Supplementary Data for A Pipeline for Computational Design of Novel RNA-like Topologies

Swati Jain[1], Alain Laederach[2], Silvia B. V. Ramos[3], and Tamar Schlick[1,4,5,*]

[1]Department of Chemistry, New York University, New York, NY, USA
[2]Department of Biology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA
[3]Department of Biochemistry and Biophysics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA
[4]Courant Institute of Mathematical Sciences, New York University, NY, USA
[5]NYU-ECNU Center for Computational Chemistry at NYU Shanghai, Shanghai, China
[*]Corresponding author: schlick@nyu.edu

S1 Complete details of Materials and Methods

S1.1 RAG 2D and 3D tree graphs

The RAG approach represents any 2D RNA structure as a planar, undirected, connected *tree* graph [1]. Each vertex of the graph corresponds to a junction, a bulge or internal loop (with more than one nucleotide in each strand), a hairpin loop, or the $5'$ and $3'$ ends of the RNA sequence. Each edge in the graph corresponds to a helix (double-stranded stem) connecting the two loop vertices. This graph is a 2D representation of connectivity of the 2D structural elements of an RNA molecule (Figure 2b in the main paper).

The 2D tree graphs described above do not account for number of nucleotides in the loops nor the number of base pairs in the helices. Thus, many RNAs with the same loop connectivity but different number of nucleotides are associated with the same planar tree graph. To incorporate size and build 3D objects for our sampling approach RAGTOP [2], we convert this 2D graph into a 3D graph (Figure 2c in the main paper). The 3D graph includes additional vertices for the two ends of each helix, and for single-nucleotide internal loops and bulges that were omitted in the 2D tree representation. Loop vertices are connected by edges to the proximal-end helical vertices, and each helical vertex is also connected to its corresponding conjugate vertex representing the other end of the same helix. The lengths of the graph edges are scaled according to the number of residues in the loop strands or the number of base pairs in helices [2]. A 3D tree graph can also be constructed from a given RNA 3D structure (Figure 2d in the main paper) by using the coordinates of the C1$'$ sugar atom, C6 pyrimidine atom, and C8 purine atom, as specified in [2]. Building 3D graphs from 3D structures allows us to compare predicted graphs to graphs derived from known structures, and score 3D atomic models using our statistical potential.

S1.2 Classification of RNA-like and non RNA-like graph topologies

Representing RNA 2D structures in the form of tree graphs allows us to study and analyze RNA structure using methods of graph theory. We use graph enumeration methods to generate possible tree graphs for a given number of vertices [3, 4]. The topology of an RNA tree graph is represented by a Laplacian matrix and its corresponding eigenvalues. The second-smallest eigenvalue, λ_2, measures the overall compactness of the graph [5]. We use λ_2 to label each graph topology uniquely. The RAG database currently characterizes different tree graph topologies up to 13 vertices [6]. Tree graph topologies are given RAG IDs in increasing order of λ_2. Some of these generated topologies correspond to known RNA 2D and 3D structures, whereas others are hypothetical graph topologies.

Based on the features extracted from the Laplacian spectrum, we applied the clustering algorithm PAM (Partitioning Around Medoids) to cluster all RAG topologies. The graphs associated with known RNA structures are classified as "existing RNA". The remaining, hypothetical graphs are classified as "RNA-like", or "non RNA-like" as trained by known RNAs [4]. A recent assessment showed that this classification has merit (based on the percentage of predicted RNA-like and non RNA-like topologies that were subsequently solved experimentally), but of course is not perfect [6]. Figure 3 in the main paper shows a sample of topologies from the RAG resource, classified as existing (red), RNA-like (blue), and non RNA-like (black).

S1.3 Graph partitioning and RAG-3D database

To study the 2D structure submotifs of an RNA structure, we partition the 2D tree graph into subgraphs [7]. Graph partitioning algorithms rely on the second eigenvector of the Laplacian matrix L to partition the graph into topologically distinct subgraphs. To maintain junction connectivity, a junction vertex and the vertices connected to it are kept together (*i.e.*, a junction is not broken apart). Graph partitioning extracts all possible subgraphs for a given 2D tree graph, and for each subgraph the corresponding RAG ID is identified. If the 3D tree graph and the tertiary structure is available for a given 2D graph, then the corresponding 3D tree graph and the atomic fragment can also be obtained for each 2D tree subgraph. Figure 4 in the main paper shows the partitioning of the structure of the TPP riboswitch (PDB ID: 3D2G), and its various subgraphs and corresponding atomic fragments.

Graph partitioning was applied to ≈1500 representative RNA structures (obtained from the PDB as of March 2014) to create a database of RNA structures and substructures called *RAG-3D* [8]. This database can be used to search for matching subgraphs and substructures for any given RNA structure. The RAG-3D database catalogs atomic fragments associated with 51 different RAG IDs which we use in our fragment assembly procedure to design sequences for RNA-like topologies, as described next.

S1.4 Fragment assembly procedure for design

Essentially, we partition the RNA-like graph topologies into subgraphs, and then obtain atomic fragments for each of these subgraphs from the RAG-3D database. Next, we piece together subgraphs by fragment assembly of corresponding atomic fragments to build 3D models and predict sequences to fold onto the given target. We have recently reported development of this fragment assembly based approach, *F-RAG*, to build atomic models for candidate RNA 3D tree graph topologies for RNA structure prediction [9]. We use a modified F-RAG procedure in this paper for designing sequences for target RNA graph topologies. Our automated procedure requires as input the following pieces (see Figure 1 in the main paper):

1. Target graph topology for design, with the order of the vertices specified in the 5′ to the 3′ direction (e.g., 8_9 graph in Figure 1 in the main paper).

2. Adjacency matrix A corresponding to the target graph. The adjacency matrix (with vertices listed from the 5′ to the 3′ end) is defined as an n x n matrix (n = number of vertices), where a 1 is placed in each (i, j) element if vertices i and j are connected to one another (see Figure 1 in the main paper for an example).

3. Number of subgraphs of the target graph, along with their RAG IDs and the vertices that belong to each subgraph. Two subgraph are shown for the 8_9 target in Figure 1 in the main paper.

4. List of RNA substructures corresponding to each of the target's subgraph RAG IDs (in the RAG-3D database), along with their 2D structure, 3D graph, and atomic fragments. As Figure 1 in the main paper shows, tRNA-Phe is an example of an atomic fragment corresponding to the 5_3 subgraph of the 8_9 target.

5. Loop number, along with a given 2D structure, 3D graph, and atomic fragment of a specific motif, if an internal loop or a hairpin needs to be restricted to the given motif.

The design algorithm employs a recursive procedure for each subgraph of the target graph (starting from the 5′ direction) to generate atomic coordinates for that subgraph. For each subgraph, the number and connectivity of RNA loops (hairpin loops, internal loops, and junctions) from the 5′ to the 3′ direction is compared to the number and connectivity of loops in each atomic fragment, and any mismatched fragments are eliminated. The remaining fragments for each subgraph are superimposed on the partially built atomic model from previous subgraphs (as part of the recursive procedure) using the residues in the common helix connecting the two subgraphs. For specified motif design, the atomic fragment for the specified motif is used for that loop. The number and identity of the bases in the atomic fragments are left unchanged. Unpaired residues at the 5′ or 3′ ends of the sequence are removed. Once atomic coordinates are generated for each subgraph, a 3D tree graph is calculated corresponding to the atomic model (as described in subsection titled "RAG 2D and 3D tree graphs"). This 3D graph is scored using our knowledge-based statistical potential (initially developed for RNA structure prediction): this potential contains terms for bend and torsion angles for internal loops and radius of gyration, and was recently updated to account for special k-turn geometries [10]. Atomic models with corresponding sequences, 2D structures, 3D tree graph, and scores are produced as output of the fragment assembly.

The fragment assembly we use for design is similar to the F-RAG procedure developed for RNA structure prediction, as detailed in [9], except for these differences: the input is the target RAG topology along with corresponding adjacency matrix (instead of a specific sequence and 2D structure); 3D tree graph vertices are not used for docking fragments; and the number and identity of fragment residues are not changed (to provide flexibility in the design process).

S1.5 Selecting candidate sequences for target graph

The above fragment assembly procedure is applied for different orientations of the target graph. Each orientation represents a different 5′ to 3′ order of the RNA loops for the same 2D target graph topology (see Figure S1 for different orientations used for 6 RNA-like motif objectives in this paper). Fragment assembly produces a large number of atomic models and sequences, as there are multiple candidates

available for each RAG ID in our RAG-3D database. To sort through the large number and identify candidate sequences quickly, we combine resulting atomic models and sequences from every orientation of the target graph and order them by increasing score. Any model with large chain breaks is removed, and the top 200 models with unique sequences are retained for further analysis. These top 200 sequences are clustered based on the type of RNA origin of atomic fragments.

To further narrow the pool of candidate sequences, we subject the top 200 unique sequences selected above to RNAfold (available with the Vienna RNA Package 2.3.3) [11] and NUPACK [12, 13, 14] for *in silico* 2D structure prediction, with default parameters. That is, for each sequence, we identify the 2D graph topologies of the minimum energy and the centroid 2D structure produced by RNAfold, and the minimum energy structure produced by NUPACK. We consider a design successful if our sequence folds onto the same RAG topology with both RNAfold (either the minimum energy or the centroid structure) and NUPACK, regardless of whether this was the intended fold or not (since another fold might be produced). Note that the 2D structures predicted by RNAfold and NUPACK can be different, as we are concerned with the RAG topology and not the specifics of the 2D structure. For all our top 200 candidate sequences, we classify the number of sequences that fold onto different RAG topologies, and select the sequences that are predicted to fold onto the target graph for further study.

S1.6 Mutation analysis for robustness and increased yield

As an additional step to test the robustness of the successfully designed sequences (by the criteria above) that fold onto the target graph, we subject them to manual mutations. Specifically, we use EteRNA's [15] puzzle-maker interface to perform mutations on our designed sequences. EteRNA uses the same software for 2D structure prediction (RNAfold and NUPACK) as done here, with a useful graphical interface and real time feedback. To test the robustness of our candidate sequences, we mutate helical regions to disrupt the 2D structure (e.g., changing a G of a G-C base pair to an A) and then follow with a compensatory (but different) mutation (changing the C to U to get back a base pair) to revert to the same 2D structure. We select the top sequences for a target RNA-like topology based on their score, most productive clusters, and the above robustness test.

We also subject sequences that do not fold onto the target topology to manual mutation. This was performed as a proof of concept to determine whether systematic mutations might be advantageous by identifying residues using tree graphs differences. To identify target residues for mutations, we compare the 2D tree graphs of our designed 2D structure and the 2D tree graphs of the 2D structures computed by EteRNA. This allows us to identify specific helices and loops to be added or eliminated. We perform those mutations one at a time in the EteRNA graphical interface. This process is repeated until the EteRNA predicted 2D tree graph matches the target RAG topology (see Figure S2 for an example). Note that we aim to design a specified tree graph topology and not a 2D structure; therefore variations in 2D structure are allowed as long as the tree graph topology matches the target topology.

