

RNA2DMut: a web tool for the design and analysis of RNA structure mutations

WALTER N. MOSS

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

ABSTRACT

With the widespread application of high-throughput sequencing, novel RNA sequences are being discovered at an astonishing rate. The analysis of function, however, lags behind. In both the *cis*- and *trans*-regulatory functions of RNA, secondary structure (2D base-pairing) plays essential regulatory roles. In order to test RNA function, it is essential to be able to design and analyze mutations that can affect structure. This was the motivation for the creation of the RNA2DMut web tool. With RNA2DMut, users can enter in RNA sequences to analyze, constrain mutations to specific residues, or limit changes to purines/pyrimidines. The sequence is analyzed at each base to determine the effect of every possible point mutation on 2D structure. The metrics used in RNA2DMut rely on the calculation of the Boltzmann structure ensemble and do not require a robust 2D model of RNA structure for designing mutations. This tool can facilitate a wide array of uses involving RNA: for example, in designing and evaluating mutants for biological assays, interrogating RNA–protein interactions, identifying key regions to alter in SELEX experiments, and improving RNA folding and crystallization properties for structural biology. Additional tools are available to help users introduce other mutations (e.g., indels and substitutions) and evaluate their effects on RNA structure. Example calculations are shown for five RNAs that require 2D structure for their function: the MALAT1 mascRNA, an influenza virus splicing regulatory motif, the EBER2 viral noncoding RNA, the Xist lncRNA repA region, and human Y RNA 5. RNA2DMut can be accessed at <https://rna2dmut.bb.iastate.edu/>.

Keywords: RNA; structure; EBV; influenza; MALAT1; ncRNA

INTRODUCTION

The field of RNA biology is being revolutionized by next-generation sequencing techniques. RNA-seq generates massive quantities of novel RNA sequences and has shown that the transcriptomes of organisms are more extensive and complex than previously imagined. For example, while <2% of the human genome encodes protein, >75% is transcribed (Clark et al. 2013). Many of these noncoding (nc)RNA transcripts are differentially expressed under varying conditions, including disease states such as cancer (Huarte 2015), highlighting their potential function. Furthermore, much of this ncRNA shows evidence of evolutionary conservation, specifically the conservation of base-pairing that makes up RNA secondary (2D) structure. Indeed, one of the hallmark features of ncRNA is its propensity for forming stable and conserved RNA structure (Washietl et al. 2005; Gruber et al. 2010; Mathews et al. 2010), which mediates various interactions important to ncRNA function. Likewise, 2D structure also plays roles in regulating coding messenger (m)RNAs, where structured *cis*-regulatory elements affect diverse processes

such as post-transcriptional editing (Nishikura 2016), mRNA maturation (e.g., splicing [Warf and Berglund 2010]), subcellular localization (Martin and Ephrussi 2009), and translation (Mignone et al. 2002, 2005; Liu et al. 2010).

Recent advances in RNA structure probing, which are linked to next-generation sequencing readout of probing sites, are emerging as powerful tools for capturing snapshots of the RNA “structurome” (Wan et al. 2011; Ding et al. 2015; Lu et al. 2016; Zubradt et al. 2017). Here, the loops and helices that make up 2D structure can be biochemically identified transcriptome-wide. As these techniques become more widely applied, the identification of novel RNA sequences and 2D structures will likely expand rapidly. To gain functional information on this vast quantity of sequence/structure information requires the ability to design mutations that can specifically affect structure, even when robust 2D models are unavailable. This was the motivation for the creation of the RNA2DMut tool, which generates all possible single-point mutations for a sequence and assesses their

Corresponding author: wmoss@iastate.edu

Article is online at <http://www.rnajournal.org/cgi/doi/10.1261/rna.063933.117>.

© 2018 Moss This article is distributed exclusively by the RNA Society for the first 12 months after the full-issue publication date (see <http://rnajournal.cshlp.org/site/misc/terms.xhtml>). After 12 months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

impact on the predicted 2D structural ensemble. This approach makes use of a metric known as the ensemble diversity (ED), which is calculated from the Boltzmann 2D structural ensemble (McCaskill 1990; Freyhult et al. 2005). Here, the Boltzmann ensemble data are accessed via calculating a partition function, where the thermodynamic states considered are the different RNA 2D structures in the ensemble. The ED metric is found by calculating the average “distance” between structures in the ensemble; here, distance is measured based on the number of base pairs that are different between structures, weighted by their calculated probabilities. Low ED suggests a tight ensemble centered on one dominant conformation, while high ED suggests alternative folds or a lack of defined structure (Martin 2014).

Designing mutations to RNAs using statistical measures of structure is attractive, as this captures information on not only the most stable predicted fold, but also alternative folds that may be populated. This is particularly significant as the minimum free energy (MFE) structure predicted by most algorithms may not be the best representation of the native fold. This is due to limits in prediction accuracy that arise from the energy model, as well as ignoring protein interactions, tertiary interactions, and other factors in the native cellular environment. When the MFE model structure fails to predict the native fold correctly the native fold is, however, usually predicted by a near-energy suboptimal fold. Thus, evaluating RNA mutation effects on the structural ensemble can better reflect how those mutations could alter the native fold (Mathews et al. 2010). For example, an approach for mutating RNA structure based on the ensemble defect, which is similar to ensemble diversity, but measures the distance of the ensemble from a defined structure (e.g., the MFE prediction), was previously used to disrupt a *cis*-regulatory structure in influenza virus. Two point mutations were able to completely abolish folding of this element, leading to *in vivo* effects on RNA splicing and viral fitness (Jiang et al. 2016). Another study evaluated the effect of all possible point mutations on ensemble defect, and led to insights into the structure of the small subunit ribosomal (r)RNA (Martin 2014).

Currently, there are a number of useful web tools available for RNA structure prediction: e.g., the Vienna RNA secondary structure server (Hofacker 2003), the RNAstructure Web Servers (Bellaousov et al. 2013), and the mFold Server (Zuker 2003). Here, we present a complementary web tool for aiding in the design and evaluation of RNA mutations that considers the effect of all possible point mutations on the statistical properties of RNA structure. In addition to identifying mutations with high likelihood of disrupting native base pairs, the RNA2DMut web tool can also identify mutations that abolish native structures by stabilizing alternative conformations. It can aid in the design of mutations to stabilize a desired (e.g., native) conformation by stabilizing native pairs, or by destabilizing pairs in possible alternative conformations. This tool has utility in biological assay design (generating mutants for the functional evaluation of RNA structure) as

well as applications to biophysical analyses of RNA (e.g., aiding in NMR or crystallography). Additional sequence manipulation tools are made available to assist in the design of sequential insertions/deletions/substitutions and in evaluating sets of mutated sequences. Thus, the RNA2DMut web tool allows a user to design any possible mutation to an RNA sequence and evaluate its impact on 2D folding.

RESULTS AND DISCUSSION

RNA2DMut (<https://rna2dmult.bb.iastate.edu/>) is a web-based server for generating all possible mutations of an input RNA and evaluating their effect on the Boltzmann structural ensemble, as measured by the ensemble diversity (ED) metric. The ED calculates the average distance across the structural ensemble and can be used to evaluate the effects of mutations on structures, even without a robust starting reference structure (e.g., the MFE fold). This makes RNA2DMut especially valuable for the analysis of novel sequences, where a well-established structure model might not be available.

RNA2DMut is a user-friendly tool that automates the folding of mutants (using the popular RNAfold algorithm [Hofacker 2003; Lorenz et al. 2011]), organizes data and generates publication-quality figures via the VARNA visualization applet for RNA 2D structure (Darty et al. 2009). RNA2DMut can facilitate the design of mutations to disrupt the native structure ensemble, favor alternative folds, or favor the native fold. This is useful in aiding in assay design for testing functions of RNA structure, as well as biophysical assays that require single tight conformations of RNA. For example, for RNA 3D structure determination it is essential to have conformational uniformity in the sample. RNA2DMut can be used to exclude alternate competing conformations that can exist in solution. This is particularly important when evaluating the effect of mutant sequences that are generated to improve crystallization properties of RNA for example, or in tightening up dynamic regions of RNA in NMR.

This tool can be used to evaluate the potential effects of rationally designed mutations, as well as gain basic insights into how an RNA of interest folds by identifying key interactions that stabilize native/alternative conformations. User-defined mutations can be evaluated using the RNA2DMut Sequence Evaluation tool. The Sequence Manipulation tools included on the RNA2DMut site also facilitate this process by allowing users to insert, delete, or replace sequences systematically across a target RNA. Thus, using this platform, a researcher can identify mutational hot-spots and design/evaluate combinations of point mutations or other, more complex, manipulations to RNA. They can also reanalyze previously generated mutants to assist in the rationalization of experimental results or to evaluate the effects of discovered sequence variants: e.g., indels and single nucleotide polymorphisms—particularly in regions prone to SNP-induced conformational switching (Halvorsen et al. 2010).

Five examples are used to illustrate the utility of RNA2DMut: The *mascRNA* (58 nt) embedded within the human MALAT1 lncRNA, an influenza virus splicing regulatory motif (63 nt), the EBER2 ncRNA (173 nt) from Epstein–Barr virus, the *repA* region of the mouse *Xist* lncRNA (490 nt), and finally, the human *Y RNA 5* (hY5, 83 nt) ncRNA are used to illustrate the use of the RNA2DMut Sequence Manipulation and Evaluation tools.

