

Decoding RNA triple helices: identification from sequence and secondary structure

Margherita A.G. Matarrese ^{1,*,#}, Michela Quadrini ^{2,#}, Nicole Luchetti ¹, Federico Di Petta ², Daniele Durante ³,
Monica Ballarino ³, Letizia Chiodo ¹, Luca Tesei ^{2,*}

¹Department of Engineering, Università Campus Bio-Medico di Roma, Via Àlvaro del Portillo, 21, 00128 Rome, Italy

²School of Sciences and Technology, University of Camerino, Via Madonna delle Carceri, 7, 62020 Camerino, Italy

³Department of Biology and Biotechnologies “Charles Darwin”, Sapienza University of Rome, Piazzale Aldo Moro, 5, 00185 Roma, Italy

*Corresponding authors. Margherita A.G. Matarrese, E-mail: m.matarrese@unicampus.it; Luca Tesei, E-mail: luca.tesei@unicam.it.

[#]Margherita A.G. Matarrese and Michela Quadrini contributed equally to this work.

Abstract

The discovery of long non-coding RNAs (lncRNA) has revealed additional layers of gene-expression control. Specific interactions of lncRNAs with DNA, RNAs, and RNA-binding proteins enable regulation in both cytoplasmic and nuclear compartments; e.g. a conserved triple-helix motif is essential for MALAT1 stability and oncogenic activity. Here, we present a secondary-structure-based framework to annotate and detect RNA triple helices. First, we extend the dot-bracket formalism with a third annotation line that encodes Hoogsteen contacts. Second, we introduce TripleMatcher, which searches for a triple-helix pattern, filters candidates by C1'–C1' distance thresholds, and merges overlaps into region-level zones. Using telomerase RNAs and RNA-stability elements with experimentally established triple helices (8 RNAs), TripleMatcher localized all annotated regions (structure-wise detection 8/8); geometric filtering removed most spurious candidates and improved precision (positive predictive value from 0.42 to 0.81) and overall accuracy (F₁ from 0.42 to 0.62) while maintaining sensitivity. Benchmarking eight predictors showed that pseudoknot-aware methods most reliably reproduce the local architecture required for detection, aligning secondary-structure quality with downstream triple-helix recovery. Applied prospectively, the framework identified candidate regions directly from predicted secondary structures and scaled to a screen of 4160 RNAs, where distance filtering reduced 150 990 (median per molecule: 108 [20–270]) raw candidates to 97 geometrically feasible regions across seven molecules, including human telomerase complexes. Together, the notation and TripleMatcher provide a concise route from secondary structure to a small, interpretable set of triple-helix candidates suitable for targeted experimental validation.

Keywords non-coding RNA, long non-coding RNA, RNA pattern search, RNA secondary structure, RNA structure prediction

Introduction

RNA molecules adopt diverse secondary structures, such as stem-loops, bulges, G-quadruplexes, and pseudoknots, that regulate biosynthesis, stability, localization, and molecular interactions [1–3]. Experimental and computational approaches have linked structures to function [4–7]; chemical probing, nuclear magnetic resonance (NMR), and comparative analysis have clarified folding and dynamics [8]. Yet, capturing how these structures adapt to varying cellular conditions, and how they modulate interactions with RNA or protein partners remains a significant challenge.

The role of secondary structures is particularly crucial for non-coding RNAs (ncRNAs), whose functionality extends beyond their linear nucleotide sequences and lack of protein-coding capacity. Long non-coding RNAs (lncRNAs) can act as scaffolds, bringing proteins and RNAs into proximity within nuclear and cytoplasmic compartments

[9]. Deciphering the intricate folding patterns of these RNAs is essential for understanding their biological functions and could open avenues for therapeutic interventions in various diseases.

A significant example is the polyadenylated nuclear (PAN) RNA from *Kaposi's sarcoma-associated herpesvirus* (KSHV); PAN RNA is abundant during the lytic phase, representing a large fraction of the cell's polyadenylated RNA [10, 11]. Its expression and nuclear retention element (ENE) form a triple-helix with the poly(A) tail via U•A–U interactions, protecting PAN from exonucleolytic decay [12–15]. Another paradigmatic case is the Metastasis-Associated Lung Adenocarcinoma Transcript 1 (MALAT1), a multifunctional lncRNA with diverse roles in health and disease [16]. Conserved structural domains mediate interactions and nuclear localization [16, 17], and a conserved triple-helix underlies its stability and nuclear accumulation [18, 19].

Received: August 26, 2025. **Revised:** December 8, 2025. **Accepted:** January 5, 2026

© The Author(s) 2026. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

Beyond lncRNAs, triple helices are widespread in viral RNAs, riboswitches, catalytic ribozymes, and telomerase [17]. First observed by Felsenfeld *et al.* as U•A–U Hoogsteen triples and later confirmed in tRNA and telomerase RNA [20, 21, 22], they contribute to RNA stability and regulation. In telomerase RNA, conserved U•A–U triples stabilize the pseudoknot required for catalytic activity [23–26], and disrupting these triples impairs telomerase function [24, 27, 28].

Despite their biological significance, triple helices remain difficult to detect experimentally, primarily relying on crystallographic and NMR studies, with computational methods largely absent. Furthermore, standard notations for RNA secondary structure do not easily represent triple helices, complicating computational analyses [29, 30].

Here, we present TripleMatcher, a computational framework for identifying and characterizing RNA triple helices from sequence and secondary structure. First, we extend the dot-bracket notation to encode Hoogsteen interactions. Then, we implement a four-stage workflow: (i) structural characterization from experimentally determined 3D RNA structures; (ii) secondary-structure prediction; (iii) development and validation of TripleMatcher, a search tool to detect putative triple-helix regions; (iv) identification and reliability assessment of triple helices. To our knowledge, TripleMatcher is the first tool specifically designed to identify RNA–RNA–RNA major-groove triple helices, whereas previous computational methods focus on RNA–DNA triplex formation [31–33].

Our ultimate goal is to enable the identification of tertiary structural motifs directly from nucleotide sequence via predicted secondary structure. Integrating these higher-order interactions may help bridge the gap between experimental and predicted RNA structural analyses, particularly for lncRNAs, where experimentally resolved structures are limited.

Materials and methods

Validation dataset

We selected RNAs with experimentally determined structures containing triple helices, focusing on telomerase RNAs, and two lncRNAs, MALAT1 and PAN (Fig. 1 and Supplementary Table S1). Major groove interactions were prioritized while minor groove interactions and intermolecular triple helices (e.g. NEAT1 [34]) were excluded. As an experimental design choice, we restricted the analysis to unimolecular RNAs so that experimental structures and predicted secondary structures could be treated in a consistent way.

MALAT1 is a nuclear lncRNA involved in gene regulation and cancer progression [19, 35]. Its 3' end forms a bipartite triple-helix composed of two runs of U•A–U base triples, interrupted by a C•G–C triplet and a C–G doublet that induce a “helical reset,” realigning the strands and preventing steric clashes [19]. Two A-minor interactions engage adjacent G–C base pairs [19, 35]. The core region is in the 4PLX PDB [19, 36].

PAN RNA from KSHV forms two unimolecular triple helices [15]. The “PAN core triple-helix” features a shortened apical P2 helix, designed to promote triple-helix formation and stability, with mild protection against exonucleolytic degradation [14, 15]. The engineered “GCPAN triple-helix” adds a GC clamp at the ENE base, anchoring the A-rich sequence to the lower base-paired region and increasing exonuclease resistance [15]. High-resolution PDBs for these unimolecular conformations were derived from 3P22 and 6X5N.

Telomerase RNA (TER) provides the scaffold for ribonucleo-protein assembly and the template for telomere elongation. TER is conserved and contains essential features, including a conserved triple-helix in the pseudoknot that supports catalytic activity and stability [24, 37–39]. In *Kluyveromyces lactis* (*K. lactis*), the pseudoknot junction forms a triple-helix stabilized by C•G–C and U•A–U base triples, with bound divalent cations [38]. In *Homo sapiens*, the wild-type pseudoknot (2K95) includes a U41 bulge (U177 in [39]) that modulates catalysis; the deletion mutant (2K96/1YMO) lacks U177 but shares the network of base triples and a minor-groove A•G–C triple.

For all RNAs in the validation set, complete 3D models were generated by combining the available high-resolution segments with computationally reconstructed regions. The triple-helix core was always taken from experimentally resolved structures; only surrounding elements were modeled, if needed, to restore continuity, with ModeRNA [40] and RNAComposer [41] (Fig. 1A).

Triple helices characterization

Secondary structure validation

Experimental secondary structure was obtained from the corresponding reference publication. When PDB structures were available, we extracted sequence and base-pairing information using RNAView through RNAPdb [42] and manually adjusted the output to match the published models (Fig. 1A). We annotated all base pairs and third-strand nucleotides forming the triple-helix and, when applicable, marked the major-groove face involved in Hoogsteen base pairing (Fig. 1B).