S1.7 Experimental structure probing via SHAPE

As a proof of principle, we used SHAPE-MaP (Selective 2′-hydroxyl acylation analyzed by primer extension) with next-generation sequencing [16] to probe the structures of two candidate design sequences, as marked in Figure 6 in the main paper with black stars, namely 1b and 1d. The synthetic RNAs were transcribed using a T7 RNA polymerase system using as a template synthetic DNAs linked to a T7 promoter sequence

similarly as in [17], except that a T7 high yield RNA kit (New England Biolabs) was used. The 7_4 initial design sequence 1b (5′-TAATACGACTCACTATAGGCCTTCGGGCCAAGGTCCCGCGTACAAGACGCG GTCGATAGACGACATATATACGCGTGGATATGGCACGCGAGTTTCTACCGGGCACCGTAAATGT CCGACTATGTCGCACTAACAGACCTCGATCCGGTTCGCCGGATCCAAATCGGGCTTCGGTCCG GTTC-3′) and 7_4 k-turn design sequence 1d (5′-TAATACGACTCACTATAGGCCTTCGGGCCAAGGT CCCGCGTACAAGACGCGGTCGATAGGGAGGACATATATACGCGTGGATATGGCACGCGAGTTT CTACCGGGCACCGTAAATGTCCGATTATGTCCGCGAAACCTAACAGACCTCGATCCGGTTCGC CGGATCCAAATCGGGCTTCGGTCCGGTTC-3′) containing the hairpin adapters for SHAPE as described in [18] were synthesized (IDT) and used as a template for the *in vitro* transcription reaction. The obtained RNAs were then separated from the free nucleotides using Sephadex G50 columns (GE Healthcare Life Science) followed by a second cleanup step using MEGAClear (Thermo Fisher) columns. A minimum of 2 pmol of RNA was used for each modification reaction as described in [16] and [19] with nominal modifications. Briefly, the RNA was diluted in water, denatured at 95°C for 1 min, and snap cooled on ice. After the addition of folding buffer (100 mM KCl, 10 mM MgCl2, 100 mM HEPES, pH 8.0 final concentration), the RNA was incubated for 30 minutes at 37°C. The folded-RNA was either mixed with DMSO (negative control) or 2-methylnicotinic acid imidazolide (NAI) [20] diluted in DMSO to a final concentration of 25 mM (modified sample), both of them were incubated for 10 min at 37°C to allow full reaction.

The RNA was then reverse transcribed using SuperScript II (Life Technologies) with universal structure cassette cDNA primers [18] under error-prone conditions for all samples as previously described [21, 16], followed by cleanup with a G50 column. Then a secondary PCR was performed to add TruSeq barcodes sequences to each sample. All samples were purified using Ampure XP beads and the DNA concentrations were determined via Qubit fluorometric quantitation. The final quality of the libraries was evaluated using a Bioanalyzer. The libraries were then run on a MiSeq (Illumina) as paired end, 2x150 read multiplex run. The obtained sequences were then used to evaluate mutations using SHAPEMapper pipeline and obtain SHAPE data [21, 22]. The structures were folded based on normalized SHAPE reactivities using RNAfold (`http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi`) [23] and RNAstructure (`https://rna.urmc.rochester.edu/RNAstructure.html`) [24], both with default parameters for the pseudo-free energy term, as described in [25]. Structures were visualized in VARNA [26].

To further verify our results, we also predicted the 2D structures for sequences 1b and 1d with shuffled variations of the real SHAPE data. We used the statistical computing package R (version 3.1.1) [27] to generate 100 different shuffled variations of the SHAPE data for each of the two sequences (using the function sample() that can generate random reordering of the data). We then predict the 2D structures using the shuffled SHAPE data with RNAfold (available with the Vienna RNA Package 2.3.3), with pseudo-energy parameters as described in [25], and implemented in RNAfold as described in [23]. We identified the RAG topologies of the predicted minimum energy 2D structures, and analyzed the different topologies obtained. The results are shown in Figures S9 and S10 for 7_4 initial design sequence 1b and 7_4 k-turn design sequence 1d, respectively.

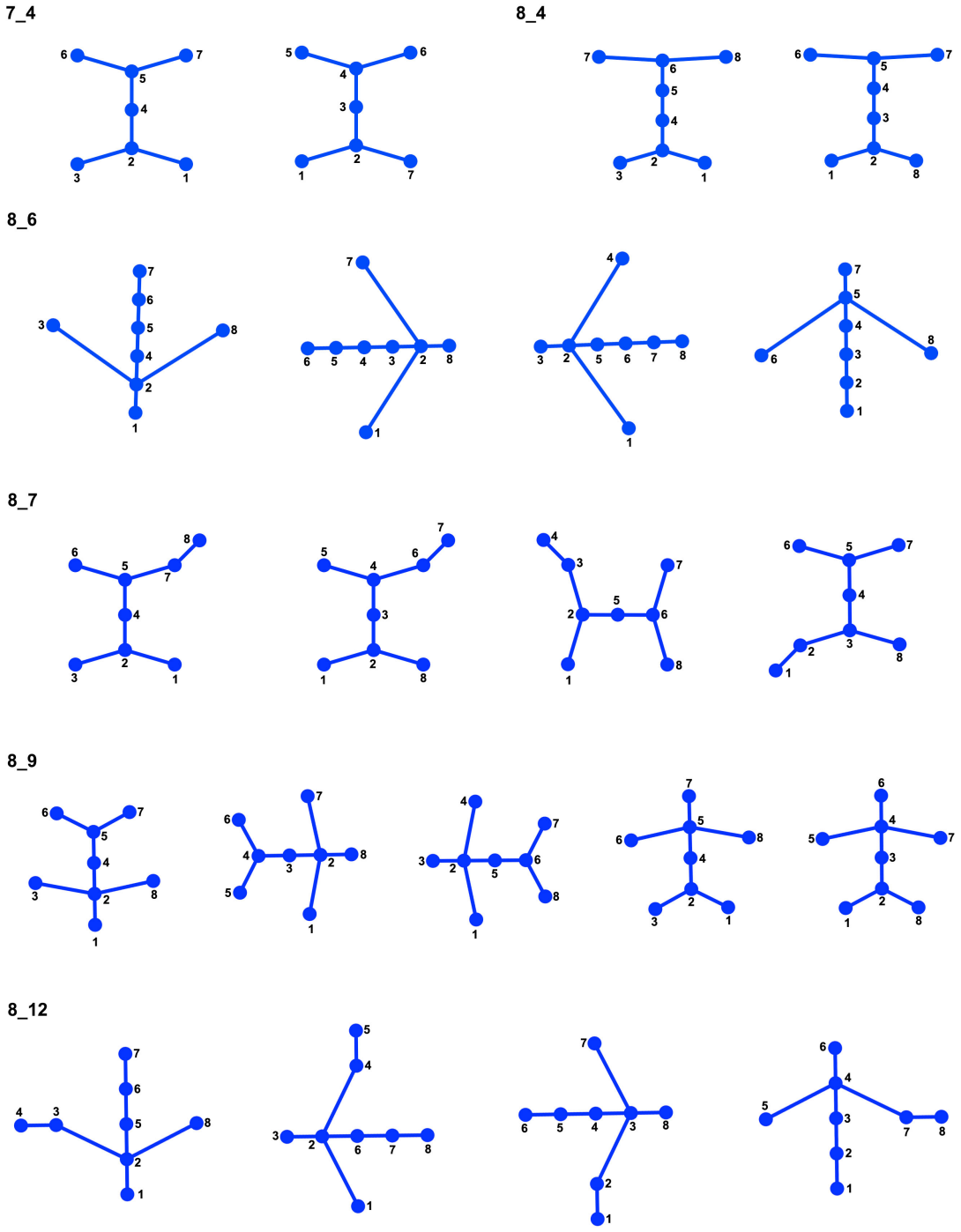


Figure S1: **Six target RNA-like topologies and their different orientations.** Each orientation represents a different order of loops from the 5′ to the 3′ direction, with the vertices numbered from the 5′ to the 3′ end.

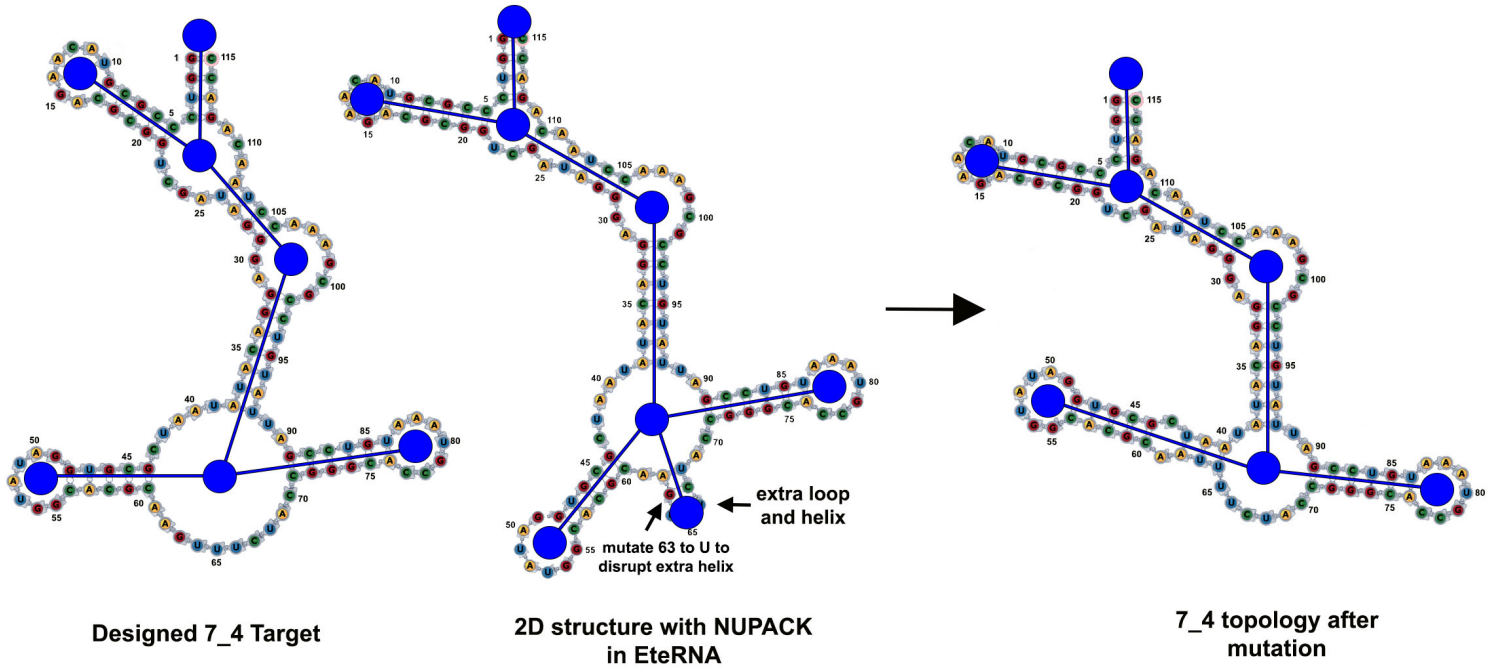


Figure S2: **Example of manual mutation with EteRNA.** An example of a designed sequence for RNA-like motif 7_4 is shown. The 2D structure of the designed sequence with NUPACK in EteRNA has 8_9 RAG topology with an extra helix and hairpin loop. Mutating residue 64 → U disrupts the extra helix.