Input

RNA2DMut works through a web-based user interface (Fig. 1) that accepts several inputs. For the Sequence Mutation tool, users input their sequence of interest, which can be as long as 600 nucleotides (nt). This length was selected because calculation accuracy drops swiftly for longer sequences and, additionally, the largest RNA with a determined high-resolution structure, the ribosome, has the majority (99%) of base pairs spanning <600 nt (Doshi et al. 2004; Deigan et al. 2009). Any IUPAC nt is allowed in the input sequence, however, structure is only evaluated for “canonical” pairings between A and U, G and C, or G and U (wobble pairing).

Next, users are allowed to enter a mutational constraint on the residues. Each position in the constraint string, corresponding to the same position in the RNA sequence, can be constrained. A dot, “.”, represents a character that is unconstrained and can be allowed to mutate to all other bases. An “x” forbids changes at that nucleotide, while a “Y” or an “R” constrains mutations to be only pyrimidines or purines, respectively. All possible mutations allowed using constraints are generated and evaluated. If no constraint is given, all positions are allowed to vary.

When RNA2DMut generates 2D structure figures, the default is to show the “ensemble centroid” 2D structure (described below). Users are able to input a “dot-bracket” structure, where matched brackets “()” define base pairs and dots “.” represent unpaired bases. Dot-bracket notation is similar to RNA structure annotations used in Stockholm format alignments. The dot-bracket structure should match the input sequence in length and define pairs to be repre-

The figure consists of two side-by-side screenshots of the RNA2DMut web interface. The left screenshot shows the 'Create RNA2DMut Submission' form. At the top, it says 'IOWA STATE UNIVERSITY Moss Lab - RNA2DMut'. Below that are navigation tabs: 'Sequence Mutation', 'Sequence Evaluation', 'Sequence Manipulation', 'About', and 'Contact Us'. There are also links for 'How to Cite' and 'Related Links'. The main heading is 'Create RNA2DMut Submission'. Below this is a 'Use Default Data' button. The 'Email' field is empty. The 'RNA Sequence' section has a text area for pasting the sequence and a 'Paste RNA Sequence above or upload .fasta file below.' instruction. There is an 'RNA Sequence File' section with a 'Browse...' button (no file selected) and an 'Upload' button. A note says 'Files must be less than 2 MB. Allowed file types: fasta.' The 'Temperature' section has an 'Optional' dropdown set to 'In degrees C'. The 'Constraint Mask' section has an 'Optional' dropdown. The '2D Structure' section has an 'Optional' dropdown. The 'Expires' section has a dropdown set to 'One day'. A 'Submit' button is at the bottom. The right screenshot shows the submission status. It says 'IOWA STATE UNIVERSITY Moss Lab - RNA2DMut' and '2017-11-27 12:21:35'. A green box indicates 'Status: Complete' with a 'Process Log' link. Below this, it says 'Expires: One day' and 'Completion Time: Monday, November 27, 2017 - 10:54'. There are download links for 'outfile1: fileuBIUEq_0.txt', 'outfile2: fileDL47v6_0.txt', 'epsfilemin: filep5S3uM_0.eps', and 'epsfilemax: filelOqbxS_0.eps'. There are also 'minimage' and 'maximage' sections, each with a 2D RNA structure diagram. The 'minimage' diagram shows a structure with blue and red highlights. The 'maximage' diagram shows a structure with blue and red highlights. At the bottom, there is a 'Varna Command File' section with a download link 'fileDrNq0u_0.txt'.

FIGURE 1. Example screenshots of the RNA2DMut web interface. (Left panel) Shown are the input fields for the RNA2DMut Sequence Mutation tool. Fields for entering in sequence data (pasted in or uploaded), temperature, a constraint mask for making mutations, and a place to define the 2D structure used to generate image files are shown. Users can also enter in an e-mail address and have results sent upon completion of the calculation. At the top are additional tabs for other RNA2DMut tools: the Sequence Evaluation tool (evaluates the energy and ensemble diversity of input sequences) and the Sequence Manipulation tools (used to generate sequential indels or substitutions, scramble or reverse complement sequences). Additional links to an “About” page (which has a user guide for RNA2DMut), contact information; citation information and links to useful sites are included. (Right panel) Shown is an example output page with the two text (.txt) outfile download links. Below these are links to download two 2D structure image (.eps) files and, below these, the images of the 2D structures embedded in the page. Here “minimage” shows the mutated positions with the minimum ED relative to WT, while “maximage” shows those mutated sites that have the maximum ED relative to WT. The blue and red color intensity maps define the magnitude of ED change from WT. Command line input is also generated for users to manipulate 2D models within VARNA.

sented. Adding this bracket structure only changes the 2D drawing and *does not* affect the calculated structural data, which are derived from the unconstrained structural ensemble.

The user is next able to define a calculation temperature in degrees Celsius. This option rescales the energy parameters used in the calculation and could be potentially useful in the evaluation of a mutant’s effects at different temperatures.

The default calculation temperature is 37°C (human body temperature).

The RNA2DMut Sequence Evaluation and Manipulation tools are accessed by separate tabs (Fig. 1, left panel) and can accept sequence data as raw nt data or FASTA files. Given one or more sequences (in FASTA format) the Evaluation tool will return data on the structure, energy, and ensemble diversity for each RNA. There are multiple Sequence Manipulation tools available: The insertion tool allows a user to insert a sequence every X nt. The deletion tool deletes a user-defined sequence fragment every X nt. The substitution tool replaces a user-defined sequence fragment every X nt. The scramble tool randomizes an input sequence, generating a user-defined number of shuffled results. Finally, the reverse complement tool simply reverse complements the input sequence.

Output

After RNA2DMut Sequence Mutation calculations are complete, four output files are generated. An example output page is shown in Figure 1. The first file (“outfile1”; Fig. 1, right panel) contains all mutant structure data. Examples are shown in the [Supplemental File outfile1 data](#): (Column A) The mutant name is given, where each mutant is given a unique sequential number ranging from Mutant_0 (the wild-type [WT] sequence) and onward. (Column B) The sequence of the RNA being evaluated, each having a unique point mutation spanning all positions defined in the WT input sequence. (Column C) The MFE structure is given in dot-bracket notation. (Column D) The predicted MFE change in the Gibb’s folding energy (ΔG) is given in kcal/mol. (Column E) The ensemble centroid 2D structure is given in dot-bracket notation. The ensemble centroid structure is the conformation with the smallest base pair distance to the other structures in the ensemble and is thus, a good representation of the overall set. Indeed, centroid pairs are sometimes better able to identify native base pairs than the MFE fold (Ding et al. 2005; Mathews et al. 2010). (Column F) Finally, the ED metric is given. This is the average base pair distance between the set of structures in the calculated 2D structure Boltzmann ensemble; it measures how clustered the structures are in structure space. Low ED scores indicate a single, highly similar cluster of structures, while high ED scores suggest multiple divergent conformations or a lack of structure (Martin 2014). The data in this file are tab-delimited and can be pasted into spreadsheets for manipulation and analysis.

The second file (“outfile2”; Fig. 1, right panel) contains information on the ED for every mutation at a given residue. Examples are shown in [Supplemental File outfile2 data](#). Each row holds data for a particular site: (Column A) The nt number is given, followed by (Column B) the WT base and (Column C) the WT ED, then each possible point mutation (Columns D, G, and J), the mutant name (Columns E, H, an K, which match the names for mutants in outfile1)

and its associated ED (Columns F, I, and L) is given. The data in this file are tab-delimited and can be pasted into spreadsheets for manipulation and analysis.

Two .eps image files are generated that contain 2D structure figures annotated with the minimal and maximal calculated ED, respectively, at each residue (“minimage” and “maximage”; Fig. 1, right panel). It is important to note that the minimal ED change figure shows the difference in ED of the WT minus the mutant sequence; this is done to have positive values for the intensity map used in the image generation process. Examples of minimal and maximal ED annotated structures are shown in Figures 2–5. Files are generated as vector-graphic .eps formatted images by the program VARNA (Darty et al. 2009). A file containing the VARNA command-line scripts appears at the end of the output (Fig. 1, right panel bottom). Unless the user specifies a structure in the input, the default 2D structure shown is the WT sequence ensemble centroid structure.

The RNA2DMut Sequence Evaluation tool generates a single output file that is similar to the RNA2DMut outfile1, only each line of the result file is for a separate input sequence. The Sequence Manipulation tools each generate FASTA files with the input sequence at the top and mutants following below.

Application examples of RNA2DMut

Example 1: analysis of the MALAT1 mascRNA

The mascRNA (MALAT1-associated small cytoplasmic RNA) is embedded in the 3’ end of the MALAT1 lncRNA, where it adopts a tRNA-like structure (Wilusz et al. 2008). This fold is specifically recognized by RNase P, an enzyme used to mature pre-tRNA sequences, and cleaved. It is then further processed by RNase Z to release the mascRNA as an independent ncRNA. As the mascRNA tRNA-like fold is essential for RNase P recognition and processing, it is interesting to identify point mutations that could affect this fold, as well as gain insights into the structural characteristics of the mascRNA. The mascRNA was submitted to RNA2DMut with no constraints on mutations and calculations were run at 37°C.

At the top of Figure 2 is part of the first RNA2DMut output file ([Supplemental File](#)). Here, below the WT sequence (Mutant_0), the mutant sequences are sorted from maximal to minimal ED; only the top five highest and lowest ED mutants are shown. The ED of the WT mascRNA (0.79) is $\sim 8\times$ lower than the average value of mutants (6.36). The MFE and ensemble centroid structures are identical and correctly predict the tRNA-like mascRNA base pairs. This suggests that the WT mascRNA sequence is “tuned” to fold into one dominant conformation, which is well represented by the calculated MFE structure. Mutations to almost any residue are capable of increasing the ED (Fig. 2, bottom left): indeed, of the 174 mutants evaluated, 147 (84%) had higher ED than the WT sequence ([Supplemental File](#)).

