RNAmutants: A web server to explore the mutational landscape of RNA secondary structures

Jerome Waldispühl[1,2,†],Srinivas Devadas[2,3],Bonnie Berger[1,2,*],Peter Clote[4,†,*]

[1] Department of Mathematics, MIT, Cambridge, MA 02139, USA
[2] Computer Science and Artificial Intelligence Laboratory, MIT, Cambridge, MA 02139, USA
[3] Electrical Engineering and Computer Science, MIT, Cambridge, MA, 02467, USA
[4] Department of Biology, Boston College, Chestnut Hill, MA 02467, USA,

Running title: RNAmutants web server

Key words: RNA, mutation, partition function, secondary structure, sampling, deleterious mutation, purifying selective pressure.

[†] Joint first authors.

[⋆] Corresponding authors: Email: clote@bc.edu, Phone: (617) 552-1332, Fax: (617) 552-2011 and bab@mit.edu, Phone: 617-253-1827, Fax: 617-258-5429.

Abstract

The history and mechanism of molecular evolution in DNA has been greatly elucidated by contributions from genetics, probability theory and bioinformatics – indeed, mathematical developments such as Kimura's neutral theory, Kingman's coalescent and efficient software such as *BLAST*, *ClustalW*, *Phylip*, etc. provide the foundation for modern population genetics. In contrast to DNA, the function of most non-coding RNA depends on tertiary structure, experimentally known to be largely determined by secondary structure, for which latter, dynamic programming can efficiently compute the the minimum free energy secondary structure. For this reason, understanding the effect of pointwise mutations in RNA secondary structure could reveal fundamental properties of structural RNA molecules and improve our understanding of molecular evolution of RNA. The web server *RNAmutants* provides several efficient tools to compute the ensemble of low energy secondary structures for all k-mutants of a given RNA sequence, where k is bounded by a user-specified upper bound. As we have previously shown, these tools can be used to predict putative deleterious mutations and to analyze regulatory sequences from the hepatitis C and human immunodeficiency genomes.

Availability: Web server `http://bioinformatics.bc.edu/clotelab/RNAmutants/`, downloadable binaries `http://rnamutants.csail.mit.edu/`.

Introduction

Understanding the molecular evolution of DNA has proven essential to modern biology. One of the main fields that has contributed to our understanding of molecular evolution is population genetics, in its modern form founded by R. Fisher and S. Wright (2; 3) in the early part of last century, when they posed and partially solved the question of expected time (number of generations) for gene allele fixation or extinction, known subsequently as the (discrete) Fisher-Wright problem. This difficult problem of probability theory was solved using various techniques, including the Fokker-Planck single-variable diffusion equation (2; 3; 4; 5), the coalescent (6; 7), and a direct analysis of Markov chains (8). The Fisher-Wright model forms the foundation of Kimura's widely accepted *neutral* theory of molecular evolution, now a a cornerstone of modern genetics (9).

A mutation in a protein coding gene may be deleterious depending on whether it causes a change of the coded amino acid. A measure of selective pressure on protein coding genes is the term K_a/K_s (also known as dN/dS), which is the ratio of the rate of non-synonymous substitutions (K_a) to synonymous substitutions in a protein coding region (CDS). In contrast, a mutation in a non-protein coding RNA gene may be deleterious if the underlying functional structure is changed. At present there is no widely adopted measure of selective pressure in non-coding RNA genes; however, as explained in (1), *RNAmutants* can be used to quantify the deleterious nature of pointwise mutations in non-coding RNA genes. The rationale for the consideration of mutational effects on RNA secondary structure is explained in the next paragraph.

The function of structural non-coding RNA (ribozymes (10), riboswitches (11), precursor microRNA (12), selenocysteine insertion sequence (SECIS) elements (13), transfer RNA, etc.) depends on tertiary structure, which Banerjee et al. (14) have shown experimentally to largely depend on secondary structure. Secondary structure can be predicted using dynamic programming energy minimization (15); indeed, Mathews et al. (16) have shown that the minimum free energy (MFE) structure, as determined in *mfold* (17) or *RNAfold* (18), includes 73% of the base-pairs in the native[1] secondary structure, on average, when tested on RNA sequences of length 700 nt.

Computational tools like *mfold* of R. Zuker (19), *Vienna RNA Package* of I.L. Hofacker et al. (20), *RNAStructure* of D.H. Mathews and D.H. Turner (21), *Sfold* of Y. Ding et al. (22; 23), *RNAbor* of E. Freyhult et al. (24; 25) and *RNAsat* of J. Waldispühl and P. Clote (26) probe the landscape of secondary structures of a given RNA sequence. RNA sequence/structure alignment tools like *Dynalign* by D.H. Mathews and D.H. Turner (27), FOLDALIGN by J.H. Havgaard et al. (28), *MSARI* of A. Coventry et al. (29), *RNAz* of S. Washietl et al. (30), etc. can be considered to be the RNA analogue of *BLAST* and *ClustalW*, whereby conservation of secondary structure base pairing is taken into account.