Table S1: Clusters obtained for the top 200 sequences in the initial non-specific and k-turn design for the 7_4 RNA-like motif. Number of sequences that fold onto the same RAG topologies ("Same") by both RNAfold and NUPACK are shown, along with the sequences that fold onto the target motif 7_4 (shown in bold), and other RAG topologies (with at least 5 sequences) obtained in the design. Only clusters that have non-zero "Same" sequences in either the initial non-specific or the k-turn design are shown.

Cluster	Fragment 1 (4_2)	Fragment 2 (4_2)	Initial Design			K-turn Design				
			Same	**7_4**	8_9	Same	**7_4**	8_9	8_7	6_4
1	ribosomal RNA	Purine riboswitch	10	4	0	14	5	0	3	0
2	ribosomal RNA	transfer RNA	8	3	2	18	5	5	2	1
3	transfer RNA	transfer RNA	6	0	0	1	0	0	0	0
4	Glycine riboswitch	transfer RNA	5	1	4	2	0	1	0	0
5	Glycine riboswitch	Tetracycline aptamer	4	4	0	-	-	-	-	-
6	Glycine riboswitch	Purine riboswitch	3	3	0	-	-	-	-	-
7	ribosomal RNA	Hammerhead ribozyme	3	1	0	8	3	0	1	1
8	SAM-III riboswitch	Purine riboswitch	3	1	0	5	0	0	0	4
9	M-box RNA	transfer RNA	3	1	2	2	0	2	0	0
10	Glycine riboswitch	Hammerhead ribozyme	2	2	0	1	1	0	0	0
11	M-box RNA	Purine riboswitch	2	1	0	3	2	0	1	0
12	SAM-III riboswitch	transfer RNA	2	0	1	1	0	0	0	0
13	ribosomal RNA	P4-P6 ribozyme	1	1	0	1	0	0	0	0
14	ribosomal RNA	Tetracycline aptamer	1	1	0	-	-	-	-	-
15	M-box RNA	Tetracycline aptamer	1	1	0	-	-	-	-	-
16	transfer RNA	Purine riboswitch	1	0	0	2	0	0	0	0
17	ribosomal RNA	ribosomal RNA	1	0	0	1	0	0	0	0
18	transfer RNA	HDV ribozyme	1	0	0	-	-	-	-	-
19	transfer RNA	P4-P6 ribozyme	1	0	0	-	-	-	-	-
20	Prohead RNA	Purine riboswitch	-	-	-	3	2	0	1	0
21	Prohead RNA	transfer RNA	0	0	0	2	1	1	0	0
22	SAM-III riboswitch	Hammerhead ribozyme	0	0	0	1	0	0	0	1
23	transfer RNA	Hammerhead ribozyme	-	-	-	1	0	0	0	0

Table S2: Clusters obtained for the top 200 sequences in the initial non-specific and k-turn design for the 8_4 RNA-like motif. Number of sequences that fold onto the same RAG topologies ("Same") by both RNAfold and NUPACK are shown, along with the sequences that fold onto the target motif 8_4 (shown in bold), and other RAG topologies (with at least 5 sequences) obtained in the design. Only clusters that have non-zero "Same" sequences in either the initial non-specific or the k-turn design are shown.

Cluster	Fragment 1 (5_2)	Fragment 2 (4_2)	Initial Design				K-turn Design				
			Same	**8_4**	9_9	9_7	Same	**8_4**	9_9	9_7	9_4
1	Glycine riboswitch	transfer RNA	15	2	9	0	25	10	9	0	5
2	ribosomal RNA	transfer RNA	5	0	0	2	8	3	1	1	0
3	ribosomal RNA	Purine riboswitch	5	0	0	4	6	3	0	2	0
4	Glycine riboswitch	Purine riboswitch	3	3	0	0	25	21	0	3	0
5	Glycine riboswitch	Tetracycline aptamer	3	3	0	0	2	2	0	0	0
6	Glycine riboswitch	Hammerhead ribozyme	2	0	0	0	8	8	0	0	0
7	Glycine riboswitch	P4-P6 ribozyme	1	0	0	0	3	3	0	0	0
8	ribosomal RNA	Hammerhead ribozyme	1	0	0	1	2	2	0	0	0
9	Glycine riboswitch	SRP	1	0	0	0	-	-	-	-	-
10	ribosomal RNA	Tetracycline aptamer	1	0	0	1	-	-	-	-	-
11	Glycine riboswitch	ribosomal RNA	0	0	0	0	5	0	0	0	0
12	ribosomal RNA	P4-P6 ribozyme	-	-	-	-	1	1	0	0	0

Table S3: Clusters obtained for the top 200 sequences in the initial non-specific and k-turn design for the 8_6 RNA-like motif. Number of sequences that fold onto the same RAG topologies ("Same") by both RNAfold and NUPACK are shown, along with the sequences that fold onto the target motif 8_6 (shown in bold), and other RAG topologies (with at least 5 sequences) obtained in the design. Only clusters that have non-zero "Same" sequences in either the initial non-specific or the k-turn design are shown.

Cluster	Fragment 1 (4_1)	Fragment 2 (5_3)	Initial Design			K-turn Design				
			Same	**8_6**	6_5	Same	**8_6**	8_1	7_1	7_5
1	Mytonic dystrophy RNA	transfer RNA	23	11	2	11	7	2	0	0
2	SAM-I riboswitch	transfer RNA	20	0	14	7	0	0	1	4
3	T-box riboswitch	transfer RNA	11	3	0	7	1	4	0	0
4	ribosomal RNA	transfer RNA	5	1	0	14	7	1	2	0
5	M-box riboswitch	transfer RNA	3	0	0	10	6	1	0	1
6	SARS genome	transfer RNA	1	0	0	8	3	0	1	0
7	SRP	transfer RNA	-	-	-	11	11	0	0	0
8	Group I and II intron	transfer RNA	-	-	-	8	1	2	1	0

Table S4: Clusters obtained for the top 200 sequences in the initial non-specific and k-turn design for the 8_7 RNA-like motif. Number of sequences that fold onto the same RAG topologies ("Same") by both RNAfold and NUPACK are shown, along with the sequences that fold onto the target motif 8_7 (shown in bold), and other RAG topologies (with at least 5 sequences) obtained in the design. Only clusters that have non-zero "Same" sequences in either the initial non-specific or the k-turn design are shown.

Cluster	Fragment 1 (5_2)	Fragment 2 (4_2)	Initial Design				K-turn Design			
			Same	**8_7**	6_2	7_3	Same	**8_7**	6_2	10_14
1	TPP riboswitch	Purine riboswitch	24	7	13	3	9	0	7	0
2	TPP riboswitch	transfer RNA	10	0	3	2	8	0	2	0
3	ribosomal RNA	transfer RNA	7	1	0	0	9	0	0	2
4	TPP riboswitch	Hammerhead ribozyme	3	1	2	0	2	0	2	0
5	TPP riboswitch	P4-P6 ribozyme	2	1	1	0	1	0	1	0
6	TPP riboswitch	Tetracycline aptamer	1	1	0	0	-	-	-	-
7	TPP riboswitch	SRP	1	0	1	0	1	0	1	0
8	ribosomal RNA	SRP	1	0	0	0	1	0	0	0
9	TPP riboswitch	ribosomal RNA	1	0	0	0	1	0	0	0
10	SRP	transfer RNA	1	0	0	0	0	0	0	0
11	SRP	ribosomal RNA	1	0	0	0	0	0	0	0
12	SRP	Purine riboswitch	1	0	0	0	0	0	0	0
13	ribosomal RNA	P4-P6 ribozyme	1	0	0	0	0	0	0	0
14	ribosomal RNA	Purine riboswitch	0	0	0	0	4	0	0	3
15	ribosomal RNA	Hammerhead ribozyme	-	-	-	-	2	0	0	2

Table S5: Clusters obtained for the top 200 sequences in the initial non-specific and k-turn design for the 8_9 RNA-like motif. Number of sequences that fold onto the same RAG topologies ("Same") by both RNAfold and NUPACK are shown, along with the sequences that fold onto the target motif 8_9 (shown in bold), and other RAG topologies (with at least 5 sequences) obtained in the design. Only clusters that have non-zero "Same" sequences in either the initial non-specific or the k-turn design are shown.

Cluster	Fragment 1 (4_2)	Fragment 2 (5_3)	Initial Design		K-turn Design		
			Same	**8_9**	Same	**8_9**	7_9
1	ribosomal RNA	transfer RNA	19	9	39	14	2
2	transfer RNA	transfer RNA	10	0	1	0	0
3	M-box riboswitch	transfer RNA	7	4	2	0	0
4	SAM riboswitch	transfer RNA	6	2	8	0	5
5	Glycine riboswitch	transfer RNA	3	0	0	0	0
6	Prohead RNA	transfer RNA	2	0	5	2	0
7	SAM riboswitch	ribosomal RNA	1	0	1	0	0
8	Purine riboswitch	transfer RNA	-	-	2	1	0
9	ribosomal RNA	ribosomal RNA	-	-	1	0	0

Table S6: Clusters obtained for the top 200 sequences in the initial non-specific and k-turn design for the 8_12 RNA-like motif. Number of sequences that fold onto the same RAG topologies ("Same") by both RNAfold and NUPACK are shown, along with the sequences that fold onto the target motif 8_12 (shown in bold), and other RAG topologies (with at least 5 sequences) obtained in the design. Only clusters that have non-zero "Same" sequences in either the initial non-specific or the k-turn design are shown.

