

# CSC482B Project Proposal: RNA\_LZW (title subject to change)

Quinn Gieseke, Claire Champernowne

February 20, 2021

## Motivation

Detecting RNA secondary structures is a process which is integral to understanding the functions of non-coding RNA sequences. Though there are existing bioinformatics tools used for detection, these programs, though not without their limitations. Specifically, RNAz 1.0 uses a sliding window approach in its analysis of a sequence [1], meaning there may be structures with loops existing outside the scope of the window.

The Lempel-Ziv-Welch(LZW) algorithm is a universal lossless compression algorithm used for a variety of applications such as Unix file compression. The algorithm compresses data by using a dictionary to assign common sequences of characters to a fixed-length code (typically 12-bit) [2].

Given the LZW algorithm's dictionary approach to pattern-matching, we hypothesize it could be used as a novel method of detecting new RNA secondary structures which are undetectable by the sliding window method used in RNAz 1.0.

## Objectives

Given the time limitations on this project, we intend to focus solely on detection of possible structures without commenting on the structure's validity in nature. Our goal is to detect RNA secondary structures which may be impossible to detect using RNAz 1.0 using our novel method of detection, RNA\_LZW.

In addition, we wish to achieve detection in a reasonable time frame, which will be assessed by comparing RNA\_LZW's processing time to that of other RNA secondary structure detection tools.

## Methods

code probably

## References

- [1] <https://psb.stanford.edu/psb-online/proceedings/psb10/gruber.pdf>
- [2] [https://www2.cs.duke.edu/courses/spring03/cps296.5/papers/welch\\_1984\\_technique\\_for.pdf](https://www2.cs.duke.edu/courses/spring03/cps296.5/papers/welch_1984_technique_for.pdf)