World Academy of Science, Engineering and Technology
International Journal of Mathematical and Computational Sciences
Vol:14, No:12, 2020

# An Overview of Bioinformatics Methods to Detect Novel Riboswitches Highlighting the Importance of Structure Consideration

**Authors :** Danny Barash

**Abstract :** Riboswitches are RNA genetic control elements that were originally discovered in bacteria and provide a unique mechanism of gene regulation. They work without the participation of proteins and are believed to represent ancient regulatory systems in the evolutionary timescale. One of the biggest challenges in riboswitch research is that many are found in prokaryotes but only a small percentage of known riboswitches have been found in certain eukaryotic organisms. The few examples of eukaryotic riboswitches were identified using sequence-based bioinformatics search methods that include some slight structural considerations. These pattern-matching methods were the first ones to be applied for the purpose of riboswitch detection and they can also be programmed very efficiently using a data structure called affix arrays, making them suitable for genome-wide searches of riboswitch patterns. However, they are limited by their ability to detect harder to find riboswitches that deviate from the known patterns. Several methods have been developed since then to tackle this problem. The most commonly used by practitioners is Infernal that relies on Hidden Markov Models (HMMs) and Covariance Models (CMs). Profile Hidden Markov Models were also carried out in the pHMM Riboswitch Scanner web application, independently from Infernal. Other computational approaches that have been developed include RMDetect by the use of 3D structural modules and RNAbor that utilizes Boltzmann probability of structural neighbors. We have tried to incorporate more sophisticated secondary structure considerations based on RNA folding prediction using several strategies. The first idea was to utilize window-based methods in conjunction with folding predictions by energy minimization. The moving window approach is heavily geared towards secondary structure consideration relative to sequence that is treated as a constraint. However, the method cannot be used genome-wide due to its high cost because each folding prediction by energy minimization in the moving window is computationally expensive, enabling to scan only at the vicinity of genes of interest. The second idea was to remedy the inefficiency of the previous approach by constructing a pipeline that consists of inverse RNA folding considering RNA secondary structure, followed by a BLAST search that is sequence-based and highly efficient. This approach, which relies on inverse RNA folding in general and our own in-house fragment-based inverse RNA folding program called RNAfbinv in particular, shows capability to find attractive candidates that are missed by Infernal and other standard methods being used for riboswitch detection. We demonstrate attractive candidates found by both the moving-window approach and the inverse RNA folding approach performed together with BLAST. We conclude that structure-based methods like the two strategies outlined above hold considerable promise in detecting riboswitches and other conserved RNAs of functional importance in a variety of organisms.

**Keywords :** riboswitches, RNA folding prediction, RNA structure, structure-based methods
**Conference Title :** ICSRD 2020 : International Conference on Scientific Research and Development
**Conference Location :** Chicago, United States
**Conference Dates :** December 12-13, 2020