Understanding the effect of pointwise mutations on RNA secondary structure reveals fundamental properties of structurally important RNA and may suggestmay suggest potentially deleterious mutations in RNA viral pathogens. Designed explicitly for this purpose, the algorithm *RNAmutants* (1) allows users to analyze the low energy ensemble of mutant RNA sequences and structures. Given an RNA sequence $\mathbf{s}$ of length n, an upper bound K for the number of mutations allowed, a desired number N of secondary structures samples to be generated, and a temperature $0 \leq T \leq 100$ in degrees Celsius, *RNAmutants* computes the following for all $k \leq K$ simultaneously: *(i)* the minimum free energy structure MFE_k^T, its free energy and the Boltzmann partition function Z_k^T, over all secondary structures of all k-point mutants, *(ii)* a plot of the ensemble free energy $-RT \ln Z_k^T$,

[1] as inferred from the X-ray structure or by comparative sequence analysis

as a function of k, and *(iii)* a collection of N RNA mutant sequences and their secondary structures, as sampled using the partition function. By comparing low energy structures from mutant RNA with the consensus structures from the Rfam database (31), one can infer putative deleterious mutations, as performed in (1).

Definitions & Methods

Definitions

Given RNA sequence $\mathbf{s} = \mathbf{s}_1, \ldots, \mathbf{s}_n$, for all $0 \leq k \leq n$, let Z_k^T denote the Boltzmann partition function at absolute temperature T for the collection of all secondary structures on all k-point mutants; i.e.

$$Z_k^T = \sum_{\mathbf{s}', d_H(\mathbf{s},\mathbf{s}')=k} \sum_{\mathcal{S}} e^{-E(\mathcal{S})/RT} \tag{1}$$

where the first sum is taken over all k-point mutants $\mathbf{s}' = \mathbf{s}'_1, \ldots, \mathbf{s}'_n$ of $\mathbf{s} = \mathbf{s}_1, \ldots, \mathbf{s}_n$, and the second sum is taken over all secondary structures $\mathcal{S}$ of the (fixed) k-point mutant. Similarly, let mfe_k^T denote the k-point mutant $\mathbf{s}' = \mathbf{s}'_1, \ldots, \mathbf{s}'_n$ of $\mathbf{s}$ whose secondary structure has least free energy over all k-point mutants of $\mathbf{s}$, and let MFE_k^T denote its secondary structure. In the sequel, mfe_k^T is called the k-superoptimal mutant and MFE_k^T is called the k-superoptimal secondary structure. Finally, we let Z_k, mfe_k, MFE_k denote the corresponding values at default temperature $T = 37°$ C.

Partition Function and Superoptimal Structures

In (32), we introduced a novel algorithm to compute the partition function Z_k^T for all k-point mutants of a given RNA sequence at absolute temperature T, with respect to the Nussinov energy model (33). In contrast to the Nussinov energy model, where each base pair contributes energy term of -1, the widely accepted Turner energy model (34) includes negative, stabilizing free energy terms for *stacked* base pairs as well as positive, destabilizing free energy terms for hairpins, bulges, internal loops and multiloops. With the exception of multiloops, for which an affine approximation is applied, these free energy parameters were obtained from UV absorption (optical melting) experiments first pioneered by Tinoco's Lab (35) and systematically carried out by Turner's Lab (36; 34). For instance, at 37° C, Turner's rules assign stacking free energy of -2.24 kcal/mol to $\begin{array}{l} 5'\texttt{-AC-}3' \\ 3'\texttt{-UG-}5' \end{array}$ and of -3.26 kcal/mol to $\begin{array}{l} 5'\texttt{-CC-}3' \\ 3'\texttt{-GG-}5' \end{array}$.

In (37), Waldispühl et al. developed a general algorithm AMSAG, applicable both to RNA and transmembrane protein structure prediction. Subsequently, Clote et al. (32) designed an algorithm to computer the partition function Z_k^T with respect to the Nussinov energy model (33), and applied AMSAG to determine the k-superoptimal secondary structures with respect to an energy model intermediate between the Nussinov and Turner models. Recently, Waldispühl et al. (1) created a unified framework for simultaneously computing k-superoptimal secondary structures MFE_k^T as well as the partition functions Z_k^T with respect to the full Turner energy model. The resulting program, *RNAmutants*, was then applied to the analysis of regulatory portions of the hepatitis C and human immunodeficiency viral genomes. Of particular interest is the determination of putative deleterious mutations, many of which were validated in prior experimental work.