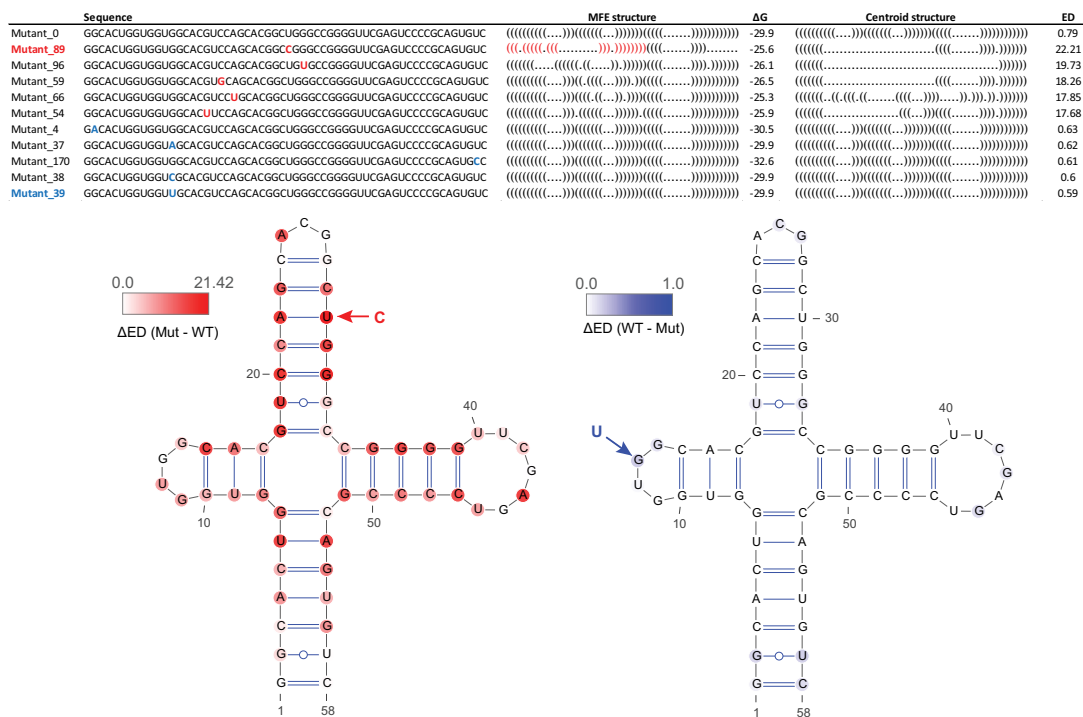


FIGURE 2. Summary of results for the MALAT1 mascRNA. (*Top*) Partial output for the calculation taken from outfile1 (Supplemental File) that shows the wild-type (WT) result, followed by the top five most maximizing and minimizing mutations affecting the ensemble diversity (ED). Maximizing point mutations are annotated in red on the sequence, while minimizing are in blue. Mutant_89, the most maximizing mutant, has minimum free energy (MFE) base pairs predicted to deviate from the WT structure, highlighted in red on the dot-bracket notation structure. (*Bottom left*) The 2D model of the WT ensemble centroid structure annotated with the maximal predicted change in ED from mutant (Mut) to WT at each base, represented by a red heat map. The most disruptive mutation (Mutant_89) is indicated with a red base and its location is shown by the red arrow. (*Bottom right*) The same 2D model for the mascRNA, only now annotated with the minimal possible ED mutations, where a blue intensity map (comparing the change in ED from WT to Mut) is used. The minimal ED mutant (Mutant_39) is indicated with a blue base and its location with a blue arrow.

The second mascRNA hairpin (nt 18–34; Fig. 2) is particularly susceptible to disruption; all of the top five maximal ED mutations occur in this hairpin. The mutation predicted to be most disruptive to the ED is the one that changes U30 to C (Mutant_89; Fig. 2), which leads to a 28× increase in ED. This is due to residues in this hairpin having a high potential for forming alternative base pairs. Changing U30 to C, for example, destabilizes the native hairpin while stabilizing alternative conformations: such as the predicted mutant MFE structure shown in Figure 2 (top, annotated in red). The Mutant_89 MFE fold is not, however, well represented by the centroid structure (only four MFE base pairs appear in the centroid), which suggests that this mutation (as also indicated by its high ED) greatly perturbs the native structural ensemble and that additional alternative conformations are possible. If one were, for example, interested in making mutations to identify the key residues for forming the structure required for RNase P processing, Mutant_89 (as well as the other highly disruptive mutants) would be an excellent choice.

Although, the ED of the WT mascRNA is already quite low, a small number of mutants (27 sequences) had lower ED. Figure 2 (bottom right) shows the location of the mutations that most minimized the ED (Supplemental File) mapped on the WT centroid structure. In addition to being

sparse, the mutations that reduce ED do so to a small degree. It is, however, possible to discover interesting features of the mascRNA here. The minimal ED mutations buried in stems can convert GU “wobble” pairs into thermodynamically more stable Watson–Crick pairs: U19 is mutated to a C to form a CG pair, while G33 is mutated to an A to form a UA pair and, similarly, G2 is mutated to an A to form an AU pair, while U57 is mutated to an C to form a GC pair (Supplemental File). Interestingly, all but one mutation to the terminal GC pair (formed by G1 and C58) slightly reduces the ED; this is due to terminal base stacking effects taken into account in RNAfold, which stabilize mutations here. Mutations in loops that reduce ED notably involve all but one of the mascRNA single-stranded G and C bases. Mutations of these residues decrease the ED by forbidding possible alternative conformations that place these bases in relatively stable (vs. AU) GC pairs.

The loop mutations that reduce ED are a good example of how RNA2DMut could facilitate high-resolution structural studies of RNA. In RNA crystallography, for example, it is common to try different mutant sequences to facilitate crystallization: by tightening RNA structure or by enhancing RNA–RNA quaternary contacts that promote crystal formation. The ED-reducing loop mutations calculated for the

	Sequence	Centroid structure	ED
Mutant_0	GGUCUGAAAAUGAUCUUCUGAAAAUUUGCAGGCCUUAUCAGAAACGAAUGGGGUGCAGAUG	(((((.....)).....(((((((.....)))).....)))).....))	2.63
Jiang et al.	GGUCUGAAAAUGAUCUUCUGAAAAUUUGCAGGCCUUAUCAGCAACCAUUGGGGUGCAGAUG	(((((.....)).....(((((((.....)))).....)))).....))	5.39
Mutant_111	GGUCUGAAAAUGAUCUUCUGAAAAUUUGCAGGCCUUAUCAGAAACCAUUGGGGUGCAGAUG	(((((.....)).....(((((((.....)))).....)))).....))	19.31
Mutant_99	GGUCUGAAAAUGAUCUUCUGAAAAUUUGCAGGCCUUAUCAGCAACCAUUGGGGUGCAGAUG	(((((.....)).....(((((((.....)))).....)))).....))	19.26
Mutant_23	GGUCUGAAAAUGAUCUUCUGAAAAUUUGCAGGCCUUAUCAGAAACGAAUGGGGUGCAGAUG	(((((.....)).....(((((((.....)))).....)))).....))	18.94
Mutant_105	GGUCUGAAAAUGAUCUUCUGAAAAUUUGCAGGCCUUAUCAGAAACGAAUGGGGUGCAGAUG	(((((.....)).....(((((((.....)))).....)))).....))	18.81
Mutant_74	GGUCUGAAAAUGAUCUUCUGAAAAUUUGCAGGCCUUAUCAGAAACGAAUGGGGUGCAGAUG	(((((.....)).....(((((((.....)))).....)))).....))	18.80
Mutant_129	GGUCUGAAAAUGAUCUUCUGAAAAUUUGCAGGCCUUAUCAGAAACGAAUGGGGUGCAGAUG	(((((.....)).....(((((((.....)))).....)))).....))	1.36
Mutant_89	GGUCUGAAAAUGAUCUUCUGAAAAUUUGCAGGCCUUAUCAGAAACGAAUGGGGUGCAGAUG	(((((.....)).....(((((((.....)))).....)))).....))	1.24
Mutant_91	GGUCUGAAAAUGAUCUUCUGAAAAUUUGCAGGCCUUAUCAGAAACGAAUGGGGUGCAGAUG	(((((.....)).....(((((((.....)))).....)))).....))	1.08
Mutant_90	GGUCUGAAAAUGAUCUUCUGAAAAUUUGCAGGCCUUAUCAGAAACGAAUGGGGUGCAGAUG	(((((.....)).....(((((((.....)))).....)))).....))	1.03
Mutant_113	GGUCUGAAAAUGAUCUUCUGAAAAUUUGCAGGCCUUAUCAGAAACGAAUGGGGUGCAGAUG	(((((.....)).....(((((((.....)))).....)))).....))	0.72

