Computational Approaches to RNA Structure Prediction

Samuel Jackson, University Of Aberystwyth

Abstract—The abstract goes here.

Index Terms—RNA, literature review, dynamic programming, context free grammars, bioinformatics, computational biology.

I. INTRODUCTION

Ribonucleic acid (RNA) is a fundamental biological macromolecule that play an important part in many biological processes. RNA folding, like DNA and protein folding can be classified into three different categories: primary, secondary, and tertiary structure. The primary structure refers to the sequence of building blocks of RNA. These are the nucleotides guanine, uracil, adenine, and cytosine, commonly abbreviated to the first letter of each molecule (G, U, A, C). While DNA molecules frequently form the eponymous double helix between two strands, RNA molecules are single stranded and form complex structures by folding upon themselves.

A. RNA secondary structure

The way in which nucleotides in an RNA strand form hydrogen bonds with one each is referred to as the RNA's secondary structure. In RNA molecules bonds are usually formed between the base pairs A-U, C-G, and U-G. The patterns of the structures that are created through base-pairing can be classified into a number of different sub structures. Reference [1] provides a comprehensive introduction. Commonly encountered structural motifs frequently used in RNA secondary structure prediction are:

- **Base pair stacks** - The most common structural element. Formed by an RNA strand folding on itself and forming hydrogen bonds between complementary bases. Bonds in base pair stacks form between two parts of the RNA each running in an anti-parallel direction to one another.
- **Hairpin loops** - A collection of unpaired nucleotides at the terminus of a base pair stack. So called because the strand loops back and binds with itself.
- **Symmetric and asymmetric loops** - a collection of unpaired nucleotides between two base pair stacks. Symmetrical if the number of nucleotides on each side is equal, asymmetrical if not.
- **Bulges** - Similar to loops but with one side having no unpaired nucleotides.
- **Junctions** - The point at which multiple base pair stacks meet is referred to as a junction.
- **Pseudoknots** - Pseudoknots are formed between the unpaired nucleotides on a hairpin loop with the unpaired nucleotides on an adjacent strand. So called because

the structure shows some resemblance to a mathematical knot. The properties of pseudoknots and similar structures propose a particular challenge to prediction due to the complex, interwoven, long range base pairing.
- **Kissing hairpins** - Similar to pseudoknots, but directly between the unpaired nucleotides of two hairpin loops.

B. RNA tertiary structure

II. RNA SECONDARY STRUCTURE PREDICTION

A. Single Sequence Methods

1) Thermodynamic models and Free Energy Minimisation:

2) Stochastic Context Free Grammars:

B. Comparative Methods

III. RNA TERTIARY STRUCTURE PREDICTION

ACKNOWLEDGMENT

REFERENCES

[1] J. Nowakowski and I. Tinoco, "Rna structure and stability," in *Seminars in virology*, vol. 8, no. 3. Elsevier, 1997, pp. 153–165.