# Hardness of RNA Secondary Structure Design

**Rosalía Aguirre-Hernández[1], Holger H Hoos[1], Anne Condon[1]**

**Keywords:** ribonucleic acids, secondary structure, RNA Designer

## 1 Introduction.

Ribonucleic acids (RNA) are macromolecules that play fundamental roles in many biological processes and their structure is essential for their biological function. This work is focused on the design of RNA strands that is predicted to fold to a given secondary structure, according to a standard thermodynamic model such as that of Mathews et al. [5]. This problem is relevant because it will facilitate the characterization of biological RNAs by their function and the design of new ribozymes that can be used as therapeutic agents [2]. There are also applications in nanobiotechnology in the context of building self-assembling structures from small RNA molecules [4].

One solution to the RNA secondary structure design problem is provided by Hofacker et al. [3], the implementation of which is included in the Vienna RNA Secondary Structure Package. A more recent stochastic local search algorithm, the RNA Designer of Andronescu et al. [1] shows a better performance.

The purpose of this work is to understand better the factors that make RNA structures hard to design. Such understanding provides the basis for improving the performance of RNA Designer and for characterising its limitations. We will describe a modification of the RNA Designer that improves the performance of the algorithm. Furthermore, it is not known whether there is a polynomial time algorithm for RNA secondary structure design. Therefore, to gain insights into the practical complexity of the problem, we present a scaling analysis to investigate the hardness of the problem on random RNA structures using the improved RNA Designer.

## 2 Algorithm.

The RNA Designer is a stochastic local search (SLS) procedure that uses a hierarchical decomposition of the given structure [1]. A structure is split into two substructures that do not contain multiloops. Notice that it is necessary to connect the two free ends created by the split such that both resulting substructures have exactly two free ends. To create structural boundary conditions at the split points that are similar to those of the original structure, this connection is achieved by merging the free ends with those of a static cap structure, which is a small hairpin loop of size four (five paired, four unpaired and five paired bases); furthermore, two unpaired bases are added to the two remaining free ends of each substructure if it contains a bulge directly after the first base pair.

We found that there are some structures difficult to design by using this approach. This is the case, for example, for structures in which two loops are separated by a very short stem. But it is possible to improve the performance of the algorithm by introducing a dynamic cap structure and dynamic dangling ends in order to create structural boundary conditions at the split points that are exactly the same to those of the original structure. The number of paired bases in the hairpin loop (cap structure), added to one of the substructures will

[1]Department of Computer Science, University of British Columbia, Vancouver, BC V6T 1Z4,Canada. E-mail: {rosalia,hoos,condon}@cs.ubc.ca

depend on the number of paired bases at the beginning of the other substructure. On the other hand, we will add one dangling end to the $5'$ ($3'$) end of a substructure if, and only if, its adjacent base in the original structure is a free base. Figure 1 (a) shows the performance correlation between the two versions of the algorithm for 60 structures of length 75. Notice that there are two outliers, which correspond to two structures for which the dynamic cap structure and dynamic dangling ends are crucial.

## 3 Scaling analysis.

In order to investigate the empirical complexity of solving RNA secondary structure design problems with the improved version of RNA Designer and with the Vienna algorithm, we performed a scaling analysis with random structures of length 50, 75, 100, 150 and 200. As can be seen from Figure 1 (b), the median expected run-time of both, RNA Designer and the Vienna algorithm scales polynomially with the size of the random structures. In future work, we will extend our analysis to biologically more realistic structures.

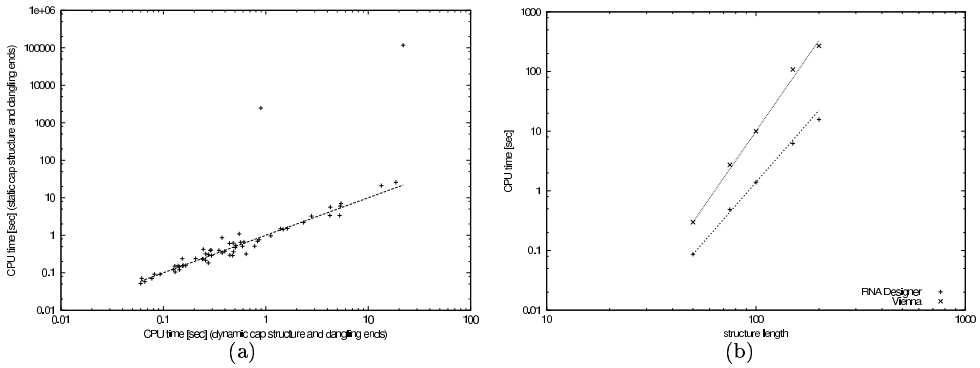


Figure 1: (a) Correlation between improved and original RNA Designer. (b) Median expected run-time over test-set of 1000 instances for length 50-75, and 100 instances for size 100-150.

# References

[1] Andronescu, M., Fejes, A. P., Hutter, F., Condon, A. and Hoos, H. H. A New Algorithm for RNA Secondary Structure Design. *Journal of Molecular Biology.* To appear.

[2] Breaker, R. R. 1996. Are engineered proteins getting competition from RNA? *Curr. Opin. Biotech.* 7:442–448.

[3] Hofacker, I. L., Fontana, W., Stadler, P. F., Bonhoeffer, L. S., Tacker, M. and Schuster, P. 1994. Fast folding and comparison of RNA secondary structures. *Chemical Monthly.* 125:167–188.

[4] Jaeger, L., Westhof, E. and Leontis, N. B. 2001. TectoRNA: modular assembly units for the construction of RNA nano-objects *Nucleic Acid Research.* Vol. 29, No. 2, 455–463.

[5] Mathews, D. H., Sabina, J., Zuker, M. and Turner, D. H. 1999. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *Journal of Molecular Biology.* 288:911–940.