Using an RNA Secondary Structure Partition Function to Determine Confidence in Base Pairs Predicted by Free Energy Minimization

David H. Mathews[1]

Keywords: RNA Partition Function, RNA Secondary Structure, Statistical Mechanics

1 Introduction.

RNA plays many diverse roles in Biology, including catalyzing peptide bond formation [4, 10], catalyzing RNA splicing [2], localizing protein [11], and flagging development [5, 6]. New roles are being found for RNA, and the completion of whole genome projects provides the opportunity to find many new functional non-coding RNA sequences [3].

To understand the detailed mechanism of action of an RNA sequence, the structure of that RNA must be determined. Secondary structure, the sum of canonical base pairs, is usually determined by the comparative analysis of homologous sequences. In the absence of homologous sequences, free energy minimization by dynamic programming can be used to predict the structure of a single sequence with an average of 73% sensitivity for known pairs [7, 8]. This accuracy is sufficient to serve as a starting point for building an alignment for comparative sequence analysis or as an aid for designing RNA sequences, but improvements in the accuracy of base pair predictions would clearly be useful.

The predicted minimum free energy (MFE) structure provides a single best guess for the secondary structure, but it assumes that the secondary structure is at equilibrium, that there is a single conformation for the RNA, and that the thermodynamic parameters for evaluating conformation free energies are without error. One method to represent other possible or competing structures is to sample suboptimal secondary structures with free energies similar to the lowest free energy structure [12]. Another method to demonstrate a diversity of structures, pioneered by McCaskill, is to determine the pairing probabilities of all possible base pairs using a partition function calculated with dynamic programming [9].

2 Results.

A partition function calculation for RNA secondary structure is presented that uses a current set of nearest neighbor parameters for conformational free energy at 37 ºC [7, 8]. The calculation includes free energy increments for the coaxial stacking of helices, but remains $O(N^3)$ in time, where N is the number of nucleotides. The calculation is rapid, e.g. the base pairing probabilities for a 433 nucleotide *Tetrahymena* group I intron can be calculated in 19 seconds with a Pentium 4, 3.06 GHz processor.

For a diverse database of RNA sequences with known secondary structure [8], base pairs in the predicted minimum free energy structure that are predicted by the partition function to have high base pairing probability have a significantly higher positive predictive value for known base pairs.

[1] Center for Human Genetics and Molecular Pediatric Disease, Aab Institute of Biomedical Sciences, University of Rochester Medical Center, 601 Elmwood Avenue, Box 703, Rochester, NY 14642

For example, the average positive predictive value, 65.8% is increased to 90.7% when only base pairs with 99% or above probability are considered.

The recursions were written to allow constraints on base pairing determined by experiments, such as enzymatic cleavage, flavin mononucleotide cleavage, or chemical modification. The quality of base pair predictions are increased by the addition of experimentally determined constraints. For example, the percentage of highly probable pairs (greater than or equal to 95%) for the Dog SRP RNA increases from only 9.9% to 57.0% by including experimentally determined constraints in the calculation [1].

3 Summary.

The partition function calculation presented here does not replace the method of RNA secondary structure prediction by free energy minimization, but provides adjunct information that can be used to infer confidence in predicted base pairs. These data can be superimposed on predicted secondary structures using color annotation to quickly demonstrate high probability base pairs.

References

[1] Andreazzoli, M. and Gerbi, S.A. 1991. Changes in 7SL RNA conformation during the signal recognition particle cycle. *EMBO J*, 10: 767-777.

[2] Doudna, J. and Cech, T. 2002. The chemical repertoire of natural ribozymes. *Nature*, 418: 222-228.

[3] Eddy, S.R. 2001. Non-coding RNA genes and the modern RNA world. *Nature Reviews*, 2: 919-929.

[4] Hansen, J.L., Schmeing, T.M., Moore, P.B., and Steitz, T.A. 2002. Structural insights into peptide bond formation. *Proc. Natl. Acad. Sci. USA*, 99: 11670-11675.

[5] Lagos-Quintana, M., Rauhut, R., Lendeckel, W., and Tuschl, T. 2001. Identification of novel genes coding for small expressed RNAs. *Science*, 294: 853-858.

[6] Lau, N.C., Lim, L.P., Weinstein, E.G., and Bartel, D.P. 2001. An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans. Science*, 294: 858-862.

[7] Mathews, D.H., Disney, M.D., Childs, J.L., Schroeder, S.J., Zuker, M., and Turner, D.H. In preparation. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure.

[8] Mathews, D.H., Sabina, J., Zuker, M., and Turner, D.H. 1999. Expanded sequence dependence of thermodynamic parameters provides improved prediction of RNA Secondary Structure. *J. Mol. Biol.*, 288: 911-940.

[9] McCaskill, J.S. 1990. The equilibrium partition function and base pair probabilities for RNA secondary structure. *Biopolymers*, 29: 1105-1119.

[10] Nissen, P., Hansen, J., Ban, N., Moore, P.B., and Steitz, T.A. 2000. The structural basis of ribosomal activity in peptide bond synthesis. *Science*, 289: 920-930.

[11] Walter, P. and Blobel, G. 1982. Signal recognition particle contains a 7S RNA essential for protein translocation across the endoplasmic reticulum. *Nature*, 299: 691-698.

[12] Zuker, M. 1989. On finding all suboptimal foldings of an RNA molecule. *Science*, 244: 48-52.