Cluster	Fragment 1 (6_5)	Fragment 2 (3_1)	Initial Design			K-turn Design		
			Same	**8_12**	7_7	Same	**8_12**	7_7
1	SAM I riboswitch	ribosomal RNA	6	0	1	13	2	4
2	SAM I riboswitch	Group I and II intron	3	0	3	1	1	0
3	SAM I riboswitch	GlmS ribozyme	3	0	3	1	1	0
4	SAM I riboswitch	snRNA	2	0	0	0	0	0
5	SAM I riboswitch	IRES domain	1	0	0	2	0	0
6	SAM I riboswitch	SRP	1	0	0	0	0	0
7	SAM I riboswitch	HIV/Virus RNA	1	0	1	0	0	0
8	SAM I riboswitch	RNA aptamer	0	0	0	1	1	0
9	SAM I riboswitch	T-box leader RNA	0	0	0	1	1	0
10	SAM I riboswitch	transfer RNA	0	0	0	1	1	0
11	SAM I riboswitch	H/ACA box RNA	0	0	0	1	0	1

Table S7: Distribution of sequences resulting from the design of RNA-Like motif 7_4 as predicted by RNAfold and NUPACK only when they coincide. The target RNA-like motif is highlighted in bold.

RAG ID	No. of sequences	
	Initial Design	K-turn Design
7_4	**24**	**19**
8_9	9	9
8_7	4	8
7_2	4	3
6_4	3	7
8_6	3	1
8_2	2	2
8_12	2	-
7_9	1	3
9_4	1	1
9_5	1	1
7_1	1	-
7_7	1	-
8_16	1	-
12_351	1	-
7_6	-	4
8_3	-	2
9_22	-	2
6_2	-	1
8_15	-	1
9_17	-	1
10_27	-	1

Table S8: Distribution of sequences resulting from the design of RNA-Like motif 8_4 as predicted by RNAfold and NUPACK only when they coincide. The target RNA-like motif is highlighted in bold.

RAG ID	No. of sequences	
	Initial Design	K-turn Design
9_9	9	10
8_4	**8**	**53**
9_7	8	6
10_13	2	1
8_22	2	-
9_1	2	-
10_17	1	1
6_2	1	-
8_6	1	-
9_15	1	-
9_26	1	-
10_32	1	-
9_4	-	5
12_375	-	3
8_9	-	2
7_3	-	1
8_1	-	1
8_2	-	1
9_6	-	1

Table S9: Distribution of sequences resulting from the design of RNA-Like motif 8_6 as predicted by RNAfold and NUPACK only when they coincide. The target RNA-like motif is highlighted in bold.

RAG ID	No. of sequences	
	Initial Design	K-turn Design
6_5	16	2
8_6	**15**	**36**
7_1	4	5
5_2	4	-
6_1	3	2
7_3	3	2
8_1	2	10
7_2	2	3
6_3	2	1
7_6	2	1
5_1	2	-
7_7	2	-
8_3	2	-
7_5	1	5
8_8	1	1
9_16	1	1
9_13	1	-
8_10	-	3
9_5	-	2
6_2	-	1
8_11	-	1

Table S10: Distribution of sequences resulting from the design of RNA-Like motif 8_7 as predicted by Mfold and RNAfold only when they coincide. The target RNA-like motif is highlighted in bold.

RAG ID	No. of sequences	
	Initial Design	K-turn Design
6_2	20	13
8_7	**11**	**0**
7_3	5	3
7_5	2	2
9_17	2	1
9_12	2	-
9_14	2	-
10_14	1	7
9_5	1	2
7_2	1	1
12_332	1	1
8_2	1	-
9_30	1	-
9_44	1	-
10_27	1	-
12_169	1	-
13_942	1	-
11_26	-	2
11_34	-	2
6_3	-	1
7_6	-	1
9_32	-	1
10_26	-	1

Table S11: Distribution of sequences resulting from the design of RNA-Like motif 8_9 as predicted by Mfold and RNAfold only when they coincide. The target RNA-like motif is highlighted in bold.

RAG ID	No. of sequences	
	Initial Design	K-turn Design
8_9	**15**	**17**
8_2	3	4
8_1	3	2
7_2	3	1
8_12	3	-
9_5	3	-
9_25	3	-
7_4	2	2
7_5	2	2
8_6	2	2
7_9	1	7
6_1	1	1
6_5	1	1
8_15	1	1
7_1	1	-
7_3	1	-
8_5	1	-
9_1	1	-
9_13	1	-
8_7	-	4
6_3	-	2
8_3	-	2
9_17	-	2
10_51	-	2
6_4	-	1
8_11	-	1
8_16	-	1
10_13	-	1
10_27	-	1
10_30	-	1
11_112	-	1

Table S12: Distribution of sequences resulting from the design of RNA-Like motif 8_12 as predicted by Mfold and RNAfold only when they coincide. The target RNA-like motif is highlighted in bold.

RAG ID	No. of sequences	
	Initial Design	K-turn Design
7_7	8	5
6_4	4	-
6_2	3	-
6_5	2	2
8_12	**-**	**7**
7_2	-	2
7_6	-	2
9_13	-	2
8_2	-	1

Table S13: Sequence identities of one representative sequence from each cluster that fold onto the target topology for the initial non-specific 7_4 design.

	Cluster 1	Cluster 2	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9	Cluster 10	Cluster 11	Cluster 13	Cluster 14	Cluster 15
Cluster 1	100.00	-	-	-	-	-	-	-	-	-	-	-	-
Cluster 2	46.96	100.00	-	-	-	-	-	-	-	-	-	-	-
Cluster 4	48.48	49.22	100.00	-	-	-	-	-	-	-	-	-	-
Cluster 5	47.58	50.43	70.08	100.00	-	-	-	-	-	-	-	-	-
Cluster 6	55.74	51.20	66.17	62.32	100.00	-	-	-	-	-	-	-	-
Cluster 7	46.36	68.75	44.85	43.51	46.88	100.00	-	-	-	-	-	-	-
Cluster 8	73.68	52.21	49.24	50.41	48.85	45.90	100.00	-	-	-	-	-	-
Cluster 9	48.55	50.00	67.86	49.30	48.23	52.31	48.51	100.00	-	-	-	-	-
Cluster 10	45.97	47.50	69.17	70.40	67.18	68.55	45.86	49.64	100.00	-	-	-	-
Cluster 11	69.03	46.88	50.34	47.06	48.59	43.97	69.05	71.43	52.55	100.00	-	-	-
Cluster 13	45.00	43.80	48.78	45.95	44.09	45.69	51.82	45.65	48.31	49.61	100.00	-	-
Cluster 14	45.83	53.33	48.18	67.72	47.73	47.46	49.17	43.24	48.48	41.96	42.86	100.00	-
Cluster 15	48.06	47.24	46.21	67.67	46.38	50.00	48.12	74.07	51.88	71.64	45.31	66.15	100.00

Table S14: Sequence identities of one representative sequence from each cluster that fold onto the target topology for the k-turn specific 7_4 design.

	Cluster 1	Cluster 2	Cluster 7	Cluster 10	Cluster 11	Cluster 20	Cluster 21
Cluster 1	100.00	-	-	-	-	-	-
Cluster 2	46.72	100.00	-	-	-	-	-
Cluster 7	46.46	64.57	100.00	-	-	-	-
Cluster 10	50.00	52.85	66.91	100.00	-	-	-
Cluster 11	74.79	51.49	46.85	52.17	100.00	-	-
Cluster 20	79.13	53.85	48.80	48.51	74.44	100.00	-
Cluster 21	50.00	55.38	50.81	46.15	46.58	70.25	100.00

Table S15: Sequence identities of one representative sequence from each cluster that fold onto the target topology for the initial non-specific 8_4 design.

	Cluster 1	Cluster 4	Cluster 5
Cluster 1	100.00	-	-
Cluster 4	72.22	100.00	-
Cluster 5	74.29	72.46	100.00

Table S16: Sequence identities of one representative sequence from each cluster that fold onto the target topology for the k-turn specific 8_4 design.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 12
Cluster 1	100.00	-	-	-	-	-	-	-	-
Cluster 2	50.34	100.00	-	-	-	-	-	-	-
Cluster 3	48.15	71.43	100.00	-	-	-	-	-	-
Cluster 4	72.79	49.67	65.97	100.00	-	-	-	-	-
Cluster 5	74.47	49.67	47.26	69.93	100.00	-	-	-	-
Cluster 6	73.91	49.33	50.39	67.35	73.38	100.00	-	-	-
Cluster 7	72.39	47.95	51.94	69.85	69.34	72.59	100.00	-	-
Cluster 8	47.55	72.34	67.38	49.29	48.30	70.42	48.92	100.00	-
Cluster 12	48.51	68.84	72.09	49.64	45.99	45.07	65.87	72.09	100.00

Table S17: Sequence identities of one representative sequence from each cluster that fold onto the target topology for the initial non-specific 8_6 design.

	Cluster 1	Cluster 3	Cluster 4
Cluster 1	100.00	-	-
Cluster 3	67.97	100.00	-
Cluster 4	44.83	44.36	100.00

Table S18: Sequence identities of one representative sequence from each cluster that fold onto the target topology for the k-turn specific 8_6 design.

	Cluster 1	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8
Cluster 1	100.00	-	-	-	-	-	-
Cluster 3	53.49	100.00	-	-	-	-	-
Cluster 4	55.26	47.37	100.00	-	-	-	-
Cluster 5	55.75	54.62	52.99	100.00	-	-	-
Cluster 6	57.89	74.60	56.14	59.38	100.00	-	-
Cluster 7	47.95	52.55	48.55	47.62	51.56	100.00	-
Cluster 8	48.92	51.82	50.76	51.20	51.16	72.79	100.00

Table S19: Sequence identities of one representative sequence from each cluster that fold onto the target topology for the initial non-specific 8_7 design.

	Cluster 1	Cluster 3	Cluster 4	Cluster 5	Cluster 6
Cluster 1	100.00	-	-	-	-
Cluster 3	48.98	100.00	-	-	-
Cluster 4	67.81	49.31	100.00	-	-
Cluster 5	69.63	48.00	72.73	100.00	-
Cluster 6	69.72	50.66	73.91	69.12	100.00

Table S20: Sequence identities of one representative sequence from each cluster that fold onto the target topology for the initial non-specific 8_9 design.

	Cluster 1	Cluster 3	Cluster 4
Cluster 1	100.00	-	-
Cluster 3	42.28	100.00	-
Cluster 4	46.40	50.38	100.00

Table S21: Sequence identities of one representative sequence from each cluster that fold onto the target topology for the k-turn specific 8_9 design.

	Cluster 1	Cluster 6	Cluster 8
Cluster 1	100.00	-	-
Cluster 6	58.27	100.00	-
Cluster 8	52.59	50.34	100.00

Table S22: Sequence identities of one representative sequence from each cluster that fold onto the target topology for the k-turn specific 8_12 design.

