Title

Comparing prediction algorithms for RNA secondary structures with pseudoknots

Authors

Dan Du,[1] Khodeza Begum,[2] and Ming-Ying Leung[1,2,3]

[1]Computational Science Program, [2]Border Biomedical Research Center, and [3]Department of Mathematical Sciences, The University of Texas at El Paso, El Paso, TX

Abstract

The secondary structures of Ribonucleic Acid (RNA) encompass two fundamental categories: stem-loops and pseudoknots. Both patterns play significant roles in vital biological processes, including gene expression and regulation. Pseudoknots are now widely acknowledged as prevalent motifs with diverse and crucial functions. Despite the substantial contributions of computational RNA secondary structure predictions to our understanding of RNA's molecular mechanisms, accurately predicting pseudoknots remains a computationally intensive challenge. Over the past two decades, tools such as PKnots and pKiss, utilizing thermodynamic free energy minimization, have been pivotal in predicting RNA structures, particularly those involving pseudoknots. Recent advancements, like the application of deep learning technologies in tools such as SPOT-RNA and UFold, showcase the ongoing progress in this field. Our current work involves developing an assessment scheme for evaluating four distinct RNA sequence secondary structure prediction methods that possess the capability to predict pseudoknots. The predictions for the 398 known pseudoknot structures in the PseudoBase++ (pseudobaseplusplus.utep.edu) database demonstrate that PKnots and pKiss achieve higher prediction accuracies, whereas SPOT-RNA and UFold exhibit superior runtime efficiency and accommodate longer sequence lengths. These results are consistent with predictions made for an experimentally confirmed pseudoknot from SARS-CoV-2. This pseudoknot is 78 nucleotides in length, situated within the ORF1ab gene, which is expressed as a result of frameshifting.