Using dynamic programming, *RNAmutants* computes mfe_k^T, MFE_k^T, and Z_k^T for all values of $0 \leq k \leq K$ in worst-case time $O(n^3K^2)$ and space $O(n^2K)$. From statistical mechanics, it is known that the expected internal energy $\langle E_k \rangle$ of all k-point mutants and their secondary structures is equal to RT^2 times the partial derivative of $\ln Z_k^T$, and hence can be approximated using the difference $Z_k^{T+1} - Z_k^T$ (32). Ensemble free energy $-RT \ln Z_k^T$ can be computed as well and plotted as a function of k. Similarly, other thermodynamic parameters (heat capacity, etc.) can be obtained from the partition function.

Web server

Input

The web server `http://bioinformatics.bc.edu/clotelab/RNAmutants` runs on a Linux cluster with head and file server nodes, and 25 compute nodes, including 6 Dell Power Edge 1750, 2 x Intel Xeon P4 (2.80 GHz), 2 GB RAM, 11 Dell Power Edge 1750, 2 x Intel Xeon P4 (2.80 GHz), 4 GB RAM, and 8 Dell Power Edge 1950, 2 x Intel Xeon E5430 Quad core (2.80 GHz), 16 GB RAM.

The input form for *RNAmutants* is shown in Figure 1. The user must submit an RNA sequence, either by pasting in the space provided, or by uploading a file. As well, the user must enter a valid email address,[2] an upper bound for the number of pointwise mutations, the desired number of sampled structures, and optionally the temperature in degrees Celsius. Input for each job is saved under a unique anonymized job ID, sent to the user's email address, thus allowing the user to retrieve information from old runs. As long as the user's browser is open, updates to the results page will be made; however, for long runs, the user will receive an email with job ID and link to the completed results page.

Output

If K denotes the user-specified upper bound for number of mutations, then *RNAmutants* computes for each $k \leq K$ the k-superoptimal sequence mfe_k, secondary structure MFE_k, and free energy E_k, where we recall that the superoptimal secondary structure MFE_k is that which has lowest free energy over all secondary structures of all k-point mutants of the input RNA sequence. Additionally, *RNAmutants* computes for each $k \leq K$ the Boltzmann partition function $Z_k = \sum_{\mathcal{S}} e^{-E(\mathcal{S})/RT}$, and using this computes a sample of structures from the low energy ensemble, following a technique similar to (but distinct from) that of Ding and Lawrence (38). *RNAmutants* output of mfe_k, MFE_k, E_k is depicted in Figure 2, while sampled sequence/structure pairs are given in Figure 3.

By writing scripts to post-process the output, a number of interesting results can be obtained, as exemplified in Figures 4-6. Figure 4 was generated using *RNAplot* and *RNAfold* from the *Vienna RNA Package* (18), using the 51 nt portion of the $3'$ untranslated region from murine beta-galactoside binding protein mRNA, with NCBI accession code MUSGBPA (31). This figure shows the Rfam consensus structure (31), the minimum free energy structure and the 20-superoptimal structure. The upper triangular portion of the left panel of Figure 5 shows the base pairing frequencies over all sampled structures for the 88 nt hepatitis delta virus ribozyme with EMBL accession code `X85253.1/682-769`, while the lower triangular portion shows the base pairs in the minimum free energy structure. (We follow the dot plot conventions of *Vienna RNA Package*.) The right

[2] This email may be bogus; however, for long jobs, which cannot be done interactively, the results will be sent to the email address provided.

panel shows superoptimal and ensemble free energy (y-axis), plotted as a function of number of pointwise mutations (x-axis). Figure 6 displays the mutational profile of the 48 nt HAR1 region, an important region of the novel RNA gene HAR1F (39), expressed in Cajal-Retziusneurons in the developing human neocortex, a gene believed to show significant evolutionary acceleration.

Conclusion

RNAmutants is a novel application which computes, for each $k \leq K$, *(i)* the minimum free energy structure MFE_k^T, free energy E_k^T, and the Boltzmann partition function Z_k^T, over all secondary structures of all k-point mutants, *(ii)* a plot of the ensemble free energy $-RT \ln Z_k^T$, as a function of k, and *(iii)* a collection of RNA mutant sequences and their secondary structures, as sampled using the partition function. Since *RNAmutants* runs in worst-case $O(n^3K^2)$ time, where n is the length of input RNA sequence, and K is an upper bound for the number of mutations, the web server cannot provide computational resources for large values of n, K. In such cases, the user should download executable code, which can be retrieved from the web server. *RNAmutants* allows the user to estimate the impact of mutations on the structure of functional RNA, and better understand the evolutionary process of RNA molecules.