Augmented dot-bracket notation for Hoogsteen pairs

Dot-bracket notation [30] represents base pairs and pseudoknots, but not non-canonical interactions. To encode Hoogsteen contacts, we propose a third annotation line where unpaired third-strand nucleotides are marked with lowercase letters (e.g. *z*, *x*, *y*, *w*, *v*), and the facing nucleotides in the Watson–Crick–Franklin (WCF) base-paired region are marked with the matching uppercase letters (*Z*, *X*, *Y*, *W*, *V*). When a third-strand nucleotide contacts both bases of a pair, the uppercase symbol appears twice; all other positions carry a dash (-). This scheme is flexible and can accommodate other multi-nucleotide interactions. All augmented notations for the Validation Dataset are reported in Supplementary Table S4.

Three-dimensional atomic distance assessment

For each base triple, we measured the Euclidean distance (in Å) between the C1' atom of the third-strand nucleotide and the C1' atoms of each nucleotide in the corresponding WCF base pair. To quantify triple-helix geometric consistency, we computed a localized RSI_2 [43] from the distribution of these C1'–C1' distances. Lower RSI_2 values indicate high geometric regularity of the triple-helix, while higher values reflect structural divergence relative to typical base-pair spacing.

Secondary structure prediction

For each RNA in the validation dataset, we predicted secondary structures using eight folding tools: CentroidFold [44], IPknot++ [45], Mfold [46], pKiss [47], RNAfold [48], RNASHAPES [47], RNAstructure [49], and vsfold5 [50].

Predicted structures were compared against experimentally refined secondary structures using standard evaluation metrics: true positive

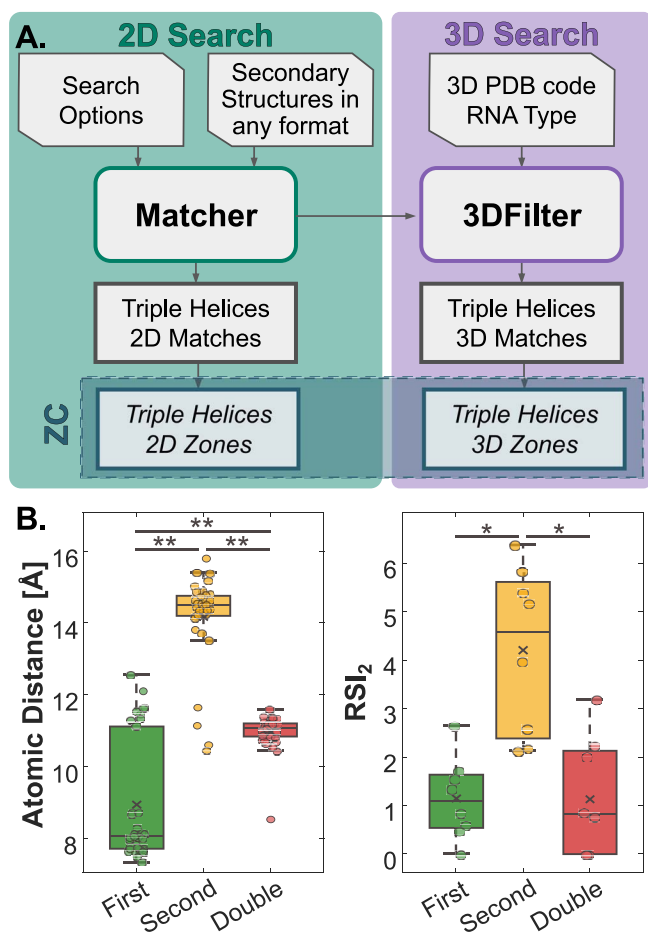


Figure 2 (A) Schematic representation of the TripleMatcher tool. The *Matcher* scans RNA primary and secondary structures to report 2D-matches; the *3DFilter* applies C1'–C1' distance thresholds to keep only geometrically feasible triples (3D-matches); the *ZoneCombiner* (ZC) merges overlapping matches into non-overlapping zones. (B) C1'–C1' distance and RSI₂ distributions for *First* (distance between the third-strand and the interacting WCF base on the major-groove face), *Second* (distance between the third-strand and the opposite WCF base), and *Double* (distance in canonical WCF pairs). The separation between categories motivated the *3DFilter* cutoff (default 11 Å) used to discard geometrically infeasible candidates. Each point denotes one base pair.

Triple helices identification

TripleMatcher architecture

We developed *TripleMatcher*, a Java-based tool composed of two modules: the *Matcher* and the optional *3DFilter* [56]. The *Matcher* scans secondary structures to detect regions consistent with major-groove triple-helix motifs, defined as a stretch of canonical WCF base pairs (e.g. A–U or C–G) facing a run of unpaired nucleotides (commonly U or C) capable of forming Hoogsteen interactions. This pattern is specific to RNA–RNA–RNA triple helices and is not intended to capture other tertiary motifs. The *Matcher* operates directly on sequence and standard secondary structure; 3D information is not required at this stage. When a candidate is found, the tool produces an augmented dot-bracket notation that marks Hoogsteen interactions for visualization. Figure 2A shows the architecture of the tool.

The *Matcher* uses a dynamic-programming algorithm to identify 2D-matches, defined as pairs of regions in the secondary structure:

one unpaired segment acting as a potential third strand and one continuous WCF stack (see [Supplementary Table S2](#)). Both regions must satisfy constraints on minimum length, pairing continuity, and the number of tolerated mismatches or indels, as specified by the user-defined options (see [Table 1](#)). By default, submatches are not reported to limit the number of 2D-matches generated, but can be enabled with `-a` option.

When a 3D structure is available, the *3DFilter* checks spatial feasibility of each 2D-match by measuring the atomic distances between the third strand nucleotides and the corresponding WCF base pairs (see [Supplementary Table S3](#)). Matches exceeding the empirically calibrated thresholds are discarded. A 3D-match is therefore a 2D-match that satisfies these geometric constraints.

Finally, both 2D- and 3D-matches can be aggregated into non-overlapping zones using the independent *ZoneCombiner* (ZC) module, which groups nearby matches within the same RNA molecule. This reduces redundancy and highlights broader structural regions likely to host a triple-helix, even when detected as several nearby matches by the *Matcher* or *3DFilter*.

TripleMatcher validation

We evaluated *TripleMatcher* by comparing its predicted 2D-matches against the experimentally annotated base triples for the validation dataset. Let T denotes the set of annotated base triples (consisting of a WCF base pair and one third-strand nucleotide forming a Hoogsteen interaction). For each base triple $t \in T$, we define a true positive (TP) if there exists at least one predicted 2D-match in which all three nucleotide positions involved in t are present. If no such match exists for t , we record a false negative (FN). Conversely, any predicted 2D-match that does not fully cover any base triple in T is counted as a false positive (FP). We do not define true negatives (TNs), since we cannot exhaustively list every RNA segment incapable of forming a triple-helix.

We quantified the *localization accuracy* (LA) by comparing the centers of the predicted and annotated regions. Let the RNA length be N . For the annotated triple-helix r and a predicted 2D-match m , let U_r, U_m be the sets of nucleotide indices belonging to the third strand and D_r, D_m the sets of indices belonging to the WCF double strand. For $x \in \{r, m\}$, we define the normalized centers:

$$c_{3rd}(x) = \frac{1}{|U_x|} \sum_{i \in U_x} i, \quad c_{WCF}(x) = \frac{1}{|D_x|} \sum_{j \in D_x} j.$$

The LA score of m relative to r is then

$$LA(m, r) = 1 - \frac{1}{2} \left(|c_{3rd}(m) - c_{3rd}(r)| + |c_{WCF}(m) - c_{WCF}(r)| \right)$$

so that $LA(m, r) \in [0, 1]$; $LA = 1$ corresponds to perfect alignment of predicted and annotated centers, and smaller values indicate poorer localization.

Since the *Matcher* does not test spatial feasibility, we re-scored its output after applying the optional *3DFilter*. A 3D-match can contain fewer base triples than the original 2D-match or be excluded entirely if no geometrically valid triples remain. On the *3DFilter*, we applied the same base-triple-level criteria, defining: (i) TP as a retained predicted base triple overlapping an annotated one; (ii) FN as an annotated base triple with no retained match; and (iii) FP as a retained predicted triple with no corresponding annotation.