	Cluster 1	Cluster 2	Cluster 3	Cluster 8	Cluster 9	Cluster 10
Cluster 1	100.00	-	-	-	-	-
Cluster 2	79.41	100.00	-	-	-	-
Cluster 3	85.40	80.30	100.00	-	-	-
Cluster 8	77.21	90.18	77.86	100.00	-	-
Cluster 9	83.09	84.68	83.46	83.87	100.00	-
Cluster 10	81.62	81.89	80.74	81.10	85.16	100.00

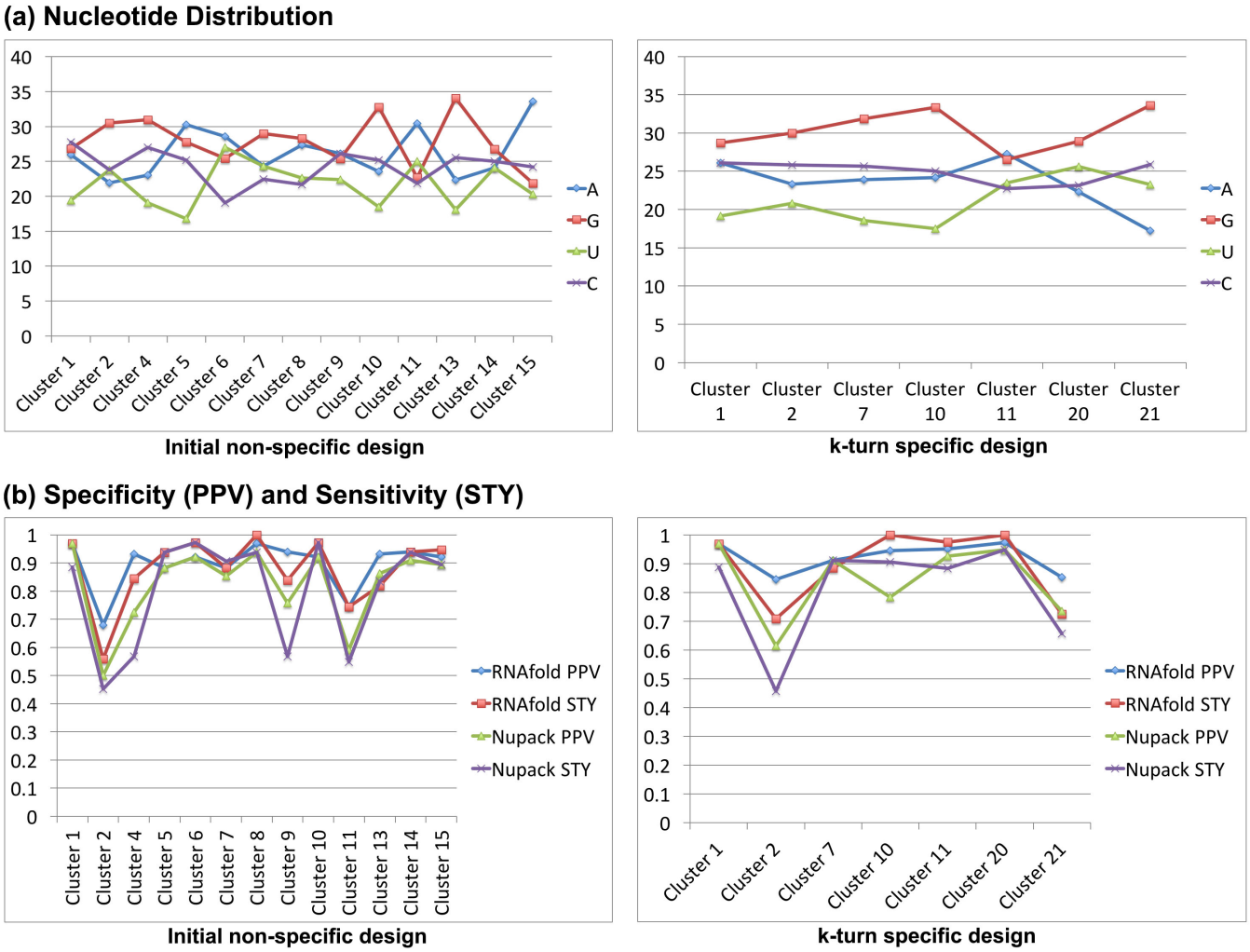


Figure S3: Nucleotide distributions, specificity (PPV), and sensitivity (STY) values between the designed and the predicted 2D structures of one representative sequence from each cluster that fold onto the target topology for the initial non-specific and k-turn specific 7_4 designs. PPV and STY values are calculated for only cannonical base pairs (without pseudoknots). PPV is the percentage of base pairs in the predicted 2D structure that are found in the designed 2D structure; STY is the percentage of base pairs in the designed 2D structure that are found in the predicted 2D structure.

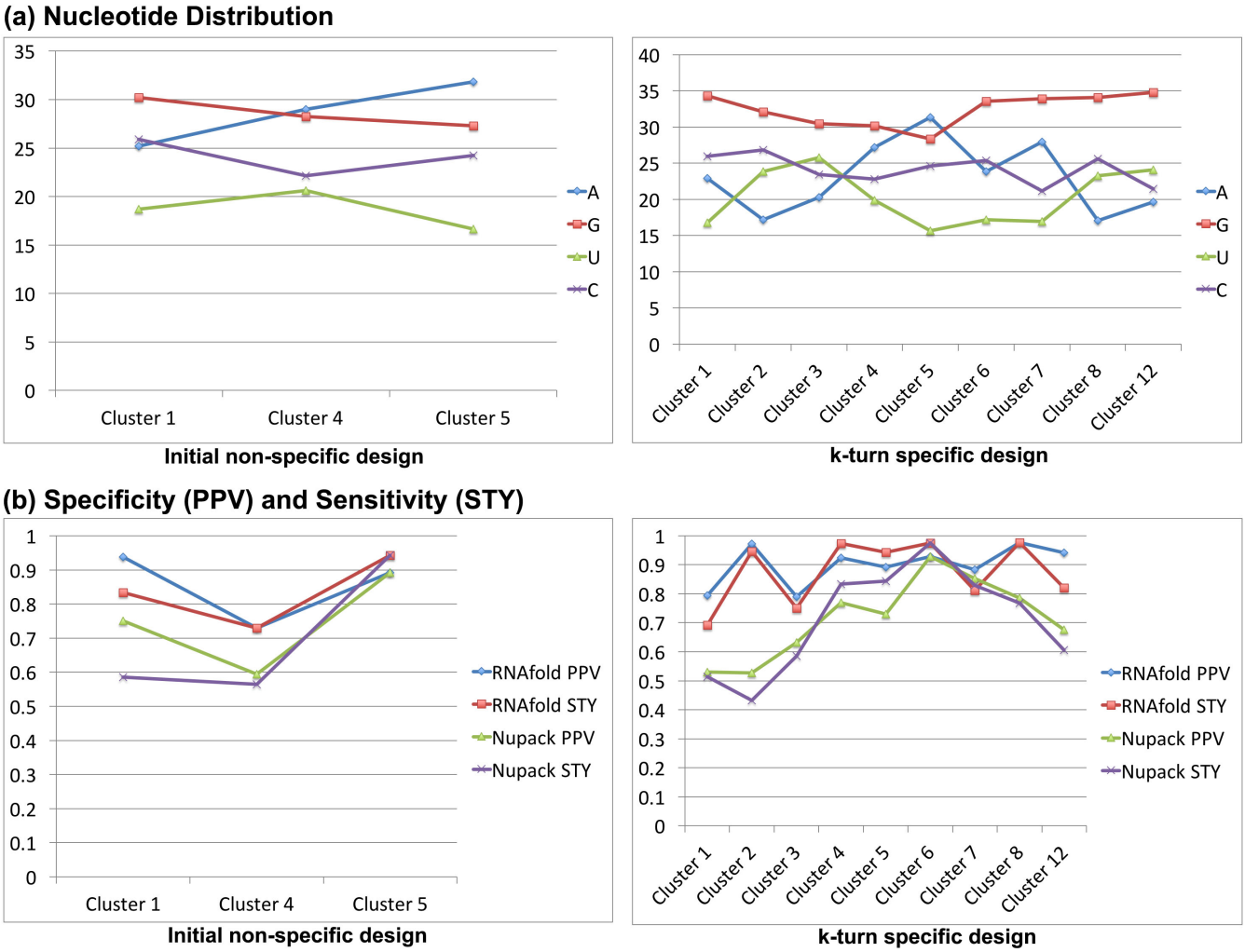


Figure S4: Nucleotide distributions, specificity (PPV), and sensitivity (STY) values between the designed and the predicted 2D structures of one representative sequence from each cluster that fold onto the target topology for the initial non-specific and k-turn specific 8_4 designs. PPV and STY values are calculated for only cannonical base pairs (without pseudoknots). PPV is the percentage of base pairs in the predicted 2D structure that are found in the designed 2D structure; STY is the percentage of base pairs in the designed 2D structure that are found in the predicted 2D structure.

