MethodsX 11 (2023) 102275

Contents lists available at ScienceDirect

MethodsX

journal homepage: www.elsevier.com/locate/methodsx

Discovery of RNA secondary structural motifs using sequence-ordered thermodynamic stability and comparative sequence analysis

Jake M. Peterson, Collin A. O'Leary, Evelyn C. Coppenbarger, Van S. Tompkins, Walter N. Moss*

Roy J. Carver Department of Biophysics, Biochemistry and Molecular Biology, Iowa State University, Ames, IA 50011, USA

ARTICLE INFO

Method name:
Cobretti

Keywords:
RNA structure
Motif discovery
Bioinformatics
Sequence analysis
Covariance analysis

ABSTRACT

Major advances in RNA secondary structural motif prediction have been achieved in the last few years; however, few methods harness the predictive power of multiple approaches to deliver in-depth characterizations of local RNA motifs and their potential functionality. Additionally, most available methods do not predict RNA pseudoknots. This work combines complementary bioinformatic systems into one robust discovery pipeline where:

- RNA sequences are folded to search for thermodynamically favorable motifs utilizing ScanFold.
- Motifs are expanded and refolded into alternate pseudoknot conformations by Knotty/ Iterative HFold.
- All conformations are evaluated for covariance via the cm-builder pipeline (Infernal and R-scape).

Specifications table

Subject area:	Bioinformatics
More specific subject area:	Structural bioinformatics of RNA
Name of your method:	Cobretti
Name and reference of original method:	R.J. Andrews, J. Roche, W.N. Moss, ScanFold: an approach for genome-wide discovery of local RNA structural elements-applications to Zika virus and HIV, PeerJ 6 (2018) e6136. https://doi.org/10.7717/peerj.6136 [1]
Resource availability:	*Software requirements:* Biopython 1.70 (https://biopython.org/wiki/Download) Python 3.6.5 (https://www.python.org/downloads/release/python-365/) Perl (https://www.perl.org/get.html) ViennaRNA (https://github.com/ViennaRNA/ViennaRNA) [2] ScanFold (https://github.com/moss-lab/ScanFold) [1] Knotty (https://github.com/HosnaJabbari/Knotty) [3] Iterative HFold (https://github.com/HosnaJabbari/Iterative-HFold) [4]

(*continued on next page*)

* Corresponding author.
E-mail address: wmoss@iastate.edu (W.N. Moss).

https://doi.org/10.1016/j.mex.2023.102275
Received 15 December 2022; Accepted 28 June 2023
Available online 29 June 2023

RNAFramework (https://github.com/dincarnato/RNAFramework) [5]
Infernal (https://github.com/EddyRivasLab/infernal) [6]
cm-builder (https://github.com/dincarnato/labtools) [7]
R-scape 2.0.0.k (https://github.com/EddyRivasLab/R-scape) [8]
Software for high-throughput operations:
UNIX-based high-performance computing node
SLURM Workload Manager (https://github.com/SchedMD/slurm)
NCBI BLAST+ (https://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST/) [9]
Cobretti (https://github.com/moss-lab/Cobretti)

Introduction

The impetus for the Cobretti python script was to automate the process of predicting regions of RNA likely to contain functional structures. This provides a streamlined pipeline, moving from sequence to structure prediction to conservation analysis to reveal regions of RNA with a propensity to form functional structures.

Compared to random RNAs of the same nucleotide frequency, functional RNAs are by and large more thermodynamically stable [10]. Evolutionary pressures demur loss of function, leading to increasingly ordered RNA structures over time [1]. Based on these findings, ordered thermodynamic stability can be measured to determine if a given RNA structure has undergone selection [1]. This is the basis for ScanFold, previously used to uncover unusually stable local motifs in human immunodeficiency virus (HIV-1) [1], Zika virus (ZIKV) [1], Epstein-Barr virus (EBV) [11], severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) [12], and influenza virus [13].

