

Supplementary Information

High throughput *in vivo* mapping of RNA accessible interfaces to identify functional sRNA binding sites

Mihailovic, et al.

Supplementary Methods

Total RNA extraction

Total RNA was extracted from a sample of 1-5 mL of culture. Briefly, 1 mL of room temperature Trizol (Invitrogen) was used to resuspend each cell pellet upon which samples were left at room temperature for 5 min. Next, 200 μ L of 24:1 chloroform: isoamyl alcohol were added to each sample and samples were inverted vigorously for 15 seconds prior to centrifugation at 13,000 xg for 10 min at 4°C. The aqueous phase was then transferred to 2mL DNase/RNase-free microcentrifuge tubes. Equal volumes of 0.8 M sodium citrate and 1.2 M sodium chloride (~ 300 μ L) were added to each sample and inverted for 1 min prior to overnight precipitation at -20°C. The following day, samples were centrifuged at 13,000 xg at 4°C for 30 min. The liquid was removed carefully as to not disturb the formed RNA pellet. RNA pellets were then washed twice with 75% ethanol (10 min 13,000 xg centrifugations at 4°C). After the second wash, liquid was completely removed from the RNA pellet and sample was left to dry in a laminar flow hood for 10 min. Once dry, RNA pellets were resuspended in 30 μ L nuclease-free water and RNA concentration was measured via spectrophotometry.

Total RNA samples (diluted to 200 ng/ μ L) were treated with RNase-free DNase I (PI-90083 Thermo-Fisher Scientific) following manufacturer protocol. After DNase I treatment, 10 μ L of GlycoBlue (AM9516 Life Technologies) was added to an equivalent volume solution of 100% isopropanol and DNase-treated RNA (total 55 μ L) and precipitated at -20°C overnight. The following day, samples were pelleted (15 min at 13,000 xg) and liquid was removed carefully. Next, samples were washed with chilled 95% ethanol. After the second wash, liquid was completely removed from the RNA pellet and sample was left to dry in a laminar flow hood for 10 min. Once dry, RNA pellets were resuspended in 50 μ L nuclease-free water and stored at -80°C. Finally, the quality of RNA was evaluated by using a bioanalyzer (Agilent) at the Genomic Sequencing and Analysis Facility (GSAF at UT Austin) to confirm that no significant degradation had occurred.

PCR Amplification and in vitro Transcription (IVT) for EMSA

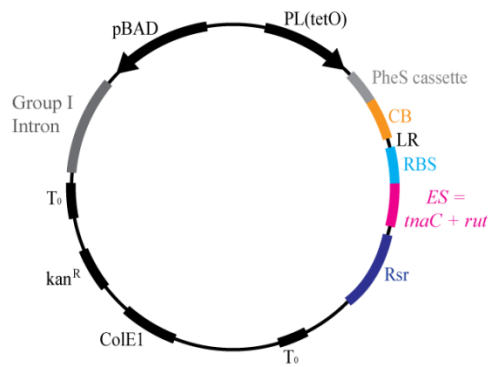
Primers were designed to amplify sRNAs and mRNAs (Supplementary Table 5) (with an upstream T7 promoter) from previously-designed plasmids or genomic DNA (*E. coli* BW25113). Specifically, 4 random nucleotides (GACT) were added upstream of the T7 promoter sequence (TAATACGACTCACTATAGGAG). PCR product sizes were verified on a 1.5 % agarose gel prior to PCR clean up (GE Illustra GFX PCR).

mRNAs were *in vitro* transcribed using the MEGA Script T7 kit following the manufacturer instructions. Specifically 1-1.5 pmol of DNA template was used per reaction, and each reaction was incubated for 6 hours. Following DNase digestion, one of two reliable RNA recovery methods was performed. If an RNA Clean and Concentrator-5 kit (Zymo Research) was used, 30 μ L of nuclease-free water was added to the IVT reaction prior to following manufacturer instruction manual. In the case that RNA was recovered using a chloroform-based method, 115 μ L of nuclease-free water and 15 115 μ L of Ammonium Acetate Stop Solution (from MEGA Script T7 kit) were added and mixed. Next, two volumes of 25:24:1 phenol: chloroform: isoamyl alcohol were added to each sample and samples were inverted for one minute, then centrifuged at 13,000 xg at 4°C for 5 min. The top aqueous phase was transferred to two equivalent volumes of 24:1 chloroform: isoamyl alcohol and samples were centrifuged at 13,000 xg at 4°C for 5 min. The top phase was again transferred, this time to microcentrifuge tubes containing 1 mL isopropyl alcohol and 1.5 μ L GlycoBlue each, and left to precipitate overnight at -20°C.

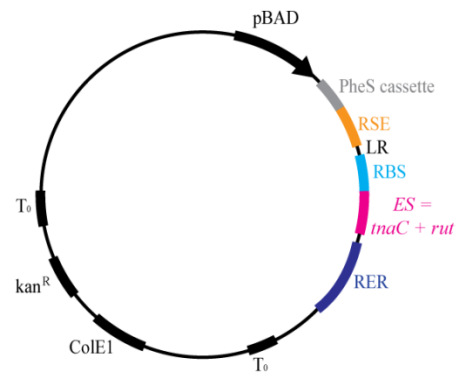
The following day, samples were centrifuged at 13,000 xg at 4°C for 15 min. The liquid was removed carefully as to not disturb the formed RNA pellet. One mL of 95% ethanol was added to each sample and the samples were centrifuged at 13,000 xg at 4°C for 5 min. The liquid was again removed carefully as to not disturb the RNA pellet. One mL of 75% ethanol was added to each sample and the samples were centrifuged for at 13,000 xg at 4°C for 5 min. The liquid was again removed and samples left to dry in a solvent hood for 2 hours. Once dry, 30 µL of nuclease-free water was added to re-suspend the RNA. RNA concentration was measured via spectrophotometry. To validate transcripts (length and lack of spurious transcripts), 3 µL of each RNA was mixed with 7 µL of nuclease-free water and 5 µL of RNA Loading Buffer II (NEB) and denatured for 5 min at 70°C. Samples were loaded into an 8% urea gel (SequaGel, National Diagnostics) and run at 100 V for 3 hours. The gel was then stained using Sybr Green II (Thermo-Fisher Scientific) for 30 min and imaged under Sybr Green setting in a gel imager (ChemiDoc XRS +, BioRad).

sRNAs were prepared as were mRNAs with slight modifications to support internal P-32 labeling. Specifically, 0.5-1.5 µL UTP [α -32P] (3000Ci/mmol 10mCi/ml, 250 µCi, PerkinElmer) was added to replace unlabeled UTP from the MEGA Script IVT kit. RNA samples were re-suspended in 20 µL of nuclease-free water and, if RNA was recovered using the chloroform-based method, free NTPs removed using DTR Gel Filtration Cartridges (EdgeBio) following manufacturer instructions.

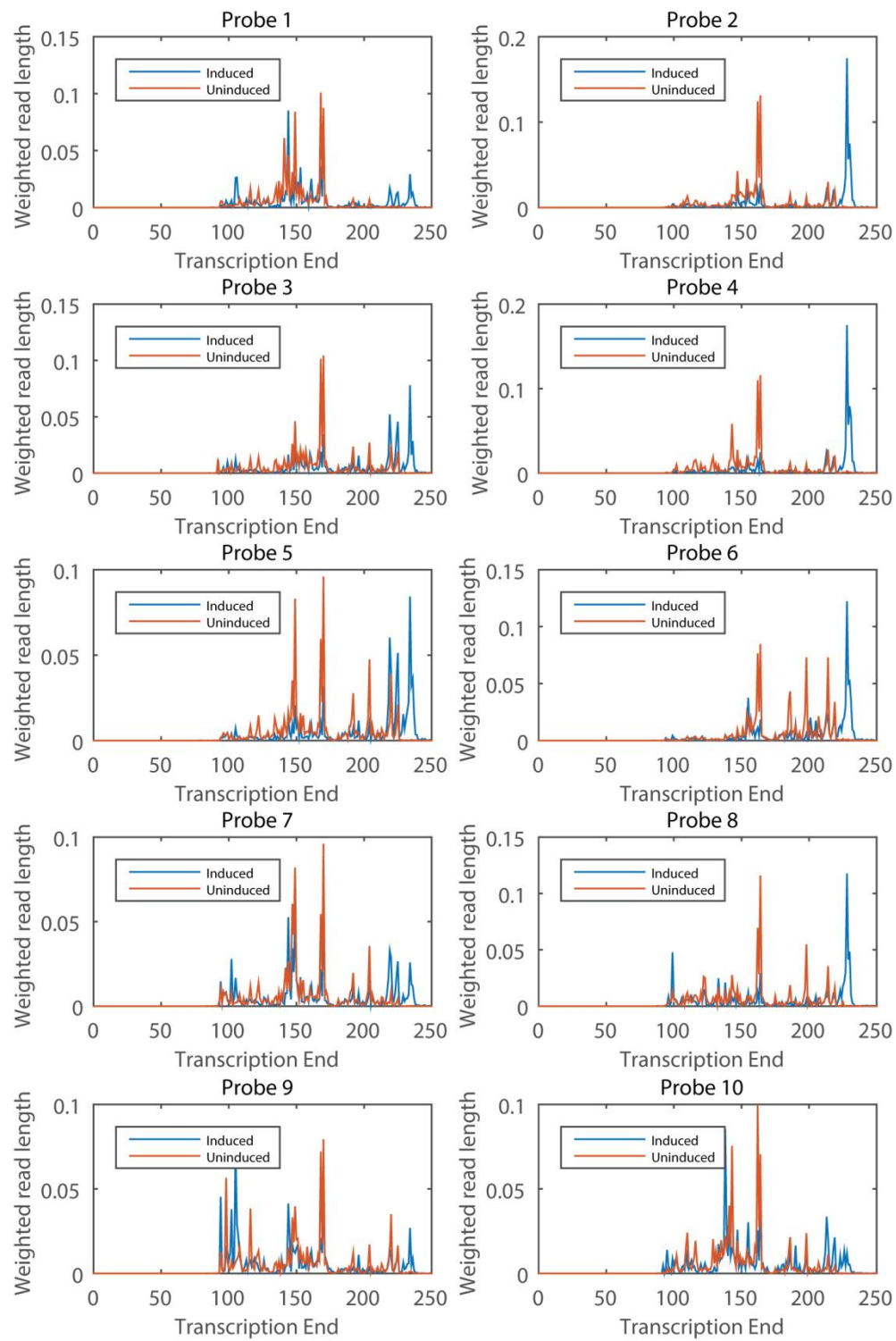
a. For heterologously-expressed taRNA

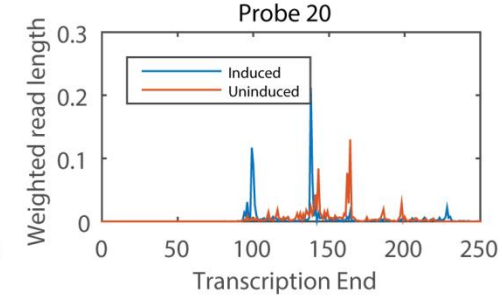
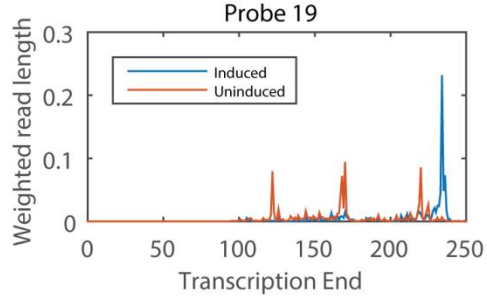
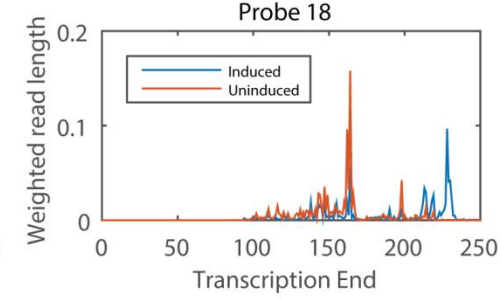
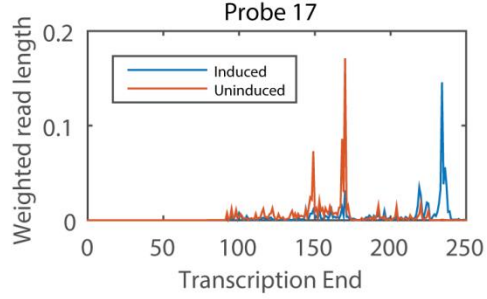
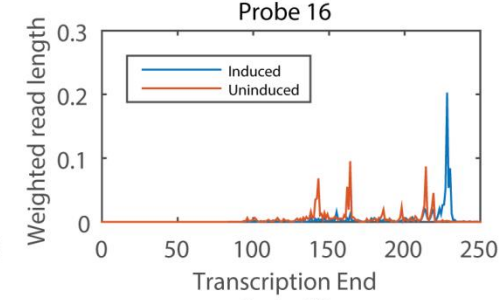
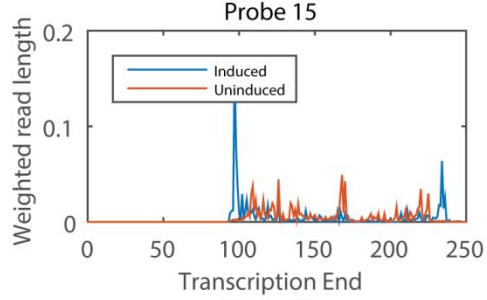
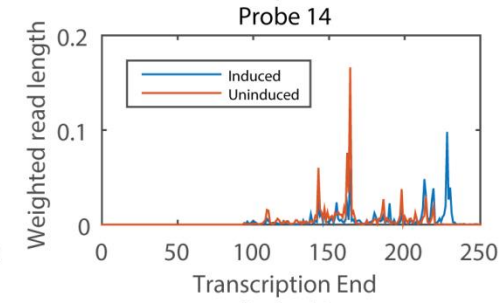
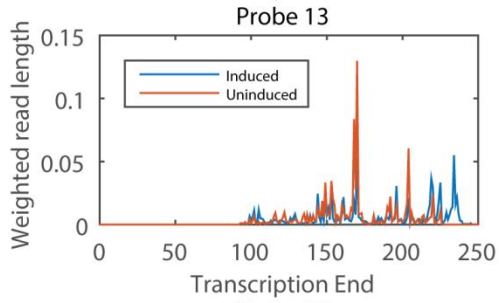
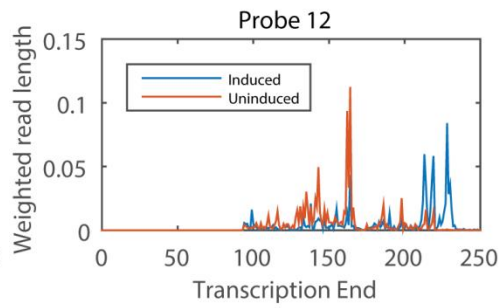
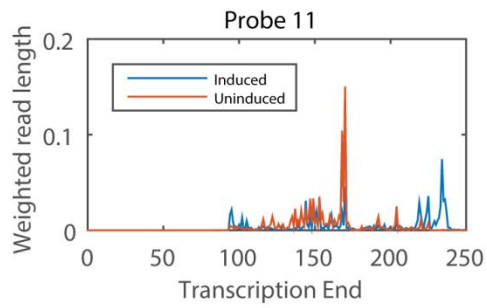


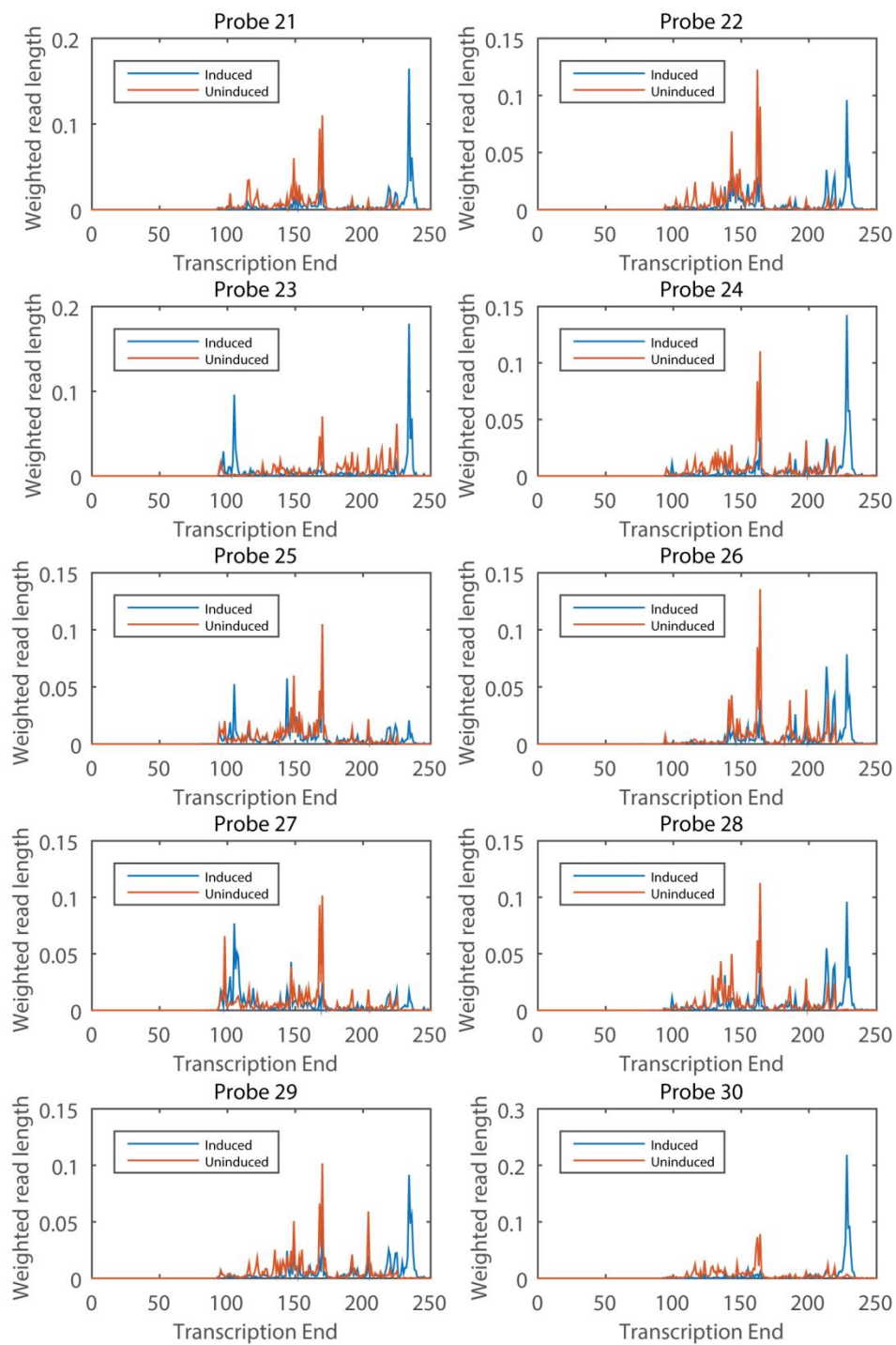
b. For native taRNA



Supplementary Figure 1. INTERFACE plasmid maps (a). Plasmid map of O-INTERFACE (for heterologously expressed target RNAs). (b) Plasmid map INTERFACE (for native target RNAs).