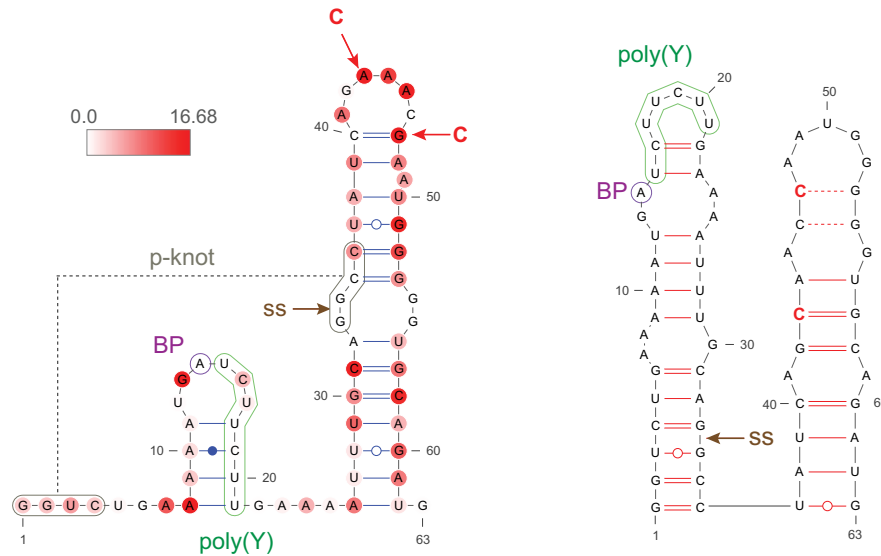


FIGURE 3. Summary of results for the influenza virus splice site regulatory structure. (Top) Results from outfile1 (Supplemental File) are shown and annotated similarly as in Figure 2 (the MFE energy and structure are omitted for space). The Jiang and colleagues double point-mutation evaluated using the RNA2DMut Evaluation tool is added in the second position. Centroid predicted base pairs that differ from the literature model are annotated in red in the Jiang and colleagues structure, as well as Mutants_111 and 99, which are combined in the Jiang and colleagues double point-mutant. (Bottom left) The most ED maximizing mutations annotated on the literature 2D model. Annotations are similar to Figure 2. In addition, the branch point (BP) and poly(Y) (pyrimidine) tract are annotated in purple and green, respectively. The 3' splice site (ss), used to generate M2 mRNA, is indicated with a brown arrow. A pseudoknot helix from an alternative conformation is indicated in gray and connected with a dashed line. (Bottom right) The alternative fold stabilized by the double point-mutation (red bases) used by Jiang and colleagues. The pairs that differ from the literature model are colored red and two MFE pairs that are missing from the centroid structure model are indicated with dashed lines. Annotations for functional elements are the same as the structure to the left.

mascRNA would be ideal candidate mutations for such a process: They minimize the potential for alternative (presumably non-native) conformations that could contaminate/inhibit crystals while also introducing novel functional groups that could promote stabilizing RNA–RNA crystal contacts.

Example 2: analysis of an influenza virus splicing regulatory motif

The influenza A virus (IAV) genome is comprised of 8 negative-sense (–)RNA fragments, which encode >11 proteins via alternative splicing of coding (+)RNAs. IAV splicing is regulated by a number of varied mechanisms, including the manipulation of RNA folding (Dubois et al. 2014). Previous analyses of IAV, predicted stable structures throughout viral RNAs (Gultyaev et al. 2007; Moss et al. 2011). The presence of structured RNA is common between IAV strains and also occurs in influenza B and C viruses (Gultyaev et al. 2007; Dela-Moss et al. 2014), suggesting common structural strategies

for post-transcriptional regulation of viral gene expression. Functional analyses of IAV structured regions have confirmed the importance of 2D structure to the virus (Soszynska-Jozwiak et al. 2015; Gultyaev et al. 2016). For example, an IAV structured region occurs in the pre-mRNA used to generate the M2 ion channel protein; it incorporates the 3' splice site along with the branch-point (BP) and poly-pyrimidine (poly-Y) tract sequences into a compact structure, which has the potential to shift from a hairpin to a pseudoknot (p-knot) conformation, which buries the splice site in the p-knot helix (formed by nt 1–4 and 33–36; Fig. 3, bottom left). This structured region was extensively characterized using *in silico* predictions, comparative sequence/structure analysis, biochemical probing and high-resolution NMR studies (Moss et al. 2011, 2012a, 2012b; Chen et al. 2015). The functional significance of structure in this region was assessed using a mutational strategy based on properties of the RNA structural ensemble. Here, mutations that increased ensemble defect were identified and incorporated

accessibility for splicing. Interestingly, the p-knot conformation of this RNA also sequesters the splice site in a helix (Moss et al. 2012b). It is important to note here that p-knot conformations are not considered by RNA2DMut because these structures are not allowed in RNAfold (or in most folding algorithms, due to their complexity). This is one important limitation of RNA2DMut that needs to be considered; individual p-knot helices can appear in the ensemble, but never simultaneously, thus one or both p-knotted helices may be underrepresented.

Previous analyses of this sequence found that natural mutations accumulated to stabilize strains that replicated in higher temperature environments (e.g., human lung vs. the swine lung or avian gut). RNA2DMut was run at 33°C, 37°C, and 41°C (the approximate temperatures of the human and pig lungs, and avian gut, respectively) and the resulting data appear in the [Supplemental File](#) (influenza intron outfile2). Looking at WT vs. Mutant_111, for example, increasing temperatures increase the ED of both WT and mutant; however, the difference in ED, comparing mutant to WT, is reduced from 7.3× (at 33°C) to 5.8× (at 37°C). Interestingly, though the ED increased with temperature for most sequences, the magnitude of change varied between mutants and, in some cases, ED was predicted to *decrease*. The varied behavior of point mutants at different temperatures suggests that RNA2DMut could also be a useful tool for evaluating the potential effects of sequence variations over different temperatures.

Example 3: analysis of the EBER2 viral ncRNA

Epstein–Barr virus (EBV) is a ubiquitous human herpes virus that is implicated in a variety of different cancers and autoimmune disorders (Raab-Traub 2007; Chen 2011; Iwakiri 2016). During latent infection, the double-stranded DNA viral genome resides in the nucleus of infected cells. During infection, the viral ncRNAs (EBER1 and 2) are transcribed by RNA polymerase III and accumulate to great abundance: up to 10^7 copies per cell (Lerner et al. 1981). EBERs alone can promote tumor growth (Yamamoto et al. 2000) and are implicated in pathogenesis (Iwakiri 2016); however, molecular mechanisms for their functions have yet to be fully characterized. Recent advances have been made in the analysis of the ribonucleoprotein (RNP) composition of EBER2 and its function. The EBER2 RNP includes host proteins SFPQ, RBM14, NONO, and PAX5 (Lee et al. 2015, 2016). EBER2 possesses a secondary structure that is likely important in recruiting interacting proteins. Likewise, *intermolecular* base-pairing between EBER2 and another EBV transcript generated from the *terminal repeat* (TR) region plays an essential role in loading the PAX5 transcription factor onto the EBV genome (Lee et al. 2015). The maximum predicted extent of the TR-interacting region is annotated in green on Figure 4.

Results for calculations on EBER2 are summarized in Figure 4 and detailed results are in the [Supplemental File](#). The WT ED of EBER2 (39.55) is higher than the previous

two examples. The WT ensemble centroid resembles the EBER2 model structure based on biochemical probing and sequence analysis (Glickman et al. 1988; Moss et al. 2014); only failing to predict the short terminal stem loop in the large hairpin and adding an isolated base pair to the shorter 5' hairpin loop (Fig. 4, top). Analyzing the mutations, only 34% of the 426 point mutations evaluated increased the ED ([Supplemental File](#)). The mutations that increase ED cluster in the basal EBER2 stem (Fig. 4, bottom left). Mutations that reduce ED, however, are common, high in magnitude and occur throughout the long 3' hairpin (Fig. 4, bottom center), suggesting this region could be dynamic.

The most ED-reducing mutation occurs at A83, which changes this to a C. C83 in Mutant_155 stabilizes a hairpin (nt 81–104) in an alternative conformation that totally disrupts the WT model 3' hairpin. The WT long hairpin structure is replaced with four shorter hairpins in the Mutant_155 ensemble centroid fold (Fig. 4, bottom right). This conformation binds up the last three C residues of the TR interaction site in an adjacent hairpin loop. To interact with the TR transcript, the hairpin formed from nt 25 to 61 must unwind. In the WT structure model, the terminal nt are unpaired in a flexible loop and are accessible for initiation of annealing to the TR transcript. Mutant_155, by sequestering these bases in a helix, could inhibit TR-binding and, potentially, PAX5 loading. In addition, the drastic change in the long 3' hairpin structure, which may be a “hub” for RNP interactions, could alter protein binding or intermolecular RNA interactions: e.g., the long WT loop region (nt 106–125; Fig. 4 bottom left and center) is occluded in the Mutant_155 centroid model (Fig. 4, bottom right).

Interestingly, the hairpin stabilized by C83 in Mutant_155 (nt 81–103) as well as part of the adjacent hairpin (nt 105–132) are both identical to hairpins predicted in an early model for EBER2 based on the structures of the adenovirus-associated RNAs (VAs) (Rosa et al. 1981; Iwakiri 2016). Like EBERs, VAs are small nonpolyadenylated ncRNAs; notably, EBERs can partially compensate for lack of VAI in adenovirus replication (Iwakiri 2016), suggesting overlapping interactions/functions. It is possible that Mutant_155, as well as other mutations that “lock in” VA-like conformations ([Supplemental File](#)), are identifying an alternative conformation of EBER2 that is biologically significant. For example, the VA-like hairpins could be bound and stabilized by proteins expressed during particular times of infection, altering EBER2 structural equilibrium. Further work is required to test this; however, the EBER2 example is an excellent one for showcasing how RNA2DMut is more than a tool for mutational design—it can also give insights into RNA structure and help to generate biological hypotheses.