The Cobretti pipeline begins by folding sequences of interest using ScanFold [1]. ScanFold breaks up sequence analysis into two parts: ScanFold-Scan and ScanFold-Fold [1]. ScanFold-Scan uses a scanning nucleotide window approach that folds the native sequence and calculates the minimum free energy (MFE) using RNAfold. RNAfold is a component of the ViennaRNA package that utilizes a heuristic modeling approach based on the Turner thermodynamic energy models [2]. The native MFE is compared to MFEs of 100 randomly shuffled sequences of the same nucleotide composition, resulting in a z-score (the number of standard deviations the native MFE is above or below average, an approach adapted from Clote et. Al. [10]). ScanFold-Scan then "steps" the analysis window one nucleotide down the input sequence, repeating z-score calculations for each step until the entire sequence has been analyzed (i.e., scanned). Subsequently, ScanFold-Fold compiles base pairings across overlapping ScanFold-Scan windows to identify the most favorable pairing state [11]. For a default scan (single nucleotide step, window size optimized to 120 nucleotides [1,12,14]), each nucleotide (outside of the extreme 5′ and 3′ ends) will appear in 120 separate MFE models. Base pairs across overlapping windows are used to generate a weighted consensus structure; the weight is determined by the recurrence and window z-score for each base pair configuration. The result is base pairs that most contribute to sequence-ordered stability. These models overcome inaccuracies in MFE-only models and have been shown to identify robust, highly probable, and experimentally supported secondary structures [1,11–13,15–27].

A major limitation of ScanFold is that the underlying folding algorithms forbid crossing/non-nested base pairings, which prevent the discovery of pseudoknots [2]. Pseudoknots are valuable functional targets, forming highly stable structures that are frequently involved in biologically significant processes [28]. To remedy this shortcoming, ScanFold motif regions can be extended on their 5′ and 3′ ends by Cobretti and refolded using two different pseudoknot prediction programs: Knotty and Iterative HFold.

Knotty [3] uses free energy minimization and a novel dynamic programming algorithm to find pseudoknots within a given sequence. It does not include ScanFold structure as an input, and therefore provides an unrestrained prediction for a given sequence. This approach is beneficial for identifying pseudoknot/hairpin [28] or multibranch/hairpin [29] hybrids (as seen in influenza) where conformational composition fluctuates depending on environmental conditions.

Iterative HFold [4] is a heuristic pseudoknot prediction algorithm that can use ScanFold motifs as a constraint. The program attempts to build a pseudoknot around the motif using four different prediction methods and outputs the lowest free energy structure, both ScanFold constrained and unconstrained [4]. A lack of a pseudoknot in the Iterative HFold lowest free energy structure, corroborated with ScanFold or Knotty, may suggest a predominant multibranch loop or hairpin conformation.

Conservation of RNA secondary structures is regarded as an indicator of evolutionary constraint and potential function [8,30]. Functional, structured RNAs are expected to have some level of structural conservation that can lead to statistically observable co-evolution across base paired residues—covariation [31]. To check RNA secondary structures from ScanFold, Knotty and Iterative HFold (with and without ScanFold constraints) for covariance, Cobretti uses cm-builder [7] and R-scape (RNA Structural Covariation Above Phylogenetic Expectation) [8].

The cm-builder pipeline [7], implemented within the RNA Framework toolkit [5], uses the program Infernal (Inference of RNA Alignment) [6] to generate homologous alignments of motif sequences of interest against an NCBI BLAST database. This database can be obtained manually using the BLAST website, or through the BLAST+ command line module [9]. Infernal [6] cannot analyze non-nested (pseudoknotted) structures, so Cobretti breaks down motifs into nested substructures and trims unpaired nucleotides from the 5′ and 3′ ends to reduce noise [32].

Significant covariation in RNA sequence/structure alignments can be identified via the program R-scape [8,31]. R-scape estimates the statistical power of alignments to determine if there is enough variation to detect covariation [31,33]. High-power covarying alignments (> 0.10) support a conserved RNA structure, whereas low-power results indicate inconclusive evidence of covariation and high-power alignments with no observed covariance suggests evolutionary selection incompatible with a structured RNA [31]. One caution to this approach is that deep evolutionary conservation is required to observe statistical covariance in functional RNA motifs; not all functional RNAs (e.g., long ncRNAs [8]) have sufficient depth for this approach to work.