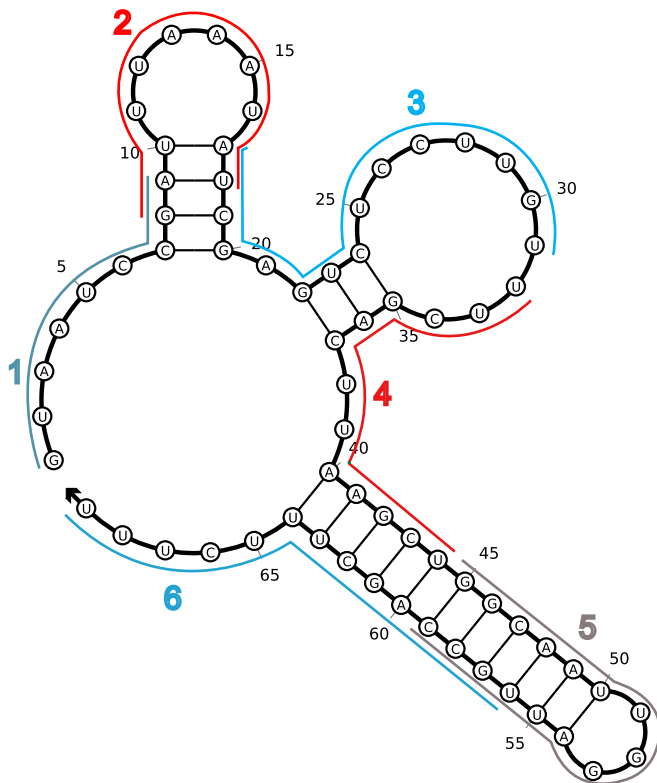




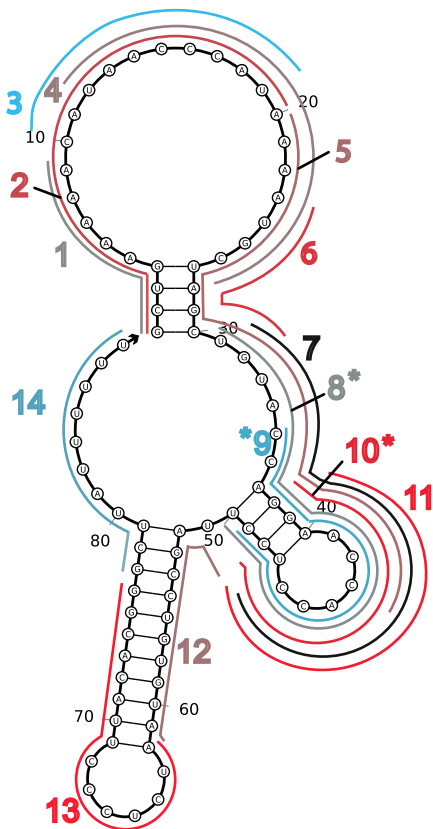


Supplementary Figure 2. O-INTERFACE transcription end traces for 30 new group I intron regions probed.

