucleic Acids Res. 2022;50(9):5251–62. https://doi.org/10.1093/nar/gkac261 PMID: 35524574

57. Zuber J, Mathews DH. Estimating uncertainty in predicted folding free energy changes of RNA secondary structures. RNA. 2019;25(6):747–54. https://doi.org/10.1261/rna.069203.118 PMID: 30952689

58. Spasic A, Berger KD, Chen JL, Seetin MG, Turner DH, Mathews DH. Improving RNA nearest neighbor parameters for helices by going beyond the two-state model. Nucleic Acids Res. 2018;46(10):4883–92. https://doi.org/10.1093/nar/gky270 PMID: 29718397

59. Zucker M, Mathews D, Turner D. Algorithms and thermodynamics for rna secondary structure prediction. Barciszewski J, Clark BFC (eds). A Practical Guide in RNA Biochemistry and Biotechnology. Kluwer Academic NATO ASI Series edition, Dordrecht. 1999.

60. Rice GM, Leonard CW, Weeks KM. RNA secondary structure modeling at consistent high accuracy using differential SHAPE. RNA. 2014;20(6):846–54. https://doi.org/10.1261/rna.043323.113 PMID: 24742934

61. Beckmann IK, Waldl M, Will S, Hofacker IL. 3D feasibility of 2D RNA-RNA interaction paths by stepwise folding simulations. RNA. 2024;30(2):113–23. https://doi.org/10.1261/rna.079756.123 PMID: 38071473

62. Lou F, Clote P. Thermodynamics of RNA structures by Wang-Landau sampling. Bioinformatics. 2010;26(12):i278-86. https://doi.org/10.1093/bioinformatics/btq218 PMID: 20529917

63. Bindewald E, Afonin K, Jaeger L, Shapiro BA. Multistrand rna secondary structure prediction and nanostructure design including pseudoknots. Therapeutic RNA Nanotechnology. Jenny Stanford Publishing; 2021. p. 95–120.

64. Mann M, Wright PR, Backofen R. IntaRNA 2.0: enhanced and customizable prediction of RNA-RNA interactions. Nucleic Acids Res. 2017;45(W1):W435–9. https://doi.org/10.1093/nar/gkx279 PMID: 28472523

65. Fukunaga T, Hamada M. RIblast: an ultrafast RNA-RNA interaction prediction system based on a seed-and-extension approach. Bioinformatics. 2017;33(17):2666–74. https://doi.org/10.1093/bioinformatics/btx287 PMID: 28459942

66. Lai D, Meyer IM. A comprehensive comparison of general RNA-RNA interaction prediction methods. Nucleic Acids Res. 2016;44(7):e61. https://doi.org/10.1093/nar/gkv1477 PMID: 26673718

67. Meyer IM, Miklós I. Co-transcriptional folding is encoded within RNA genes. BMC Mol Biol. 2004;5:10. https://doi.org/10.1186/1471-2199-5-10 PMID: 15298702

68. Schärfen L, Neugebauer KM. Transcription regulation through nascent RNA folding. J Mol Biol. 2021;433(14):166975. https://doi.org/10.1016/j.jmb.2021.166975 PMID: 33811916

69. Andronescu M, Bereg V, Hoos HH, Condon A. RNA STRAND: the RNA secondary structure and statistical analysis database. BMC Bioinformatics. 2008;9:340. https://doi.org/10.1186/1471-2105-9-340 PMID: 18700982

70. Stephen C, Palmer D, Mishanina TV. Opportunities for riboswitch inhibition by targeting co-transcriptional RNA folding events. Int J Mol Sci. 2024;25(19):10495. https://doi.org/10.3390/ijms251910495 PMID: 39408823

71. Widom JR, Nedialkov YA, Rai V, Hayes RL, Brooks CL 3rd, Artsimovitch I, et al. Ligand modulates cross-coupling between riboswitch folding and transcriptional pausing. Mol Cell. 2018;72(3):541-552.e6. https://doi.org/10.1016/j.molcel.2018.08.046 PMID: 30388413

72. Zhang H, Zhang C, Li Z, Li C, Wei X, Zhang B, et al. A new method of RNA secondary structure prediction based on convolutional neural network and dynamic programming. Front Genet. 2019;10:467. https://doi.org/10.3389/fgene.2019.00467 PMID: 31191603

73. Su C, Weir JD, Zhang F, Yan H, Wu T. ENTRNA: a framework to predict RNA foldability. BMC Bioinformatics. 2019;20(1):373. https://doi.org/10.1186/s12859-019-2948-5 PMID: 31269893

74. Yu B, Lu Y, Zhang QC, Hou L. Prediction and differential analysis of RNA secondary structure. Quant Biol. 2020;8(2):109–18. https://doi.org/10.1007/s40484-020-0205-6

75. Spasic A, Assmann SM, Bevilacqua PC, Mathews DH. Modeling RNA secondary structure folding ensembles using SHAPE mapping data. Nucleic Acids Res. 2018;46(1):314–23. https://doi.org/10.1093/nar/gkx1057 PMID: 29177466

76. Isambert H, Siggia ED. Modeling RNA folding paths with pseudoknots: application to hepatitis delta virus ribozyme. Proc Natl Acad Sci U S A. 2000;97(12):6515–20. https://doi.org/10.1073/pnas.110533697 PMID: 10823910

77. Aalberts DP, Hodas NO. Asymmetry in RNA pseudoknots: observation and theory. Nucleic Acids Res. 2005;33(7):2210–4. https://doi.org/10.1093/nar/gki508 PMID: 15831794