The Cobretti python script automates many of these previously manual processes (as well as file conversions between programs, building and running SLURM scripts, user-friendly data outputs), minimizing setup by the end-user and reducing troubleshooting processes.

Method details

Install software and verify dependencies

The software listed in the "Resource availability" section have been ordered to ensure all dependencies are met prior to installation. However, software requirements are always changing, so best practice is to check program dependencies at their source.

For high-throughput operations, it is recommended to have all programs and dependencies available on a high-performance computing (HPC) node along with a SLURM queuing system (https://slurm.schedmd.com/), NCBI BLAST+ [9], and local BLAST databases. The most up-to-date Cobretti code, user's guide, and example files are available on the Moss Lab GitHub page (https://github.com/moss-lab/Cobretti).

ScanFold identifies and extracts thermodynamically stable secondary structures

ScanFold has been optimized to find highly stable structured RNA regions and can be run using default settings (no probing data, 120 nucleotide windows, –global_refold flag on, -2 z-score structures) for most applications. There are experimental approaches which probe RNA secondary structure in vivo or in vitro and these experiments yield per nucleotide structure reactivity values which can help to inform RNA folding algorithms. Existing probing data can be used as a constraint to inform the final model but are not required; purely in silico ScanFold results (especially -2 z-score motifs) correlate well with SHAPE and DMS reactivities [12,19]. A 120 nucleotide window size is sufficient for most structured RNAs (e.g., most known human cis-regulatory RNA structures are < ~150 nucleotides [21]), but the size can be reduced or increased depending on user needs. Window size can be increased to 150 nucleotides with little decline in accuracy, although it is not recommended to exceed 200 nucleotides [1,12]. Adding the –global_refold flag extends folding around ScanFold identified motifs (i.e., base pairs that are thermodynamically favored but not necessarily significantly stable) and is recommended for reasonably sized targets. All significantly stable (-2 z-score) motifs are extracted in dot-bracket notation (.dbn) by ScanFold for subsequent analysis. This parameter is not typically altered, but can be rewritten within the ScanFold code to extract alternate motifs (as we have done previously [19]).

For high-throughput operations, Cobretti is capable of scanning the current working directory for all FASTA format (.fa or .fasta) files, building ScanFold shell scripts with default settings for each FASTA, and running all scripts in parallel in a SLURM queue enabled HPC environment. Cobretti will then find the ScanFold extracted motifs for all runs and copy them to a centralized location for analysis and further processing.

Predicting alternate conformations within highly structured regions

Cobretti extends the 5′ and 3′ ends of all ScanFold motifs and runs each motif through Knotty and Iterative HFold. The default 30 nucleotide extension by Cobretti allows for pseudoknot predictions localized around the ScanFold motif region without adding excessive noise or using excessive computational power. All resulting structures are then compiled into one text file for user analysis and broken down into individual motif files for further evaluation.

Covariation identifies evolutionarily interesting secondary structures

Cobretti nests and trims motifs of interest, batching motifs together into cm-builder shell scripts and running each script. Trimming of unpaired nucleotides in nested substructures generally improves the number of sequences Infernal can align, which in turn improves R-scape's ability to identify covariance [32]. Cobretti is set up to create BLAST+ databases in parallel with the ScanFold runs, although they are not used until this step. BLAST+ is used in blastn mode with the nt database, providing 2500 target sequences with only the top hit for each sequence (-max_hsps 1). While this could result in very large genomic databases, Cobretti only saves the accession numbers and aligned sequences for each target sequence. Once all cm-builder alignments are complete, Cobretti runs R-scape and compiles all covariance results into a PDF and a tab-separated file for analysis. The R-scape settings include a two-set statistical test (-s, required for testing the proposed substructure) and multiple FastTree alignments (—ntree 10, improves E-value repeatability).

