M. Shel Swenson, Georgia Institute of Technology, Atlanta, GA, USA
Emily Rogers and Christine E. Heitsch, Georgia Institute of Technology, Atlanta, GA, USA

Towards an Understanding of SHAPE-directed RNA Structure Prediction Accuracy

RNA secondary structure prediction remains an open problem in computational biology. Data from recently emerging high-throughput structure probing technologies, such as the SHAPE method, have been used in the framework of thermodynamic optimization to predict RNA secondary structure. We investigate the factors influencing the accuracy of SHAPE data-directed predictions via stochastic simulations. We find that the accuracy of data-directed predictions is broadly correlated with the accuracy of undirected predictions. Our analysis also indicates that it is possible to broadly estimate the prediction accuracy of a sequence by the similarity between the undirected and data-directed structures. At the level of individual basepairs, we find that those common to a data-directed prediction and the undirected structure are more likely to be correct than base pairs that are only in the undirected structure. Finally, we explore the potential of constraints in the form auxiliary data to improve prediction accuracy.