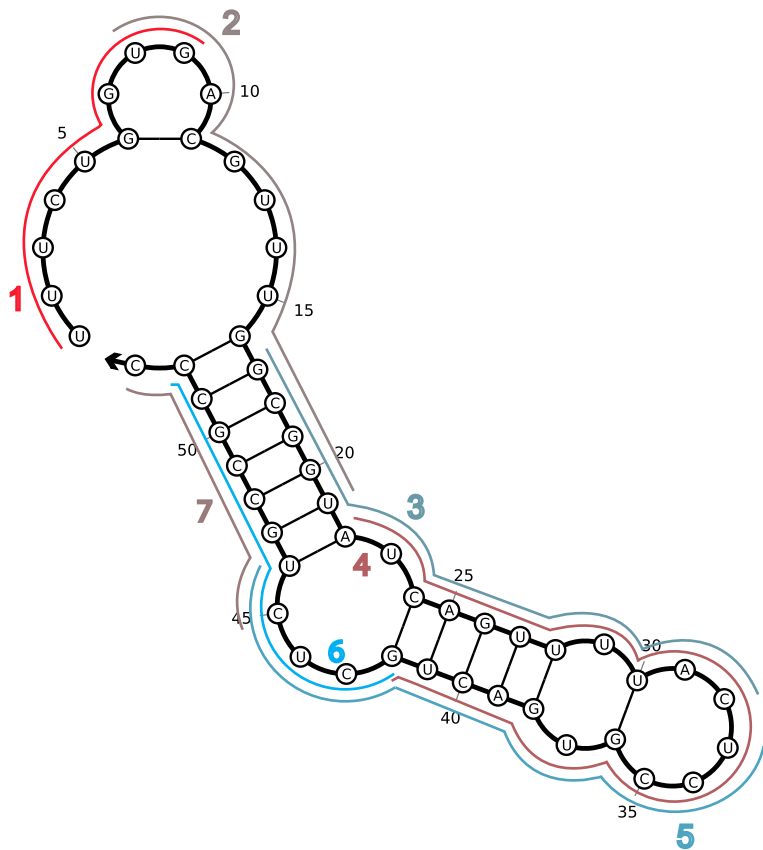
ArrS



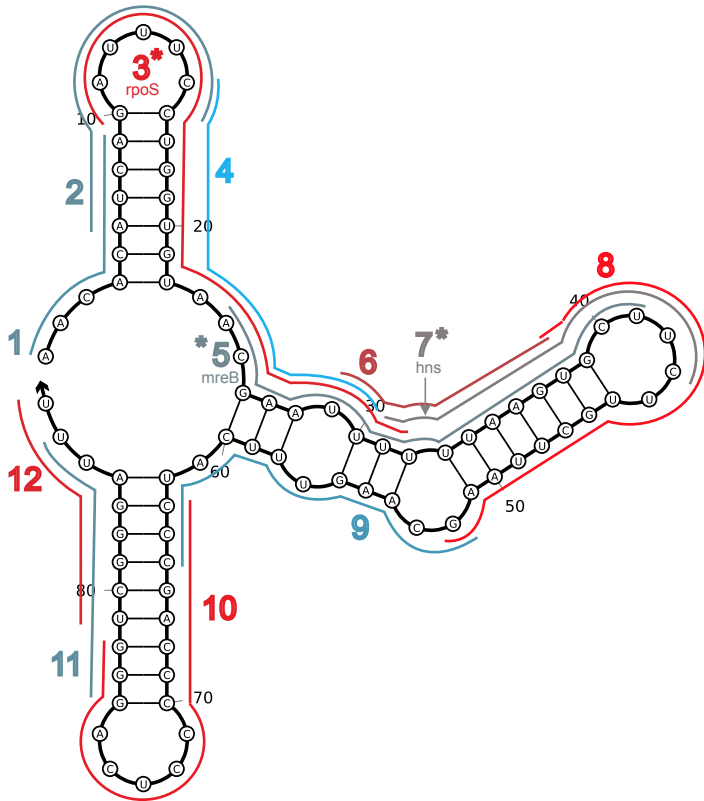
CyaR



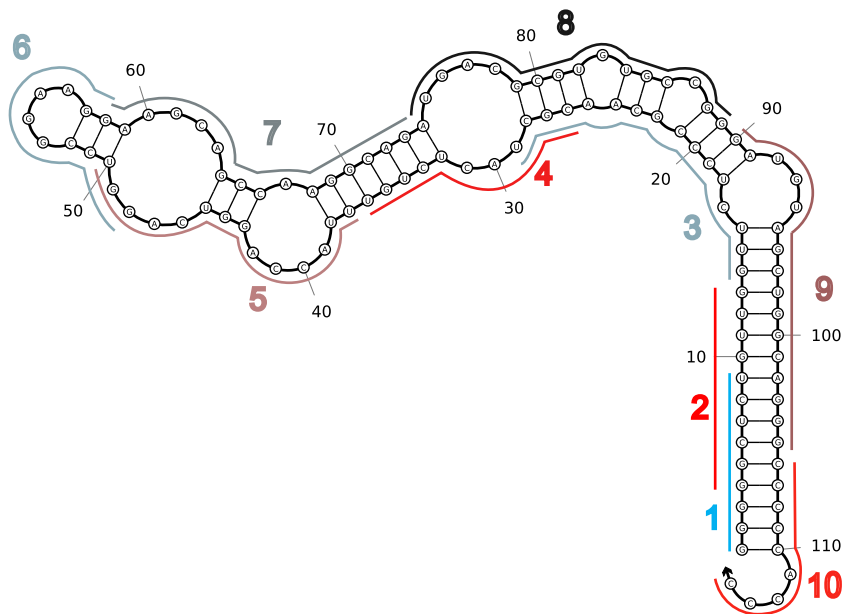
DicF



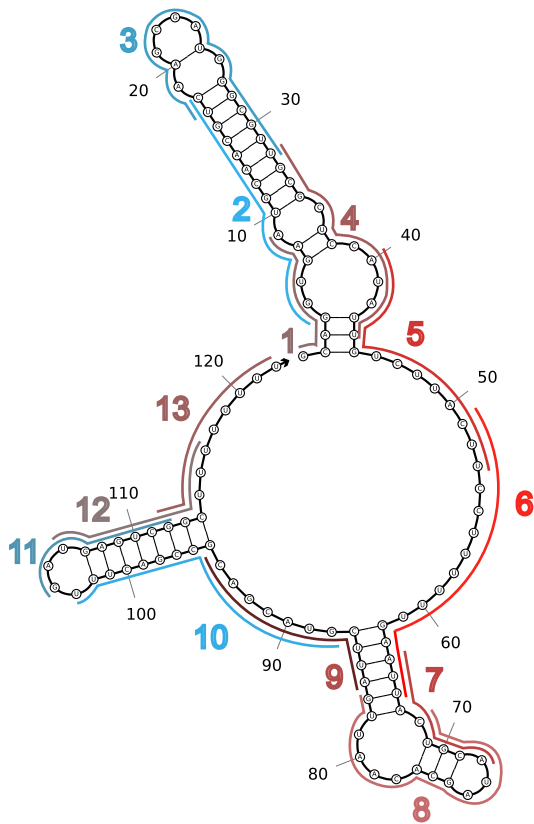
DsrA



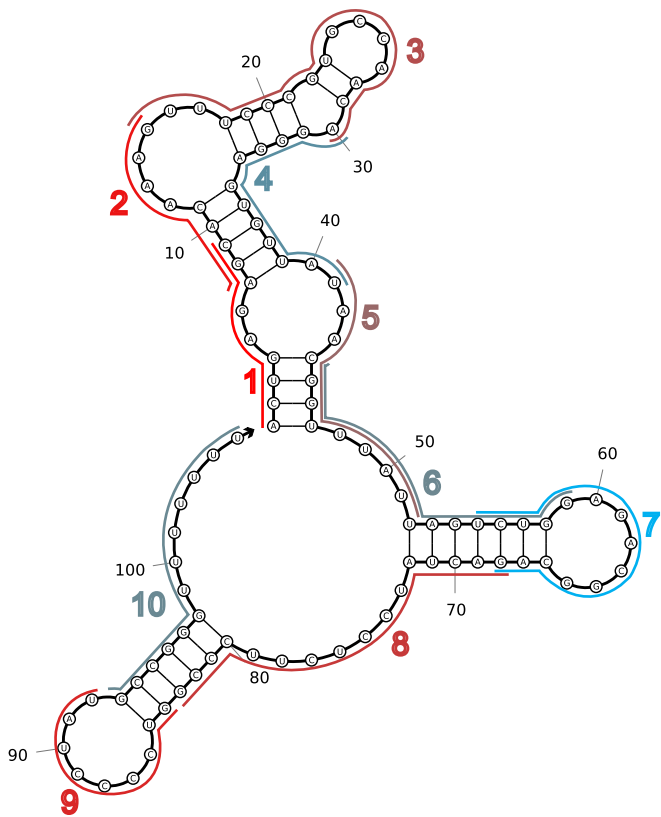
Ffs



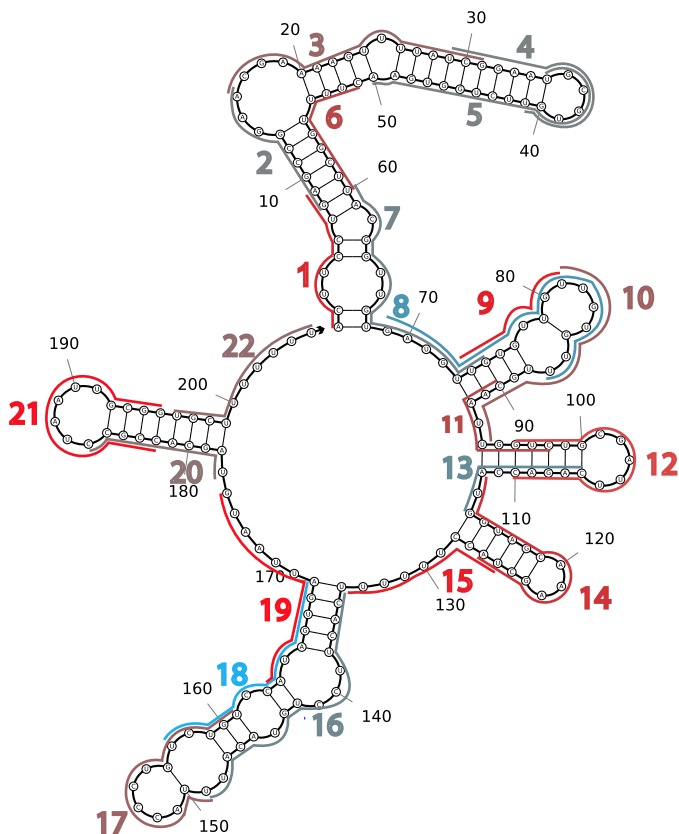
FnrS



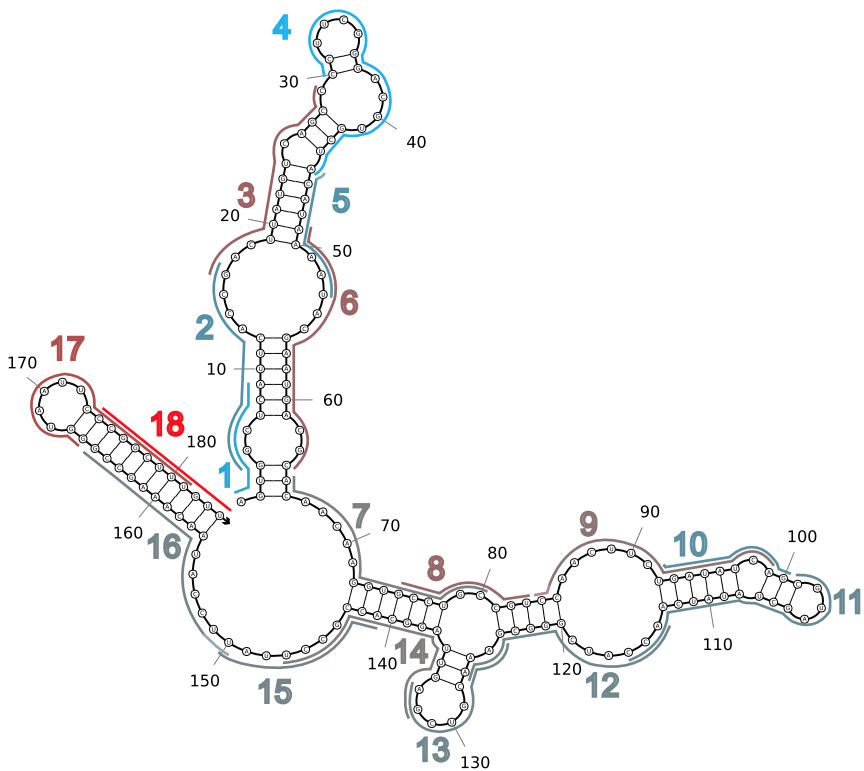
GadY



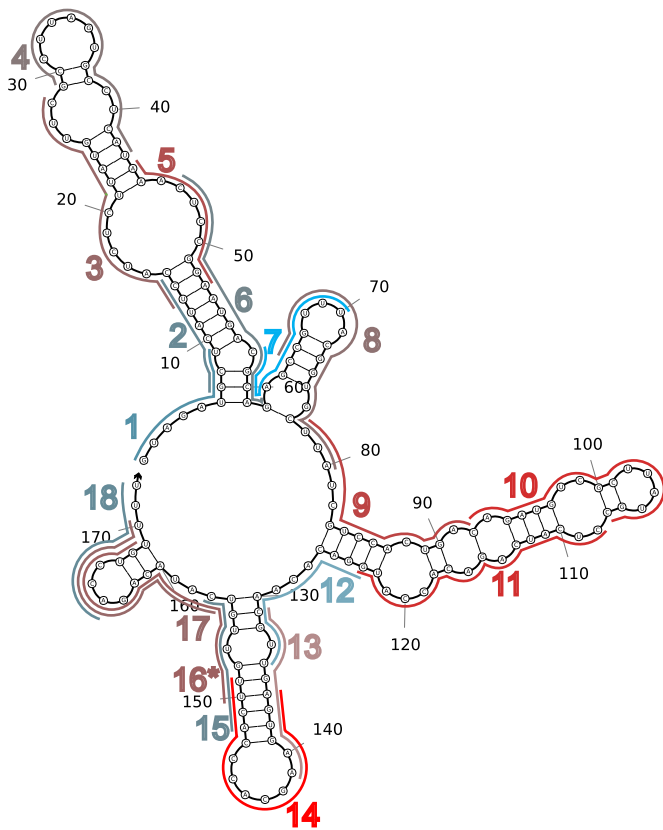
GcvB



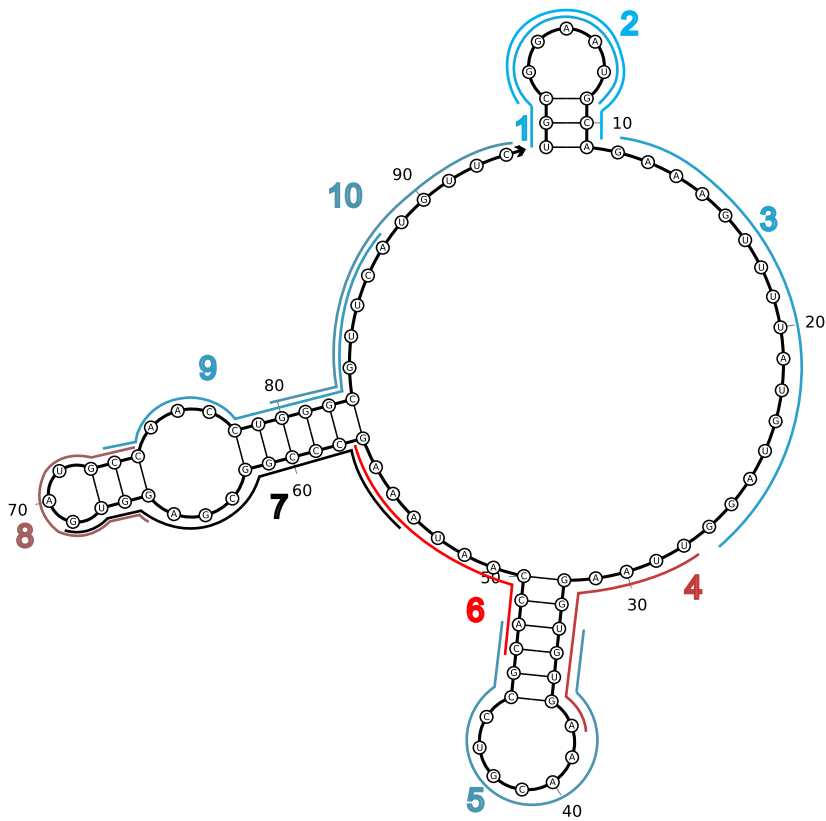
GlmY



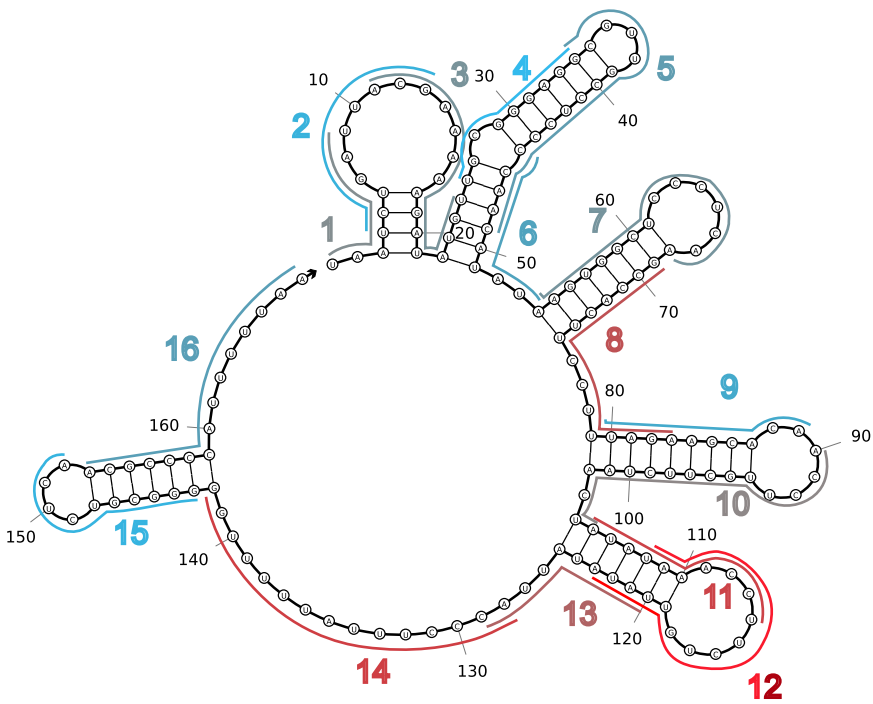
GlmZ



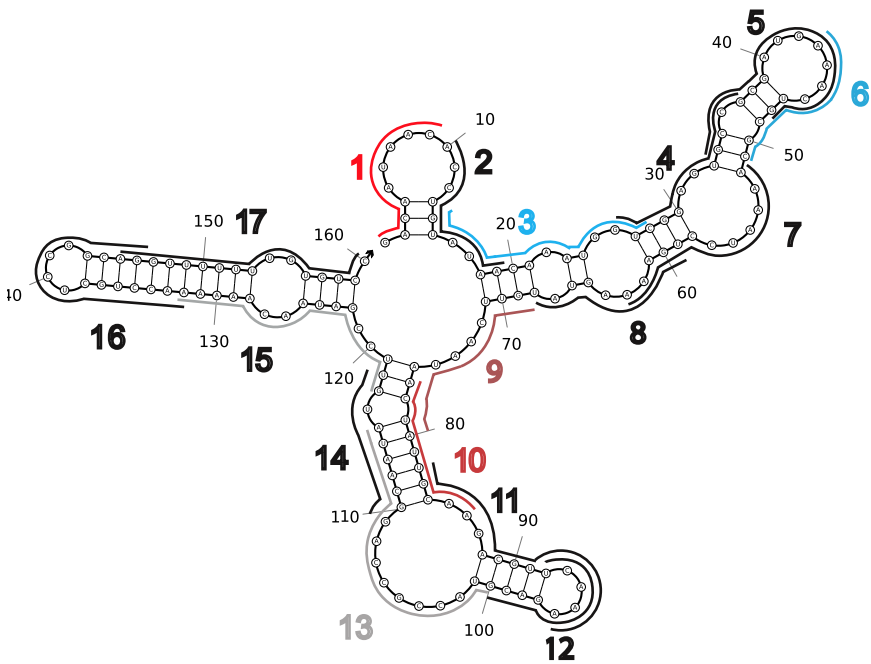
InvR