Method validation

An initial pool of functional RNAs were acquired from previous studies – SimRNA [34], AnnapuRNA [35], and RNA-Puzzles [36]. After removing duplicates and limiting analysis to single chain RNAs (i.e., single-stranded RNA, no protein interactions, with or without a small molecule/ligand interaction), there were 252 remaining structures (Supplemental File 1). From this, RNApdbee 2.0 [37] was used to determine sequence and secondary structures for each PDB using the following parameters: first model only; 3DNA/DSSR; do not include non-canonical; hybrid algorithm paired residues; no image. PDB IDs were searched on NCBI and due

Table 1
Breakdown of all highly stable structural hits across 59 published structures.

Structure Match (≥ 5 bp)	High-Power (≥ 1 bp)	High Covariance (Observed > Expected)	Published Structure Hits (ScanFold Motif Hits)				
			All Methods	ScanFold	Knotty	Iterative HFold	Iterative HFold w/ ScanFold
			32 (40)*	27 (35)	32 (40)*		
X			25 (32)	19 (21)	23 (30)	22 (29)	21 (27)
	X		22 (30)	14 (15)	20 (28)	19 (25)	20 (26)
X	X		17 (24)	11 (12)	16 (23)	14 (19)	14 (18)
	X	X	14 (19)	7 (8)	10 (14)	8 (9)	9 (10)
X	X	X	11 (16)	5 (6)	9 (13)	7 (8)	7 (8)

* Includes ScanFold motifs that did not cover published regions prior to 30 nucleotide 5′ and 3′ extensions.

diligence was taken to analyze publications for species origin and sequence descriptors. If sequences were labeled as synthetic constructs, they were removed from the testing pool. Sequences were also removed if their origin could not be determined, or if they were determined to be too ambiguous (i.e., less than a 90% match to any species-specific sequences in an NCBI BLAST search). Small RNA modifications made by researchers to increase stability for NMR/x-ray analysis were retained. Accession numbers for sequences of interest were not always available, so the top BLAST hit (blastn search limited to each selected organism with default settings) was used for the reference sequence. Reference sequences for RNA structures in minus strand sequences were reverse complemented, and all duplicate sequences were removed, leaving 59 published structure sequences for further analysis.

To reduce computational requirements and provide a comparable baseline, sequences were reduced to 1 kb centered on the region of interest. If any structure was located near the 5′ or 3′ end of a sequence, 1 kb of sequence was selected from the 5′ or 3′ end, respectively. In one case, the structure was larger than 1 kb (1NJN), so 500 nucleotides were added to each side of the region of interest. If the reference sequence was less than 1 kb in length (1A60, 1KKS, 1MNX, 4R4V), the entire sequence was used regardless of the location of the structure of interest. These sequences were then subjected to ScanFold [1], outputting all ≤ -2 z-score motifs. NCBI BLAST databases were built using the shortened 1 kb sequences with the blast-plus module on Iowa State University's Pronto HPCs (blastn, nt database, 2500 sequences).

ScanFold identified 213 total motifs with a z-score ≤ -2. Of these, 35 motifs (16.4%) were found in 27 of the 59 published structure sequences (45.8%). Another 5 motifs were within 30 nucleotides of published structure regions and were therefore included for analysis for refolding by Knotty and Iterative HFold (Table 1). Overall, 54.2% of the published structures (32 of 59) were represented.

Of the ScanFold predicted motifs, partial structure region matches (5 base pairs or more in common) were found in 19 of the 27 published structures (70.4%). After sequence extension and folding by Knotty and Iterative HFold, that number improved to 25 of 32 (78.1%). Thus, extending ScanFold with Knotty and Iterative HFold increased the likelihood of predicting a known structure.