Table 1 *Matcher* and *3DFilter* command line options and their descriptions. For further details, see [56]

| Matcher option | Description | Default value |
|-------------------------------------|--|---------------|
| -n <unpaired-nucleotide> | Nucleotide type for third strand (U, A, C, or G) | U |
| -b <canonical-base-pair> | Canonical base pair type (AU, UA, GC, or CG) | UA |
| -ml <pattern-minimum-length> | Minimum number of consecutive matches to detect | 4 |
| -st <sequence-tolerance> | Allowed mismatches/insertions/deletions in the unpaired sequence | 1 |
| -bt <base-pair-tolerance> | Allowed mismatches/insertions/deletions in the base-pair sequence | 1 |
| -pt <paired-tolerance> | Maximum number of base-paired nucleotides in unpaired region | 1 |
| -ct <consecutive-tolerance> | Maximum allowed interruptions in base-pair continuity | 1 |
| -p <pseudoknot-tolerance> | Maximum allowed bonds that are not part of a pseudoknot | Optional |
| -a <allow-all-submatches> | Find all sub-matches of exact bond matches | Optional |
| 3DFilter option | Description | Default value |
| -t <tolerance> | Tolerance in Ångström to be added to a maximum default distance 11 | 0 |

For both modules, we computed: (i) PPV, measuring the proportion of predicted base triples that are correctly identified; (ii) TPR, measuring the fraction of true base triples recovered; (iii) FM index, the geometric mean of PPV and TPR; (iv) F1-score, the harmonic mean of PPV and TPR; (v) LA, i.e. the center alignment; and (vi) *3DFilter* efficiency, defined as the percentage of FP base triples removed by the filter while retaining all TP.

TripleMatcher usage: search dataset

We applied *TripleMatcher* to a large-scale search dataset composed of RNA secondary structures from the *PhyloRNA* database (<https://bdslab.unicam.it/phyloRNA> [57]) and the *PDB* archive [58]. In total, we analyzed 4160 RNA structures, including ribosomal RNAs (2897), transfer RNAs (1208), telomerase RNAs (32), and group I and II introns (23); 2210 of these contain pseudoknots. We selected structures with 3D models available to enable full usage of the *3DFilter* module.

The *Matcher* scanned each secondary structure for patterns compatible with triple-helix motifs. Then, the *3DFilter* module was used to verify the spatial feasibility of each match.

TripleMatcher usage: predicted secondary structure

To assess the applicability of *TripleMatcher* in a fully computational setting, we applied the tool to predicted structures (Section Secondary structure prediction) per RNA in the validation dataset.

Predicted 2D-matches were evaluated against the annotated triple-helix regions using the same base-triple-level criteria described above. We computed the following metrics: (i) the number of TPs recovered per folding algorithm; (ii) the false positive rate (FPR), defined as the number of predicted base triples outside the annotated region; and (iii) structure-wise detection rate defined as the fraction of RNAs for which at least one TP was recovered.

TripleMatcher usage: predicted 3D structure

We applied *TripleMatcher* in two distinct *in silico* scenarios. In the first, we kept the experimental secondary structure and paired it with predicted 3D models to test the behavior of the *3DFilter*. In the second, we used secondary structures extracted directly from predicted 3D models to evaluate a fully computational workflow. For each RNA in the validation set, we generated models with *Rhofold+* [59], *3dRNAv2.0* [60], *RNAComposer* [41], and *Farfar2* [61], and derived their 2D structures using *RNAView* [62].

Statistical analysis

We used the Wilcoxon signed-rank test to compare atomic distances and RSI_2 values across three categories of interactions in RNA triple helices: (i) between the third strand and the interacting side of the WCF double strand (*First*), (ii) between the third strand and the opposite side of the major groove (*Second*), and (iii) between nucleotides within the canonical WCF base pair (*Double*).

For evaluation metrics (TP, FP, TPR, PPV, F1-score, FM index, LA, and *3DFilter* efficiency), per-RNA values were reported as median and interquartile range (IQR: 25th–75th percentile). Differences across groups, such as RNA type or folding tool, were assessed using the Wilcoxon rank-sum test. All statistical analyses were performed in *MATLAB* R2021b. Statistical significance was considered at $P < .05$ (*) and $P < .001$ (**).

Results

Triple helices characterization

We characterized the geometry of RNA triple helices using structural annotations and spatial distance metrics. Our augmented dot-bracket notation was used to annotate Hoogsteen interactions between a third strand and a canonical WCF double strand, allowing clear identification of triple-helix motifs within RNA secondary structures.

Figure 2B shows the geometric analysis of base triples in the validation dataset. We measured C1'–C1' distances across three categories (*First*, *Second*, and *Double*). Distances on the Hoogsteen-interacting side were significantly shorter [8.5 (8.0–11.2) Å] compared to the non-interacting side [15.0 (14.7–15.3) Å] and the WCF controls [10.4 (10.2–10.8) Å] ($P < .001$).

The corresponding RSI_2 values supported this observation. RSI_2 on the Hoogsteen-interacting face were low and similar to canonical WCF base pairs [1.7 (1.1–2.4) vs. 0.4 (0.3–2.1); $P = .38$], indicating structural regularity. In contrast, higher RSI_2 values on the non-interacting side indicated greater spatial variability [4.8 (3.8–5.9); $P = .008$].

Secondary structure prediction

Figure 3A reports per-RNA metrics and the class aggregated values (telomerase, $n=4$; RNA-stability elements, $n=4$). Performance varied across both RNA types and prediction tools. Pseudoknot-aware methods (*pKiss*, *IPknot*, *vsfold*) outperformed the others on telomerase

RNAs [MCC_{pKiss}: 0.84 (0.52–0.95)], and among the methods without pseudoknot support the best was `mfold` [MCC_{mfold}: 0.43 (0.28–0.57)]. Overall, `pKiss` was consistently top performing: on telomerase it showed (i) TPR 0.87 (0.74–0.97), (ii) TNR 0.94 (0.65–1.00), (iii) FM 0.92 (0.81–0.96), and (iv) ACC 1.00 (0.81–1.00); on RNA stability elements, metrics were: (i) TPR 0.93 (0.85–0.95), (ii) TNR 0.83 (0.68–0.94), (iii) FM 0.93 (0.81–0.96), and (iv) ACC 1.00 (0.87–1.00). Non-pseudoknot algorithms often broke the pseudoknot stem or shifted its register, reducing MCC [median across `CentroidFold`, `mfold`, `RNAfold`, `RNAshapes`, `RNAstructure`: 0.34 (0.30–0.38)] while maintaining moderate TPR through over-pairing [0.66 (0.53–0.73)].

Abstract structure similarities mirrored these results (Fig. 3B). `pKiss` yielded the highest S_{ASPR} and S_{SERNA} , followed by `IPknot`, then `vsfold` (Fig. 3B). Class aggregates were higher for telomerase [S_{ASPR} : 0.33 (0.25–0.50), S_{SERNA} : 0.26 (0.14–0.59)] than for RNA-stability elements [S_{ASPR} : 0.06 (0.02–0.18), S_{SERNA} : 0.08 (0.03–0.19); P_{ASPR} = .02, P_{SERNA} = .008]. Within telomerase, `hTER` (159 nt) showed the highest similarities [S_{ASPR} : 0.59 (0.58–0.61), S_{SERNA} : 0.58 (0.03–0.79)], whereas the shorter constructs (`1YMO`, `2M8K`, `2K95`; 47–48 nt) were lower [S_{ASPR} 0.25 (0.13–0.56), S_{SERNA} 0.18 (0.08–0.54)]. A plausible explanation is a length/context effect in the abstract-structure metrics. S_{ASPR} and S_{SERNA} measure similarity from edits on *base pairs* only. `hTER` contains many helices outside the pseudoknot core that most tools predict correctly. These correct pairs dilute the impact of local errors in the pseudoknot, giving smaller distances and thus higher similarities. The shorter telomerase constructs concentrate most pairs within the pseudoknot. Any register shift or topology error then affects a large fraction of pairs and lowers similarity.

TripleMatcher validation

On the eight validation RNAs, the `Matcher` identified a median number of 26.5 (7–72) 2D-matches per RNA, for a total of 316 2D-matches. Per-RNA median counts were TP = 4 (3–4), FP = 4.5 (2–5), and FN = 2 (1.5–5.5); the structure-wise detection rate was 8/8 (100%). Per-RNA performance was TPR = 0.5 (0.32–0.75), PPV = 0.42 (0.36–0.66), F_1 = 0.42 (0.36–0.70), FM = 0.42 (0.36–0.70), and LA = 0.97 (0.83–0.99). Class-wise summaries were comparable for telomerase vs. RNA-stability elements [TPR: 0.50 (0.27–0.57) vs. 0.60 (0.32–0.89), P = .63; PPV: 0.36 (0.20–0.47) vs. 0.57 (0.42–0.76), P = .08; F_1 : 0.42 (0.23–0.51) vs. 0.58 (0.36–0.82), P = .63; FM: 0.42 (0.23–0.52) vs. 0.58 (0.36–0.82), P = .63].