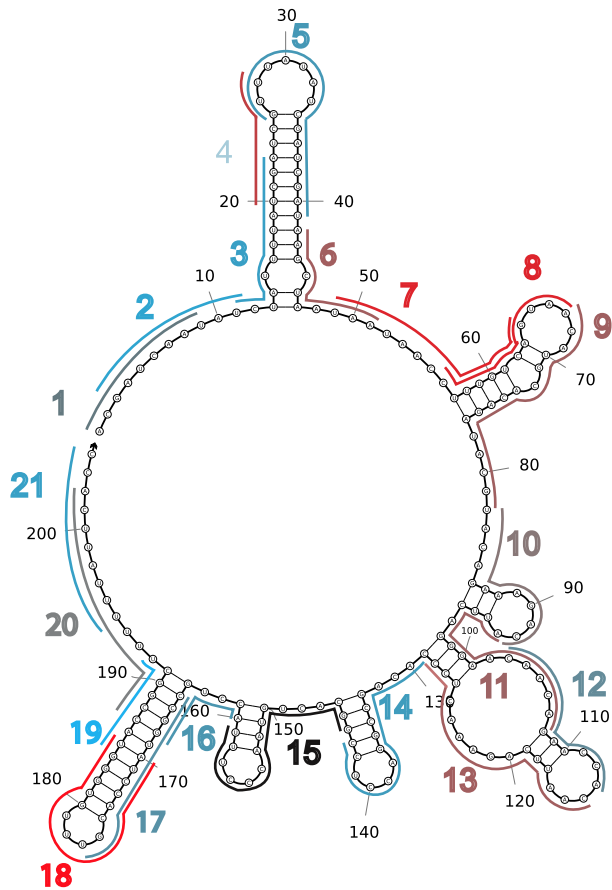
lpeX



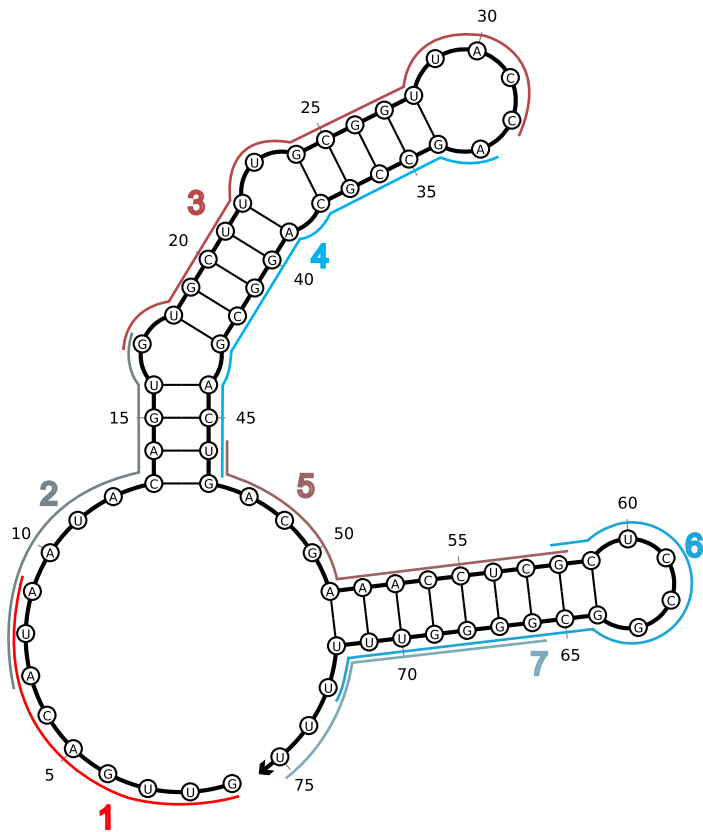
IsrB



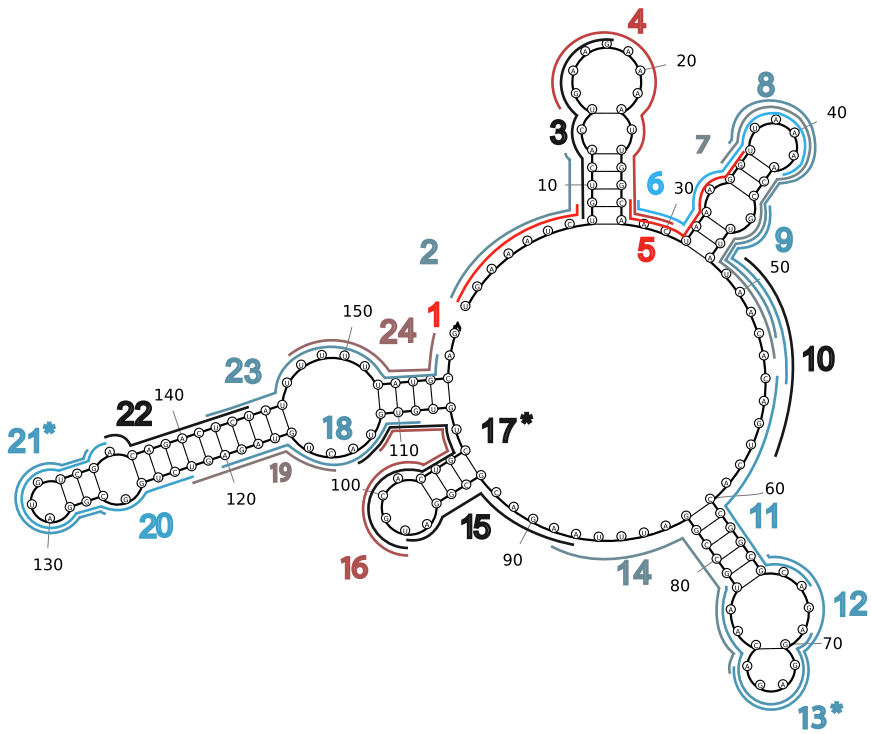
IsrC



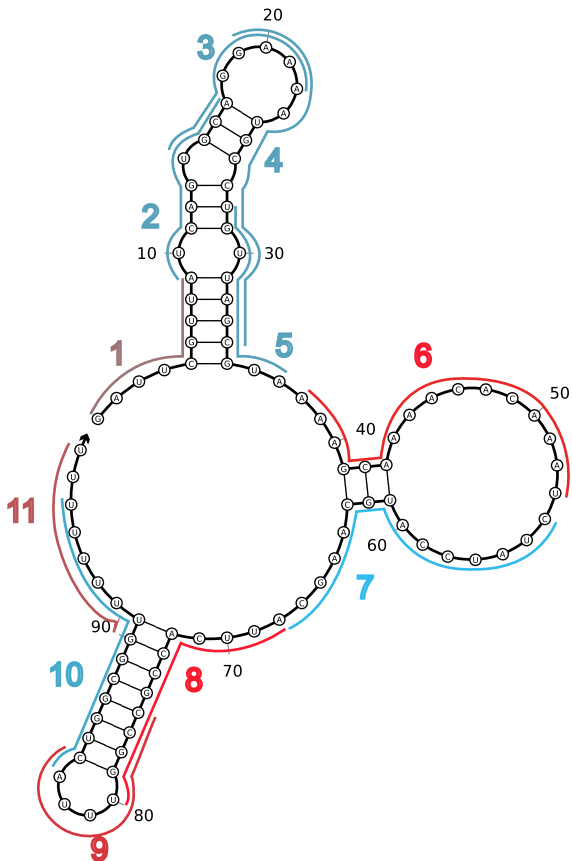
IstR-1



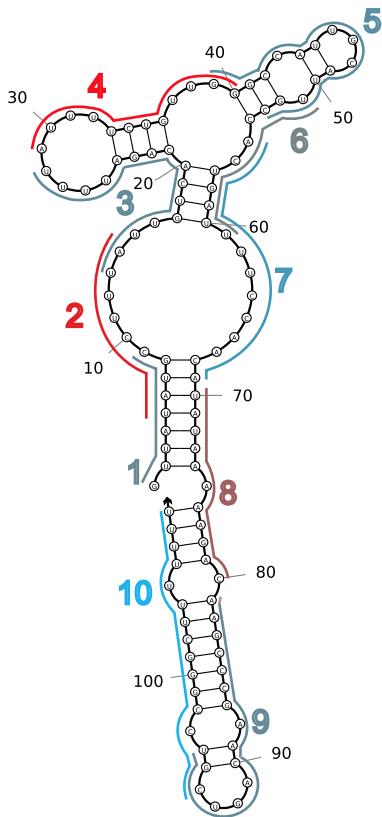
McaS



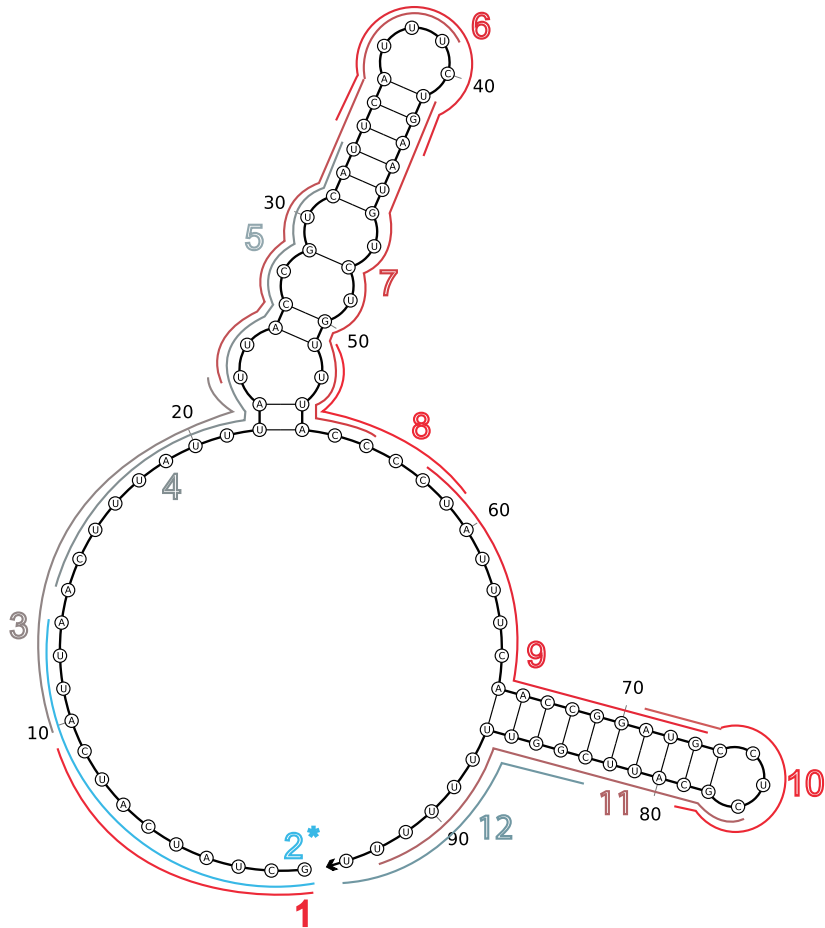
MgrR



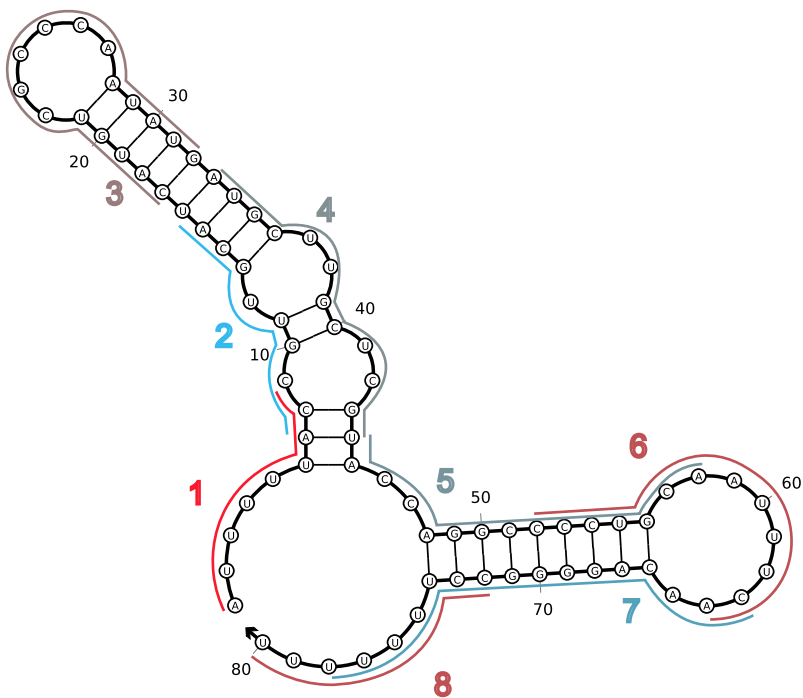
MicC



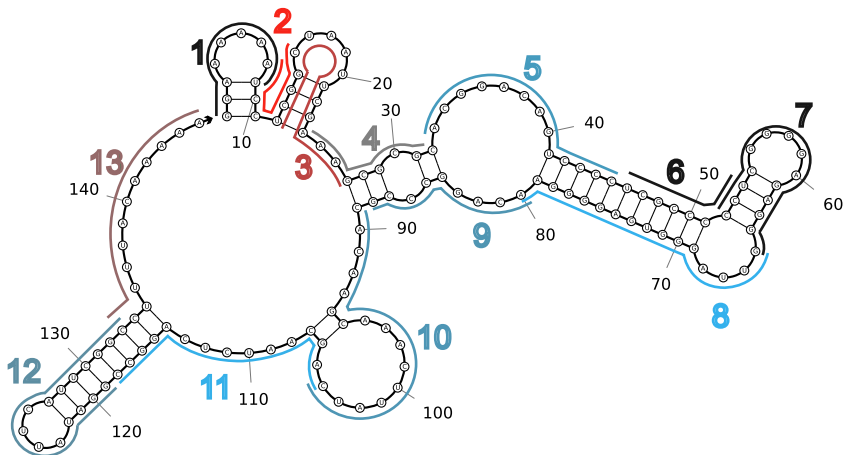
MicF



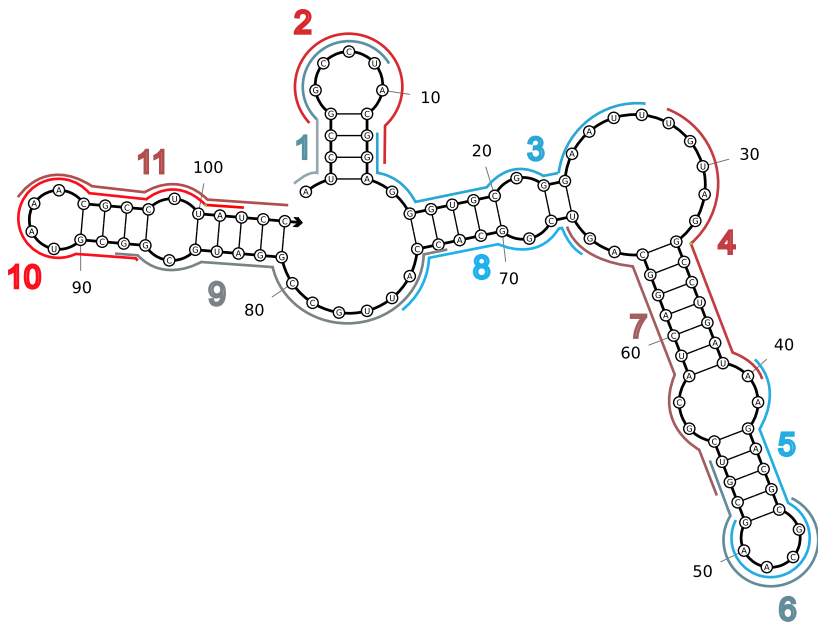
MicL



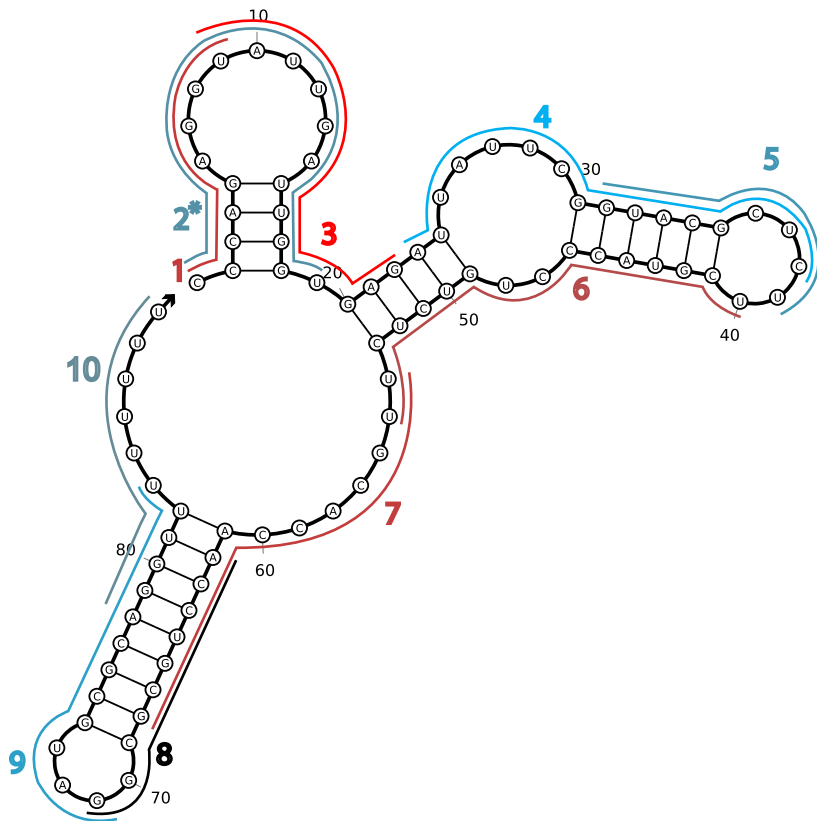
Nc2



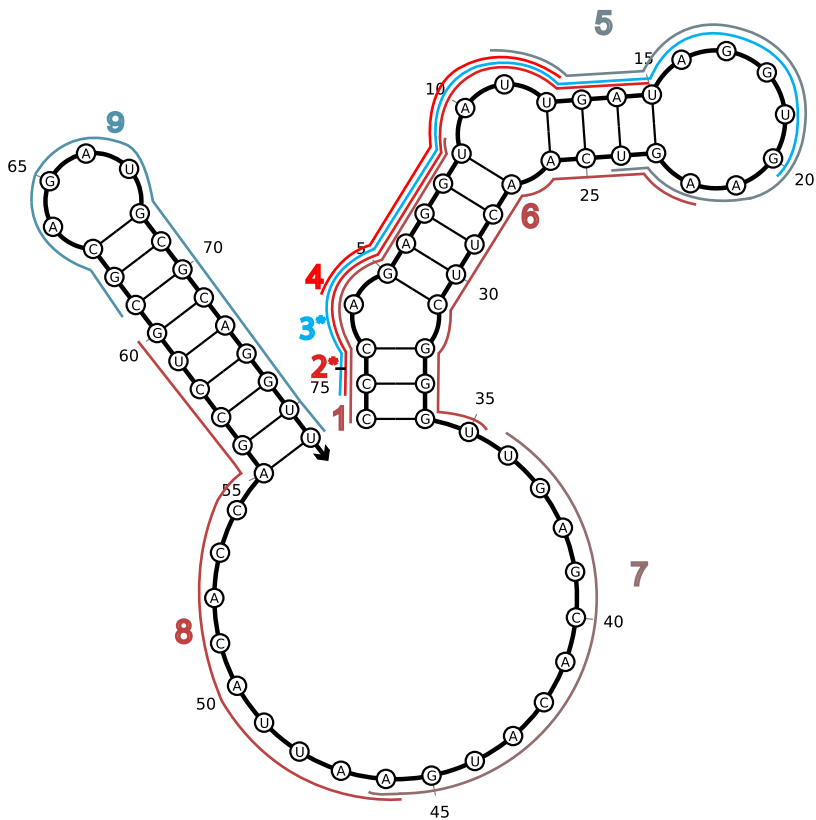
Nc5



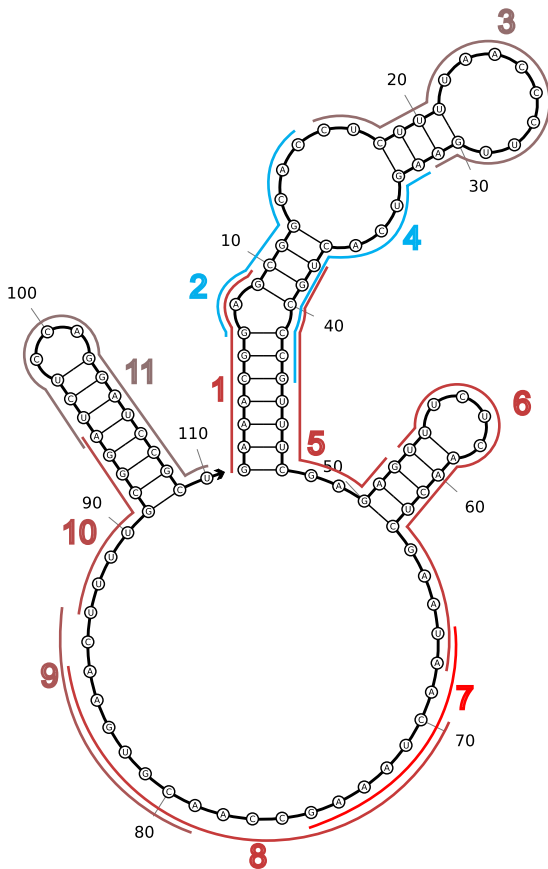