Extending and refolding all 213 ScanFold motifs using Knotty/Iterative HFold produced 852 total motifs, which further broke down into 1066 nested substructures. cm-builder and Infernal were able to align 859 (80.6%) of these nested substructures. R-scape determined that 458 substructures had at least one high-power base pair, with 189 having more observed covarying base pairs than expected. For method validation, all nested substructures were considered non-nested structures, where covariance in any nested substructure was measured as a hit for the original ScanFold motif (Table 1). Of the 40 extended ScanFold motifs, 30 had at least one high-power covarying base pair, resulting in 22 of the 59 published structures (37.3%) exhibiting some level of covariance. Further analysis of high-power motifs revealed 14 unique published structures with more observed covarying base pairs than expected. Overall, 17 unique published structures (28.8% of 59) were found common to being highly stable, a partial match to published structure, and as having at least one high-power covarying base pair, while 11 (18.6% of 59) also had more observed covarying base pairs than expected. Of note, 9 of these 11 published structures contain published pseudoknots – Knotty managed to predict 8, whereas ScanFold and Iterative HFold combined to predict 7 of the 9.

Combined, ScanFold, Knotty and Iterative HFold with and without ScanFold constraints identified 107 of the original 213 motifs as having a partial structure match (defined above). To qualitatively assess this result, the positive predictive value (PPV) and sensitivity for all 107 partial structure matches were determined. All structures were aligned and predicted structures unassociated with the published structure region were removed to reduce noise. For method validation, sensitivity was defined as the number of matching base pairs between the published and predicted structures divided by the total number of base pairs in the published structure. PPV was defined as the number of matching base pairs divided by the total number of base pairs in the predicted structure. PPV and sensitivity are measured from 0 to 1, where 1 is a perfect match.

For all the partial structure matches, the average PPV was 0.71 and average sensitivity was 0.53 (Table 2). While ScanFold produced the least number of structure matches (21), it had the highest PPV across all predicted structures. Analysis of sensitivity showed a broad range of values (0.01 to 1.00), which can be predominantly associated with sequence length. At maximum, the motifs produced in this method will be 180 nucleotides (120 from the ScanFold window and 60 from 5′ and 3′ extensions). However, some published structures are much larger than this (e.g., 1NJN at 2877 nucleotides), such that even large accurate predictions will produce poor sensitivity. To address this, sequences were binned based on the size of the predicted structure (Table 2). Indeed, as the size of the published structure decreases, the sensitivity increases, with an average sensitivity of 0.82 across all methods for structures 60 nucleotides or smaller. Filtering structures by length for average PPV showed a slight improvement, but not to the extent of sensitivity. The range of PPV values was wholly unaffected (0.24 to 1.00) by structure length.

Table 2
PPV and Sensitivity for all partial structure matches (≥ 5 bp).

		All Methods	ScanFold	Knotty	Iterative HFold	Iterative HFold w/ ScanFold
Total Motifs		107	21	30	29	27
		Average (Minimum Range)*				
PPV	All Structures	0.71 (0.24)	0.84 (0.24)	0.62 (0.26)	0.71 (0.26)	0.69 (0.24)
	≤ 180 nt	0.73 (0.24)	0.86 (0.24)	0.64 (0.26)	0.75 (0.26)	0.71 (0.24)
	≤ 120 nt	0.72 (0.24)	0.82 (0.24)	0.66 (0.26)	0.82 (0.26)	0.75 (0.24)
	≤ 60 nt	0.83 (0.24)	0.87 (0.24)	0.72 (0.27)	0.89 (0.27)	0.85 (0.24)
Sensitivity	All Structures	0.53 (0.01)	0.57 (0.01)	0.52 (0.01)	0.51 (0.01)	0.52 (0.01)
	≤ 180 nt	0.59 (0.10)	0.60 (0.11)	0.60 (0.13)	0.59 (0.10)	0.58 (0.10)
	≤ 120 nt	0.76 (0.13)	0.71 (0.13)	0.72 (0.13)	0.73 (0.13)	0.71 (0.13)
	≤ 60 nt	0.82 (0.33)	0.79 (0.33)	0.82 (0.33)	0.82 (0.33)	0.83 (0.33)

* All categories had a maximum range value of 1.00.