Acknowledgements

RNAmutants software was written by J. Waldispühl, who designed the web site `http://rnamutants.csail.mit.edu/`. The web server `bioinformatics.bc.edu/clotelab/RNAmutants` was developed by P. Clote, using a general framework created by Jason Piersampieri and Yann Ponty. Work of P. Clote was partially supported by the Deutscher Akademischer Austauschdienst for a visit to Max Planck Institute of Molecular Genetics, as well as by the foundation Digiteo - Triangle de la Physique, and by NSF grants DBI-0543506 and DMS-0817971. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

[1] Waldispuhl, J., Devadas, S., Berger, B., and Clote, P. August 2008 Efficient algorithms for probing the RNA mutation landscape. *PLoS. Comput. Biol.* **4**, e1000124.

[2] Fisher, R. (1930) The Genetical Theory of Natural Selection, Clarendon Press, Oxford.

[3] Wright, S. (1945) The differential equation of the distribution of gene frequencies. *Proc. Natl. Acad. Sci. USA* **31**, 382–389.

[4] Watterson, G. (1962) Some theoretical aspects of diffusion theory in population genetics. *Ann. Math. Statist.* **33**, 93–957 Correction: **34**,352.

[5] Ewens, W. (1963) The mean time for absorption in a process of genetic type. *J. Austral. Math. Soc.* **3**, 375–383.

[6] Kingman, J. F. C. (1982) The coalescent. *Stochastic Processes and their Applications* **13**, 235–248.

[7] Hein, J., Schierup, M., and Wiuf, C. (2004) Gene Genealogies, Variation and Evolution : A Primer in Coalescent Theory, Oxford University Press, .

[8] Buss, S. and Clote, P. (2005) Solving the fisher-wright and coalescence problems with a discrete Markov chain analysis. *Advances in Applied Probability* **36(3)**, 1175–1197.

[9] Kimura, M. (1964) Diffusion models in population genetics. *J. Appl. Prob.* **1**, 177–232.

[10] Doudna, J. and Cech, T. (2002) The chemical repertoire of natural ribozymes. *Nature* **418**, 222–228.

[11] Barrick, J., Corbino, K., Winkler, W., Nahvi, A., Mandal, M., Collins, J., Lee, M., Roth, A., Sudarsan, N., Jona, I., Wickiser, J., and Breaker, R. (2004) New RNA motifs suggest an expanded scope for riboswitches in bacterial genetic control. *Proc. Natl. Acad. Sci. USA* **101(17)**, 6421–6426.

[12] Ng, K. L. and Mishra, S. K. June 2007 De novo SVM classification of precursor microRNAs from genomic pseudo hairpins using global and intrinsic folding measures. *Bioinformatics* **23**, 1321–1330.

[13] Commans, S. and Böck, A. (1999) Selenocysteine inserting tRNAs: an overview. *FEMS Microbiology Reviews* **23**, 333–351.

[14] Banerjee, A., Jaeger, J., and Turner, D. (1993) Thermal unfolding of a group I ribozyme: The low-temperature transition is primarily disruption of tertiary structure. *Biochemistry* **32**, 153–163.

[15] Zuker, M. and Stiegler, P. (1981) Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res.* **9**, 133–148.

[16] Mathews, D., Sabina, J., Zuker, M., and Turner, H. (1999) Expanded sequence dependence of thermodynamic parameters provides robust prediction of RNA secondary structure. *J. Mol. Biol.* **288**, 911–940.

[17] Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31(13)**, 3406–3415.

[18] Hofacker, I. (2003) Vienna RNA secondary structure server. *Nucleic Acids Res.* **31**, 3429–3431.

[19] Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31(13)**, 3406–3415.

[20] Hofacker, I. (2003) Vienna RNA secondary structure server. *Nucleic Acids Res.* **31(13)**, 3429–3431.

[21] Mathews, D., Disney, M., Childs, J., Schroeder, S., Zuker, M., and Turner, D. (2004) Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc. Natl. Acad. Sci. USA* **101**, 7287–7292.

[22] Ding, Y. and Lawrence, C. (2003) A statistical sampling algorithm for RNA secondary structure prediction. *Nucleic Acids Res.* **31(24)**, 7280–7301.

[23] Ding, Y., Chan, C., and Lawrence, C. (2004) Sfold web server for statistical folding and rational design of nucleic acids. *Nucleic Acids Res.* **32(Web Server issue)**, W135–141.

[24] Freyhult, E., Moulton, V., and Clote, P. Aug 2007 Boltzmann probability of RNA structural neighbors and riboswitch detection. *Bioinformatics* **23**, 2054–2062 doi: 10.1093/bioinformatics/btm314.