OmrA



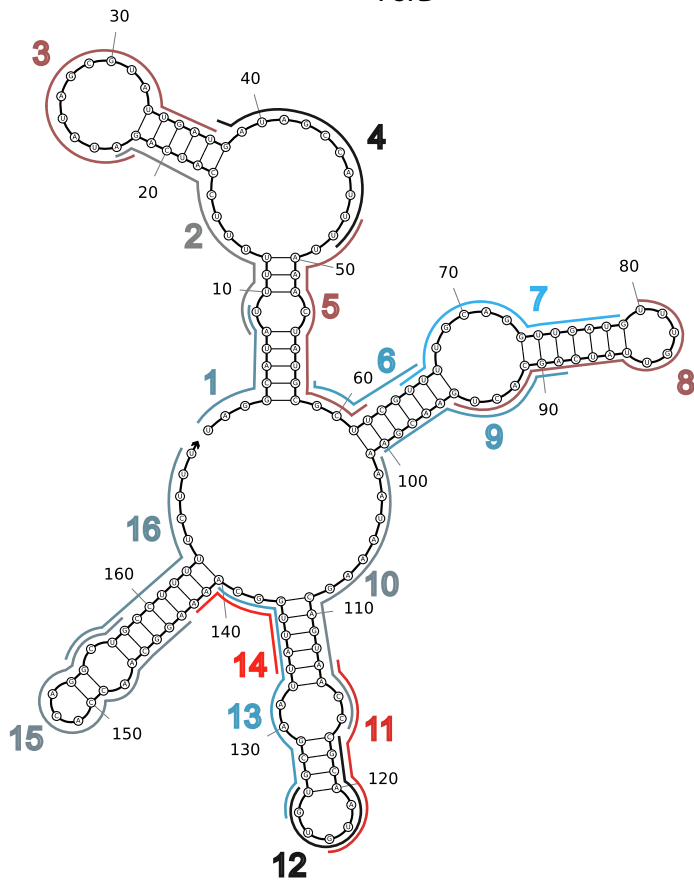
OmrB



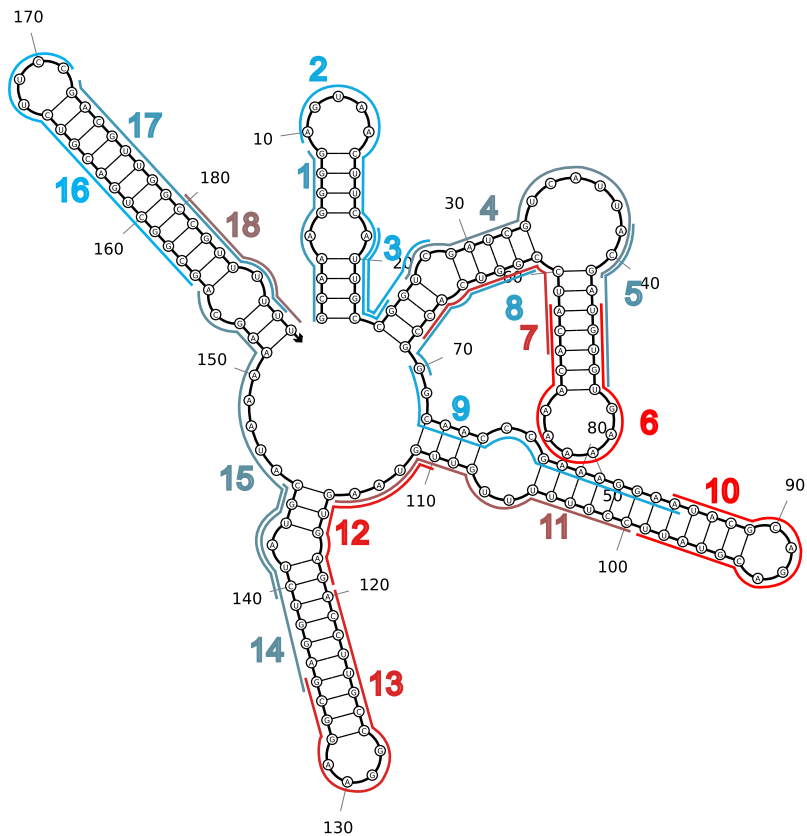
OxyS



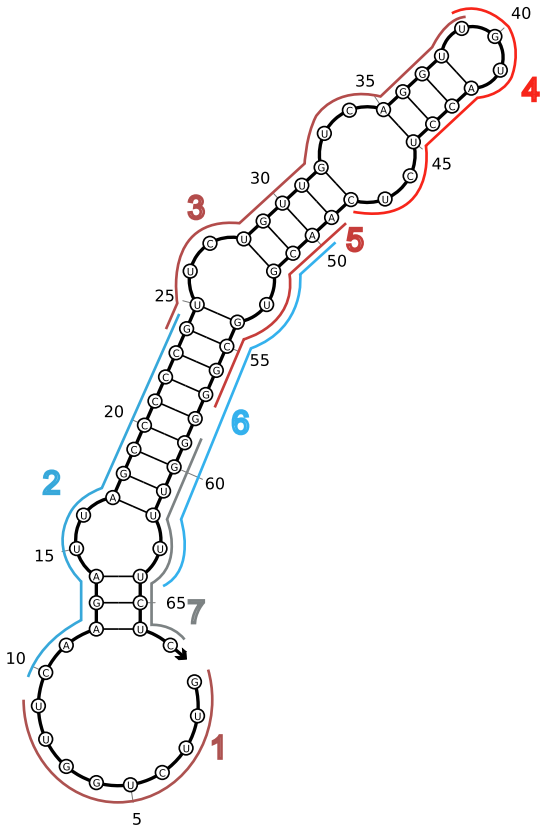
PsrD



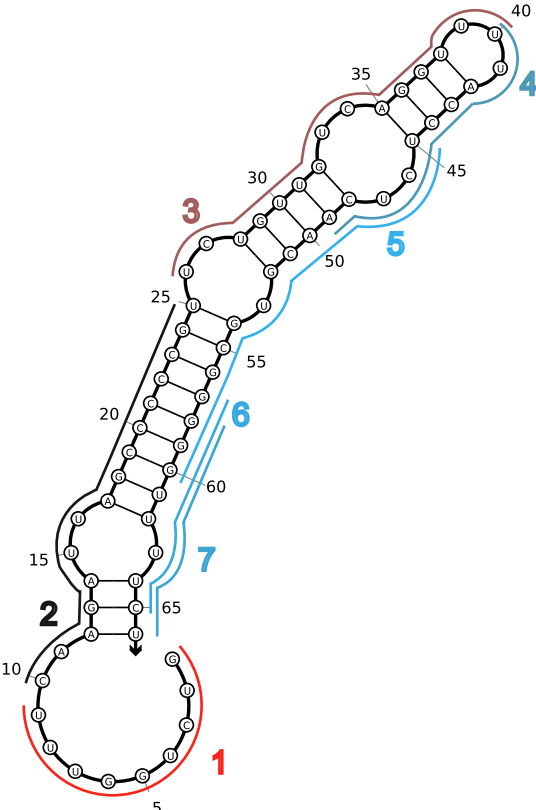
PsrN



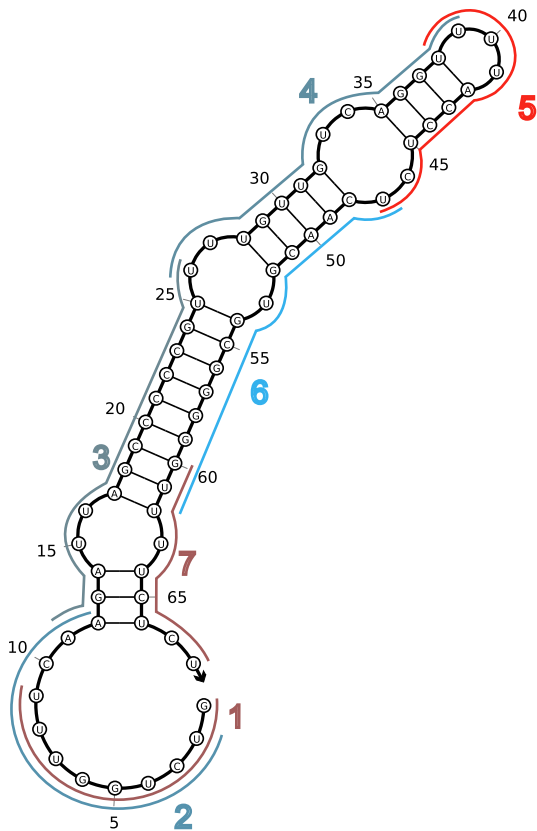
RdIA



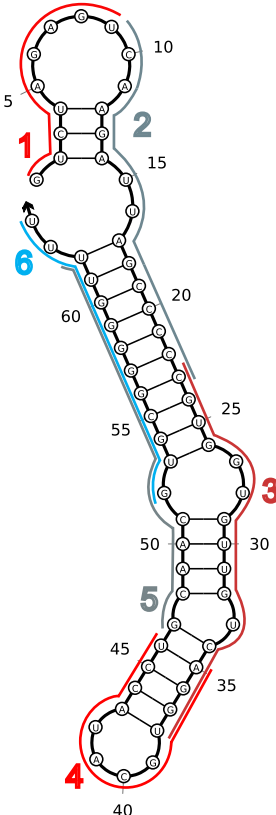
RdIB



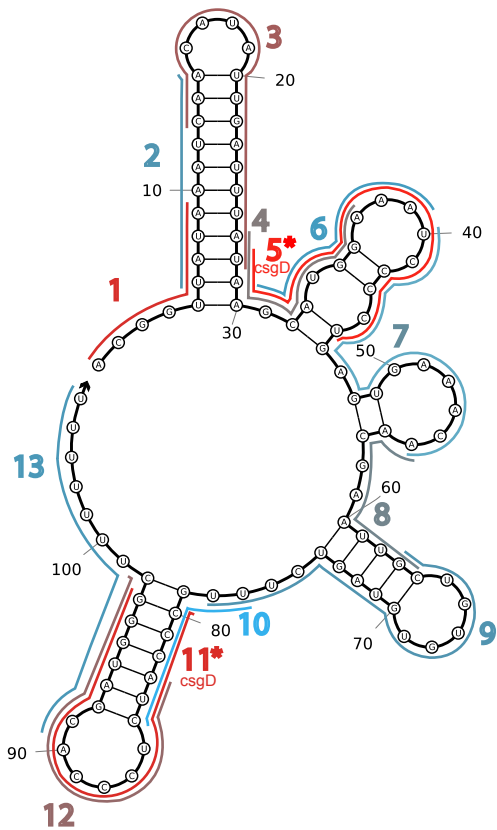
RdIC



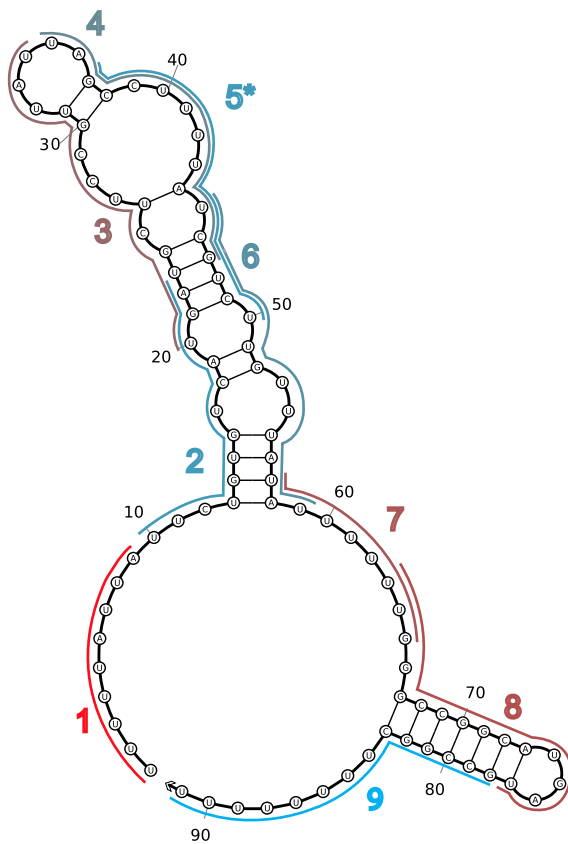
RdID



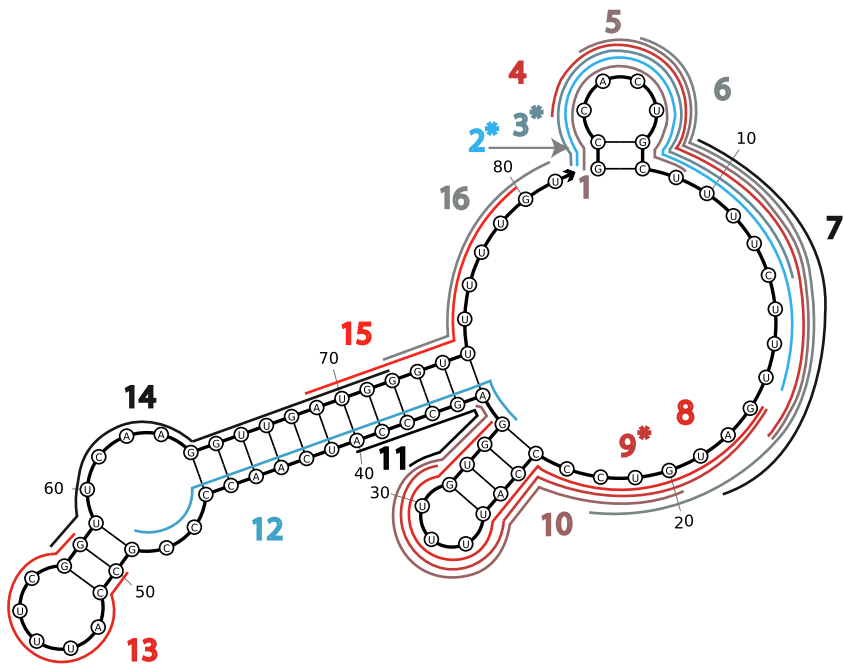
RprA



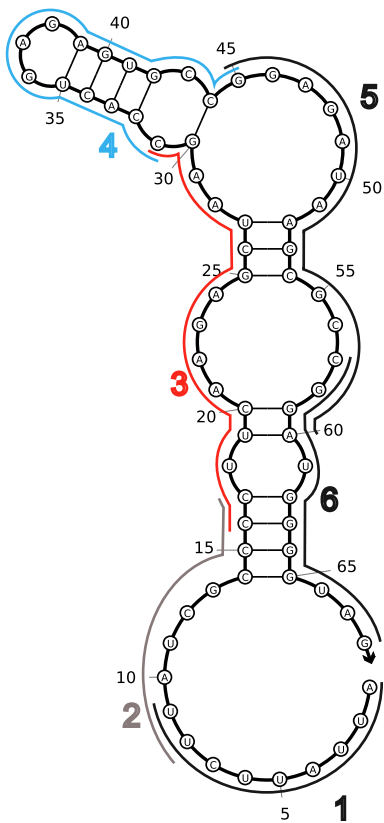
RseX



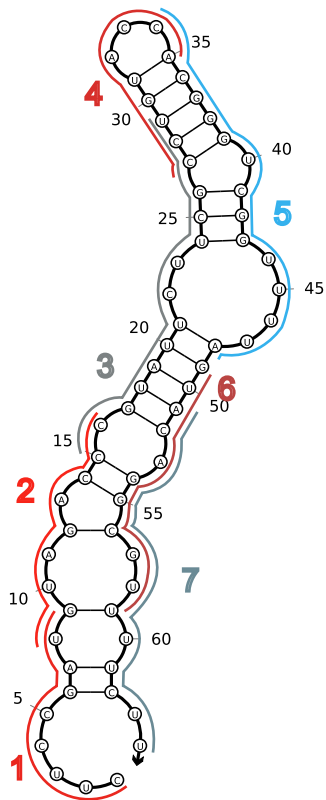