To summarize, 59 sequences of interest were tested using the Cobretti pipeline, resulting in a prediction of 213 highly structured regions. These motifs were refolded into 852 total conformations that contained 1066 nested substructures. Covariance analysis aligned 859 of those substructures, revealing 458 with at least one high-power covarying base pair, with 24 found in 17 of the published structures. 189 of the 458 high-power structures had more observed covariance than expected, 16 of which were identified in published structures. These 189 motifs represent 77 unique RNA regions, 11 of which are associated with published structured regions (13.1% of 59). Of 107 published partial structure matches, there was an average PPV of 0.71 and an average sensitivity of 0.53, and sensitivity greatly improved with reduced structure size.

Our lab continues to improve upon this method through incorporation of the recently published ScanFold 2 [38], testing alternative ScanFold and pseudoknot folding algorithms, and by finding more complementary ways to characterize RNA structure. It is our hope that by making our methods available to the broader scientific community, others can weigh in and help improve this methodological pipeline.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Jake M. Peterson: Conceptualization, Methodology, Software, Validation, Investigation, Writing – original draft, Writing – review & editing, Visualization. **Collin A. O'Leary:** Software, Validation, Investigation, Writing – review & editing. **Evelyn C. Coppenbarger:** Software, Validation. **Van S. Tompkins:** Software, Writing – review & editing. **Walter N. Moss:** Conceptualization, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Acknowledgments

We would like to thank Elena Rivas for technical assistance and early access to R-scape version 2.0.0.k, as well as members of the Moss lab for helpful discussions.

This work was supported by National Institutes of Health R01GM133810.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.mex.2023.102275.

References

[1] R.J. Andrews, J. Roche, W.N. Moss, ScanFold: an approach for genome-wide discovery of local RNA structural elements-applications to Zika virus and HIV, Peer J 6 (2018) e6136.

[2] R. Lorenz, S.H. Bernhart, C. Honer Zu Siederdissen, H. Tafer, C. Flamm, P.F. Stadler, I.L. Hofacker, ViennaRNA Package 2.0, Algorithms Mol. Biol. 6 (2011) 26.

[3] H. Jabbari, I. Wark, C. Montemagno, S. Will, Knotty: efficient and accurate prediction of complex RNA pseudoknot structures, Bioinformatics 34 (22) (2018) 3849–3856.

[4] H. Jabbari, A. Condon, A fast and robust iterative algorithm for prediction of RNA pseudoknotted secondary structures, BMC Bioinf. 15 (2014) 147.

[5] D. Incarnato, E. Morandi, L.M. Simon, S. Oliviero, RNA Framework: an all-in-one toolkit for the analysis of RNA structures and post-transcriptional modifications, Nucleic Acids Res. 46 (16) (2018) e97.

[6] E.P. Nawrocki, S.R. Eddy, Infernal 1.1: 100-fold faster RNA homology searches, Bioinformatics 29 (22) (2013) 2933–2935.

[7] I. Manfredonia, C. Nithin, A. Ponce-Salvatierra, P. Ghosh, T.K. Wirecki, T. Marinus, N.S. Ogando, E.J. Snijder, M.J. van Hemert, J.M. Bujnicki, D. Incarnato, Genome-wide mapping of SARS-CoV-2 RNA structures identifies therapeutically-relevant elements, Nucleic Acids Res. 48 (22) (2020) 12436–12452.

[8] E. Rivas, J. Clements, S.R. Eddy, A statistical test for conserved RNA structure shows lack of evidence for structure in lncRNAs, Nat. Methods 14 (1) (2017) 45–48.

[9] C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, T.L. Madden, BLAST+: architecture and applications, BMC Bioinf. 10 (2009) 421.

[10] P. Clote, F. Ferre, E. Kranakis, D. Krizanc, Structural RNA has lower folding energy than random RNA of the same dinucleotide frequency, RNA 11 (5) (2005) 578–591.

[11] R.J. Andrews, C.A. O'Leary, W.N. Moss, A survey of RNA secondary structural propensity encoded within human herpesvirus genomes: global comparisons and local motifs, PeerJ 8 (2020) e9882.