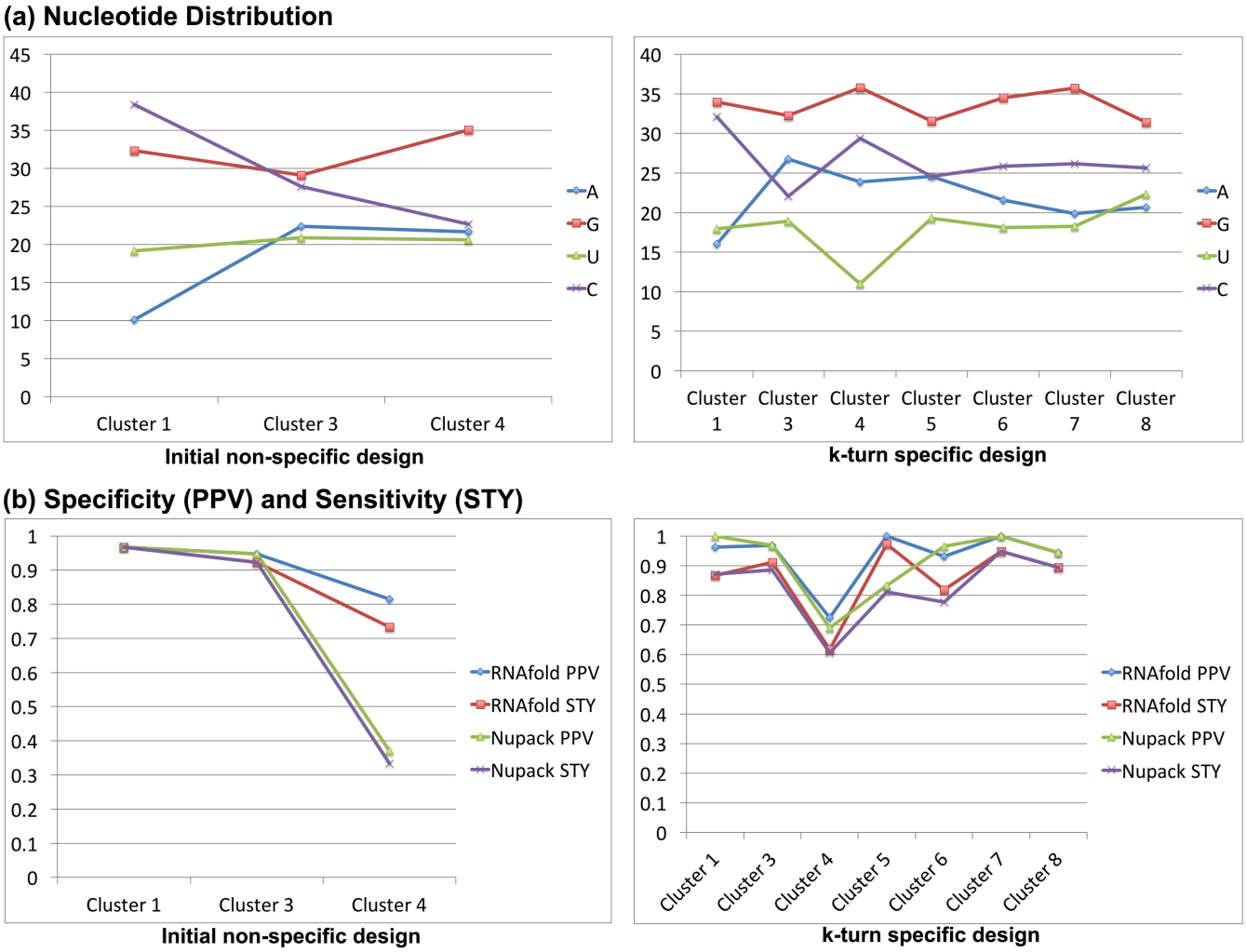


Figure S5: Nucleotide distributions, specificity (PPV), and sensitivity (STY) values between the designed and the predicted 2D structures of one representative sequence from each cluster that fold onto the target topology for the initial non-specific and k-turn specific 8_6 designs. PPV and STY values are calculated for only cannonical base pairs (without pseudoknots). PPV is the percentage of base pairs in the predicted 2D structure that are found in the designed 2D structure; STY is the percentage of base pairs in the designed 2D structure that are found in the predicted 2D structure.

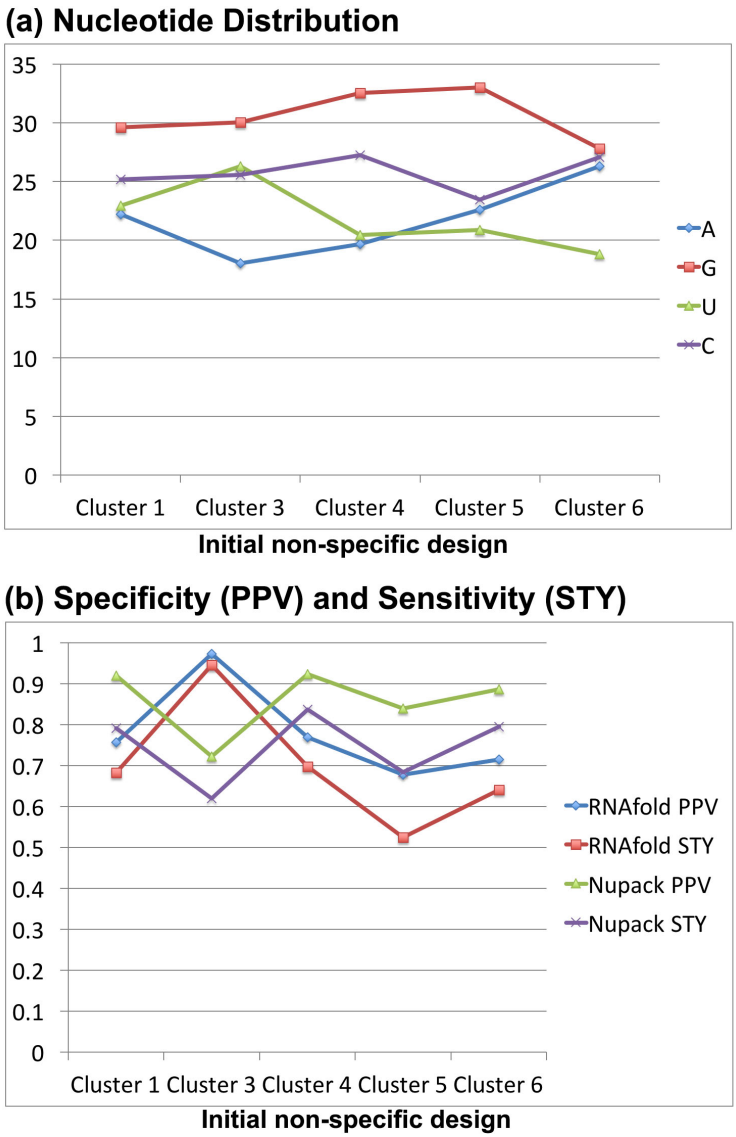


Figure S6: Nucleotide distributions, specificity (PPV), and sensitivity (STY) values between the designed and the predicted 2D structures of one representative sequence from each cluster that fold onto the target topology for the initial non-specific 8_7 design. PPV and STY values are calculated for only cannonical base pairs (without pseudoknots). PPV is the percentage of base pairs in the predicted 2D structure that are found in the designed 2D structure; STY is the percentage of base pairs in the designed 2D structure that are found in the predicted 2D structure.

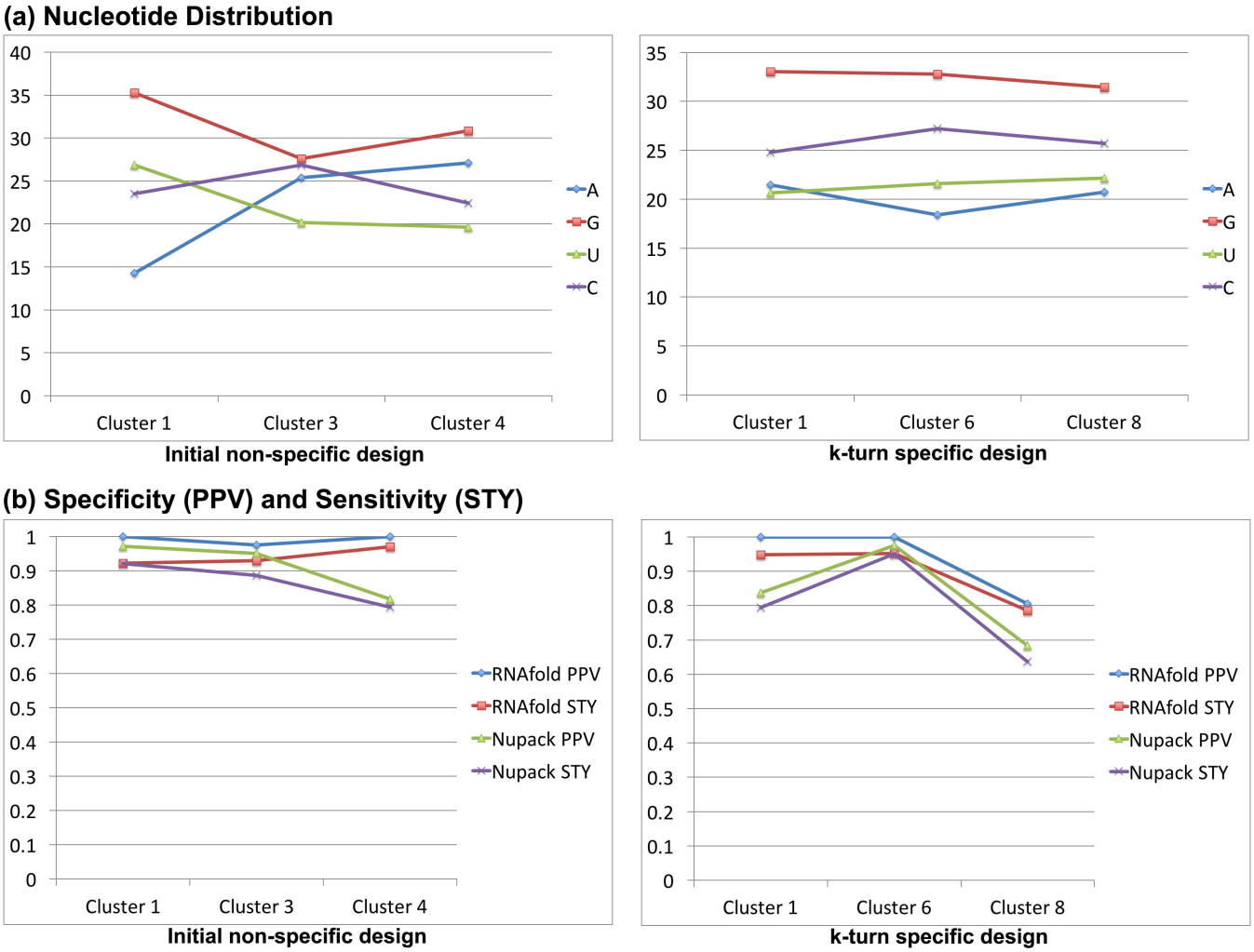


Figure S7: Nucleotide distributions, specificity (PPV), and sensitivity (STY) values between the designed and the predicted 2D structures of one representative sequence from each cluster that fold onto the target topology for the initial non-specific and k-turn specific 8_9 designs. PPV and STY values are calculated for only cannonical base pairs (without pseudoknots). PPV is the percentage of base pairs in the predicted 2D structure that are found in the designed 2D structure; STY is the percentage of base pairs in the designed 2D structure that are found in the predicted 2D structure.