After applying the `3DFilter`, with default parameters, the number of retained 3D-matches decreased to 134, with a percentage reduction of 57.6%, reducing FPs while preserving TPR. PPV increased from 0.42 (0.36–0.66) to 0.81 (0.64–0.90) (Δ = +0.32, paired Wilcoxon P = .016), F_1 from 0.42 (0.36–0.70) to 0.62 (0.57–0.78) (Δ = +0.18, $\sim P$ = .023), and FM from 0.42 (0.36–0.70) to 0.65 (0.58–0.78) (Δ = +0.19, $\sim P$ = .023). LA improved from 0.97 (0.83–0.99) to 0.99 (0.91–0.99) (Δ = +0.04, $\sim P$ = .016). The filter removed 70% (68%–82%) of FP base triples and retained all TPs in all the investigated RNAs.

Zone-level results are summarized in Fig. 3C. In the 2D aggregation (Fig. 3C, central panel), most RNAs formed a single zone; `4PLX` had two, and `MALAT1` four, reflecting two alternative single-strand segments and two separated helical segments. After 3D filtering (Fig. 3C, right panel), spurious zones were removed while the true triple-helix zones were retained. Specifically, in `2M8K`, two 2D zones collapsed to a single 3D-feasible zone with no improvement in terms of performances. `hTER` retained one zone with slightly improvement in performances, while the alternative 2D region was discarded. `4PLX`

preserved two zones, and `MALAT1` reduced from four to two zones with the same respective scores, consistent with its bipartite triple-helix. Consequently, `TripleMatcher` correctly localized all triple-helix regions in the validation set.

TripleMatcher usage: screening and identification of putative triple helices

The `Matcher` applied to the search dataset filters 1016 different PDBs, resulting in a total of 150 990 2D-matches (median per molecule: 108 [20–270]). The `3DFilter` filters 67 different PDBs (full list on `Zenodo` [55]), for a total of 97 3D-matches (~ 1 3D-match per molecule). It identifies 7 distinct molecules, of which 1 is human telomerase RNA, 2 16S (from *Escherichia coli* and *Bacillus subtilis*), and 4 23S (from *E. coli*, *Staphylococcus aureus*, *Enterococcus faecalis*, and *Saccharomyces cerevisiae*). The human telomerase (PDBs 7V99 [63], 7QXA and 7QXB [64] from PDB, 9QAX, 9QAY, and 9QAZ [65] from RCSB) contains the catalytic core, where the TERT protein interacts with the pseudoknot domain; indeed, a highly positively charged region on the surface of the TERT telomere repeat binding domain mediates intensive interactions with the major groove of pseudoknot stem P3 [63, 64] that contains the U•A–U triplets. The telomerase is indeed confirmed to contain the triple-helix. In particular, despite the three structures being almost identical to the `2K95`, `1YMO`, and the `hTER` model included in our validation dataset, they were not identified by our first screening because the triple-helix is classified as a pseudoknot in the reference papers. Interestingly, these telomerases are identified from PDB of protein–RNA complexes.

In the ribosomal RNA 16S of *B. subtilis* (7QV1), the region flagged by the `TripleMatcher` is not a triple-helix; it corresponds to a junction. We see the same in 16S rRNA from *E. coli*: a region reported in 46 PDB entries shows three-base interactions that are best interpreted as a junction rather than a major-groove triple-helix. In 23S rRNA, we also find non-triple-helix cases across 13 PDBs. In five structures (*E. coli* and *S. aureus*), a hairpin loop folds back onto its own helix (e.g. 6S12). In six structures (*E. faecalis*), a loop from one hairpin contacts a separate helix. Finally, in two *S. cerevisiae* structures, a strand within a helix folds to create a local three-base interaction. These are junction-like or loop–helix contacts, not canonical triple helices. These cases show that the `Matcher` can flag local contacts where an unpaired strand approaches a helix, even when a canonical triple-helix is not present.

TripleMatcher usage: predicted secondary structure

We applied `TripleMatcher` to the secondary structures predicted by the eight tools (Section Secondary structure prediction), using the same defaults as in the validation. Across all predictions (8 RNAs \times 8 tools = 64) predicted structures, the `Matcher` produced a total of 41 2D-matches; the fraction of predictions yielding at least one candidate triple-helix region was 25/64 (39%). Per-prediction performance over predictions with at least one candidate was: (i) TPR = 0.50 (0.24–0.66), (ii) PPV = 0.33 (0.27–0.45), (iii) F_1 = 0.43 (0.24–0.56), and (iv) FM = 0.45 (0.24–0.60). Pseudoknot-aware methods yielded candidates more often and with higher scores [detection rate 47%, F_1 = 0.43 (0.29–0.54), FM = 0.45 (0.30–0.57)] than methods without pseudoknot support [31%, F_1 = 0.39 (0.18–0.58), FM = 0.41 (0.18–0.62)].

By RNA, the proportion of tools producing at least one TP was highest for `2M8K` and `4PLX` (8/8), moderate for `1YMO` (5/8), and lower for

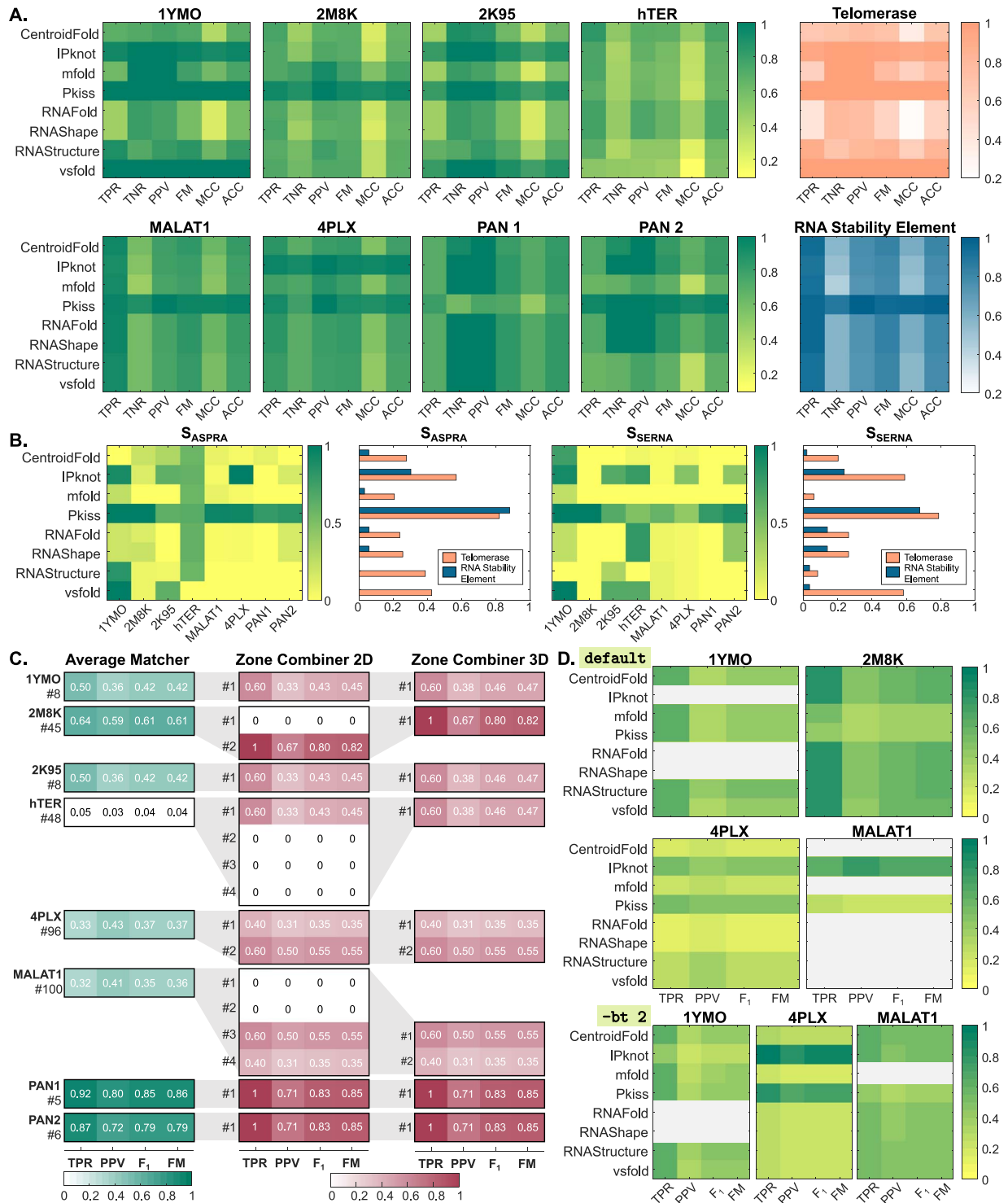


Figure 3 (A and B) Performance of RNA secondary-structure prediction algorithms on the validation dataset. Each matrix represents one RNA; rows are folding tools (CentroidFold, IPknot, mfold, pKiss, RNAfold, RNAShapes, RNAstructure, and vsfold) and columns are evaluation metrics (TPR, TNR, PPV, FM, MCC, and ACC). The two rightmost matrices show class averages (telomerase, top; RNA stability elements, bottom). Color encodes the score (0–1; light = low, dark = high). C) TripleMatcher performance by RNA and region. Rows list RNAs; columns report TPR, PPV, F₁, and FM. For each RNA, the left panel (Average Matcher) shows averages across all raw 2D-matches; the number at left (e.g. #8 for 1YMO) is the count of raw 2D-matches used. The middle panel (Zone Combiner 2D) aggregates raw matches into non overlapping candidate regions (zones #1, #2, etc.) and scores each zone over its base triple calls. Multiple zones indicate distinct candidate triple-helix loci; in MALAT1, four zones arise because two alternative single strand segments and two separated base paired segments combine into four candidate triple helices. The right panel (Zone Combiner 3D) applies C1'–C1' distance thresholds to remove geometrically infeasible triples; zones with all zeros contain no retained triples after filtering. (D) Zone Combiner 2D on predicted secondary structures (four RNAs: 1YMO, 2M8K, 4PLX, and MALAT1). Rows are folding tools; columns are base triple metrics (TPR, PPV, F₁, and FM). Gray cells indicate no detected region; metrics are undefined.