[12] R.J. Andrews, C.A. O'Leary, V.S. Tompkins, J.M. Peterson, H.S. Haniff, C. Williams, M.D. Disney, W.N. Moss, A map of the SARS-CoV-2 RNA structurome, NAR Genom. Bioinform. 3 (2) (2021) lqab043.

[13] J.M. Peterson, C.A. O'Leary, W.N. Moss, In silico analysis of local RNA secondary structure in influenza virus A, B and C finds evidence of widespread ordered stability but little evidence of significant covariation, Sci. Rep. 12 (1) (2022) 310.

[14] R.J. Andrews, L. Baber, W.N. Moss, RNAStructuromeDB: A genome-wide database for RNA structural inference, Sci. Rep. 7 (1) (2017) 17269.

[15] R.J. Andrews, L. Baber, W.N. Moss, Mapping the RNA structural landscape of viral genomes, Methods 183 (2020) 57–67.

[16] T.Y. Hou, W.L. Kraus, Analysis of estrogen-regulated enhancer RNAs identifies a functional motif required for enhancer assembly and gene expression, Cell Rep. 39 (11) (2022) 110944.

[17] D. Mitchell 3rd, S.M. Assmann, P.C. Bevilacqua, Probing RNA structure in vivo, Curr. Opin. Struct. Biol. 59 (2019) 151–158.

[18] C.A. O'Leary, R.J. Andrews, V.S. Tompkins, J.L. Chen, J.L. Childs-Disney, M.D. Disney, W.N. Moss, RNA structural analysis of the MYC mRNA reveals conserved motifs that affect gene expression, PLoS One 14 (6) (2019) e0213758.

[19] C.A. O'Leary, S. Van Tompkins, W.B. Rouse, G. Nam, W.N. Moss, Thermodynamic and structural characterization of an EBV infected B-cell lymphoma transcriptome, NAR Genom. Bioinform. 4 (4) (2022) lqac082.

[20] R. Rangan, I.N. Zheludev, R.J. Hagey, E.A. Pham, H.K. Wayment-Steele, J.S. Glenn, R. Das, RNA genome conservation and secondary structure in SARS-CoV-2 and SARS-related viruses: a first look, RNA 26 (8) (2020) 937–959.

[21] W.B. Rouse, C.A. O'Leary, N.J. Booher, W.N. Moss, Expansion of the RNAStructuromeDB to include secondary structural data spanning the human protein-coding transcriptome, Sci. Rep. 12 (1) (2022) 14515.

[22] N.N. Singh, C.A. O'Leary, T. Eich, W.N. Moss, R.N. Singh, Structural context of a critical exon of spinal muscular atrophy gene, Front. Mol. Biosci. 9 (2022) 928581.

[23] R.C.A. Tavares, G. Mahadeshwar, H. Wan, N.C. Huston, A.M. Pyle, The global and local distribution of RNA structure throughout the SARS-CoV-2 genome, J. Virol. (2020).

[24] V.S. Tompkins, W.B. Rouse, C.A. O'Leary, R.J. Andrews, W.N. Moss, Analyses of human cancer driver genes uncovers evolutionarily conserved RNA structural elements involved in posttranscriptional control, PLoS One 17 (2) (2022) e0264025.

[25] Y. Tong, Q.M.R. Gibaut, W. Rouse, J.L. Childs-Disney, B.M. Suresh, D. Abegg, S. Choudhary, Y. Akahori, A. Adibekian, W.N. Moss, M.D. Disney, Transcriptome-wide mapping of small-molecule RNA-binding sites in cells informs an isoform-specific degrader of QSOX1 mRNA, J. Am. Chem. Soc. 144 (26) (2022) 11620–11625.

[26] A. Ursu, J.L. Childs-Disney, R.J. Andrews, C.A. O'Leary, S.M. Meyer, A.J. Angelbello, W.N. Moss, M.D. Disney, Design of small molecules targeting RNA structure from sequence, Chem. Soc. Rev. 49 (20) (2020) 7252–7270.