RybB



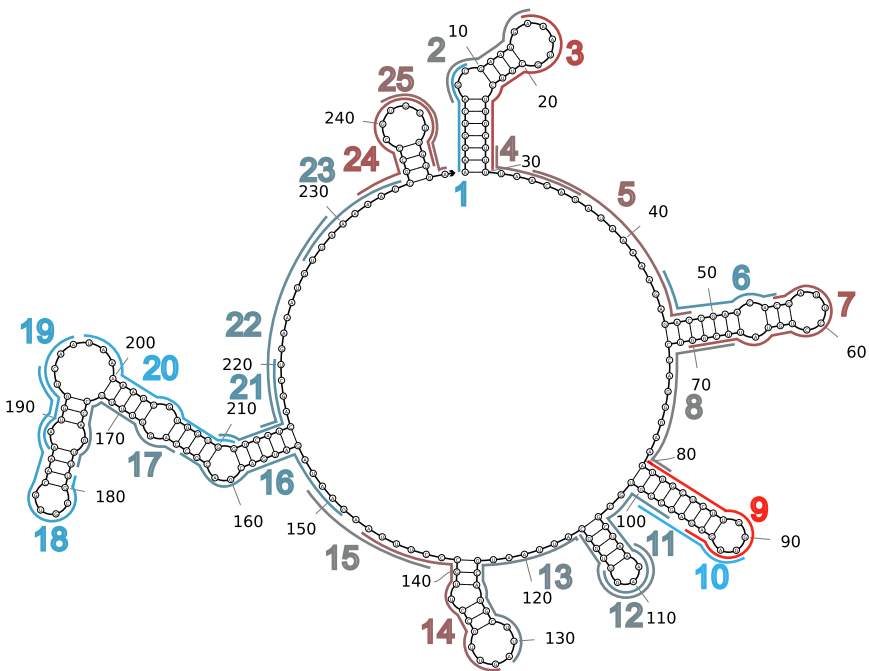
RydB



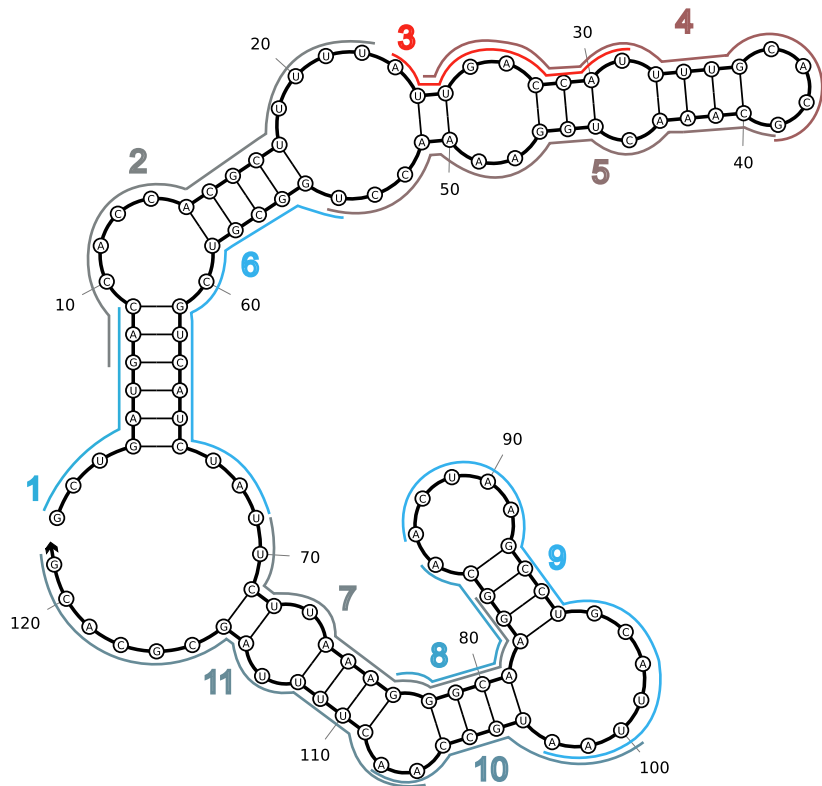
RydC



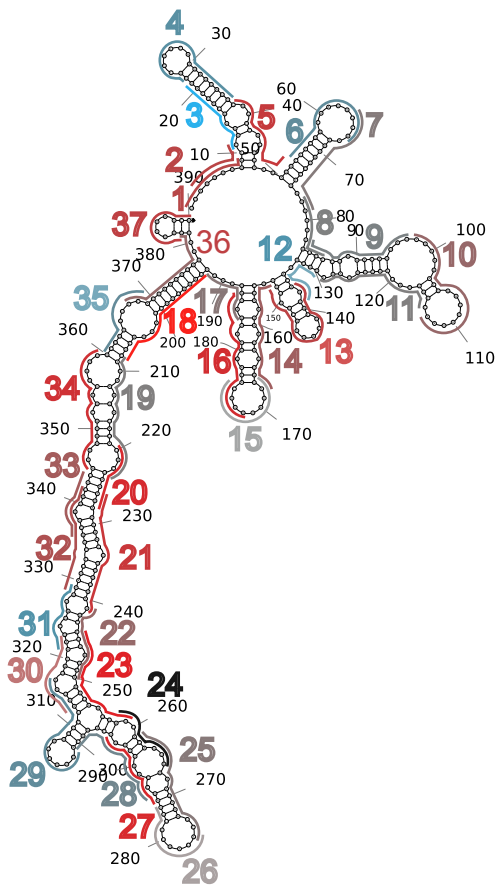
RyeA



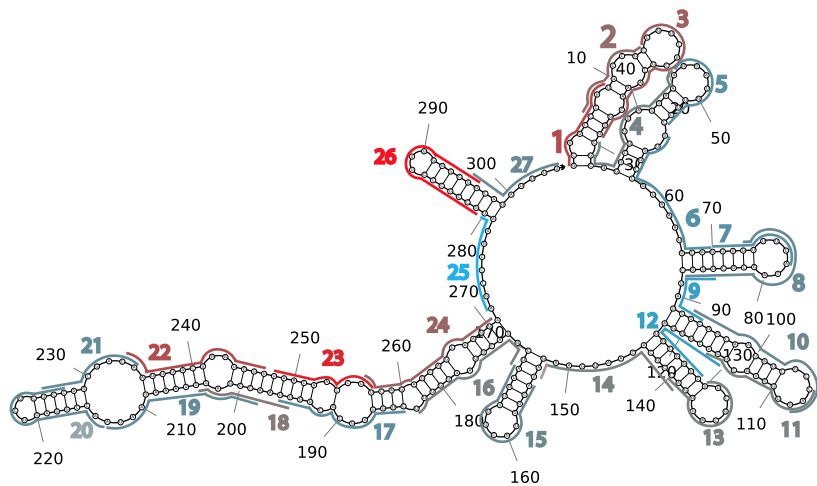
RyeB



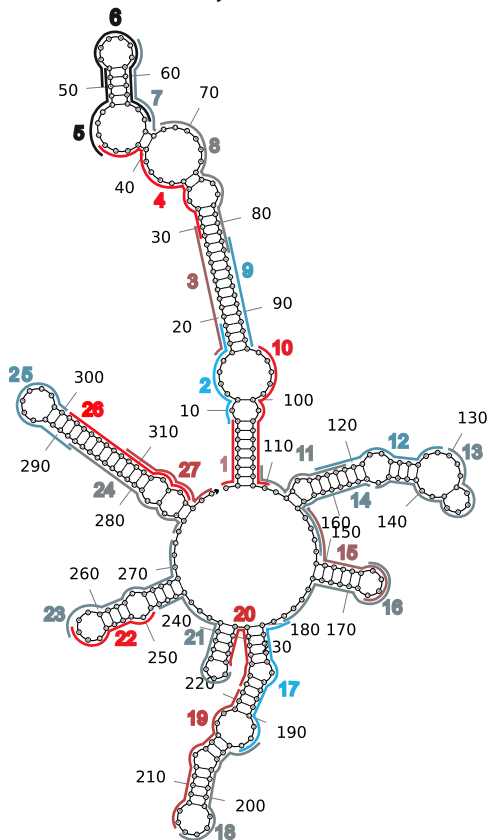
RyeF



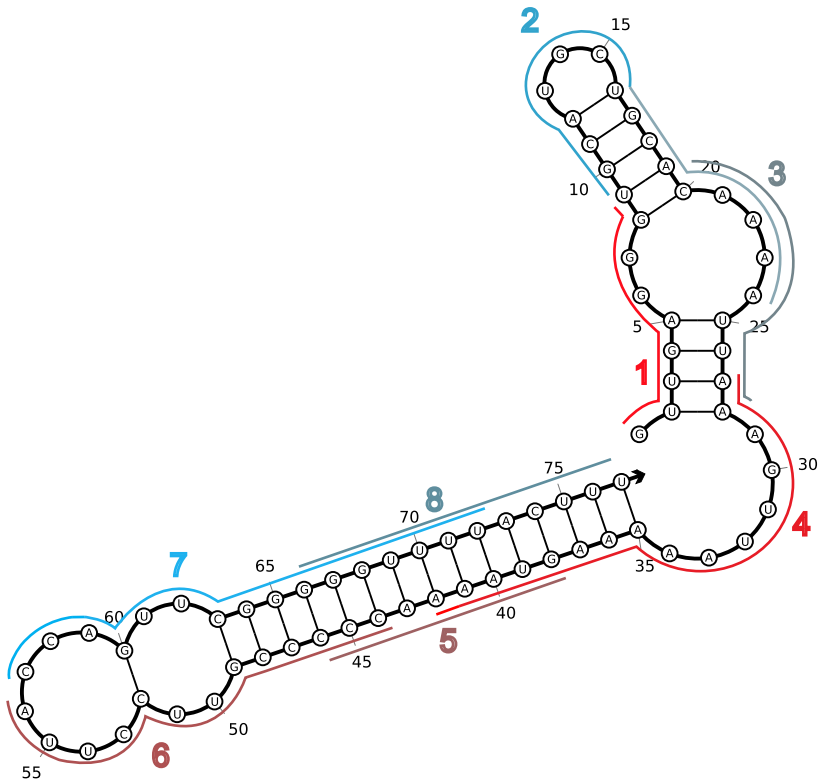
RyfA



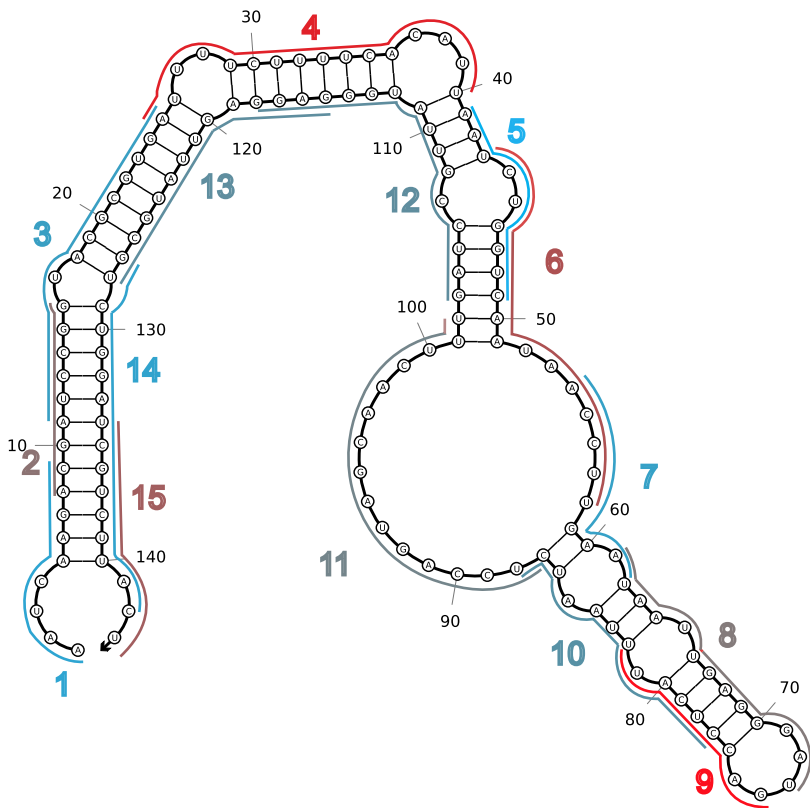
RyfB



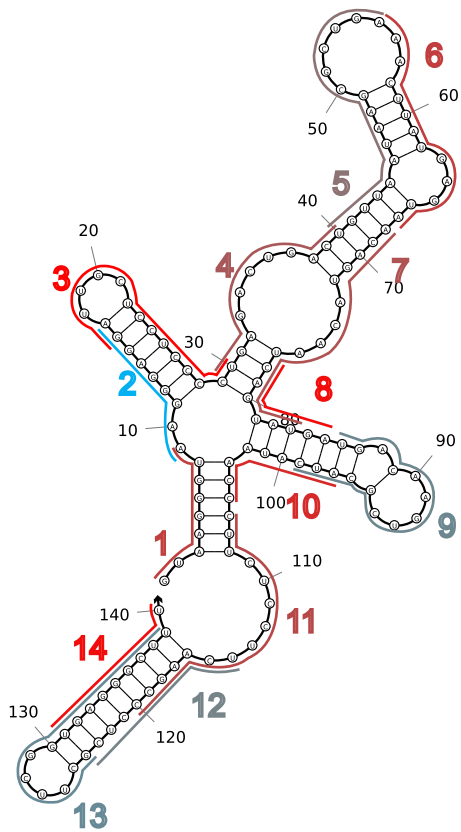
RyfC



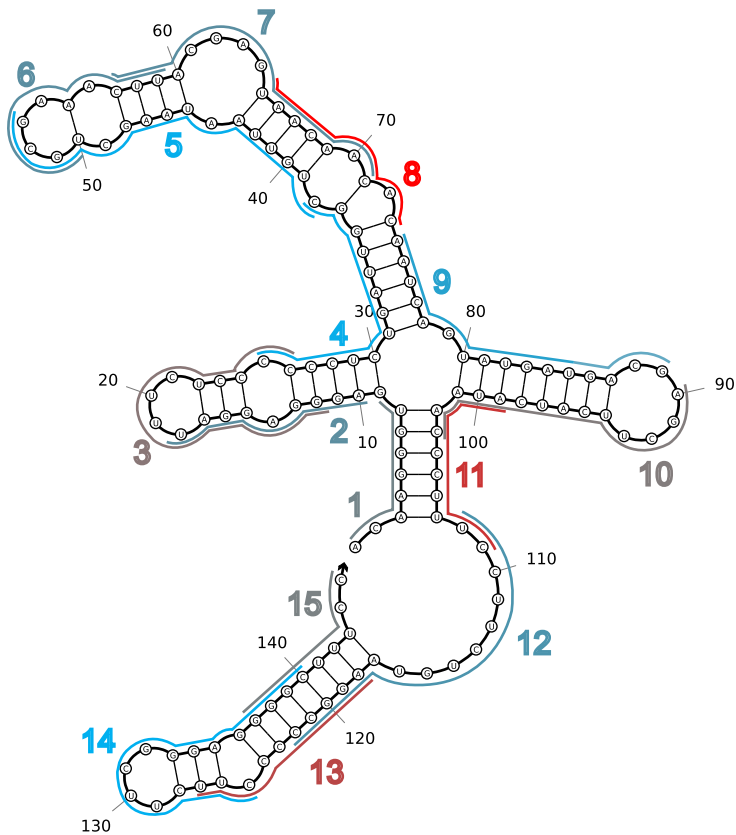
RyfD



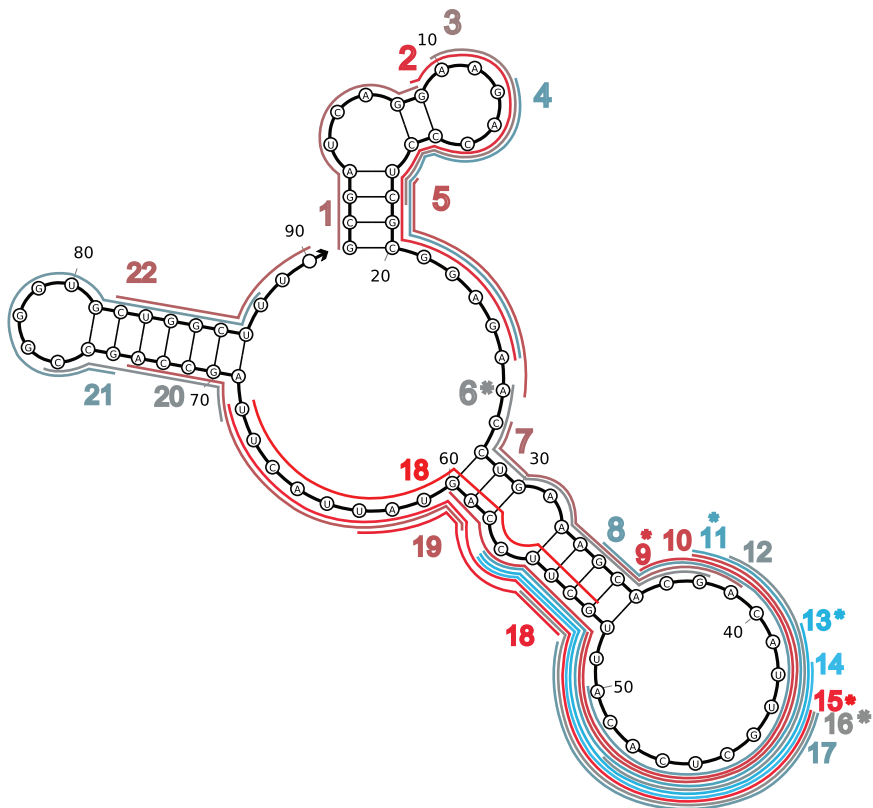
RygC



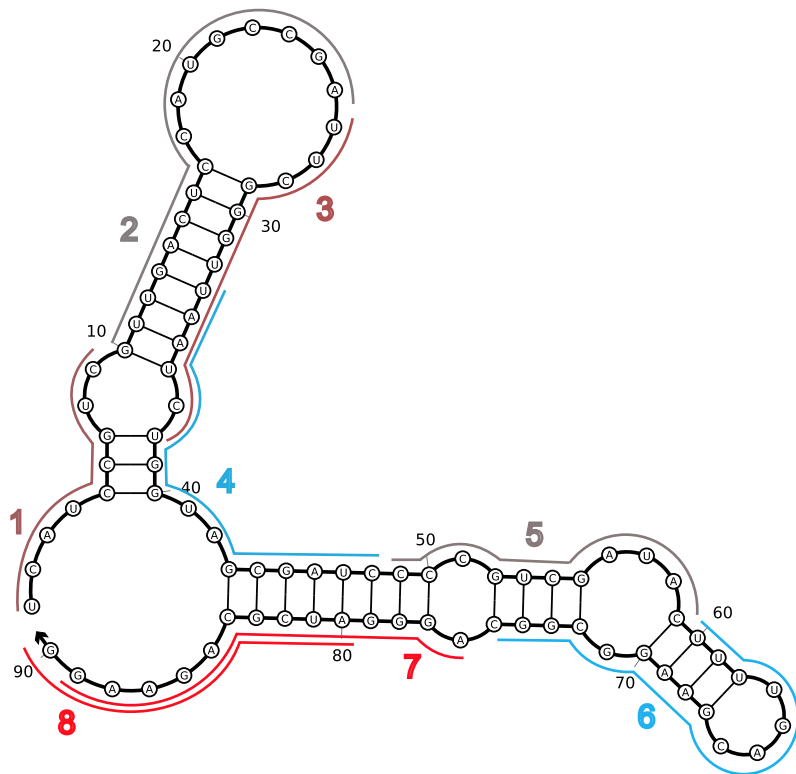
RygD (SibD)