Example 4: the Xist repeat A region

Xist is an ~18 kb lncRNA that functions by associating with one of the two X chromosomes found in female mammals

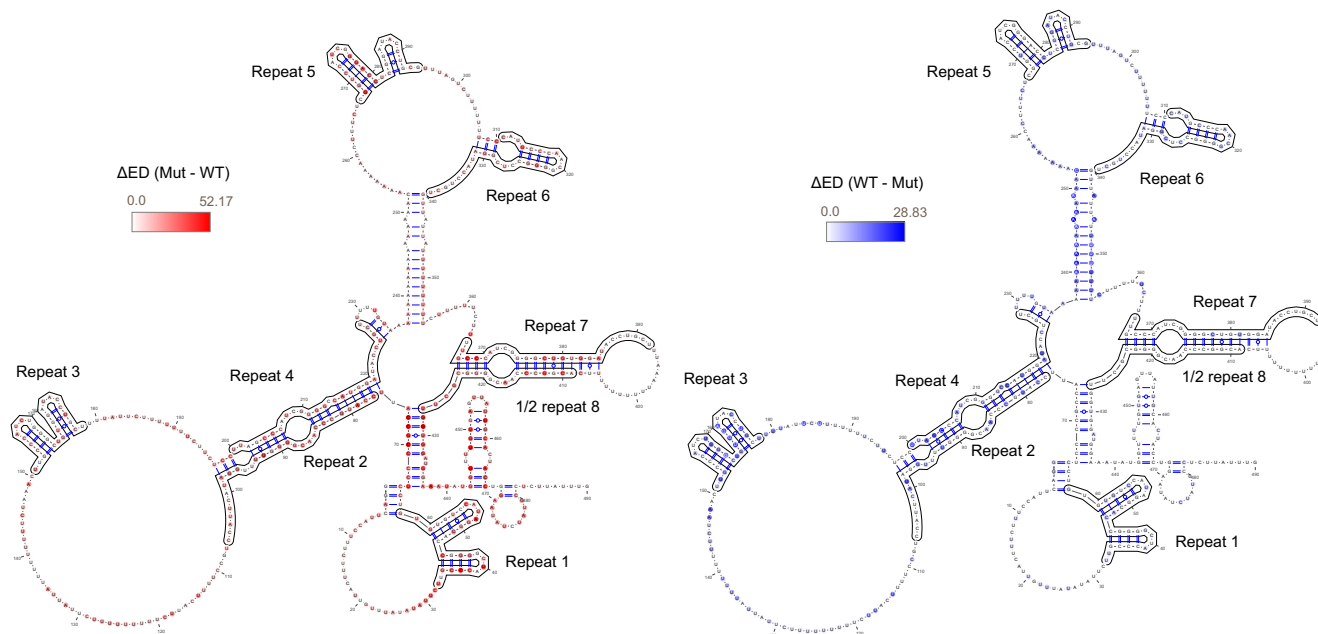


FIGURE 5. Xist repA region structure models annotated with the most maximizing and minimizing ED changes annotated in red and blue, respectively, in the *left* and *right* panels. The A repeat units are indicated with black outlines.

and stimulates chromosome inactivation. X inactivation is an essential process in maintaining healthy amounts of X-associated gene products (dosage compensation). Xist is generated by all eutherian mammals; however, except for several repeat regions, its sequence is poorly conserved. The best characterized conserved region in Xist is the repeat (rep)A region (Cerase et al. 2015), which consists of 7.5× repeat units separated by nonconserved linker sequences. Previous work analyzing repA RNA structure, using chemical probing and NMR on in vitro generated RNAs, produced alternative models: one with local intra-repeat pairing (Duszczuk et al. 2011) and a set of three with alternative inter-repeat pairing (Maenner et al. 2010). A more recent model, based on in vivo chemical probing, is comprised of elements of the previous models as well as novel interactions (Fang et al. 2015). In vivo crosslinking studies have found evidence of a complex folding landscape for repA (Lu et al. 2016) with repeats engaged in multiple pairing interactions, indicating dynamic interactions and alternative conformations.

A previously identified structured region (Fang et al. 2015; 490 nt long), which fully encompasses repA, was submitted for analysis using RNA2DMut (results appear in the Supplemental File). The ED of the WT sequence is 82.16, which is only 6.4% lower than the average of all generated mutants (87.78); this is consistent with the finding that the repA RNA structure landscape is complex. To visualize the impacts of individual mutations, however, the model derived from in vivo structure probing is shown in Figure 5. Here, ED maximizing and minimizing changes are annotated on the in vivo 2D model. More ED maximizing sites were

identified and the ED maximizing changes were of a greater magnitude than minimizing ones. In both cases, clustered regions sensitive to both kinds of change were observed. Interestingly, the clustered regions with the greatest ED changes occurred outside the A repeats, in the poorly conserved linker sequences.

The most maximizing mutations in repA are predicted to occur in the basal stem separating Repeats 1 and 2, and in a short tetraloop hairpin that occurs downstream from ½ repeat 8 (Fig. 5, left panel). Although repA linker sequences are generally poorly conserved, these structures had base pairs that were preserved in rodents, and support for the basal stem came from a (rodent specific) compensatory change (Fang et al. 2015). The most maximizing mutations in the basal stem allow alternative pairings with Repeat 1 that are predicted to form a novel hairpin (Supplemental File). The most maximizing changes in the downstream hairpin favor interactions between this sequence and the linker region between Repeats 2 and 3. The most minimizing changes occur in a stem that places Repeats 5 and 6 within a multibranch loop (Fig. 5, right panel). This long stem is formed by an upstream A-rich tract annealing with a U-rich tract downstream from Repeat 6. The ED minimizing mutations either “lock in” the modeled interaction by adding AU/UA pairs to the helix (e.g., Mutant_738; Supplemental File) or stabilize “slipped” versions of the interacting helices. For example, Mutant_721 swaps A241 with a G, which stabilizes a GC pair in a predicted helix that makes use of the U-rich stretch immediately following the WT helix (nt 263–250 paired with nt 347–361).

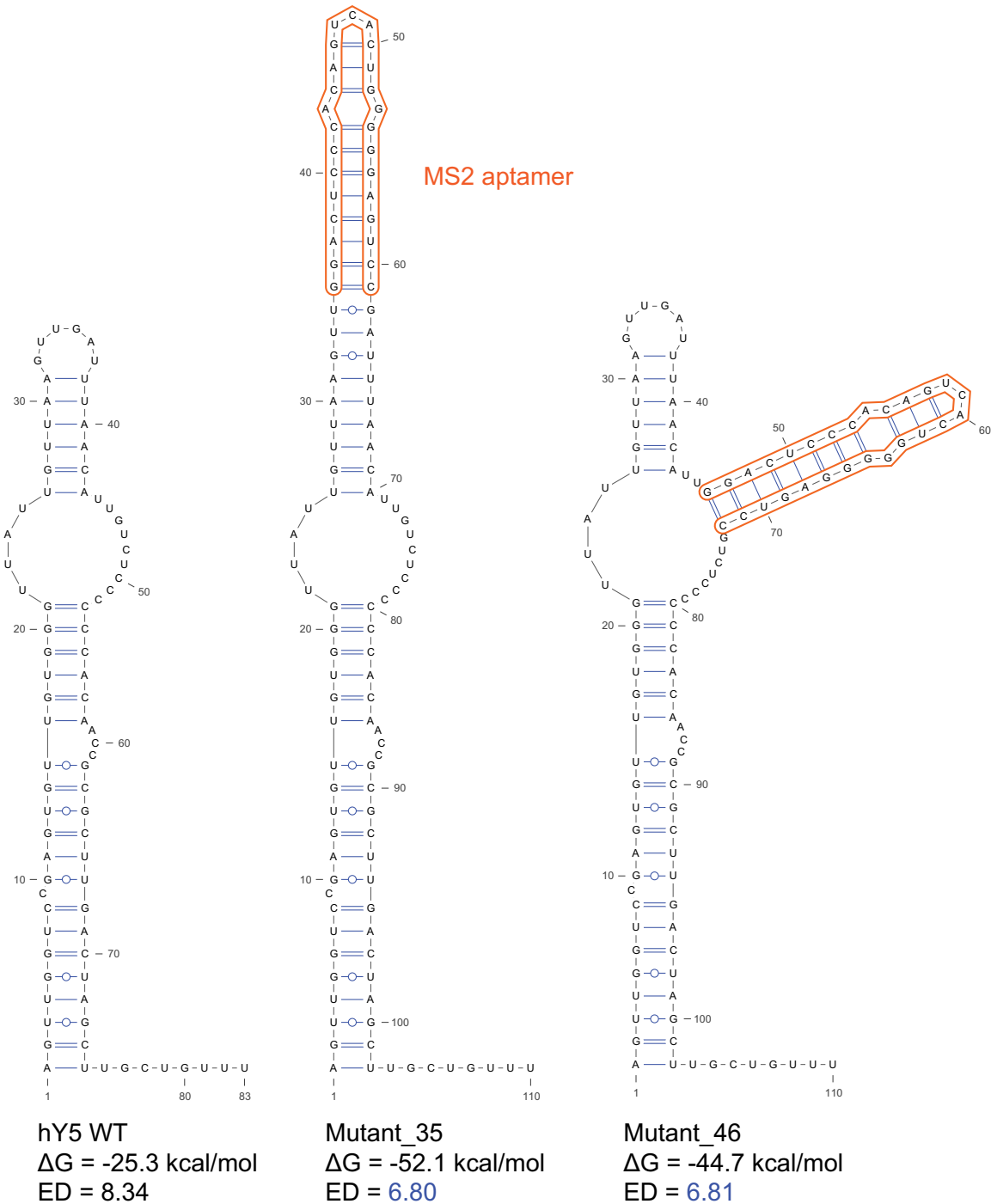


FIGURE 6. Optimized insertion sites of an MS2 aptamer into the hY5 ncRNA. The WT hY5 ensemble centroid model is shown to the *left*. Ensemble centroid models for two low ED predictions for loop insertions follow, where the MS2 aptamer sequence is indicated with orange outlines.