[27] P. Zhang, H.J. Park, J. Zhang, E. Junn, R.J. Andrews, S.P. Velagapudi, D. Abegg, K. Vishnu, M.G. Costales, J.L. Childs-Disney, A. Adibekian, W.N. Moss, M.M. Mouradian, M.D. Disney, Translation of the intrinsically disordered protein alpha-synuclein is inhibited by a small molecule targeting its structured mRNA, Proc. Natl. Acad. Sci. U. S. A. 117 (3) (2020) 1457–1467.

[28] L.I. Dela-Moss, W.N. Moss, D.H. Turner, Identification of conserved RNA secondary structures at influenza B and C splice sites reveals similarities and differences between influenza A, B, and C, BMC Res. Notes 7 (2014) 22.

[29] S.F. Priore, E. Kierzek, R. Kierzek, J.R. Baman, W.N. Moss, L.I. Dela-Moss, D.H. Turner, Secondary structure of a conserved domain in the intron of influenza A NS1 mRNA, PLoS One 8 (9) (2013) e70615.

[30] A.P. Gultyaev, M.I. Spronken, M. Richard, E.J. Schrauwen, R.C. Olsthoorn, R.A. Fouchier, Subtype-specific structural constraints in the evolution of influenza A virus hemagglutinin genes, Sci. Rep. 6 (2016) 38892.

[31] E. Rivas, Evolutionary conservation of RNA sequence and structure, Wiley Interdiscip. Rev. RNA 12 (5) (2021) e1649.

[32] R.C.A. Tavares, A.M. Pyle, S. Somarowthu, Phylogenetic analysis with improved parameters reveals conservation in lncRNA structures, J. Mol. Biol. 431 (8) (2019) 1592–1603.

[33] E. Rivas, J. Clements, S.R. Eddy, Estimating the power of sequence covariation for detecting conserved RNA structure, Bioinformatics 36 (10) (2020) 3072–3076.

[34] M.J. Boniecki, G. Lach, W.K. Dawson, K. Tomala, P. Lukasz, T. Soltysinski, K.M. Rother, J.M. Bujnicki, SimRNA: a coarse-grained method for RNA folding simulations and 3D structure prediction, Nucleic Acids Res. 44 (7) (2016) e63.

[35] F. Stefaniak, J.M. Bujnicki, AnnapuRNA: A scoring function for predicting RNA-small molecule binding poses, PLoS Comput. Biol. 17 (2) (2021) e1008309.

[36] Z. Miao, R.W. Adamiak, M. Antczak, M.J. Boniecki, J. Bujnicki, S.J. Chen, C.Y. Cheng, Y. Cheng, F.C. Chou, R. Das, N.V. Dokholyan, F. Ding, C. Geniesse, Y. Jiang, A. Joshi, A. Krokhotin, M. Magnus, O. Mailhot, F. Major, T.H. Mann, P. Piatkowski, R. Pluta, M. Popenda, J. Sarzynska, L. Sun, M. Szachniuk, S. Tian, J. Wang, J. Wang, A.M. Watkins, J. Wiedemann, Y. Xiao, X. Xu, J.D. Yesselman, D. Zhang, Y. Zhang, Z. Zhang, C. Zhao, P. Zhao, Y. Zhou, T. Zok, A. Zyla, A. Ren, R.T. Batey, B.L. Golden, L. Huang, D.M. Lilley, Y. Liu, D.J. Patel, E. Westhof, RNA-Puzzles Round IV: 3D structure predictions of four ribozymes and two aptamers, RNA 26 (8) (2020) 982–995.

[37] T. Zok, M. Antczak, M. Zurkowski, M. Popenda, J. Blazewicz, R.W. Adamiak, M. Szachniuk, RNApdbee 2.0: multifunctional tool for RNA structure annotation, Nucleic Acids Res. 46 (W1) (2018) W30–W35.

[38] R.J. Andrews, W.B. Rouse, C.A. O'Leary, N.J. Booher, W.N. Moss, ScanFold 2.0: a rapid approach for identifying potential structured RNA targets in genomes and transcriptomes, Peer J 10 (2022) e14361.