

# Assessment of Computational Methods for Predicting RNA Secondary Structures with Pseudoknots

Dan Du,<sup>1</sup> Khodeza Begum,<sup>2</sup> and Ming-Ying Leung<sup>1,2,3</sup>

<sup>1</sup>Computation Science Program, <sup>2</sup>Border Biomedical Research Center, <sup>3</sup>Department of Mathematical Sciences, The University of Texas at El Paso, El Paso, TX 79968, USA  
[ddu2@miners.utep.edu](mailto:ddu2@miners.utep.edu); [kbegum@utep.edu](mailto:kbegum@utep.edu); [mleung@utep.edu](mailto:mleung@utep.edu)

## Abstract

Ribonucleic Acid (RNA) secondary structures consist of pattern features that can be classified into two basic categories, stem-loops and pseudoknots, both of which have been implicated in important biological processes including gene expression and regulation. Pseudoknots are now recognized as a widespread motif with diverse and important functions. Although computational predictions of RNA secondary structure have already contributed substantially to the understanding of the molecular mechanisms of RNA functions, pseudoknot prediction remains as a computationally intensive challenge. A number of tools, (e.g., PKnots, pKiss) based on thermodynamic free energy minimization, have been deployed in the past two decades for predicting RNA structures with pseudoknots. Recently, new approaches (e.g., SPOT-RNA, Ufold) based on deep learning technologies have also been proposed. We are developing an assessment scheme for the above four different RNA sequence secondary structure prediction methods that have pseudoknot prediction capability by evaluating their prediction accuracies using a sample of confirmed pseudoknots in the PseudoBase++ ([pseudobaseplusplus.utep.edu](http://pseudobaseplusplus.utep.edu)) database. We also compare their sequence length limitations with the goal to develop an integrated computational method for predicting and analyzing RNA secondary structures including pseudoknots.