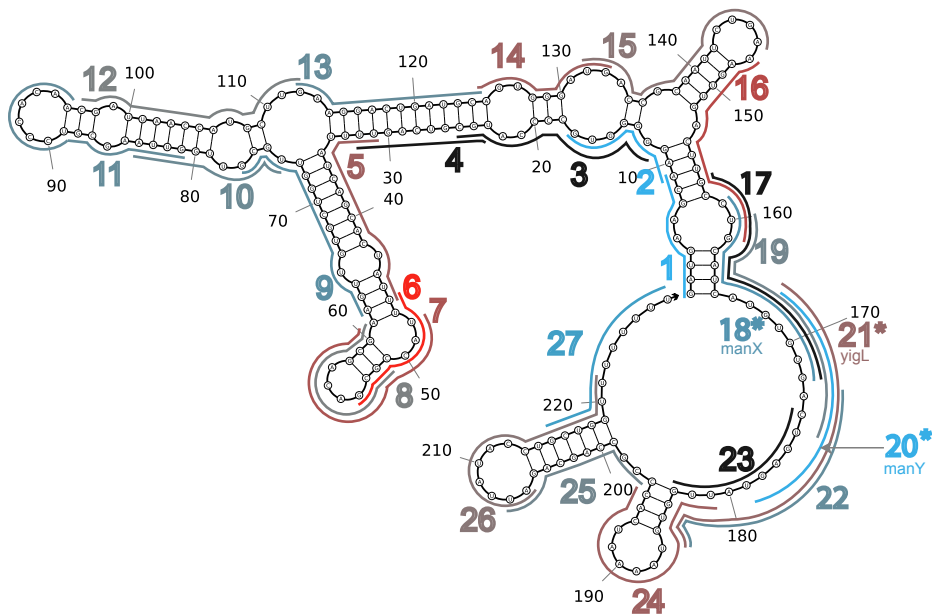
RyhB



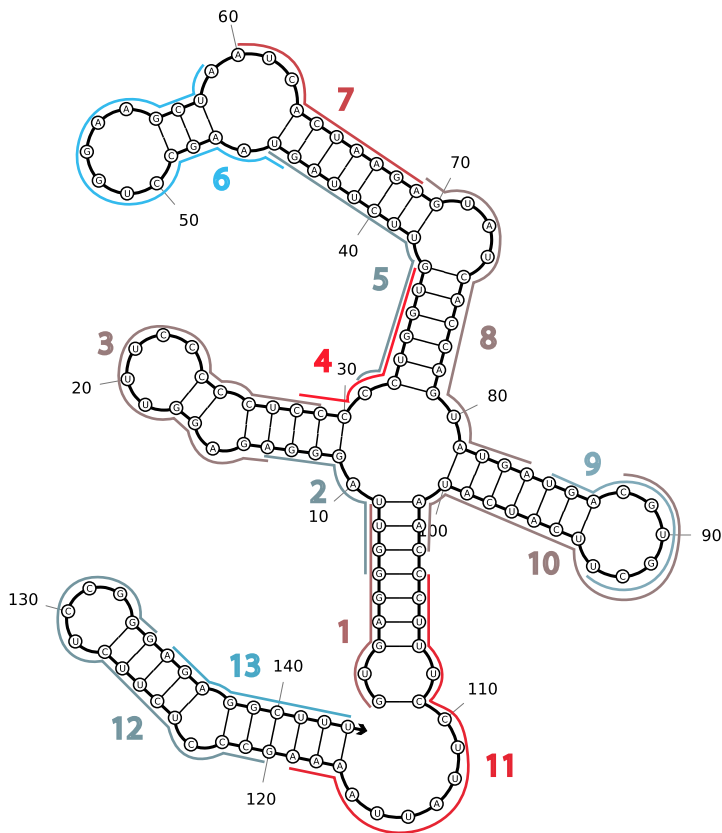
RyjB



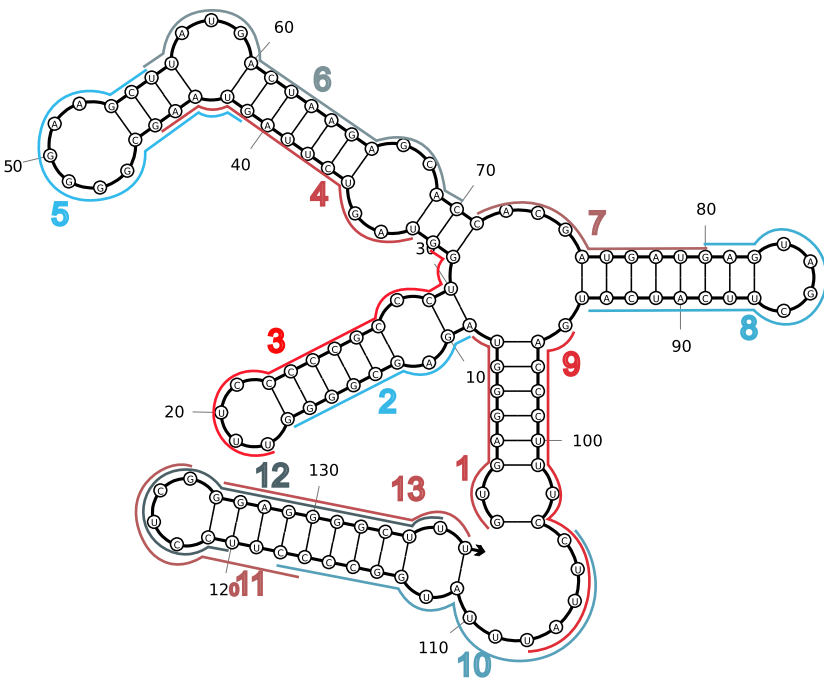
SrgS



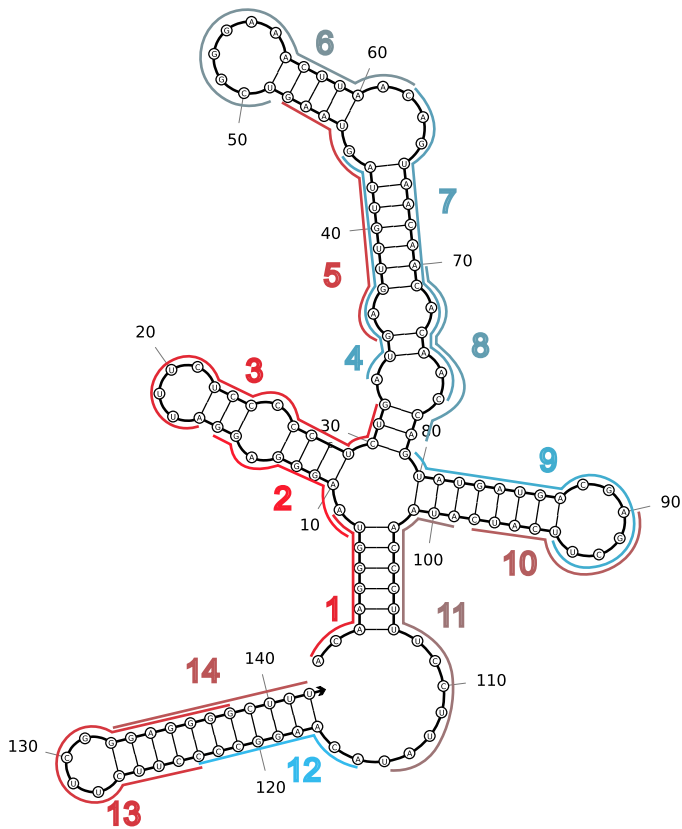
SibA



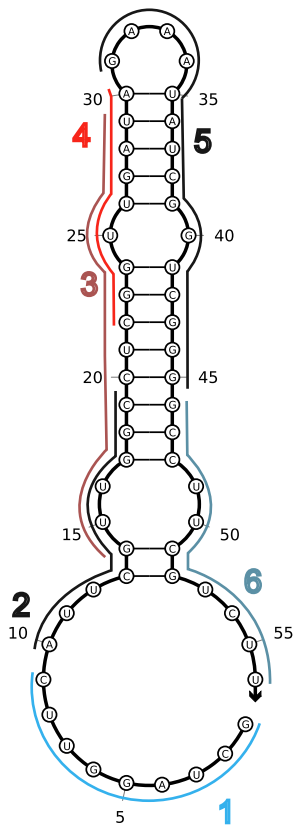
sibB



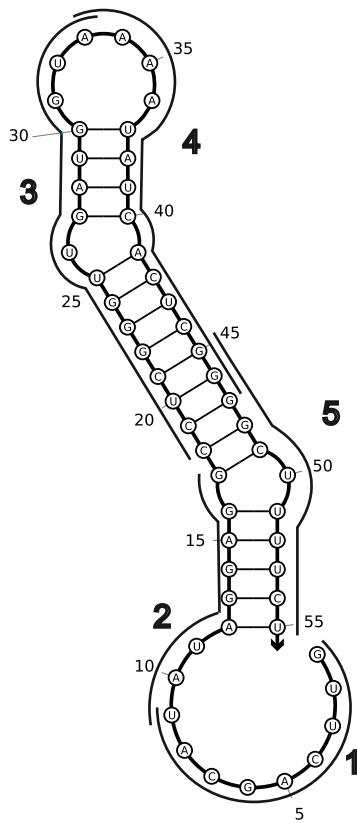
SibE



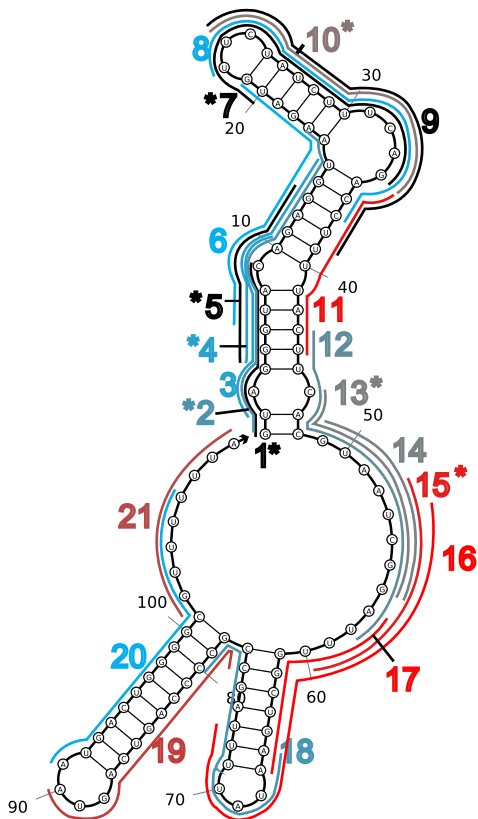
SokB



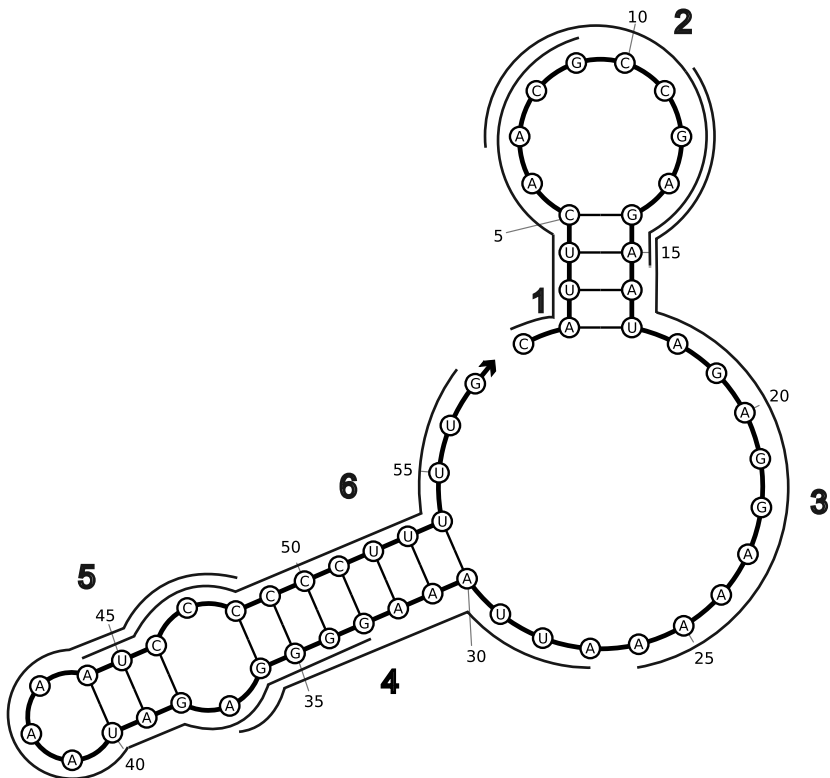
sokC



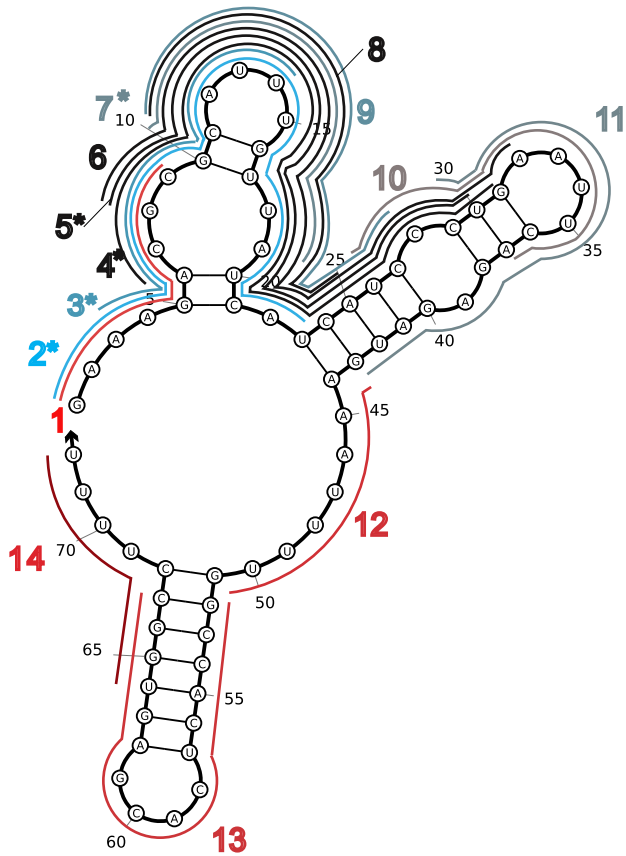
Spot 42



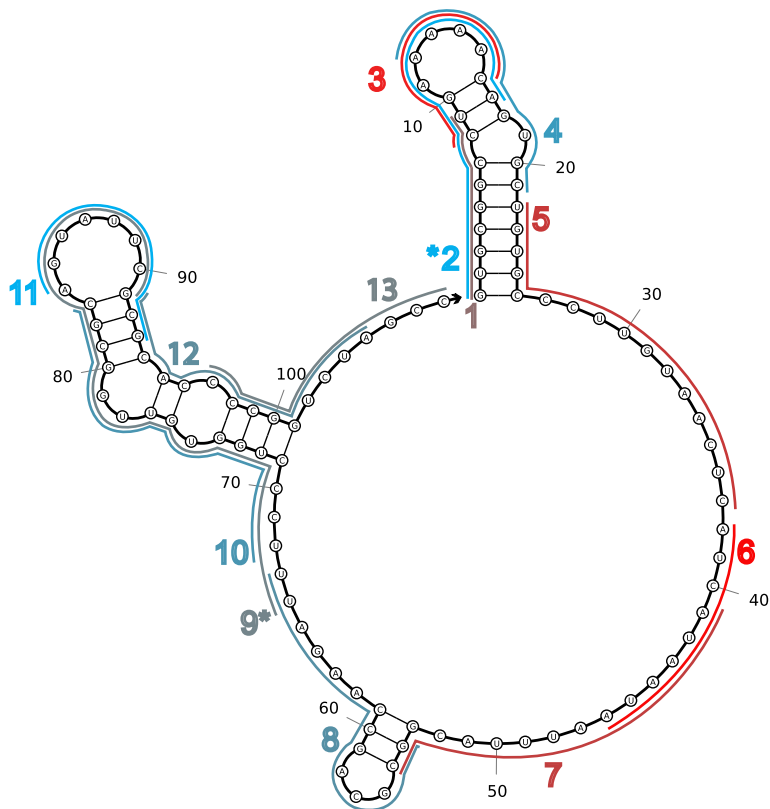
SraA



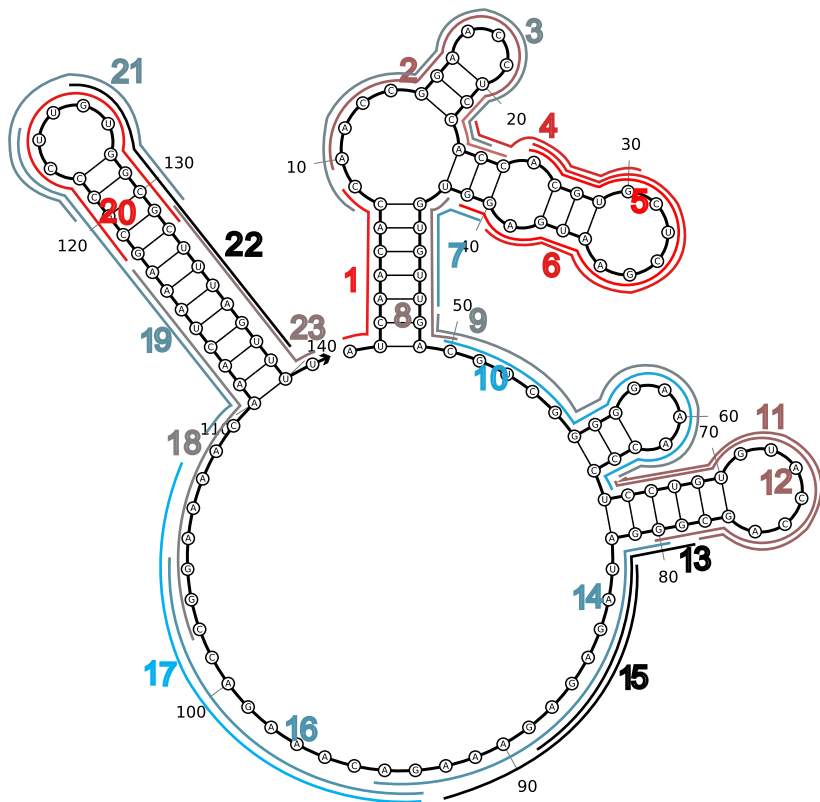
SraD (MicA)



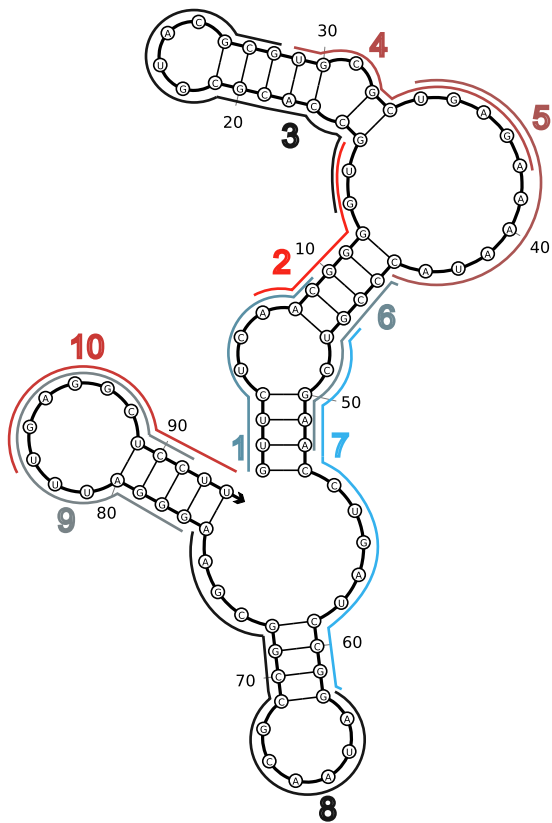
SraH (ArcZ)



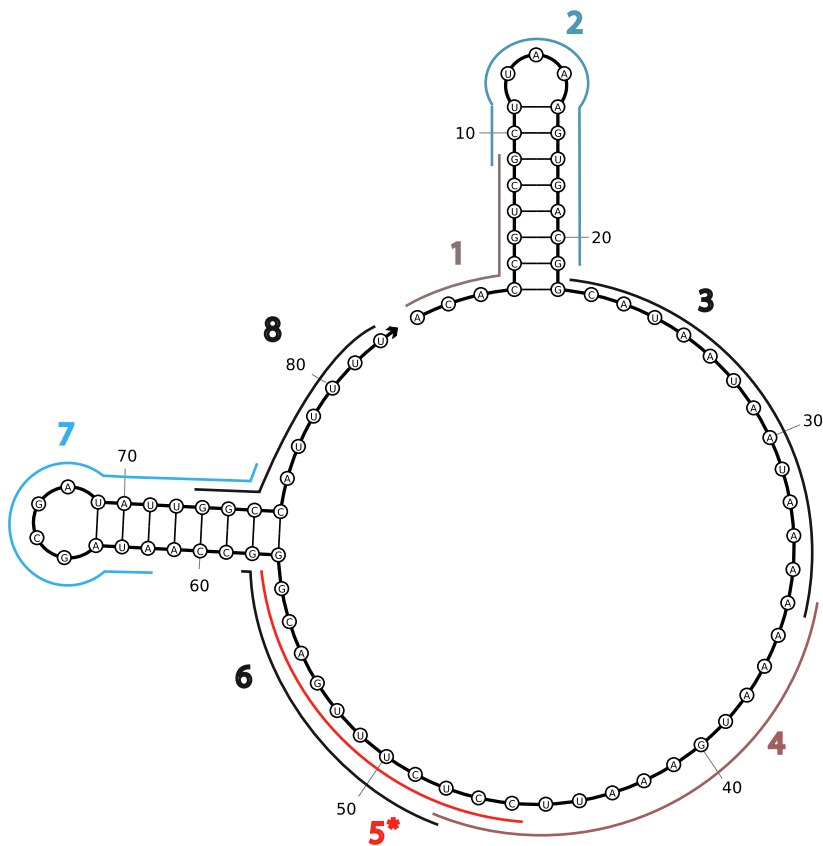
SraL (RyjA)



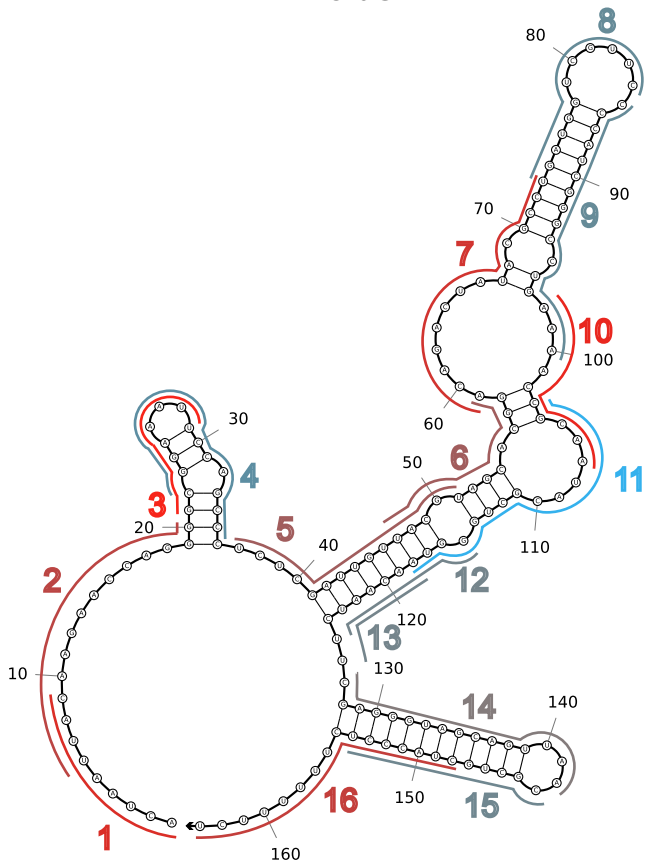
SroA



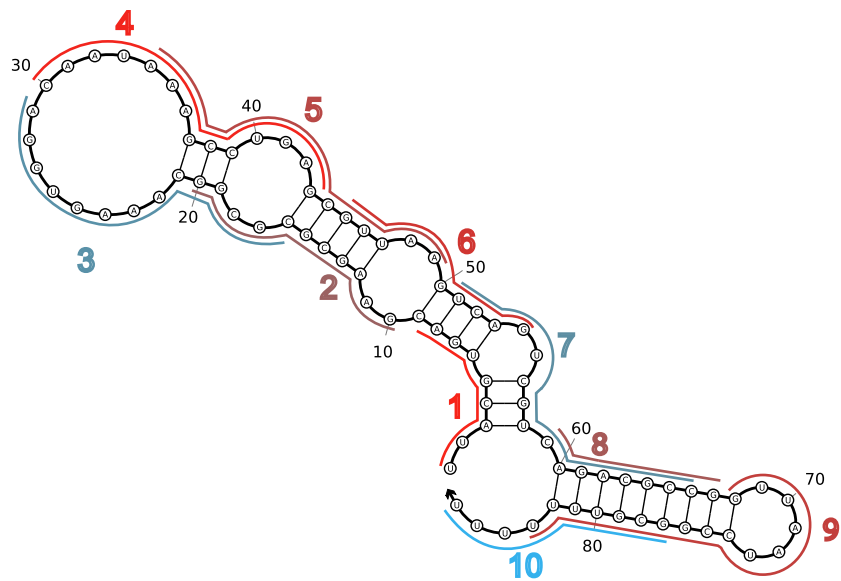
SroB (ChiX)