[25] Freyhult, E., Moulton, V., and Clote, P. July 2007 RNAbor: a web server for RNA structural neighbors. *Nucleic. Acids. Res.* **35**, W305–W309.

[26] Waldispuhl, J. and Clote, P. March 2007 Computing the partition function and sampling for saturated secondary structures of RNA, with respect to the Turner energy model. *J. Comput. Biol.* **14**, 190–215.

[27] Mathews, D. and Turner, D. (2002) Dynalign: An algorithm for finding the secondary structure common to two RNA sequences. *J. Mol. Biol.* **317**, 191–203.

[28] Havgaard, J. H., Lyngso, R. B., and Gorodkin, J. July 2005 The FOLDALIGN web server for pairwise structural RNA alignment and mutual motif search. *Nucleic. Acids. Res.* **33**, W650–W653.

[29] Coventry, A., Kleitman, D., and Berger, B. (2004) MSARI: multiple sequence alignments for statistical detection of RNA secondary structure. *Proc Natl Acad Sci U S A* **101(33)**, 12102–12107.

[30] Washietl, S., Hofacker, I., and Stadler, P. (2005) Fast and reliable prediction of noncoding RNAs. *Proc Natl Acad Sci U S A* **102**, 2454–2459.

[31] Griffiths-Jones, S., Bateman, A., Marshall, M., Khanna, A., and Eddy, S. (2003) Rfam: an RNA family database. *Nucleic Acids Res.* **31(1)**, 439–441.

[32] Clote, P., Waldispühl, J., Behzadi, B., and Steyaert, J.-M. (2005) Exploring the energy landscape of *k*-point mutagens of rna. *Bioinformatics* **21**, 4140–4147.

[33] Nussinov, R. and Jacobson, A. B. (1980) Fast algorithm for predicting the secondary structure of single stranded RNA. *Proceedings of the National Academy of Sciences, USA* **77**, 6309–6313.

[34] Xia, T., J. SantaLucia, J., Burkard, M., Kierzek, R., Schroeder, S., Jiao, X., Cox, C., and Turner, D. (1999) Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs. *Biochemistry* **37**, 14719–35.

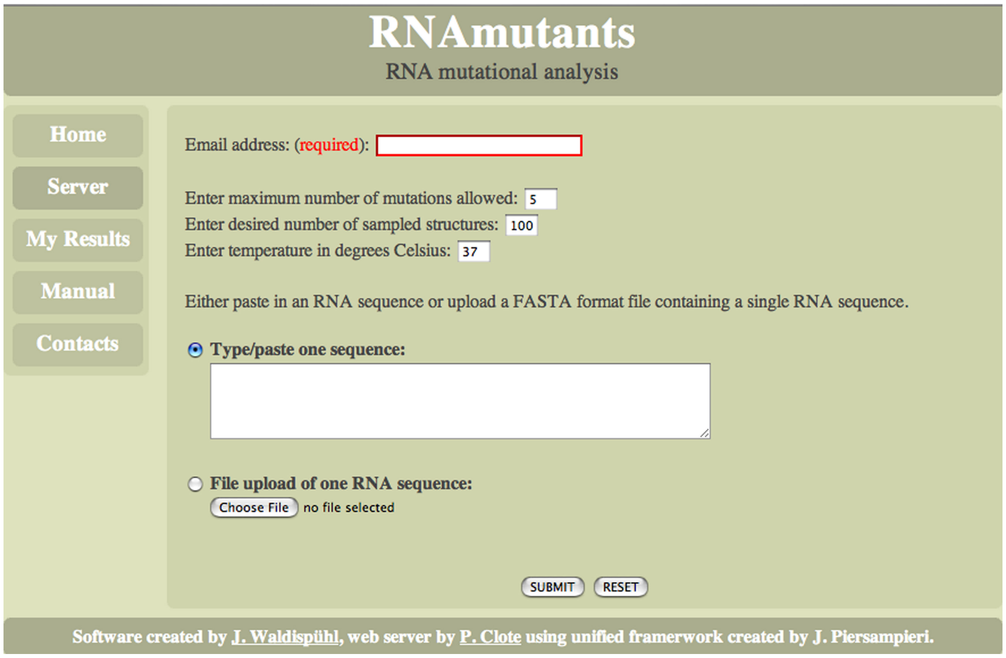
[35] I. Tinoco, J. and Schmitz, M. (2000) Thermodynamics of formation of secondary structure in nucleic acids In E. Di Cera, (ed.), Thermodynamics in Biology, pp. 131–176 Oxford University Press.