Example 5: finding the optimal MS2 aptamer insertion sites on a ncRNA

While evaluating all potential point mutations with the RNA2DMut Sequence Mutation tool can reveal useful features, it is helpful to be able to evaluate other types of mutations: e.g., insertions and deletions (indels) and sequential substitutions. To facilitate these sorts of analyses, the

RNA2DMut site includes two tools: The “Sequence Evaluation” tool, which allows users to submit multiple sequences for prediction of the MFE structure and energy, as well as the ensemble centroid structure and ED. The “Sequence Manipulation” tool allows users to generate sequential indels or substitutions at defined positions and to produce shuffled or reverse complemented sequences as

well. To highlight how the Sequence Manipulation and Evaluations tools can be used to study RNA, the example of the human Y RNA 5 (hY5) ncRNA is used. Y RNAs are essential for DNA replication in vertebrates and are also important in cellular stress response in both eukaryotes and prokaryotes (Kowalski and Krude 2015). Y RNA localization and interactions are of great interest; therefore the example task performed on hY5 was to identify the optimal site for introducing an MS2 aptamer sequence/structure. RNA aptamers are useful tools for both purification and characterization (e.g., imaging) of RNAs (Panchapakesan et al. 2017). For example, MS2 aptamers were previously introduced into the EBER1 ncRNA (Lee et al. 2012) to facilitate capture of endogenously interacting host proteins (via binding of the MS2 aptamer by MS2 bacteriophage coat protein [Stockley et al. 1995]).

Using the Sequence Manipulation (insertion) tool, the 27-nt MS2 aptamer hairpin sequence was inserted after every WT hY5 base. A total of 88 mutant sequences were generated, which were then evaluated using the Sequence Evaluation tool (hY5 results appear in the [Supplemental File](#)). In the WT hY5 structure prediction, both the MFE and ensemble centroid folds (identical to each other) match the reported model (based on in vitro biochemical probing) (van Gelder et al. 1994): This includes the bulged C9 nt, which occurs in all Y RNAs and is conserved between species (Fig. 6). The only difference is that in the in vitro model, U15 is bulged out to allow the stem formed by nt 16–21 and 53–58 to grow by two base pairs (U15 with A59 and G14 with C60). The ED of the WT hY5 sequence (8.34) is lower than the aptamer insertion mutants (average ED 10.58); most insertions (66%) increase ED. Unsurprisingly, the neutral and ED reducing insertions occurred in loops; however, not all insertion sites in loops reduced the ED to the same degree; and some even *increased* ED ([Supplemental File](#)). For example, in the terminal hairpin loop of hY5 (nt 32–37) the insertion site predicted to most reduce ED (and ΔG as well) occurs after U34 (Mutant_35; Fig. 6; [Supplemental File](#)). This insertion allows all loop nt to pair with each other and for the MS2 aptamer hairpin to form a continuous helix with them. Likewise, in the loop spanning nt 44–52, the most ED reducing insertion occurs after U45 (Mutant_46; Fig. 6; [Supplemental File](#)), suggesting that, out of all possible insertion sites in this loop, this one is least likely to result in misfolding. This example highlights how "eyeballing" a favorable insertion site may not lead to obvious effects on RNA structure. Likewise, other common sequence manipulations (e.g., deletions and replacements) may also yield surprising results. These can be accounted for using the Sequence Manipulation and Evaluation tools on the RNA2DMut site.

Potential limitations

The folding algorithm used in RNA2DMut predictions, RNAfold, has been extensively benchmarked vs. experimental data and is in the class of top performing programs for an-

alyzing the effects of single nt polymorphisms on the structural ensemble (Corley et al. 2015). Even very good algorithms, however, are not 100% predictive. Inaccuracies in the thermodynamic energy model and assumptions made during the prediction (e.g., that protein binding, 3D structure and non-nearest-neighbor energy contributions can all be ignored) all lead to errors in prediction. For MFE structure prediction, this results in approximately 70% of base pairs (on average) being correctly predicted for RNA sequences below 700 nt in length (Mathews et al. 2010). This number can vary greatly depending on the target. Pseudoknot containing RNAs (formed by non-nested or "crossing" base pairs), in particular, are poorly predicted. These motifs are forbidden in most folding algorithms (e.g., RNAfold) and must be treated differently in alternative prediction approaches (the RNA2DMut page has links to pseudoknot prediction programs). Although not accounted for in RNAfold, running RNA2DMut on pseudoknot sequences can still offer valuable information. For example, in the IAV intronic structure example, mutations that stabilize an alternative fold that competes with both the nonpseudoknotted and pseudoknotted conformations of this RNA were identified.

Increased prediction accuracy can be achieved by using complementary data from comparative sequence analysis or experimental approaches; however, even in the absence of such data, it is possible to estimate prediction accuracy. Using data from the partition function calculation, a base-pair probability can be calculated. By "counting" the occurrences of a base being unpaired or paired across the RNA 2D structural ensemble, the probability is estimated (e.g., if the same base pair is formed in almost all structures, the probability will approach 1). High pairing probabilities are correlated with higher prediction accuracies (Mathews 2004). It is worth noting that the RNAfold ensemble centroid models captured by RNA2DMut are the best representations of structural ensemble contained within the partition function and thus, are frequently better able to correctly predict pairs. Additionally, it is possible to map the base-pair probabilities onto both the MFE and ensemble centroid models using RNAfold (a link to the RNAfold server can be found on the RNA2DMut page) to visualize which regions of the model are predicted with greater certainty.

The current RNA size limit for the RNA2DMut Sequence Mutation tool is 600 nt; thus, this tool cannot be used to analyze whole lncRNAs or other long species of RNA directly. This limitation, however, can be overcome by fragmenting long sequences into smaller structural domains amenable to the RNA2DMut approach. This has the added bonus of also increasing prediction accuracy, as smaller RNAs are generally better predicted than large ones (Mathews et al. 2010). If the 2D structure of a long RNA is known, then smaller defined motifs (e.g., those contained within a helix) can be extracted and analyzed. When the domain structure is not known, then prediction approaches can be used to estimate structured domains to be analyzed. One approach is to

calculate the partition function of a long RNA and define domains as being contained within high probability helices (Siegfried et al. 2014). In this approach, limits on base-pair distance are also frequently imposed (e.g., no pairs may span >600 nt). An alternative approach is to use a scanning window to fold individual RNA fragments for structure, then reconstitute longer structural domains by concatenating overlapping windows that show propensity for structure. For example, this latter approach was used to scan the Xist lncRNA (Fang et al. 2015) and defined the repA structured domain discussed in Example 4.

RNA2DMut is only able to consider intramolecular interactions. While intermolecular pairing is not currently allowed in RNA2DMut, this tool can be used to introduce mutations that are predicted to occlude or make accessible regions of an RNA that are involved in intermolecular duplexes (e.g., miRNA binding sites or the EBER2 TR interacting site). If a binding site is known, RNA2DMut can be used to predict the effects of making point mutations on its accessibility. When binding sites are not known, several algorithms are available for predicting duplex interactions (links are presented on the RNA2DMut page).

Summary and conclusion

The ED change can be used to discover mutations that can disrupt or stabilize interactions in the structural ensemble that potentially lead to conformational shifts. From the previous examples, we observed the utility of this web tool. The examples range in size from the small, highly structured, mascRNA to the larger and more dynamically structured Xist repA sequence; thus allowing one to consider how the ED calculations behave in different starting sequences/structures. In the mascRNA, RNA2DMut found hot spots for disrupting its tRNA-like fold, which presumably would affect RNase P recognition and MALAT1 processing. Conversely, mutations in loops could forbid alternative conformations and further tighten the converged structural ensemble on the native tRNA-like conformation, which may be useful in biophysical analyses of RNA. In influenza, the utility of RNA2DMut was highlighted by independently identifying mutations highly perturbing to the 2D ensemble, which were previously found to disrupt structure/function. These previous analyses were extended by recalculating metrics at different replication temperatures of the virus and showing how these, and all other potential point mutations, change over a range of biologically relevant temperatures. In EBER2, a single-point mutation was shown to have the potential to totally disrupt a relatively large structural domain, possibly affecting RNA intermolecular interactions with proteins or other RNAs: e.g., occluding the TR interaction essential for PAX5 recruitment and stabilizing VA-like conformations. In the Xist repA region, the most volatile regions (with respect to ED changing mutations) were found to be outside the conserved repeat regions. Finally, an example

is given of how the RNA Sequence Manipulation and Evaluation tools could be used to optimize the insertion site of an MS2 aptamer hairpin into a human Y RNA.