MALAT1 (only `IPknot` and `Pkiss`, Fig. 3D). Two edge cases are informative: for `PAN1`, `Pkiss` returned a 2D-match with all predicted triples outside the annotated region; for `hTER`, `IPknot` likewise returned a 2D-match with `TP=0`, mapping away from the annotated site.

Model-specific defects in the predicted secondary structures explain several failures. In `2K95`, a single unpaired U41 breaks the adjacent paired stack in all tools; as a consequence, three annotated triples in that segment cannot be recovered, reducing the local TPs and prevent downstream detection at that site. In `PAN2`, enabling submatches (option `-a`) is required to recover a shorter, interrupted candidate with `Pkiss`; with this option, TP increases from 0 to 4 and F_1 from 0 to 0.67. For `1YMO`, relaxing the base-pair tolerance to `-bt=2` enabled `IPknot` to recover the triple-helix locus (`TP=2`); with default settings, no candidate region was detected. For `4PLX`, the same relaxation compensated a register shift for both `Pkiss` and `IPknot`, increasing TP from 5 to 8 (F_1 from 0.48 to 0.73) and from 5 to 10 (F_1 from 0.47 to 0.91), respectively. For `MALAT1`, setting `-bt=2` expanded the number of tools that reported a triple-helix candidate from 2 to 7, with a median F_1 of 0.48 (Fig. 3D).

TripleMatcher usage: predicted 3D structure

With experimental secondary structures, `TripleMatcher` performed consistently across predicted 3D models. The `3dRNA` models recovered the triple-helix in 6/8 RNAs, despite noticeable variation in backbone geometry (Fig. 4A). `RNAComposer` and `RhoFold+` recovered 6/8 RNAs, and `Farfar2` recovered all 8. When the secondary structure was extracted from the predicted 3D models, recovery dropped significantly (Fig. 4B). None of the structures generated by the `3dRNA` recovered the triple-helix region in any RNA. This indicates that inaccuracies in the predicted backbones strongly affect the `RNAView`-derived secondary structure. `RhoFold+` recovered the triple-helix in 3/4 RNAs, `RNAComposer` in 4/7, and `Farfar2` in 6/7, while `hTER` remained the most difficult case (only `Farfar2` has `TP>0`). Applying the `3DFilter` to the extracted structures reduced most spurious candidates. Only about 30% of the 2D matches were retained. None of the three `3dRNA` structures kept `TP>0`, whereas `RNAComposer`, `RhoFold+`, and `Farfar2` retained TP in all filtered cases. The `3DFilter` performs as intended once the secondary structure provides a valid triple-helix pattern, but it cannot restore patterns lost during secondary-structure extraction.

Discussion

This work advances the computational analysis of RNA triple helices at two complementary levels. First, we extend the classical dot-bracket formalism with a third annotation line that captures Hoogsteen contacts, allowing triple-helical segments to be handled by any downstream parser that can read structured strings. Second, we implement `TripleMatcher`, a detector that (i) searches secondary structures for an empirically derived triple-helix pattern, (ii) filters the raw 2D-matches with atomic-distance thresholds (`3DFilter`) to rule out sterically impossible assemblies, and (iii) merges overlapping matches into higher-level “zones” (`ZoneCombiner`). In parallel, we quantitatively characterized triple-helix geometry on experimentally resolved references, using $C1'-C1'$ distance distributions and a localized RSI_2 metric; these measurements both validate the motif and calibrate the `3DFilter` thresholds. Together, these elements provide a practical pipeline that starts from a primary sequence and ends with

a short list of three-dimensionally plausible triple-helix candidate regions.

Accuracy of the underlying secondary structures

Because `TripleMatcher` works on nucleotide sequences and secondary structure, its success is bounded by the accuracy of that input. Our benchmarking confirms that pseudoknot-aware predictors, like `Pkiss`, `IPknot`, and `vsfold`, best reconstruct the local architecture that triple helices require: a run of unpaired nucleotides positioned against a continuous WCF stack. Non-pseudoknot-aware methods, by contrast, frequently fragment the pseudoknot stem or shift its register, obliterating the geometric preconditions for Hoogsteen pairing (Fig. 3). Normalized pair-edit similarities (S_{ASPPRA} and S_{SERNA}) make this point quantitatively and further reveal a length-dependent “dilution” effect: in long RNAs such as `hTER`, even substantial local errors have modest impact on global similarity, whereas in compact constructs every mis-paired base is proportionally more damaging. From a practical standpoint, this argues for combining experimental probing data or other constraints with folding algorithms whenever one aims to mine triple helices *in silico*.

Performance of TripleMatcher on curated references

When supplied with experimentally refined secondary structures, `TripleMatcher` achieved a structure-wise detection rate of 100%, placing at least one TP match in every RNA of the validation set. The raw 2D search returned ~25 possible triple helices per molecule; geometric filtering discarded ~60% of them, halved the false-positive count, and raised PPV without loss of TPR. Zone aggregation (Fig. 3) distilled the remaining matches into one or two discrete loci per RNA, greatly easing manual inspection. In most cases, the surviving zone coincided with the annotated triple-helix; where two zones persisted (`4PLX` and `MALAT1`), the duality reflected the genuine bipartite nature of the motif.

Beyond the curated set, several rRNA hits show why geometric screening and quick visual checks are necessary. In 16S rRNA from *B. subtilis* and *E. coli*, the flagged regions are junctions, and in 23S rRNA we see loop-helix or intra-helix contacts, all mimicking the 2D pattern but not canonical major-groove triple helices.

The `hTER` example illustrates an additional benefit; the second region flagged by the `Matcher` corresponds to the P2a helix [27], not part of the triple-helix or a pseudoknot. This region accumulated more raw 2D-matches than the true triple-helix before filtering (45 vs. 3), but was discarded by the `3DFilter`. Although it is not part of the triple-helix, this apparent false positive points to a biologically important region of the `hTER` fold, as mutations in P2a/P2b reduce or abolish `hTER` activity [27].

Fully computational use-case

Applying the same pipeline to purely predicted secondary structures proved feasible but more demanding. Only 39% of the 64 tool×RNA combinations produced any candidate region, and TPs were heavily skewed toward pseudoknot-aware predictions. `2M8K` and `1YMO` remained detectable in most predictors, whereas `PAN1` and `hTER` exemplified failure modes in which a 2D-match is found but lies outside the annotated triple-helix. Several misses could be rescued by modest parameter tweaks. Allowing submatches recovered a trun-