[36] Matthews, D., Sabina, J., Zuker, M., and Turner, D. (1999) Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.* **288**, 911–940.

[37] Waldispuhl, J., Behzadi, B., and Steyaert, J. M. (2002) An approximate matching algorithm for finding (sub-)optimal sequences in S-attributed grammars. *Bioinformatics* **18**, S250–S259.

[38] Ding, Y. and Lawrence, C. E. (2003) A statistical sampling algorithm for RNA secondary structure prediction. *Nucleic. Acids. Res.* **31**, 7280–7301.

[39] Pollard, K. S., Salama, S. R., Lambert, N., Lambot, M. A., Coppens, S., Pedersen, J. S., Katzman, S., King, B., Onodera, C., Siepel, A., Kern, A. D., Dehay, C., Igel, H., Ares, Jr, M., Vanderhaeghen, P., and Haussler, D. September 2006 An RNA gene expressed during cortical development evolved rapidly in humans. *Nature* **443**, 167–172.



RNAmutants

RNA mutational analysis

Email address: (required):

Enter maximum number of mutations allowed: 5
Enter desired number of sampled structures: 100
Enter temperature in degrees Celsius: 37

Either paste in an RNA sequence or upload a FASTA format file containing a single RNA sequence.

● Type/paste one sequence:

○ File upload of one RNA sequence:
no file selected

Software created by J. Waldispühl, web server by P. Clote using unified framerwork created by J. Piersampieri.

Figure 1: Input form for *RNAmutants*.

```
# RNAmutants: part fun Z(k) and superoptimals MFE(k) for 3UTR_MUSGBPA
# 3UTR_MUSGBPA
# AGCCAGCCAGCCUGUAGCCCUCAAUAAAAGGCAGCUGCCUCUGCUCCCCAU
k Z(k) rnaSeq(k) MFE(k) Energy(k)
0 39683275.214115 AGCCAGCCAGCCUGUAGCCCUCAAUAAAAGGCAGCUGCCUCUGCUCCCCAU ((.((((..((((.((........))..)))).)))).))........... -9.900000
1 4127868682896.799805 AGgCAGCCAGCCUGUAGCCCUCAAUAAAAGGCAGCUGCCUCUGCUCCCCAU (((((((..((((.((........))..)))).)))))))........... -17.000000
2 3322405908617160.000000 AGgCAGCCAGCCUGUAGCCCUCAAUAAAAGGCgGCUGCCUCUGCUCCCCAU ((((((((.((((.((........))..)))))))))))))........... -20.400000
3 1243234054761819904.000000 AGCCAGgCAGCCUGUAGCCCUCAAUAAAgGGCAGCUGCCUCgGCUCCCCAU (((((((((((.....(((((......))))).))))))).))))...... -23.800000
4 3343267739727679952.000000 AGCCAGgCAGCCUGUAGCCCUCAAUAgAgGGCAGCUGCCUCgGCUCCCCAU (((((((((((.....((((((....)))))).))))))).))))...... -26.600000
5 1529819467602089979412 48.000000 AGggAGCCAGgCaGUAGCCCUCAAUAAAgGGCAGCUGCCUCUGCUCCCCAU .((((((.(((((((.(((((......))))).)))))))..))))))...
-30.000000
```

Figure 2: Initial portion of one output file from *RNAmutants* for 51 nt portion of the 3′ untranslated region from murine beta-galactoside binding protein mRNA, with NCBI accession code MUSGBPA. Web server displays all 51 superoptimal secondary structures, their free energy, and mutation locations. Mutated nucleotides are shown in lower case.

```
# Sampled structures from each k-Boltzmann ensemble for 3UTR_MUSGBPA
# 3UTR_MUSGBPA
# AGCCAGCCAGCCUGUAGCCCUCAAUAAAAGGCAGCUGCCUCUGCUCCCCAU
gGgggGgCcGagccggGggCcgcAgccgcGGCccCcGgCUCgGCcCCCCcg
(((((((((((((((((((((((....))))))))))))))))))))))).
gGCggcCCgcggccggcCgggggcUuuAgccCccggcCggCcGCgggCCgc
.(((((((((((((((((((((((....)))))))))))))))))))))))
AGCCgGuCgGgggGggGuCCcCgcUAAcgcGggGaccCCcCccCggCCggc
..((((((((((((((((((((((....)))))))))))))))))))))).
gcCggGgCcGggacggGggCgCcggAAggcGCccCcGuCcCgGCcCCggcg
(((((((((((((((((((((((....))))))))))))))))))))))).
gcCCgGgCgcgggGggGCggUCgAUucAcGaCcGCccCCcCgcgcCCgggc
(((((((((((((((((((((((.....)))))))))))))))))))))))
AGgggGgCAGCugaccGgCgggcggAAgcccgccgguCagCUGCcCCCCug
(((((((((((((((((((((((....))))))))))))))))))))))).
cGgggGCggGCggGgcGgCCcgccgugggcGggcCgcCCaCgcCcgCCCcc
.(((((((((((((((((((((((...)))))))))))).)))))))))))
guCCcGCggGCucccgcgCgcCcuUccAgGGCgcgcGggagccCgCgggAU
((((((((((((((((((((((((...))))))))))))))))))))))))
gGgCccCCccgggacccCCggggcggAAgccCcGggGgucCcGggggggcc
.(((((((((((((((((((((((....)))))))))))))))))))))))
```