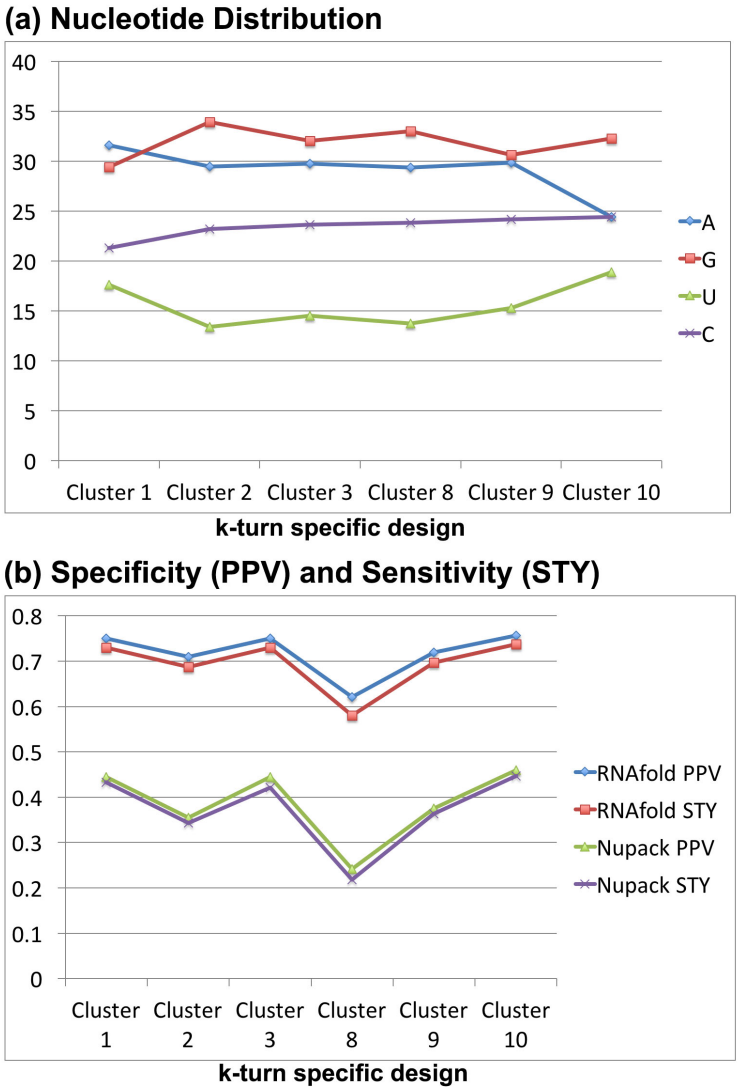
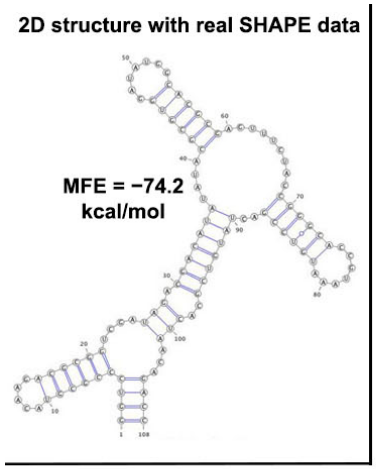


Figure S8: Nucleotide distributions, specificity (PPV), and sensitivity (STY) values between the designed and the predicted 2D structures of one representative sequence from each cluster that fold onto the target topology for the k-turn specific 8_12 design. PPV and STY values are calculated for only cannonical base pairs (without pseudoknots). PPV is the percentage of base pairs in the predicted 2D structure that are found in the designed 2D structure; STY is the percentage of base pairs in the designed 2D structure that are found in the predicted 2D structure.

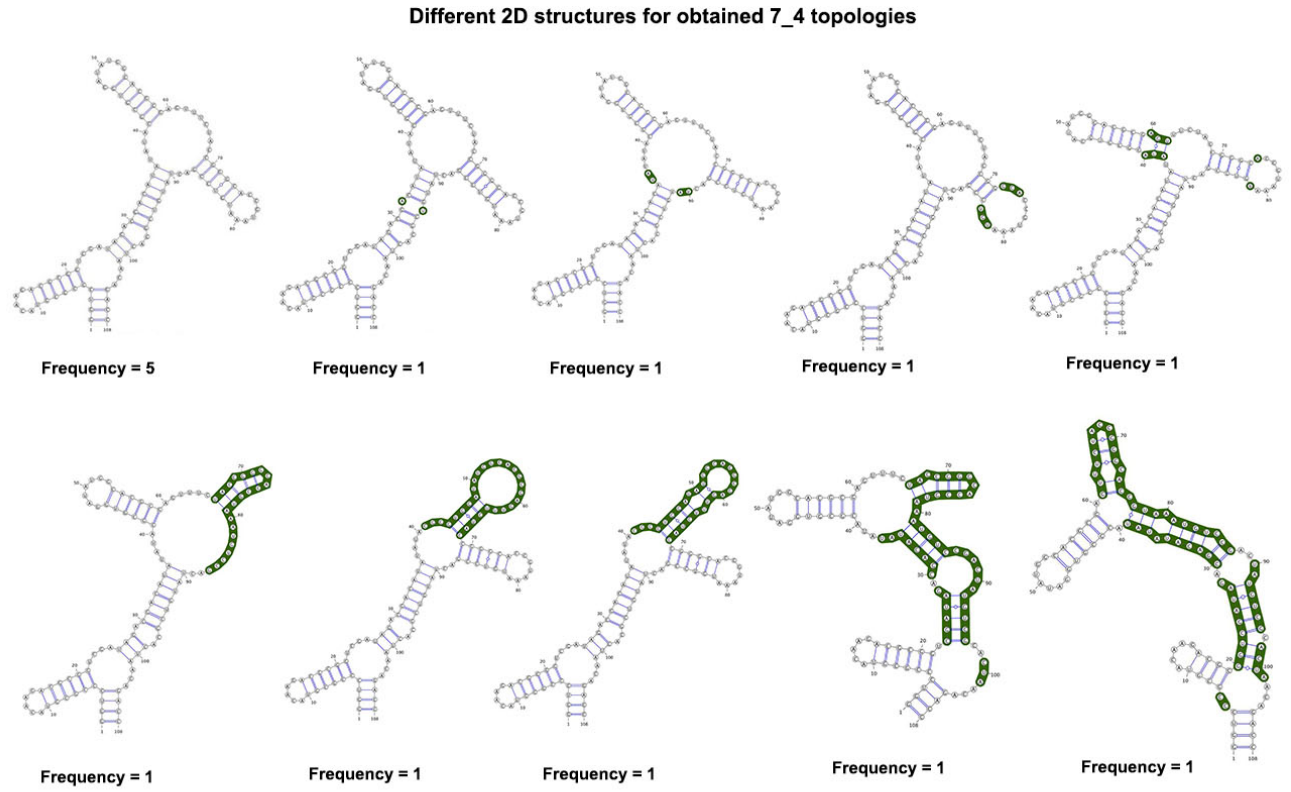
Table S23: Mutations done (using EteRNA) on the top sequences (from the most productive clusters for the target topologies) that do not fold onto the target topology to get them to fold onto the target topologies with both structure prediction programs (RNAfold and NUPACK).

Target Topology	Design	Cluster	Mutations	Reasons
7_4	Initial	1	3→G, 106→C	Add Helix 1
	k-turn	1	63→U	Remove Helix 6
		2	42→U, 76→U	Remove Helix 5
8_4	Initial	4	4→C, 72→U, 130→G, 132→C	Add Helix 1, Remove Helix 7
	k-turn	1	15→G, 25→C	Remove Helix 2
8_6	Initial	1	29→A	Add Loops on Helix 1
	k-turn	1	30→C	Remove Helix 5
		4	25→A	Remove Helix 5
8_7	Initial	1	86→U	Remove Helix 7
8_9	Initial	1	74→A	Remove Helix 5
	k-turn	1	72→A	Remove Helix 5
8_12	k-turn	1	48→U, 72→A, 86→A, 89→A	Remove Helix 2, Preserve k-turn



Obtained Topologies with 100 shuffled variations of SHAPE data

Graph	Freq	Unique 2D struct	E Range (kcal/mol)
7_4	14	10	−67.52 to −50.86
7_9	10	9	−67.44 to −57.49
7_2	9	7	−60.22 to −44.33
8_9	9	6	−68.13 to −52.59
8_6	7	7	−65.76 to −50.98
8_18	7	5	−63.26 to −50.21
7_5	4	4	−57.61 to −50.25
8_2	3	3	−57.60 to −52.48
8_3	3	2	−73.80 to −63.02
8_7	3	3	−59.63 to −55.54
8_22	3	3	−64.09 to −58.66
6_1	2	2	−59.35 to −51.63
6_2	2	2	−71.58 to −50.99
6_5	2	2	−66.07 to −53.66
7_6	2	2	−65.40 to −57.42
7_10	2	2	−58.18 to −51.57
8_15	2	2	−61.88 to −53.58
8_19	2	2	−55.16 to −51.68
9_2	2	2	−56.57 to −53.66
5_1	1	1	−53.64
6_4	1	1	−52.01
7_1	1	1	−53.61
7_3	1	1	−56.79
7_7	1	1	−58.37
7_8	1	1	−57.46
7_11	1	1	−59.93
8_4	1	1	−52.38
8_5	1	1	−51.01
8_14	1	1	−61.37
9_1	1	1	−53.20
9_14	1	1	−60.91