In conclusion, RNA2DMut is a useful tool for researchers interested in the design and analysis of RNA mutants that affect 2D structure. Programs are made available via a simple to use web interface that will facilitate the wide array of investigations, which are essential for understanding the growing world of functional RNAs.

MATERIALS AND METHODS

MALAT1 mascRNA

The mascRNA sequence and native structure model (based on homology) were obtained from the Rfam database (Rfam ID: RF01684). The input used for RNA2DMut was the mascRNA sequence, all other parameters were left blank: Defaulted to no constraint on point mutations, the output 2D structures were taken from the WT ensemble centroid and the folding temperature was 37°C.

Input sequence: GGCACUGGUGGUGGCACGUCCAGCACGG CUGGGCCGGGUUCGAGUCCCCGAGUGUC

IAV splicing regulatory motif

The IAV sequence used was from the A/Puerto Rico/1934 H1N1 (PR8) strain, which matched that used in the previous experimental analysis of this sequence/structure (Jiang et al. 2016). The sequence was obtained from the NCBI Influenza Virus Resource (Bao et al. 2008): genome segment 7 (encoding M1/M2 genes), GenBank# LC120394.1. A mask was used to constrain the calculation and avoid mutating the branch-point sequence or the 4 nt at the splice site, as well as to constrain the mutations in the poly(Y) tract to be pyrimidines. The output 2D structure was constrained to fit the hairpin model determined from *in silico*, comparative sequence, biochemical and NMR analyses (Moss et al. 2012a; Chen et al. 2015). The calculation temperature was set to three different values over three separate RNA2DMut runs: 33°C, 37°C, and 41°C.

Input sequence: GGUCUGAAAAAUGAUCUUCUUGAAAAUU UGCAGGCCUAUCAGAAACGAAUGGGGGUGCAGAUG

Input mask:xYYYYYYY.....xxxx.....

Input structure:((((.....)))....(((((((.....))))))..))

Input temps: 33, 37, and 41.

The sequence of the previously studied double-point mutant was taken from Jiang et al. (2016), placed in FASTA format and submitted for analysis by the RNA2DMut Sequence Evaluation tool.

EBER2 ncRNA

The EBER2 RNA sequence was taken from the EBV type I RefSeq genome: GenBank# NC_007605.1. A mask was used to avoid mutations to nt 34–63: the greatest possible extent of the TR interaction site (Lee et al. 2015). The 2D model was taken from previous biochemical and comparative sequence/structure analyses (Glickman et al. 1988; Moss et al. 2014). Calculations were run at the default temperature (37°C).

Input sequence: AGGACAGCCGUUGCCCUAGUGGUUUCG GACACACCGCCAACGCUCAGUGCGGUGCUACCGACCCGAG

```
GUCAAGUCCCGGGGGAGGAGAAGAGAGGCUCUCCCGCCUA  
GAGCAUUGCAAGUCAGGAUUCUCUAAUCCUCUGGGAG  
AAGGGUAUUCGGCUUGUCCGCUAUUUUU
```

```
Input mask: .....XXXXXXXXXXXXXXXXXXXX  
XXXXXXXXXX.....
```

```
Input structure: .((((((((((((.....((((.....))))))  
....))))).((((((((((((.....((((.....))))))  
))))).))))).))))).))))).))))).))))..
```

Xist repA region

The previously deduced repA structured region (Fang et al. 2015) was extracted from the mouse Xist RefSeq: accessed from the NCBI nt database (GenBank Accession NR_001463). Calculations were run at the default temperature (37°C).

```
Input sequence: GGACUUAACCUUUUCUAUUGUUUAUA  
UAUUUUGCCCAUCGGGGCCACGGAUACCUUGUGUCCU  
CCCCGCCAUUCCAUGCCCAACGGGGUUUGGAUACUUAC  
CUGCCUUUUCAUUCUUUUUUUUUCUUAUUUUUUUUU  
UCUAAACUUGCCCAUCUGGGCUGUGGAUACCUUUUA  
UUCUUUUUUUCUUCUCCUUAGCCCAUCGGGGCAUGGAU  
ACCUGCUUUUUUGUAAAAAACAACAAAACC  
UUUCUCGGUCCAUCGGGACCUCGGAUACCUUGUUAAGU  
CUUUUUUCCCAUGCCCAACGGGGCCUCGGAUACCUUGCU  
GUUAUUUUUUUUUUUUUUUUUUUUGCCCAUCGGGG  
CUGUGGAUACCUUGCUUUAAAUUUUUUUUUUCACGGGCC  
AACGGGGCCGUUGGGGAUGGAAAUUGGUUUUGGAG  
UUAAUUGCACUACCUGGAAUAUCUAGCCUUUAUUUG
```

```
Input structure: (((((((((((.....((((.....))))(((((  
))))))))))))))))))))))))))))))))))))..  
))))))))))))))))))))))))))))))))))))..  
))))))))))))))))))))))))))))))))))))..  
))))))))))))))))))))))))))))))))))))..  
))))))))))))))))))))))))))))))))))))..  
))))))))))))))))))))))))))))))))))))..  
))))))))))))))))))))))))))))))))))))..
```

Human Y RNA 5

The hY5 sequence was taken from the NCBI nucleotide database (GenBank Accession NR_001571.2). Calculations were run at the default temperature (37°C).

```
Input sequence: AGUUGGUCGAGUGUUGGGUUAAU  
GUUAAGUUGAUUAACAUUGUCUCCCCCACACCGCGC  
UUGACUAGCUUGCUGUUU
```

RNA2DMut program and web implementation

RNA2DMut and the Sequence Evaluation and Manipulation tools were written in the Perl programming language. Source code for each is available at <https://github.com/waltermoss/RNA2DMut>. The RNA2DMut Sequence Mutation and Evaluation tools make use of the RNAfold module within the ViennaRNA Package (Hofacker 2003; Lorenz et al. 2011) to calculate the MFE structure and energy, as well as the ensemble centroid fold and ED via the partition function calculation option (McCaskill 1990). Data are parsed and output into text files, as well as command-line inputs for the generation of 2D images via VARNAs (Darty et al. 2009). VARNAs is then invoked to generate .eps image files. Servers and IT support were provided by the Research IT group at Iowa State University <http://researchit.las.iastate.edu>.

Data access

RNA2DMut can be accessed at <https://rna2dmu.bb.iastate.edu/>.

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

ACKNOWLEDGMENTS

I would like to thank Levi Baber from the Iowa State University Biology Information Technology (BIT) program and Research IT department for his extensive help in implementing the RNA2DMut server. Thank you to members of the Moss Laboratory for helpful discussions and suggestions. Thank you also to the anonymous peer reviewers whose valuable suggestions made this project and paper much better. This work was supported by startup funds from the Iowa State University College of Agriculture and Life Sciences and the Roy J. Carver Charitable Trust, as well as grant 4R00GM112877-02 from the National Institutes of Health/National Institute of General Medical Sciences.

Received September 7, 2017; accepted November 25, 2017.

REFERENCES

- Bao Y, Bolotov P, Dernovoy D, Kiryutin B, Zaslavsky L, Tatusova T, Ostell J, Lipman D. 2008. The influenza virus resource at the National Center for Biotechnology Information. *J Virol* **82**: 596–601.
- Bellaousov S, Reuter JS, Seetin MG, Mathews DH. 2013. RNAstructure: Web servers for RNA secondary structure prediction and analysis. *Nucleic Acids Res* **41**: W471–W474.
- Cerese A, Pintacuda G, Tattermusch A, Avner P. 2015. Xist localization and function: new insights from multiple levels. *Genome Biol* **16**: 166.
- Chen MR. 2011. Epstein–Barr virus, the immune system, and associated diseases. *Front Microbiol* **2**: 5.
- Chen JL, Kennedy SD, Turner DH. 2015. Structural features of a 3' splice site in influenza A. *Biochemistry* **54**: 3269–3285.
- Clark MB, Choudhary A, Smith MA, Taft RJ, Mattick JS. 2013. The dark matter rises: the expanding world of regulatory RNAs. *Essays Biochem* **54**: 1–16.
- Corley M, Solem A, Qu K, Chang HY, Laederach A. 2015. Detecting riboSNitches with RNA folding algorithms: a genome-wide benchmark. *Nucleic Acids Res* **43**: 1859–1868.
- Darty K, Denise A, Ponty Y. 2009. VARNAs: interactive drawing and editing of the RNA secondary structure. *Bioinformatics* **25**: 1974–1975.
- Deigan KE, Li TW, Mathews DH, Weeks KM. 2009. Accurate SHAPE-directed RNA structure determination. *Proc Natl Acad Sci* **106**: 97–102.
- Dela-Moss LI, Moss WN, Turner DH. 2014. 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**: 22.
- Ding Y, Chan CY, Lawrence CE. 2005. RNA secondary structure prediction by centroids in a Boltzmann weighted ensemble. *RNA* **11**: 1157–1166.
- Ding Y, Kwok CK, Tang Y, Bevilacqua PC, Assmann SM. 2015. Genome-wide profiling of in vivo RNA structure at single-nucleotide resolution using structure-seq. *Nat Protoc* **10**: 1050–1066.
- Doshi KJ, Cannone JJ, Cobaugh CW, Gutell RR. 2004. Evaluation of the suitability of free-energy minimization using nearest-neighbor energy parameters for RNA secondary structure prediction. *BMC Bioinformatics* **5**: 105.