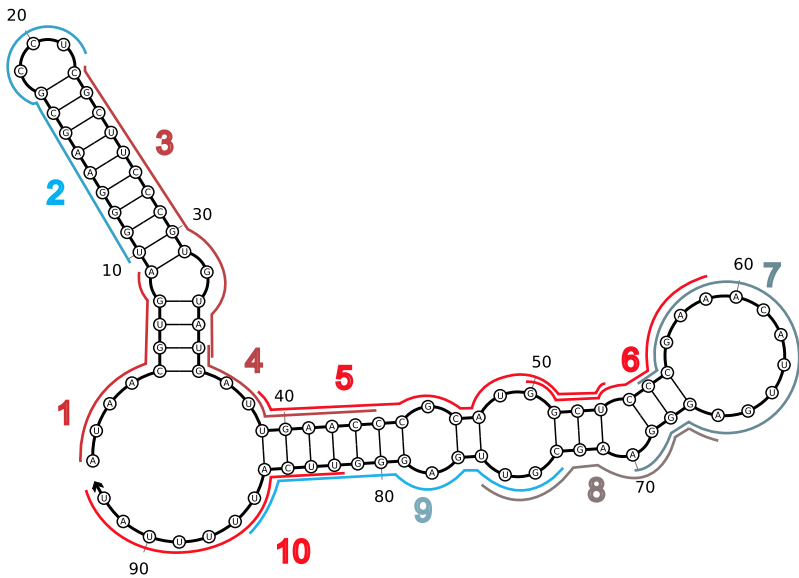
SroC



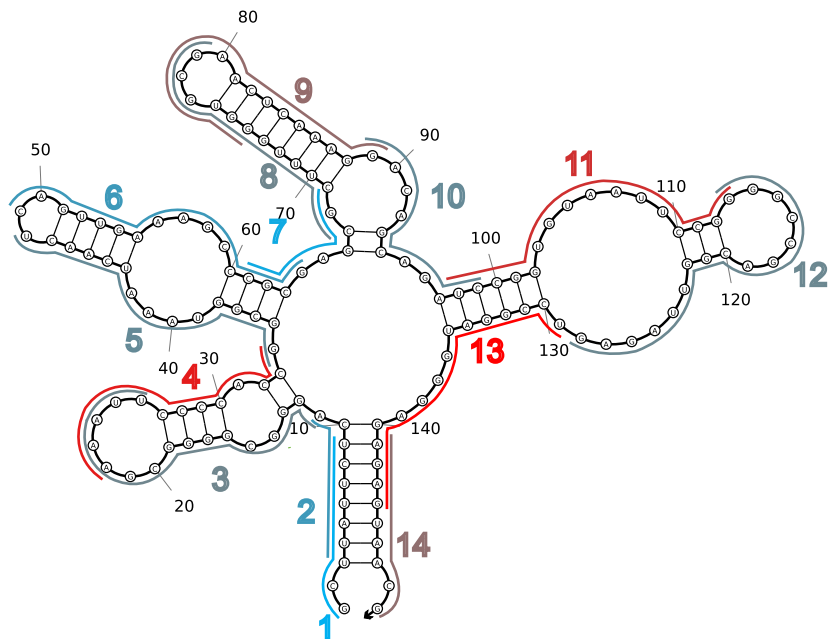
SroD



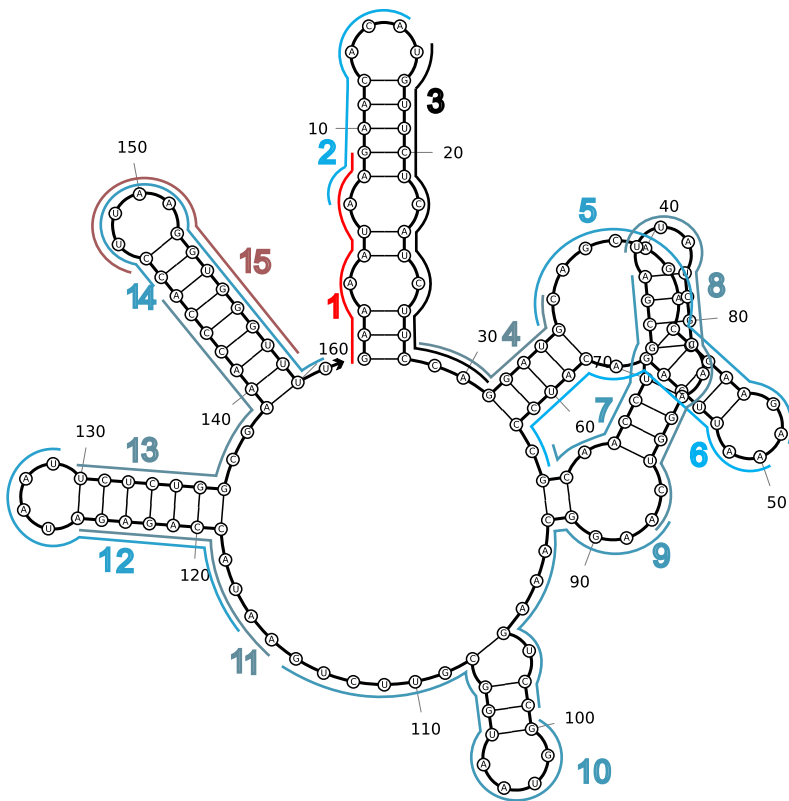
SroE



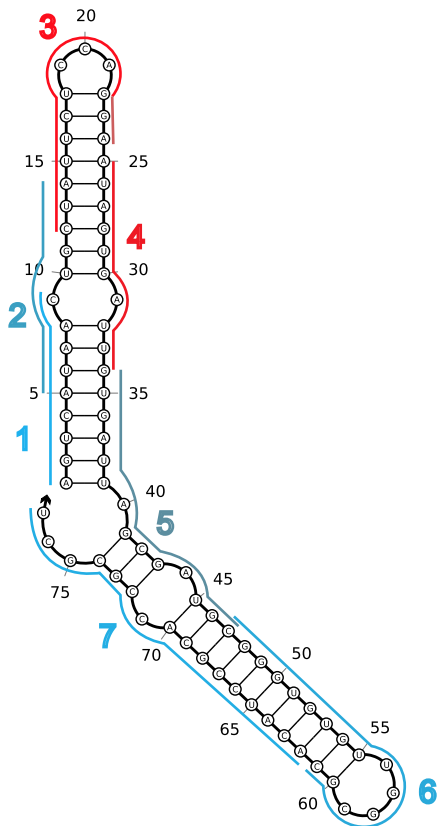
SroG



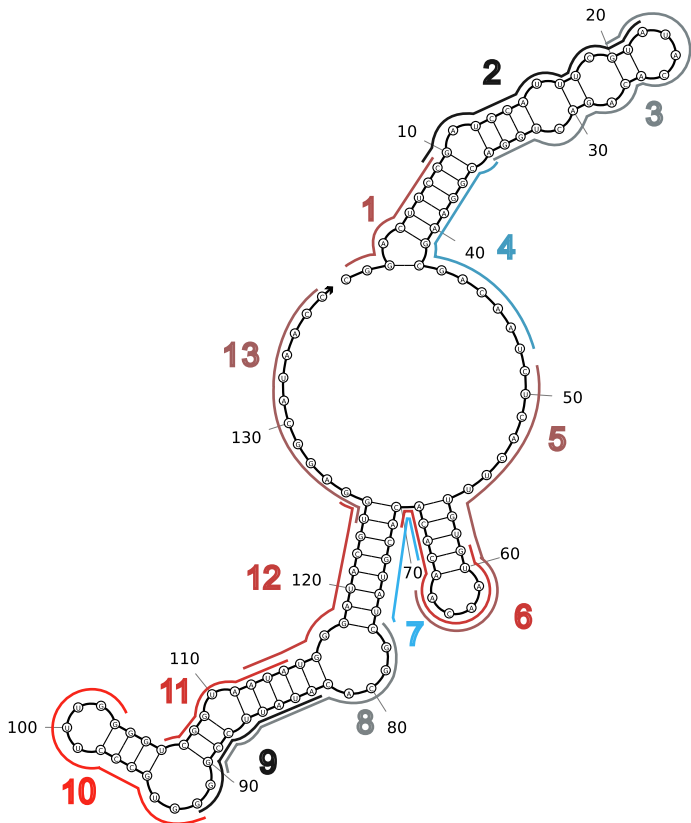
SroH



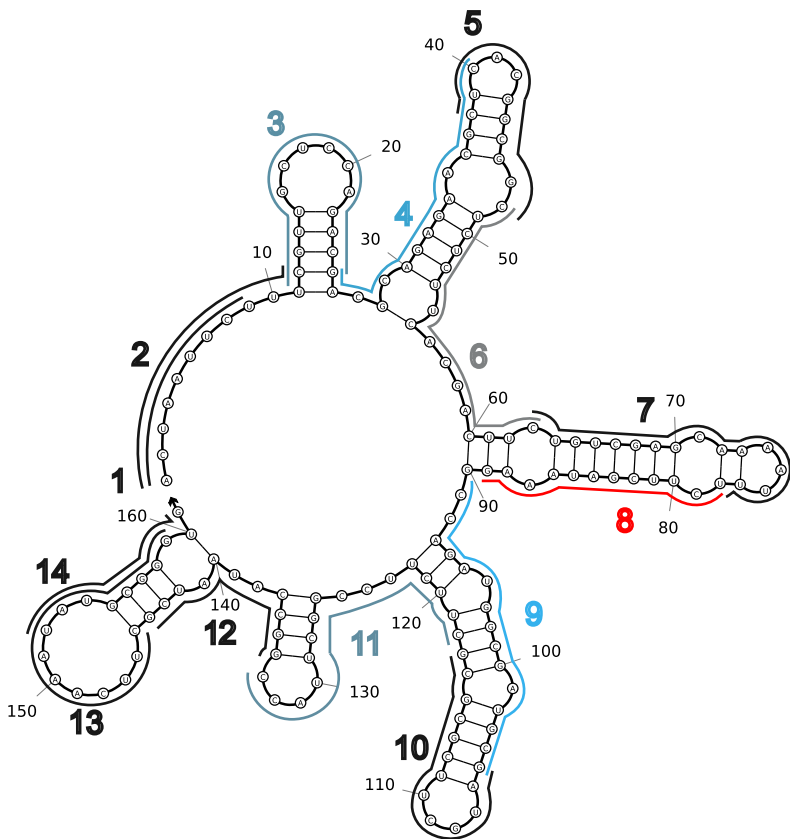
SymR



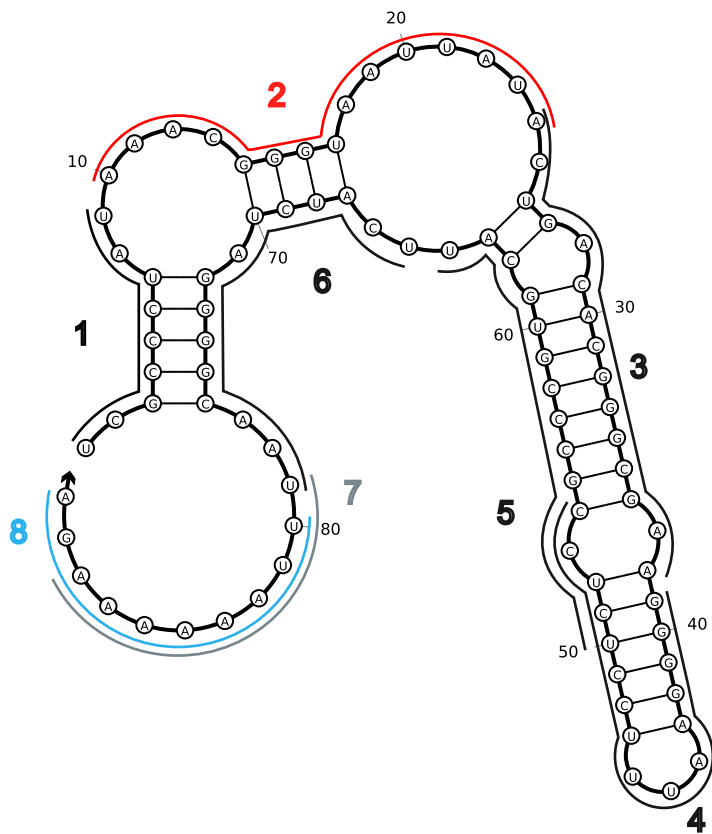
Tff



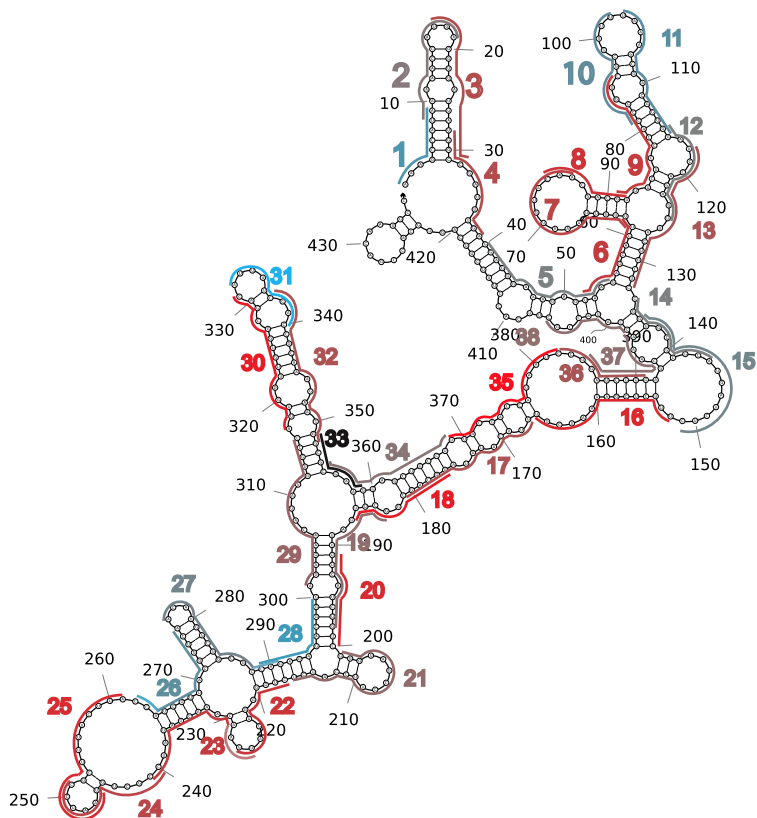
Tp2



Tpke11

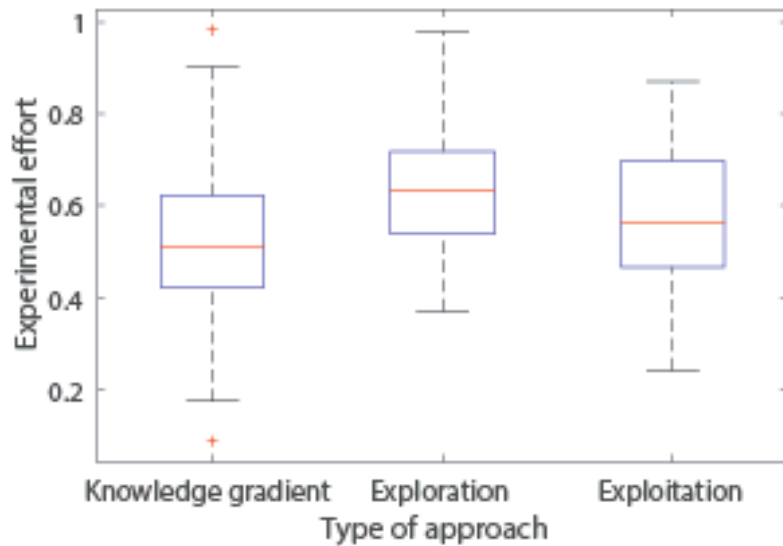


Tpke70



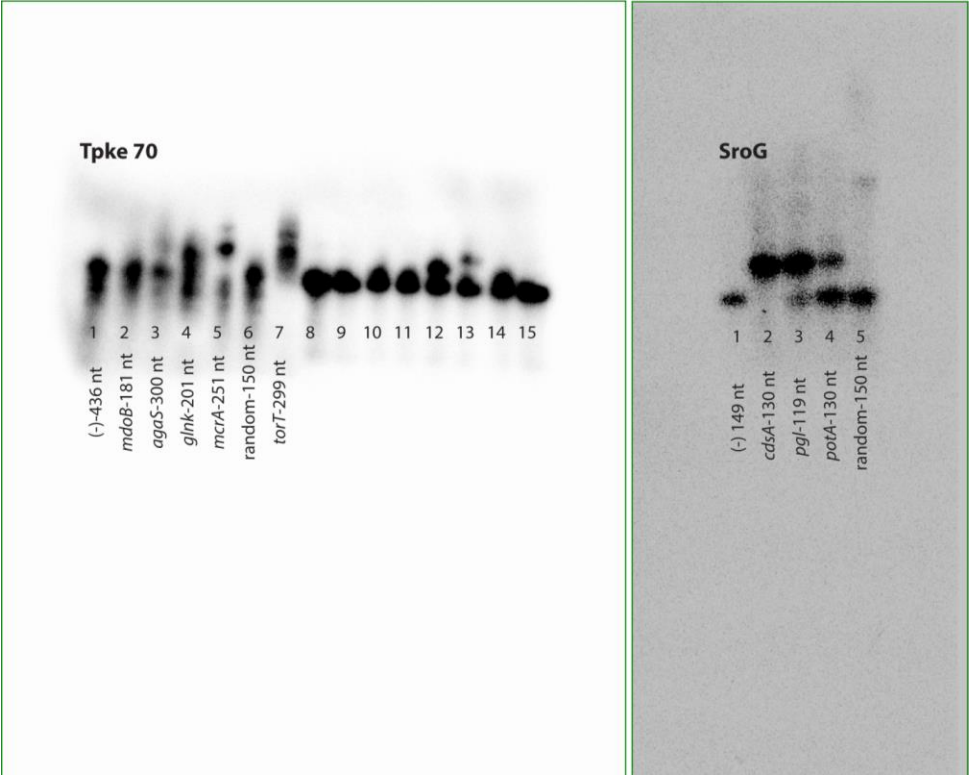
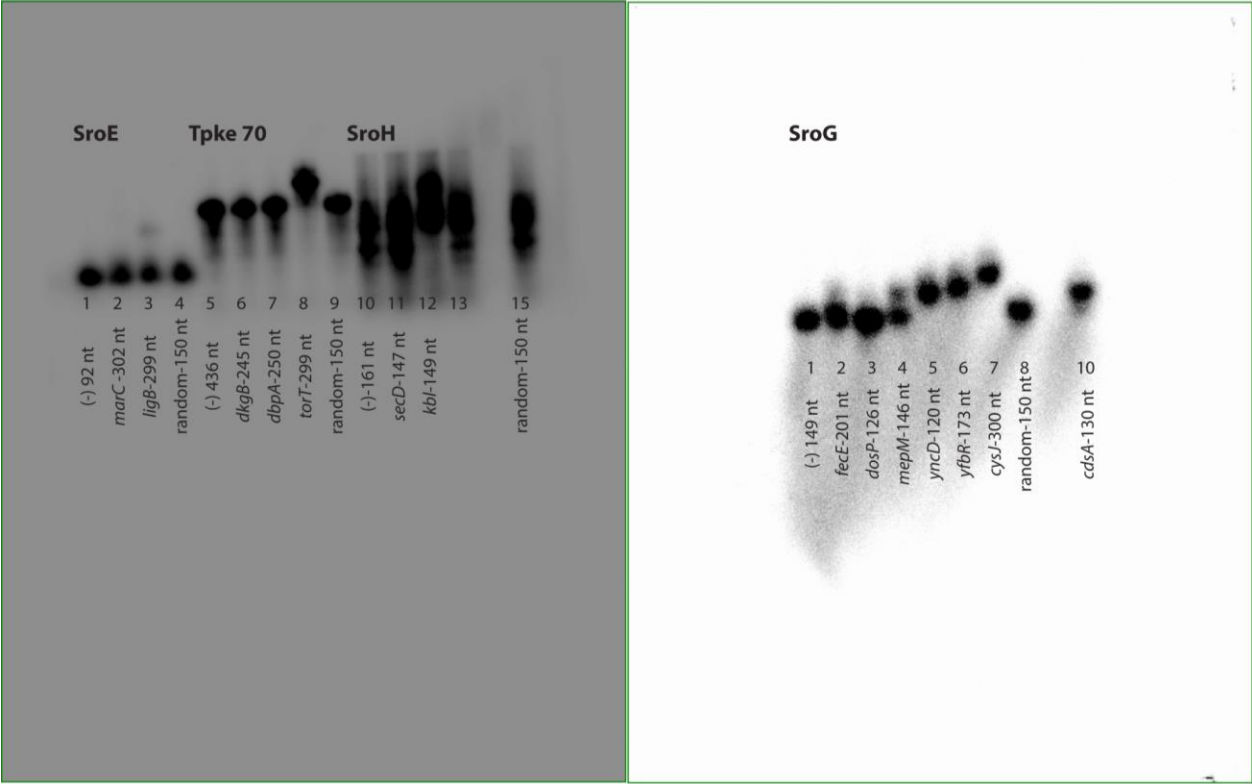
Supplementary Figure 3. INTERFACE accessibility heat maps for each sRNA molecule analyzed.

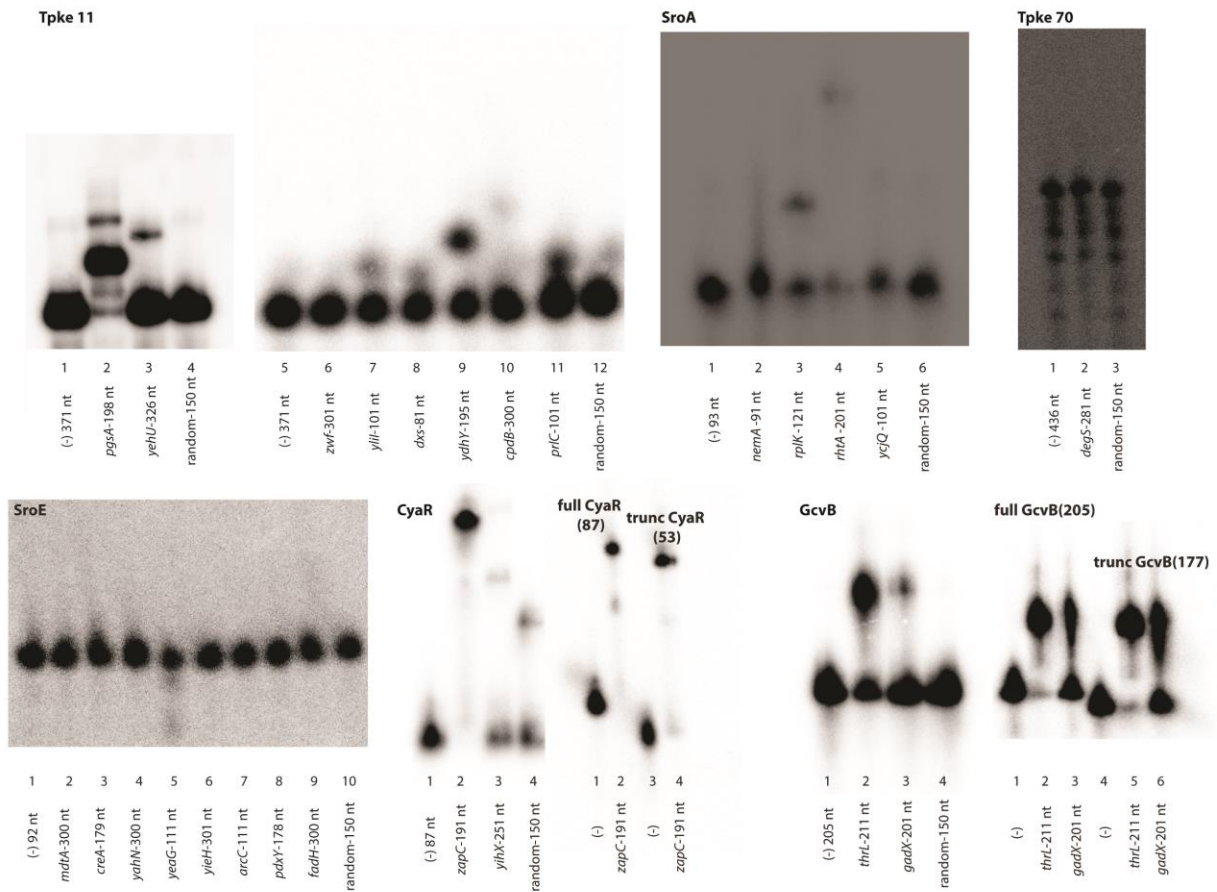
Red, gray, and blue traces along predicted secondary structures¹ represent high, mid and low accessibility, respectively. Mapped binding sites that composed the training set are indicated with asterisks. To initially test the selectivity of using INTERFACE to propose functional regions, we identified regions exhibiting 5' extreme accessibility end of 3 well-characterized sRNAs that were excluded from our training set (IstR, MicL/RyeF, and RydC). Notably, many of the selected regions agree with confirmed mRNA binding sites; for instance, the confirmed *tisB* mRNA binding site within the sRNA IstR² and the *lpp* mRNA binding site within the sRNA MicL/RyeF³ overlap 5' regions identified as highly accessible (IstR taRNA region 1 and RyeF taRNA region 5, respectively). Furthermore, a likely functional site in sRNA RydC (taRNA region 1) is consistent with a site in its *Salmonella* homologue that is confirmed to interact with the *cfa* leader⁴, suggesting the equivalent regulation in *E. coli*.



Supplementary Figure 4. Performance of target region design using an SpKG-based algorithm informed by accessibility predictions.

Center lines, box limits, whiskers, and red “+” markers of box plots represent the median, lower and upper quartiles, 1.5x interquartile range, and outliers of each group, respectively. Simulated comparisons of regional selection by the described machine learning algorithm to other two other region selection approaches— (i) "relative and random design" (machine learning exploration) or (ii) the adapted biophysical model only (exploitation)— shows that the SpKG approach coupled to a previously-established biophysical model reduced experimental INTERFACE profiling efforts by enriching the number of accessible regions assayed (p-value<0.001, 2-tailed t-test). Furthermore, using exploitation (only the biophysical model) requires less experimental effort than the exploration approach (p-value<0.001, 2-tailed t-test).





Supplementary Figure 5. Consecutive lanes of all EMSAs performed for this study. Sequential lanes of all EMSAs (corresponding to unique sRNAs) referenced in the manuscript are pictured, excluding those for GlmY (Fig. 6 in manuscript). Full-gel images are shown for the three benchmarked under-characterized sRNAs (top). Unlabeled lanes 8-15 in bottom left full-gel image were rerun for greater spatial resolution (Tpke11 lanes 5-12, equivalent order) due to the large size of Tpke11 coupled with the small sizes of some corresponding predicted targets. Length of all sRNAs are listed below negative sRNA-only control lanes (indicated by “(-)” below the corresponding well), with the exception of gels containing truncated sRNAs, in which case the length of the sRNA is indicated above corresponding wells. Identity and length of all mRNAs tested are indicated underneath respective gel lanes.

Supplementary References

1. Zadeh, J.N. et al. NUPACK: Analysis and design of nucleic acid systems. *Journal of Computational Chemistry* **32**, 170-173 (2011).
2. Darfeuille, F., Unoson, C., Vogel, J. & Wagner, E.G.H. An Antisense RNA Inhibits Translation by Competing with Standby Ribosomes. *Molecular Cell* **26**, 381-392 (2007).
3. Guo, M.S. et al. MicL, a new $\sigma(E)$ -dependent sRNA, combats envelope stress by repressing synthesis of Lpp, the major outer membrane lipoprotein. *Genes & Development* **28**, 1620-1634 (2014).
4. Fröhlich, K.S., Papenfort, K., Fekete, A. & Vogel, J. A small RNA activates CFA synthase by isoform-specific mRNA stabilization. *The EMBO Journal* **32**, 2963-2979 (2013).