

# Algorithms for RNA folding

Max Ward

School of Computer Science & Software Engineering  
The University of Western Australia

May 5, 2014

## **Abstract**

Analysis of thermodynamically based algorithms for RNA structure prediction

**Keywords:** Ribonucleic acid, structure, prediction, thermodynamic hypothesis.

**CR Classification:** J.3 Biology and genetics.

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
1.1	Motivation . . . . .	3
1.2	Relevant Algorithms . . . . .	3
1.2.1	The Zuker Algorithm . . . . .	3
1.2.2	Maximum Expected Accuracy . . . . .	3
1.2.3	Cotranscriptional folding . . . . .	3
<b>2</b>	<b>Materials and Methods</b>	<b>3</b>
2.1	Environment . . . . .	3
2.2	Data Set . . . . .	4
2.3	Algorithms . . . . .	4
2.4	Additional Software . . . . .	4
2.5	Accuracy . . . . .	4
<b>3</b>	<b>Results</b>	<b>4</b>
3.1	Accuracy . . . . .	4
3.1.1	Small RNA . . . . .	4
3.1.2	Moderate RNA . . . . .	4
3.1.3	Large RNA . . . . .	4
3.2	Time . . . . .	4
<b>4</b>	<b>Discussion</b>	<b>4</b>
<b>5</b>	<b>Conclusions</b>	<b>4</b>

## List of Figures

# 1 Introduction

## 1.1 Motivation

Ribonucleic acid (RNA) is at the core of many biological processes. Traditionally it has been described as the messenger molecule of DNA, faithfully carrying the code for protein from DNA to the site of protein synthesis. However, in a recent landmark paper, Amaral et al. [1] described our genome, and those of other eukaryotes, as being driven by an RNA machine. They noted that most of the eukaryote genome is transcribed into RNA, despite little of it coding for protein. It seems that much of our genome, originally called ‘junk DNA’, codes for functional RNA molecules. These RNAs can interact with DNA, affecting gene expression. This allows DNA to essentially regulate itself. For example, Makeyev & Maniati [2] reported that microRNAs affect the expression of genes by interfering with translation of protein. They also argued that microRNAs, and other regulatory RNAs, explain the vast differences between organisms with similar genomes. To put this idea into perspective, we share roughly 90% of our genes with the domestic Cat [4]. Mattick [3] has suggested that the process of development—from embryo to adult—is encoded in the interactions of such RNAs.

A widely held axiom is that chemical structure is tantamount to biological function. With increasingly important biological functions being associated with RNA, it is important to be able to predict its structure. The purpose of this paper is to provide a survey of some widely used RNA structure prediction algorithms. In the interest of keeping this report succinct, I review only algorithms based on a conventional thermodynamically based model. Other algorithms often use machine learned, statistical parameters; these shall not be explored here. The Zuker algorithm was the first thermodynamic algorithm to achieve usable prediction accuracy, and it forms the basis for all the methods I shall hence discuss.

## 1.2 Relevant Algorithms

### 1.2.1 The Zuker Algorithm

### 1.2.2 Maximum Expected Accuracy

### 1.2.3 Cotranscriptional folding

Explain in high level terms the three algorithms tested

# 2 Materials and Methods

## 2.1 Environment

How I tested it

## 2.2 Data Set

Strand stuff

## 2.3 Algorithms

Config and stuff

## 2.4 Additional Software

gretl etc

## 2.5 Accuracy

explain f score and why it is useful for us

# 3 Results

## 3.1 Accuracy

### 3.1.1 Small RNA

### 3.1.2 Moderate RNA

### 3.1.3 Large RNA

## 3.2 Time

# 4 Discussion

# 5 Conclusions

# References

- [1] Paulo P Amaral, Marcel E Dinger, Tim R Mercer, and John S Mattick. The eukaryotic genome as an rna machine. *Science*, 319(5871):1787–1789, 2008.
- [2] Eugene V Makeyev and Tom Maniatis. Multilevel regulation of gene expression by micrnas. *Science*, 319(5871):1789–1790, 2008.
- [3] John S Mattick. A new paradigm for developmental biology. *Journal of Experimental Biology*, 210(9):1526–1547, 2007.

- [4] Joan U Pontius, James C Mullikin, Douglas R Smith, Kerstin Lindblad-Toh, Sante Gnerre, Michele Clamp, Jean Chang, Robert Stephens, Beena Neelam, Natalia Volfovsky, et al. Initial sequence and comparative analysis of the cat genome. *Genome research*, 17(11):1675–1689, 2007.