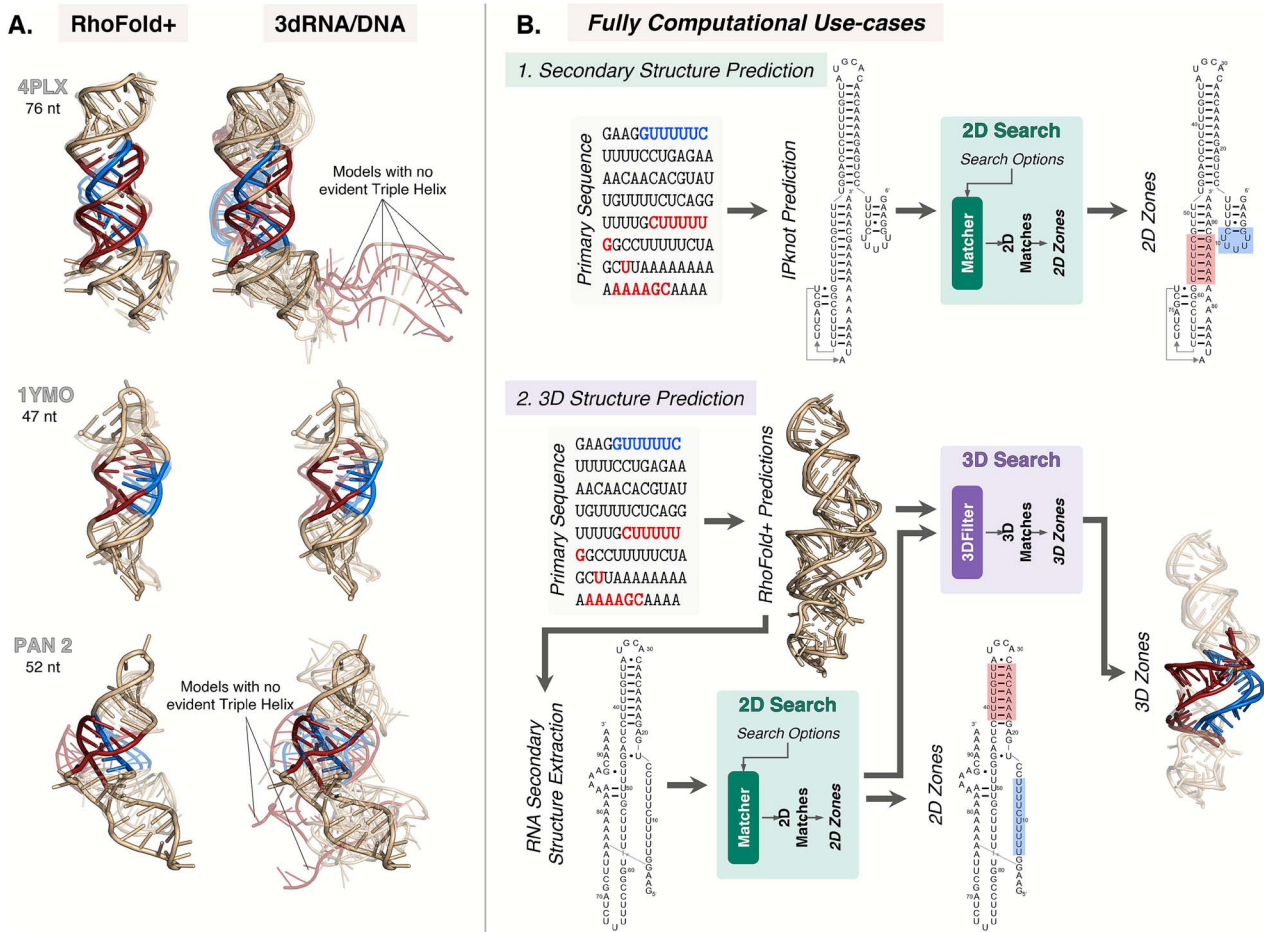


Figure 4 (A) Effect of 3D reconstruction tools on RNA triple-helix recovery. Representative 3D models for MALAT1, PAN, and 1YMO. For each RNA, RhoFold+ models (relaxed and unrelaxed) and 3dRNA models are shown, with the corresponding experimental (reference) structure superimposed and with the expected triple-helix region highlighted. 3dRNA was run with the manually refined secondary-structure and distance constraints chosen to favor a triple-helix-compatible geometry, yet most models fail to reconstruct the triple-helix, often disrupting the WCF stem or misplacing the third strand. RhoFold+, which uses only the primary sequence and does not accept explicit triple-helix constraints, generally produces folds closer to the reference, but still does not reliably recover the full triple-helix stack. The first workflow shows the standard 2D-only mode, where the Matcher operates directly on sequence and secondary structure. The second workflow shows how users can incorporate predicted 3D models: a 3D predictor (e.g. RhoFold+, 3dRNA) generates candidate structures; secondary structure is extracted (e.g. through RNApdbee with multiple annotators); the resulting dot-bracket structure is fed to the Matcher; and 3DFilter is applied only if the user wishes to assess spatial plausibility.

cated but valid candidate in PAN2; relaxing the base-pair compensated for a two-base register shift in 1YMO. Conversely, in 2K95 the apparent loss of three stacked pairs around U41 originates from the input sequence: this single nucleotide shifted the local pairing register and suppressed the three adjacent pairs used by the annotated triple-helix, thereby precluding detection at that site. This underscores how even single-nucleotide changes in the supplied sequence can strongly bias secondary-structure prediction and, in turn, downstream motif detection.

In the fully computational use-case, our results show that the main bottleneck is the quality of the secondary structure extracted from the models. When the extracted structure preserved the triple-helix pattern, the filter behaved as expected; when the pattern was lost upstream, no geometric check could recover it. Improving secondary-structure extraction from predicted 3D models is therefore likely to have a larger impact than modifying TripleMatcher itself.

Conclusion

We developed an end-to-end pipeline that starts from an RNA sequence, uses its secondary structure (experimental or predicted), and outputs a small set of candidate triple-helix regions. We validated it on two datasets with known 3D structures, and when 3D models are absent it still prioritizes likely triple-helix sites from predicted secondary structures.

Two factors primarily determine successful identification: (i) the quality of the secondary structure (notably, faithful recovery of pseudoknot topology and of a suitable run of unpaired nucleotides adjacent to a WCF stack), and (ii) sequence-intrinsic multiplicity of interaction-prone segments, which grows with RNA length and is therefore common in lncRNAs. The first can be mitigated by incorporating probing constraints or improved predictors; the second calls for stringent geometric screening and zone-level aggregation to control spurious patterns in long molecules.

TripleMatcher has the potential to suggest novel triple-helix configurations that have not been previously characterized when applied to large-scale RNA structure datasets. As seen in the case of human telomerase, even if a triple-helix is not present, it points out the presence of regions of interest for further investigations in terms of structural stability and RNA–RNA, DNA–RNA, or protein–RNA interactions.

Our rRNA screen underscores a practical caution: junctions and loop-mediated contacts can satisfy the 2D pattern; yet, they are not true triple helices. Pattern search should therefore be paired with geometric thresholds and a brief visual check, especially in long RNAs rich in junctions.

In future work, we will apply the pipeline at scale to curated ncRNA collections, with a focus on lncRNAs, and extend the analysis across species [66], prioritizing cases with established functional relevance. We will incorporate experimental constraints to improve secondary-structure fidelity and pursue targeted perturbations (e.g. CRISPR-based edits) to test whether predicted triple-helix modules influence RNA stability, localization, or macromolecular interactions *in vivo*. These studies will provide orthogonal validation of the *in silico* predictions and refine the rules that govern triple-helix formation in long RNAs.

Forthcoming RNA-Puzzles targets increasingly feature higher-order interactions, including base triples, offering a natural benchmark for TripleMatcher. Applying the pipeline to Puzzle submissions would reveal which 3D predictors maintain triple-helix geometry and where failures originate. These insights could guide improvements in RNA 3D prediction and support more reliable fully automated workflows [67, 68].

Bioinformatic prediction of RNA triple helices could support RNA-based therapeutics, since these structures appear in several disease-linked genes. A notable example is MALAT1, where small molecules targeting its 3'-end triple-helix reduce MALAT1 levels and affect branching morphogenesis in a mammary tumor organoid model [69, 70]. Triple helices are also promising targets for triplex-forming oligonucleotides, which are being explored in anticancer applications [71]. Tools able to map triple-helix candidates from sequence and secondary structure may help identify new targetable regions and guide the development of such therapeutic approaches.

Key Points

- We introduce TripleMatcher, a computational framework that detects RNA triple helices from sequence and secondary structure using rule-based structural matching on unpaired nucleotides and adjacent base-pair bonds.
- TripleMatcher enhances detection precision through 3D spatial filtering, ensuring that predicted motifs correspond to physically plausible tertiary structures.
- The framework enables large-scale screening of experimentally solved and computationally predicted RNAs, yielding a limited set of high-confidence candidate triple helices.
- By bridging sequence, secondary structure, and tertiary geometry, TripleMatcher provides a generalizable tool for exploring higher-order RNA interactions and guiding experimental validation.

Author contributions

Margherita A.G. Matarrese (Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing—original draft, Writing—review & editing), Michela Quadrini (Formal analysis, Data curation, Software, Writing—review & editing), Nicole Luchetti (Formal analysis, Data curation, Visualization, Writing—review & editing), Federico Di Petta (Formal analysis, Software, Validation, Writing—review & editing), Daniele Durante (Writing—review & editing), Monica Ballarino (Conceptualization, Funding acquisition, Writing—review & editing), Letizia Chiodo (Conceptualization, Funding acquisition, Data curation, Investigation, Supervision, Validation, Writing—original draft, Writing—review & editing), Luca Tesei (Conceptualization, Funding acquisition, Software, Validation, Supervision, Project administration, Writing—original draft, Writing—review & editing)

Supplementary material

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

Conflict of interest

None declared.