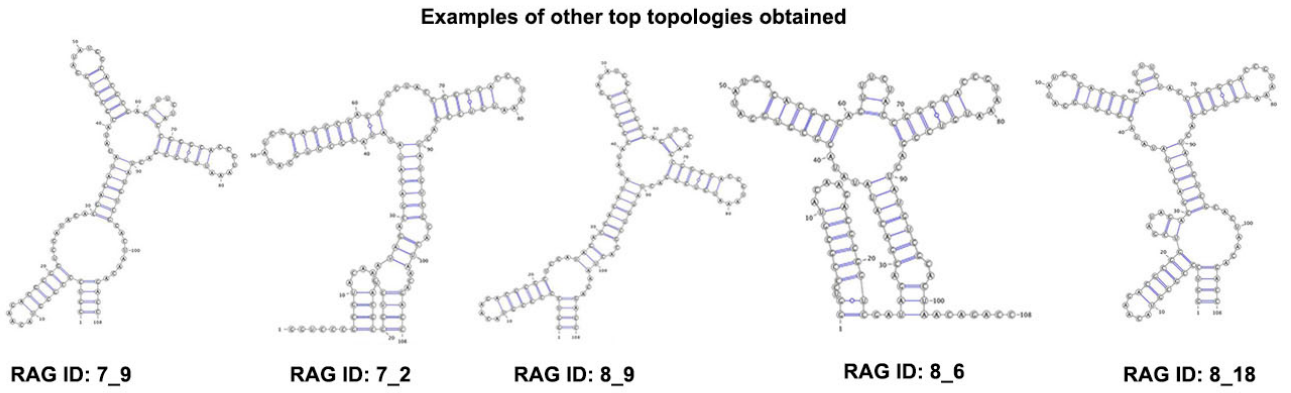
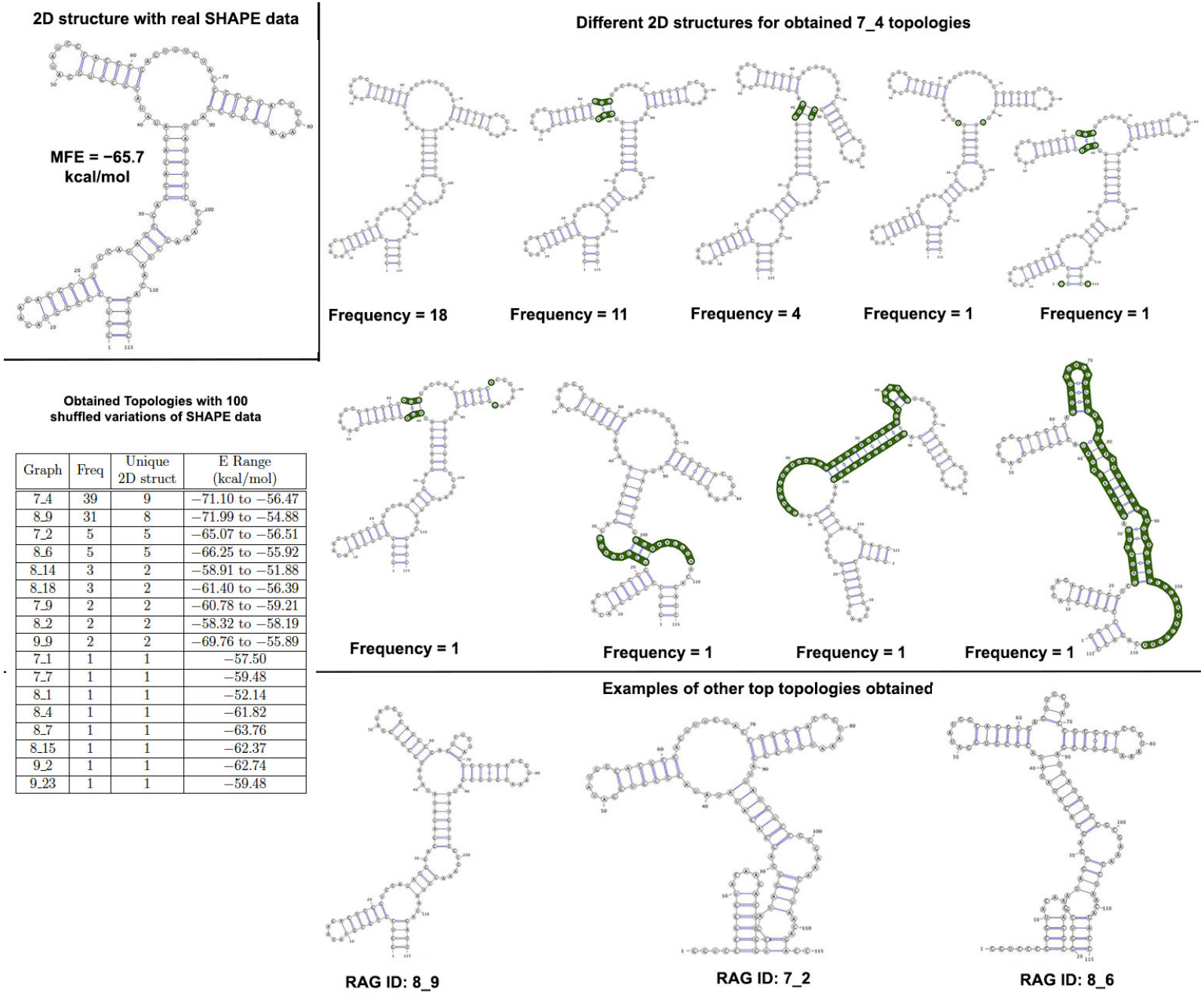


Figure S9: Different 2D structures obtained for the 7_4 topology and other top topologies obtained by RNAfold using 100 different shuffled variations of the real SHAPE data for 7_4 initial design sequence 1b. The green residues in the 7_4 topologies indicate residues that have different 2D structure as compared to the 2D structure predicted with real SHAPE data (top left).



Obtained Topologies with 100 shuffled variations of SHAPE data

Graph	Freq	Unique 2D struct	E Range (kcal/mol)
7_4	39	9	−71.10 to −56.47
8_9	31	8	−71.99 to −54.88
7_2	5	5	−65.07 to −56.51
8_6	5	5	−66.25 to −55.92
8_14	3	2	−58.91 to −51.88
8_18	3	2	−61.40 to −56.39
7_9	2	2	−60.78 to −59.21
8_2	2	2	−58.32 to −58.19
9_9	2	2	−69.76 to −55.89
7_1	1	1	−57.50
7_7	1	1	−59.48
8_1	1	1	−52.14
8_4	1	1	−61.82
8_7	1	1	−63.76
8_15	1	1	−62.37
9_2	1	1	−62.74
9_23	1	1	−59.48

Figure S10: Different 2D structures obtained for the 7_4 topology and other top topologies obtained by RNAfold using 100 different shuffled variations of the real SHAPE data for 7_4 k-turn design sequence 1d. The green residues in the 7_4 topologies indicate residues that have different 2D structure as compared to the 2D structure predicted with real SHAPE data (top left).

References

[1] Fera, D., Kim, N., Shiffeldrim, N., Zorn, J., Laserson, U., Gan, H. H., and Schlick, T. (2004) RAG: RNA-As-Graphs web resource. *BMC Bioinformatics,* **5**(1), 1.

[2] Kim, N., Laing, C., Elmetwaly, S., Jung, S., Curuksu, J., and Schlick, T. (2014) Graph-based sampling for approximating global helical topologies of RNA. *Proc Natl Acad Sci USA,* **111**(11), 4079–4084.

[3] Gan, H. H., Pasquali, S., and Schlick, T. (2003) Exploring the repertoire of RNA secondary motifs using graph theory; implications for RNA design. *Nucleic Acids Res,* **31**(11), 2926–2943.

[4] Kim, N., Shiffeldrim, N., Gan, H. H., and Schlick, T. (2004) Candidates for novel RNA topologies. *J Mol Biol,* **341**(5), 1129–1144.

[5] Fiedler, M. (1973) Algebraic connectivity of graphs. *Czechoslovak Math J,* **23**(2), 298–305.

[6] Baba, N., Elmetwaly, S., Kim, N., and Schlick, T. (2016) Predicting large RNA-Like topologies by a knowledge-based clustering approach. *J Mol Biol,* **428**(5), 811–821.

[7] Kim, N., Zheng, Z., Elmetwaly, S., and Schlick, T. (2014) RNA graph partitioning for the discovery of RNA modularity: a novel application of graph partition algorithm to biology. *PLoS ONE,* **9**(9), e106074.

[8] Zahran, M., Bayrak, C. S., Elmetwaly, S., and Schlick, T. (2015) RAG-3D: a search tool for RNA 3D substructures. *Nucleic Acids Res,* **43**(19), gkv823–9488.

[9] Jain, S. and Schlick, T. (2017) F-RAG: Generating atomic models from RNA graphs using fragment assembly. *J Mol Biol,* **429**(23), 3587–3605.

[10] Bayrak, C. S., Kim, N., and Schlick, T. (2017) Using sequence signatures and kink-turn motifs in knowledge-based statistical potentials for RNA structure prediction. *Nucleic Acids Res,* **45**(9), 5414–5422.

[11] Lorenz, R., Bernhart, S. H., Höner zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P. F., and Hofacker, I. L. (2011) ViennaRNA Package 2.0. *Algo Mol Biol,* **6**(1), 26.

[12] Dirks, R. M. and Pierce, N. A. (2003) A partition function algorithm for nucleic acid secondary structure including pseudoknots. *J Comp Chem,* **24**(13), 1664–1677.

[13] Dirks, R. M. and Pierce, N. A. (2004) An algorithm for computing nucleic acid base-pairing probabilities including pseudoknots. *J Comp Chem,* **25**(10), 1295–1304.

[14] Dirks, R. M., Bois, J. S., Schaeffer, J. M., Winfree, E., and Pierce, N. A. (2007) Thermodynamic analysis of interacting nucleic acid strands. *SIAM Rev,* **49**(1), 65–88.

[15] Lee, J., Kladwang, W., Lee, M., Cantu, D., Azizyan, M., Kim, H., Limpaecher, A., Gaikwad, S., Yoon, S., Treuille, A., Das, R., and Participants, E. (2014) RNA design rules from a massive open laboratory. *Proc Natl Acad Sci USA,* **111**(6), 2122–2127.

[16] Siegfried, N. A., Busan, S., Rice, G. M., Nelson, J. A. E., and Weeks, K. M. (2014) RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). *Nat Methods,* **11**(9), 959–965.

[17] Ball, C. B., Rodriguez, K. F., Stumpo, D. J., Ribeiro-Neto, F., Korach, K. S., Blackshear, P. J., Birnbaumer, L., and Ramos, S. B. V. (2014) The RNA-Binding Protein, ZFP36L2, Influences Ovulation and Oocyte Maturation. *PLOS ONE,* **9**(5), e97324.

[18] Wilkinson, K. A., Merino, E. J., and Weeks, K. M. (2006) Selective 2′-hydroxyl acylation analyzed by primer extension SHAPE: quantitative RNA structure analysis at single nucleotide resolution. *Nat Protocols,* **1**(3), 1610—1616.

[19] Ball, C. B., Solem, A. C., Meganck, R. M., Laederach, A., and Ramos, S. B. (2017) Impact of RNA structure on ZFP36L2 interaction with luteinizing hormone receptor mRNA. *RNA,* **23**(8), 1209–1223.

[20] Spitale, R. C., Crisalli, P., Flynn, R. A., Torre, E. A., Kool, E. T., and Chang, H. Y. (2013) RNA SHAPE analysis in living cells. *Nat Chem Biol,* **9**(1), 18–20.

[21] Smola, M. J., Rice, G. M., Busan, S., Siegfried, N. A., and Weeks, K. M. (2015) Selective 2′-hydroxyl acylation analyzed by primer extension and mutational profiling (SHAPE-MaP) for direct, versatile and accurate RNA structure analysis. *Nat Protocols,* **10**, 1643–1669.

[22] Busan, S. and Weeks, K. M. (2018) Accurate detection of chemical modifications in RNA by mutational profiling (MaP) with ShapeMapper 2. *RNA,* **24**(2), 143–148.

[23] Lorenz, R., Luntzer, D., Hofacker, I. L., Stadler, P. F., and Wolfinger, M. T. (2016) SHAPE directed RNA folding. *Bioinformatics,* **32**(1), 145–147.

[24] Xu, Z. Z. and Mathews, D. H. (2016) Experiment-Assisted Secondary Structure Prediction with RNAstructure. In Turner, D. H. and Mathews, D. H., (eds.), *RNA Structure Determination: Methods and Protocols*, pp. 163–176 Springer New York New York, NY.

[25] Deigan, K. E., Li, T. W., Mathews, D. H., and Weeks, K. M. (2009) Accurate SHAPE-directed RNA structure determination. *Proc Natl Acad Sci USA,* **106**(1), 97–102.

[26] Darty, K., Denise, A., and Ponty, Y. (2009) VARNA: Interactive drawing and editing of the RNA secondary structure. *Bioinformatics,* **25**(15), 1974–1975.

[27] R Core Team R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria (2014).