Figure 3: Initial portion of output file of 100 mutant sequence/structure pairs from *RNAmutants* for the 51 nt portion of the 3′ untranslated region from murine beta-galactoside binding protein mRNA, with NCBI accession code MUSGBPA. Mutated nucleotides are shown in lower case.

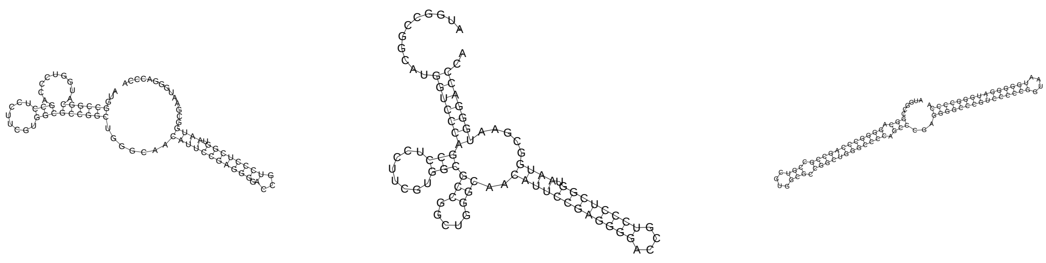


Figure 4: Rfam consensus structure (left), minimum free energy structure (middle) and 20-superoptimal structure (right) for hepatitis delta virus ribozyme with EMBL accession number `X85253.1/682-769`. Free energies Rfam data from (31); structure images produced with *RNAplot* (18).

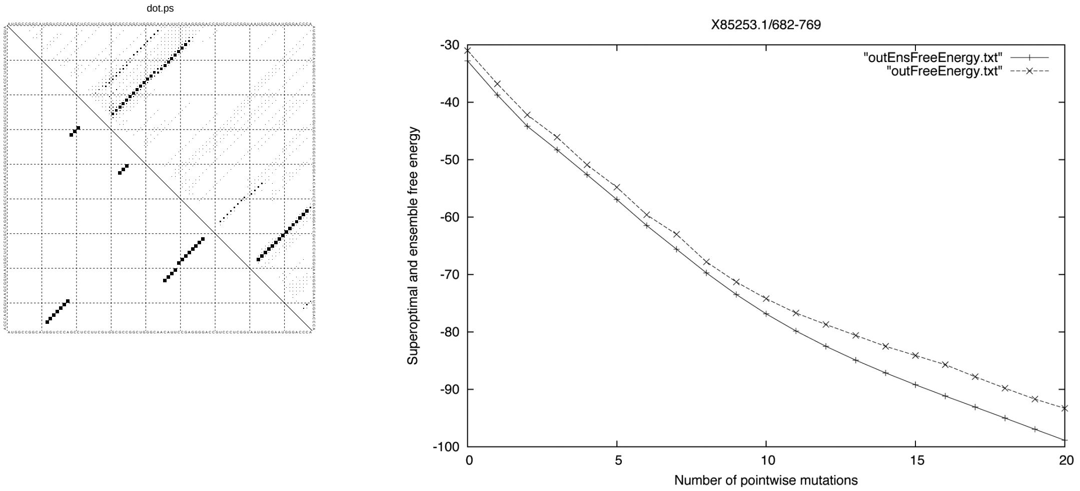


Figure 5: (Left) Base pair frequencies for sampled sequence/structure pairs for hepatitis delta virus ribozyme with EMBL accession number `X85253.1/682-769`. The upper triangular portion of the left panel represents the base pair frequencies over all 20,000 sampled structures (1000 samples for each k-point mutant, for $1 \leq k \leq 20$), while the lower triangular portion represents the minimum free energy structure of the wild type sequence. (Right) Plot of k-superoptimal and k-ensemble free energies, where the latter is defined by $-RT \ln(Z_k)$, where Z_k is the partition function over all k-point mutants.