- Dubois J, Terrier O, Rosa-Calatrava M. 2014. Influenza viruses and mRNA splicing: doing more with less. *MBio* **5**: e00070–14.
- Duszczyc MM, Wutz A, Rybin V, Sattler M. 2011. The Xist RNA A-repeat comprises a novel AUCG tetraloop fold and a platform for multimerization. *RNA* **17**: 1973–1982.
- Fang R, Moss WN, Rutenberg-Schoenberg M, Simon MD. 2015. Probing Xist RNA structure in cells using targeted structure-seq. *PLoS Genet* **11**: e1005668.
- Freyhult E, Gardner PP, Moulton V. 2005. A comparison of RNA folding measures. *BMC Bioinformatics* **6**: 241.
- Glickman JN, Howe JG, Steitz JA. 1988. Structural analyses of EBER1 and EBER2 ribonucleoprotein particles present in Epstein-Barr virus-infected cells. *J Virol* **62**: 902–911.
- Gruber AR, Findeiß S, Washietl S, Hofacker IL, Stadler PF. 2010. RNAz 2.0: improved noncoding RNA detection. *Pac Symp Biocomput* **2010**: 69–79.
- Gulyaev AP, Heus HA, Olsthoorn RC. 2007. An RNA conformational shift in recent H5N1 influenza A viruses. *Bioinformatics* **23**: 272–276.
- Gulyaev AP, Spronken MI, Richard M, Schrauwen EJ, Olsthoorn RC, Fouchier RA. 2016. Subtype-specific structural constraints in the evolution of influenza A virus hemagglutinin genes. *Sci Rep* **6**: 38892.
- Halvorsen M, Martin JS, Broadaway S, Laederach A. 2010. Disease-associated mutations that alter the RNA structural ensemble. *PLoS Genet* **6**: e1001074.
- Hofacker IL. 2003. Vienna RNA secondary structure server. *Nucleic Acids Res* **31**: 3429–3431.
- Huarte M. 2015. The emerging role of lncRNAs in cancer. *Nat Med* **21**: 1253–1261.
- Iwakiri D. 2016. Multifunctional non-coding Epstein-Barr virus encoded RNAs (EBERs) contribute to viral pathogenesis. *Virus Res* **212**: 30–38.
- Jiang T, Nogales A, Baker SF, Martinez-Sobrido L, Turner DH. 2016. Mutations designed by ensemble defect to misfold conserved RNA structures of influenza A segments 7 and 8 affect splicing and attenuate viral replication in cell culture. *PLoS One* **11**: e0156906.
- Kowalski MP, Krude T. 2015. Functional roles of non-coding Y RNAs. *Int J Biochem Cell Biol* **66**: 20–29.
- Lee N, Pimienta G, Steitz JA. 2012. AUF1/hnRNP D is a novel protein partner of the EBER1 noncoding RNA of Epstein-Barr virus. *RNA* **18**: 2073–2082.
- Lee N, Moss WN, Yario TA, Steitz JA. 2015. EBV noncoding RNA binds nascent RNA to drive host PAX5 to viral DNA. *Cell* **160**: 607–618.
- Lee N, Yario TA, Gao JS, Steitz JA. 2016. EBV noncoding RNA EBER2 interacts with host RNA-binding proteins to regulate viral gene expression. *Proc Natl Acad Sci* **113**: 3221–3226.
- Lerner MR, Andrews NC, Miller G, Steitz JA. 1981. Two small RNAs encoded by Epstein-Barr virus and complexed with protein are precipitated by antibodies from patients with systemic lupus erythematosus. *Proc Natl Acad Sci* **78**: 805–809.
- Liu B, Mathews DH, Turner DH. 2010. RNA pseudoknots: folding and finding. *F1000 Biol Rep* **2**: 8.
- Lorenz R, Bernhart SH, Höner Zu Siederdisen C, Tafer H, Flamm C, Stadler PF, Hofacker IL. 2011. ViennaRNA Package 2.0. *Algorithms Mol Biol* **6**: 26.
- Lu Z, Zhang QC, Lee B, Flynn RA, Smith MA, Robinson JT, Davidovich C, Gooding AR, Goodrich KJ, Mattick JS, et al. 2016. RNA duplex map in living cells reveals higher-order transcriptome structure. *Cell* **165**: 1267–1279.
- Maenner S, Blaud M, Fouillen L, Savoye A, Marchand V, Dubois A, Sanglier-Cianféran S, Van Dorsselaer A, Clerc P, Avner P, et al. 2010. 2-D structure of the A region of Xist RNA and its implication for PRC2 association. *PLoS Biol* **8**: e1000276.
- Martin JS. 2014. Describing the structural diversity within an RNA's ensemble. *Entropy* **16**: 1331–1348.
- Martin KC, Ephrussi A. 2009. mRNA localization: gene expression in the spatial dimension. *Cell* **136**: 719–730.
- Mathews DH. 2004. Using an RNA secondary structure partition function to determine confidence in base pairs predicted by free energy minimization. *RNA* **10**: 1178–1190.
- Mathews DH, Moss WN, Turner DH. 2010. Folding and finding RNA secondary structure. *Cold Spring Harb Perspect Biol* **2**: a003665.
- McCaskill JS. 1990. The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers* **29**: 1105–1119.
- Mignone F, Gissi C, Liuni S, Pesole G. 2002. Untranslated regions of mRNAs. *Genome Biol* **3**: reviews0004.
- Mignone F, Grillo G, Licciulli F, Iacono M, Liuni S, Kersey PJ, Duarte J, Saccone C, Pesole G. 2005. UTRdb and UTRsite: a collection of sequences and regulatory motifs of the untranslated regions of eukaryotic mRNAs. *Nucleic Acids Res* **33**: D141–D146.
- Moss WN, Priore SF, Turner DH. 2011. Identification of potential conserved RNA secondary structure throughout influenza A coding regions. *RNA* **17**: 991–1011.
- Moss WN, Dela-Moss LI, Kierzek E, Kierzek R, Priore SF, Turner DH. 2012a. The 3' splice site of influenza A segment 7 mRNA can exist in two conformations: a pseudoknot and a hairpin. *PLoS One* **7**: e38323.
- Moss WN, Dela-Moss LI, Priore SF, Turner DH. 2012b. The influenza A segment 7 mRNA 3' splice site pseudoknot/hairpin family. *RNA Biol* **9**: 1305–1310.
- Moss WN, Lee N, Pimienta G, Steitz JA. 2014. RNA families in Epstein-Barr virus. *RNA Biol* **11**: 10–17.
- Nishikura K. 2016. A-to-I editing of coding and non-coding RNAs by ADARs. *Nat Rev Mol Cell Biol* **17**: 83–96.
- Panchapakesan S, Ferguson ML, Hayden EJ, Chen X, Hoskins AA, Unrau PJ. 2017. Ribonucleoprotein purification and characterization using RNA Mango. *RNA* **23**: 1592–1599.
- Raab-Traub N. 2007. EBV-induced oncogenesis. In *Human herpesviruses: biology, therapy, and immunoprophylaxis* (ed. Arvin A, et al.). Cambridge University Press, Cambridge, UK.
- Rosa MD, Gottlieb E, Lerner MR, Steitz JA. 1981. Striking similarities are exhibited by two small Epstein-Barr virus-encoded ribonucleic acids and the adenovirus-associated ribonucleic acids VAI and VAII. *Mol Cell Biol* **1**: 785–796.
- Siegfried NA, Busan S, Rice GM, Nelson JA, Weeks KM. 2014. RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). *Nat Methods* **11**: 959–965.
- Soszynska-Jozwiak M, Michalak P, Moss WN, Kierzek R, Kierzek E. 2015. A conserved secondary structural element in the coding region of the influenza A virus nucleoprotein (NP) mRNA is important for the regulation of viral proliferation. *PLoS One* **10**: e0141132.
- Stockley PG, Stonehouse NJ, Murray JB, Goodman ST, Talbot SJ, Adams CJ, Liljas L, Valegård K. 1995. Probing sequence-specific RNA recognition by the bacteriophage MS2 coat protein. *Nucleic Acids Res* **23**: 2512–2518.
- van Gelder CW, Thijsen JP, Klaassen EC, Sturchler C, Krol A, van Venrooij WJ, Pruijn GJ. 1994. Common structural features of the Ro RNP associated hY1 and hY5 RNAs. *Nucleic Acids Res* **22**: 2498–2506.
- Wan Y, Kertesz M, Spitale RC, Segal E, Chang HY. 2011. Understanding the transcriptome through RNA structure. *Nat Rev Genet* **12**: 641–655.
- Warf MB, Berglund JA. 2010. Role of RNA structure in regulating pre-mRNA splicing. *Trends Biochem Sci* **35**: 169–178.
- Washietl S, Hofacker IL, Stadler PF. 2005. Fast and reliable prediction of noncoding RNAs. *Proc Natl Acad Sci* **102**: 2454–2459.
- Wilusz JE, Freier SM, Spector DL. 2008. 3' end processing of a long nuclear-retained noncoding RNA yields a tRNA-like cytoplasmic RNA. *Cell* **135**: 919–932.
- Yamamoto N, Takizawa T, Iwanaga Y, Shimizu N, Yamamoto N. 2000. Malignant transformation of B lymphoma cell line BJAB by Epstein-Barr virus-encoded small RNAs. *FEBS Lett* **484**: 153–158.
- Zubradt M, Gupta P, Persad S, Lambowitz AM, Weissman JS, Rouskin S. 2017. DMS-MaPseq for genome-wide or targeted RNA structure probing in vivo. *Nat Methods* **14**: 75–82.
- Zuker M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31**: 3406–3415.