Funding

This work was supported by the European Union – NextGeneration EU – National Recovery and Resilience Plan (NRRP), Mission 4, Component 2, Investment 1.1, under the PRIN 2022 PNRR call (Min. Decree No. 1409, dated 14 September 2022), project: P2022FFEWN RNA secondary structures and their relationship with function: application to non-coding RNAs (RNA2Fun), CUP: J53D23014960001.

Data availability

The modeled RNA structures (MALAT1, PAN1, PAN2, and hTER), curated secondary structures, and annotations are available on Zenodo [55], together with the predicted structures, evaluation datasets, and metric computation files. The TripleMatcher source code is open source at <https://github.com/bdslab/triplematcher> [56] with documentation and usage examples in the repository.

References

1. Buratti E, Baralle FE. Influence of RNA secondary structure on the pre-mRNA splicing process. *Mol Cell Biol* 2004;**24**:10505–14. <https://doi.org/10.1128/MCB.24.24.10505-10514.2004>
2. Svoboda P, Di A, Cara. Hairpin RNA: a secondary structure of primary importance. *Cell Mol Life Sci* 2006;**63**:901–8.
3. Vandivier LE, Anderson SJ, Foley SW. *et al.* The conservation and function of RNA secondary structure in plants. *Annu Rev Plant Biol* 2016;**67**:463–88. <https://doi.org/10.1146/annurev-arplant-043015-111754>
4. Higgs PG. RNA secondary structure: physical and computational aspects. *Q Rev Biophys* 2000;**33**:199–253.
5. Pedersen JS, Bejerano G, Siepel A. *et al.* Identification and classification of conserved RNA secondary structures in the human genome.

- PLoS Comput Biol* 2006;**2**:e33. <https://doi.org/10.1371/journal.pcbi.0020033>
6. Mathews DH. Revolutions in RNA secondary structure prediction. *J Mol Biol* 2006;**359**:526–32. <https://doi.org/10.1016/j.jmb.2006.01.067>
 7. Seetin MG, Mathews DH. RNA structure prediction: an overview of methods. In: Keiler K. (ed.) *Bacterial Regulatory RNA. Methods in Molecular Biology*, vol 905. Totowa, NJ: Humana Press, 2012;99–122. https://doi.org/10.1007/978-1-61779-949-5_8
 8. Sato K, Akiyama M, Sakakibara Y. RNA secondary structure prediction using deep learning with thermodynamic integration. *Nat Commun* 2021;**12**:941. <https://doi.org/10.1038/s41467-021-21194-4>
 9. Ribeiro DM, Zanzoni A, Cipriano A. *et al.* Protein complex scaffolding predicted as a prevalent function of long non-coding RNAs. *Nucleic Acids Res* 2018;**46**:917–28. <https://doi.org/10.1093/nar/gkx1169>
 10. Borah S, Darricarrère N, Darnell A. *et al.* A viral nuclear noncoding RNA binds re-localized poly (A) binding protein and is required for late KSHV gene expression. *PLoS Pathog* 2011;**7**:e1002300. <https://doi.org/10.1371/journal.ppat.1002300>
 11. Conrad NK. New insights into the expression and functions of the Kaposi's sarcoma-associated herpesvirus long noncoding PAN RNA. *Virus Res* 2016;**212**:53–63. <https://doi.org/10.1016/j.virusres.2015.06.012>
 12. Torabi S-F, Chen Y-L, Zhang K. *et al.* Structural analyses of an RNA stability element interacting with poly (A). *Proc Natl Acad Sci USA* 2021;**118**:e2026656118. <https://doi.org/10.1073/pnas.2026656118>
 13. Gutierrez IV, Dayton J, Harger S. *et al.* The expression and nuclear retention element of polyadenylated nuclear RNA is not required for productive lytic replication of Kaposi's sarcoma-associated herpesvirus. *J Virol* 2021;**95**:e0009621. <https://doi.org/10.1128/JVI.00096-21>
 14. Mitton-Fry RM, DeGregorio SJ, Wang J. *et al.* Poly(A) tail recognition by a viral RNA element through assembly of a triple helix. *Science* 2010;**330**:1244–7. <https://doi.org/10.1126/science.1195858>
 15. Swain M, Ageeli AA, Kasprzak W. *et al.* Dynamic bulge nucleotides in the KSHV PAN ENE triple helix provide a unique binding platform for small molecule ligands. *Nucleic Acids Res* 2021;**49**:13179–93. <https://doi.org/10.1093/nar/gkab1170>
 16. Arun G, Aggarwal D, Spector DL. MALAT1 long non-coding RNA: functional implications. *Noncoding RNA* 2020;**6**:22. <https://doi.org/10.3390/nrna6020022>
 17. McCown PJ, Wang MC, Jaeger L. *et al.* Secondary structural model of human MALAT1 reveals multiple structure-function relationships. *Int J Mol Sci* 2019;**20**:5610. <https://doi.org/10.3390/ijms20225610>
 18. Zhang X, Hamblin MH, Yin K-J. The long noncoding RNA Malat1: its physiological and pathophysiological functions. *RNA Biol* 2017;**14**:1705–14. <https://doi.org/10.1080/15476286.2017.1358347>
 19. Brown J, Bulkley D, Wang J. *et al.* Structural insights into the stabilization of MALAT1 noncoding RNA by a bipartite triple helix. *Nat Struct Mol Biol* 2014;**21**:633–40. <https://doi.org/10.1038/nsmb.2844>
 20. G Felsenfeld, David R Davies, Alexander Rich Formation of a three-stranded polynucleotide molecule. *Journal of the American Chemical Society* 1957;**79**:2023–2024.
 21. Theimer CA, Feigon J. Structure and function of telomerase RNA. *Curr Opin Struct Biol* 2006;**16**:307–18. <https://doi.org/10.1016/j.sbi.2006.05.005>
 22. Zhang Q, Kim N-K, Feigon J. Architecture of human telomerase RNA. *Proc Natl Acad Sci USA* 2011;**108**:20325–32. <https://doi.org/10.1073/pnas.1100279108>
 23. Qiao F, Cech TR. Triple-helix structure in telomerase RNA contributes to catalysis. *Nat Struct Mol Biol* 2008;**15**:634–40. <https://doi.org/10.1038/nsmb.1420>
 24. Theimer CA, Blois CA, Feigon J. Structure of the human telomerase RNA pseudoknot reveals conserved tertiary interactions essential for function. *Mol Cell* 2005;**17**:671–82. <https://doi.org/10.1016/j.molcel.2005.01.017>
 25. Tesmer VM, Ford LP, Holt SE. *et al.* Two inactive fragments of the integral RNA cooperate to assemble active telomerase with the human protein catalytic subunit (hTERT) in vitro. *Mol Cell Biol* 1999;**19**:6207–16. <https://doi.org/10.1128/MCB.19.9.6207>
 26. Mitchell JR, Collins K. Human telomerase activation requires two independent interactions between telomerase RNA and telomerase reverse transcriptase. *Mol Cell* 2000;**6**:361–71. [https://doi.org/10.1016/S1097-2765\(00\)00036-8](https://doi.org/10.1016/S1097-2765(00)00036-8)
 27. Ly H, Blackburn EH, Parslow TG. Comprehensive structure-function analysis of the core domain of human telomerase RNA. *Mol Cell Biol* 2003;**23**:6849–56. <https://doi.org/10.1128/MCB.23.19.6849-6856.2003>
 28. Chen J-L, Greider CW. Functional analysis of the pseudoknot structure in human telomerase RNA. *Proc Natl Acad Sci USA* 2005;**102**:8080–5. <https://doi.org/10.1073/pnas.0502259102>
 29. Danaee P, Rouches M, Wiley M. *et al.* bpRNA: large-scale automated annotation and analysis of RNA secondary structure. *Nucleic Acids Res* 2018;**46**:5381–94. <https://doi.org/10.1093/nar/gky285>
 30. Hofacker IL, Stadler PF. RNA Secondary Structures. In: *Reviews in Cell Biology and Molecular Medicine. Supplement 9. Nucleic Acids*. Wiley, 2006. <https://doi.org/10.1002/3527600906.mcb.200500009>
 31. Buske FA, Bauer DC, Mattick JS. *et al.* Triplexator: detecting nucleic acid triple helices in genomic and transcriptomic data. *Genome Res* 2012;**22**:1372–81. <https://doi.org/10.1101/gr.130237.111>
 32. He S, Zhang H, Liu H. *et al.* LongTarget: a tool to predict lncRNA DNA-binding motifs and binding sites via Hoogsteen base-pairing analysis. *Bioinformatics* 2015;**31**:178–86. <https://doi.org/10.1093/bioinformatics/btu643>
 33. Warwick T, Seredinski S, Krause NM. *et al.* A universal model of RNA:DNA:DNA triplex formation accurately predicts genome-wide RNA–DNA interactions. *Brief Bioinform* 2022;**23**:bbac445.
 34. Skeparnias I, Zhang J. Structural basis of NEAT1 lncRNA maturation and menRNA instability. *Nat Struct Mol Biol* 2024;**31**:1650–4. <https://doi.org/10.1038/s41594-024-01361-z>
 35. Brown JA. Unraveling the structure and biological functions of RNA triple helices. *Wiley Interdiscip Rev RNA* 2020;**11**:e1598. <https://doi.org/10.1002/wrna.1598>
 36. Protein Data Bank. *RCSB Protein Data Bank*. <https://www.rcsb.org>, 2024. (5 August 2025, date last accessed).
 37. Chen J, Blasco MA, Greider CW. Structure of stem-loop IV of human telomerase RNA. *EMBO J* 2000;**19**:621–31.
 38. Cash DD, Cohen-Zontag O, Kim N-K. *et al.* Pyrimidine motif triple helix in the *Kluyveromyces lactis* telomerase RNA pseudoknot is essential for function in vivo. *Proc Natl Acad Sci USA* 2013;**110**:10970–5. <https://doi.org/10.1073/pnas.1309590110>
 39. Kim N-K, Zhang Q, Zhou J. *et al.* Solution structure and dynamics of the wild-type pseudoknot of human telomerase RNA. *J Mol Biol* 2008;**384**:1249–61. <https://doi.org/10.1016/j.jmb.2008.10.005>
 40. Rother M, Rother K, Puton T. *et al.* ModeRNA: a tool for comparative modeling of RNA 3D structure. *Nucleic Acids Res* 2011;**39**:4007–22. <https://doi.org/10.1093/nar/gkq1320>
 41. Popena M, Szachniuk M, Antczak M. *et al.* Automated 3D structure composition for large RNAs. *Nucleic Acids Res* 2012;**40**:e112–2. <https://doi.org/10.1093/nar/gks339>
 42. Antczak M, Zok T, Popena M. *et al.* RNApdbee—A webserver to derive secondary structures from pdb files of knotted and unknotted