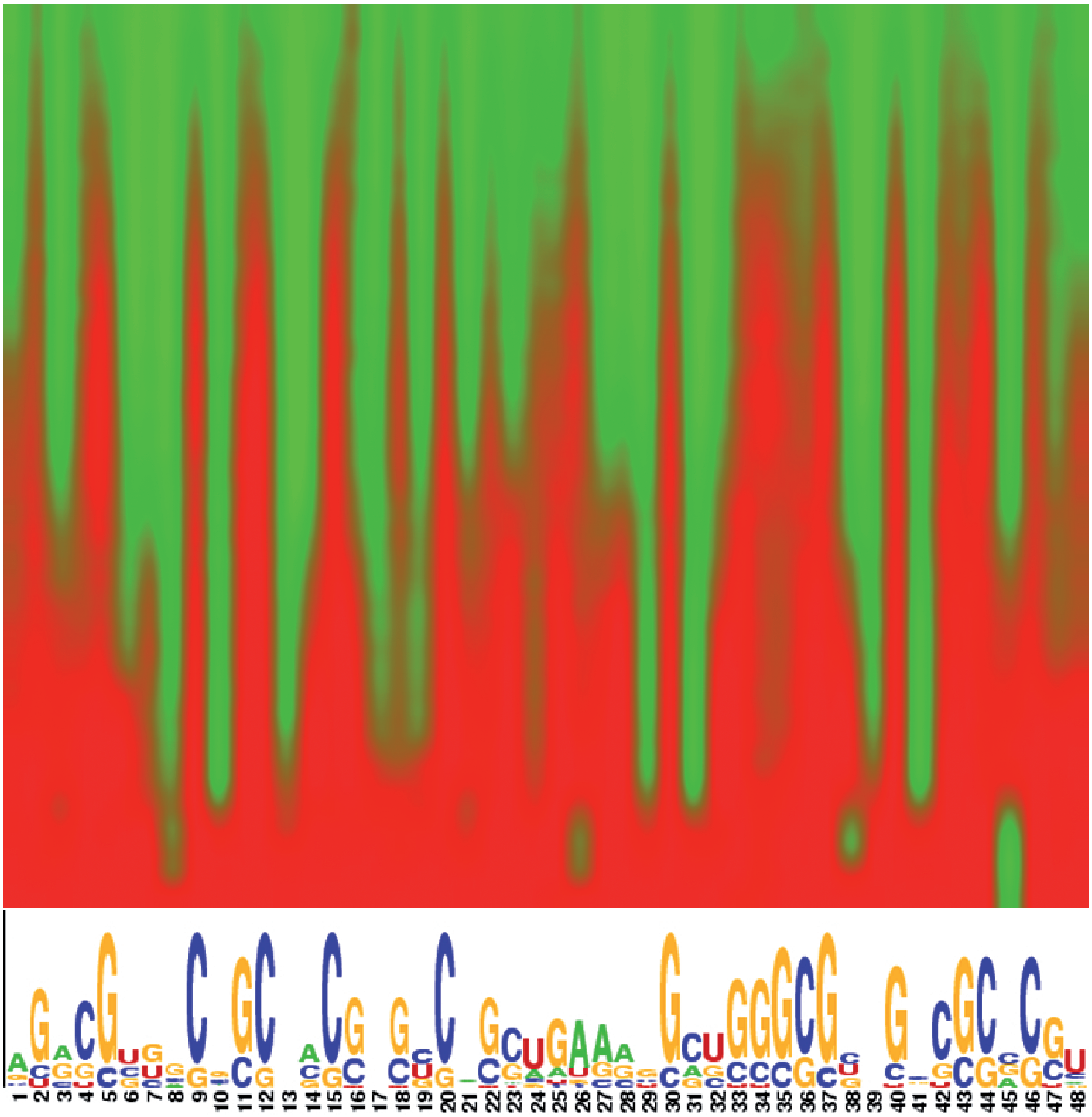


Figure 6: Mutability profile of 48 nt HAR1 region, where the number of mutations ranges from 1 to 40. HAR1 is part of the novel RNA gene HAR1F (39), expressed in Cajal-Retzius neurons in the developing human neocortex, a gene believed to show significant evolutionary acceleration. As in traffic lights, red regions are *not* mutated, while green regions are mutated from wild type nucleotide. The x-axis represents nucleotide position, as suggested by the logo plot below; the y-axis represents position-specific mutability. Mutability value of 0 corresponds to finding no mutations at that position among all samples, depicted by RGB color triple (255,0,0), while mutability value of 1 corresponds to finding every sample mutated at that position, depicted by RGB color triple (0,255,0), while fractional ratios of mutant positions are depicted by the triple $(\alpha, \beta, 0)$, where $\alpha + \beta = 255$. Python scripts that produced this PPM figure can be downloaded at the web server.