A formal language for classifying RNA secondary structures

Quadrini M., Merelli E., Tesei L.

School of Science and Technology, Computer Science Division, University of Camerino,

e-mail:{michela.quarini, emanuela.merelli, luca.tesei}@unicam.it

We introduce a formal language for representing RNA secondary structures as interactions of loops [1] towards a topological shape language [2]. A base loop is an hairpin. All the other loops, such as bulge, helix, inner loop and multiple loop, are compositions of hairpins. Two loops are sequentially connected by base pair weak interactions. We introduce a set of operators to manipulate loops and interactions between loops. The grammar of the resulting language of RNA secondary structure allows us to generate both pseudoknot free and pseudoknotted RNA secondary structures starting from the RNA sequences. Moreover, we can represent a pseudoknot free RNA secondary structure in a "canonical form", thus we have a way to tell whether two given structures differ by a loop. We will investigate the characterization of the higher order language corresponding to the loops interactions [3].

References:

1. Michael Waterman, T.F. Smith. RNA secondary structure: a complete mathematical analysis. Mathematical Biosciences 42(3), 1978.

2. Adane Mamuye, Emanuela Merelli, Luca Tesei. Towards a Shape Language for Interpreting RNA Folding. Bio-inspired Information and Communications Technologies (BICT) 2015.

3. Davide Sangiorgi, Valeria Vignudelli: Environmental bisimulations for probabilistic higher-order languages. POPL 2016