- RNAs. *Nucleic Acids Res* 2014;**42**:W368–72. <https://doi.org/10.1093/nar/gku330>
43. Matarrese MAG, Loppini A, Nicoletti M. *et al.* Assessment of tools for RNA secondary structure prediction and extraction: a final-user perspective. *J Biomol Struct Dyn* 2023;**41**:6917–36. <https://doi.org/10.1080/07391102.2022.2116110>
 44. Sato K, Hamada M, Asai K. *et al.* CENTROIDFOLD: a web server for RNA secondary structure prediction. *Nucleic Acids Res* 2009;**37**:W277–80.
 45. Sato K, Kato Y, Hamada M. *et al.* IPknot: fast and accurate prediction of RNA secondary structures with pseudoknots using integer programming. *Bioinformatics* 2011;**27**:i85–93. <https://doi.org/10.1093/bioinformatics/btr215>
 46. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 2003;**31**:3406–15. <https://doi.org/10.1093/nar/gkg595>
 47. Janssen S, Giegerich R. The RNA shapes studio. *Bioinformatics* 2015;**31**:423–5. <https://doi.org/10.1093/bioinformatics/btu649>
 48. Gruber AR, Lorenz R, Bernhart SH. *et al.* The Vienna RNA web-suite. *Nucleic Acids Res* 2008;**36**:W70–4. <https://doi.org/10.1093/nar/gkn188>
 49. Reuter JS, Mathews DH. RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics* 2010;**11**:1–9.
 50. Dawson W, Fujiwara K, Kawai G. *et al.* A method for finding optimal RNA secondary structures using a new entropy model (vsfold). *Nucleosides Nucleotides Nucleic Acids* 2006;**25**:171–89. <https://doi.org/10.1080/15257770500446915>
 51. Quadrini M, Tesei L, Merelli E. An algebraic language for RNA pseudoknots comparison. *BMC Bioinformatics* 2019;**20**:161. <https://doi.org/10.1186/s12859-019-2689-5>
 52. Quadrini M, Tesei L, Merelli E. ASPRALign: a tool for the alignment of RNA secondary structures with arbitrary pseudoknots. *Bioinformatics* 2020;**36**:3578–9. <https://doi.org/10.1093/bioinformatics/btaa147>
 53. Tesei L, Levi F, Quadrini M, Merelli E. *Alignment of RNA Secondary Structures with Arbitrary Pseudoknots Using Structural Sequences*. Preprint. Research Square, 2024. <https://doi.org/doi:10.21203/rs.3.rs-4831215/v1>.
 54. Tesei L, Levi F, Quadrini M. *et al.* SERNAlign v1.0 - structural sEquence RNA secondary structure alignment. Zenodo, 2024. <https://github.com/bdslab/seralign>, <https://doi.org/10.5281/zenodo.16751306>
 55. Matarrese MAG, Quadrini M, Luchetti N. *et al.* Data, results, and scripts for “decoding RNA triple helices: identification from sequence and secondary structure”. Zenodo, 2025. <https://doi.org/doi:10.5281/zenodo.16752003>
 56. Matarrese MAG, Quadrini M, Luchetti N. *et al.* TripleMatcher v1.0 – RNA triple helix matcher and 3D filter. Zenodo, 2025. <https://github.com/bdslab/triplematcher>. <https://doi.org/10.5281/zenodo.16733465>
 57. Quadrini M, Tesei L. PhyloRNA: a database of RNA secondary structures with associated phylogeny (DOI reference). Zenodo, 2025. <https://bdslab.unicam.it/phyloRNA/>. <https://doi.org/10.5281/zenodo.16752870>
 58. Armstrong DR, Berrisford JM, Conroy MJ. *et al.* PDBe: improved findability of macromolecular structure data in the PDB. *Nucleic Acids Res* 2020;**48**:D335–43. <https://doi.org/10.1093/nar/gkz990>
 59. Shen T, Zhihang H, Sun S. *et al.* Accurate RNA 3D structure prediction using a language model-based deep learning approach. *Nat Methods* 2024;**21**:2287–98. <https://doi.org/10.1038/s41592-024-02487-0>
 60. Wang J, Wang J, Huang Y. *et al.* 3dRNA v2. 0: an updated web server for RNA 3D structure prediction. *Int J Mol Sci* 2019;**20**:4116.
 61. Watkins AM, Rangan R, Das R. FARFAR2: improved de novo Rosetta prediction of complex global RNA folds. *Structure* 2020;**28**:963–976.e6. <https://doi.org/10.1016/j.str.2020.05.011>
 62. Yang H, Jossinet F, Leontis N. *et al.* Tools for the automatic identification and classification of RNA base pairs. *Nucleic Acids Res* 2003;**31**:3450–60. <https://doi.org/10.1093/nar/gkg529>
 63. Wan F, Ding Y, Zhang Y. *et al.* Zipper head mechanism of telomere synthesis by human telomerase. *Cell Res* 2021;**31**:1275–90. <https://doi.org/10.1038/s41422-021-00586-7>
 64. Sekne Z, Ghanim GE, Van Roon A-MM. *et al.* Structural basis of human telomerase recruitment by TPP1-POT1. *Science* 2022;**375**:1173–6. <https://doi.org/10.1126/science.abn6840>
 65. Balch S, Sekne Z, Franco-Echevarria E. *et al.* Cryo-EM structure of human telomerase dimer reveals H/ACA RNP-mediated dimerization. *Science* 2025;**389**:eadr5817. <https://doi.org/10.1126/science.adr5817>
 66. Conrad NK. The emerging role of triple helices in RNA biology. *Wiley Interdiscip Rev RNA* 2014;**5**:15–29. <https://doi.org/10.1002/wrna.1194>
 67. Miao Z, Adamiak RW, Antczak M. *et al.* RNA-puzzles round IV: 3D structure predictions of four ribozymes and two aptamers. *RNA* 2020;**26**:982–95. <https://doi.org/10.1261/rna.075341.120>
 68. Fan B, Adam Y, Adamiak RW. *et al.* RNA-puzzles round V: Blind predictions of 23 RNA structures. *Nat Methods* 2025;**22**:399–411.
 69. Donlic A, Zafferani M, Padroni G. *et al.* Regulation of MALAT1 triple helix stability and in vitro degradation by diphenylfurans. *Nucleic Acids Res* 2020;**48**:7653–64. <https://doi.org/10.1093/nar/gkaa585>
 70. Abulwerdi FA, Wenbo X, Ageeli AA. *et al.* Selective small-molecule targeting of a triple helix encoded by the long noncoding rna, malat1. *ACS Chem Biol* 2019;**14**:223–35. <https://doi.org/10.1021/acscchembio.8b00807>
 71. Li C, Zhou Z, Ren C. *et al.* Triplex-forming oligonucleotides as an anti-gene technique for cancer therapy. *Front Pharmacol* 2022;**13**:1007723. <https://doi.org/10.3389/fphar.2022.1007723>