Detection of signal beyond secondary structure from SHAPE experiment

Alexandra Pogorelskaya, Andrey Mironov, Ruslan Soldatov

Moscow State University, Faculty of Bioengineering and Bioinformatics, Russia, 119991, Moscow, 1-73 Leninskie Gory, alexapogorelskaya@gmail.com

RNA secondary structure prediction is a fundamental problem in computational biology. The most widespread approach is thermodynamic-based prediction, which uses Turner's model and Zuker's algorithm [1] in order to search for structure with minimal free energy. This approach has a number of caveats, which results in moderate quality of predictions. For instance, thermodynamic model shows compact ensemble of suboptimal structures and does not take into consideration folding kinetics.

SHAPE experiment that assess single-ribonucleotide conformation, has gained appeal recently. Its incorporation allows making more precise prediction of RNA secondary structure. Special reagent can interact with ribose 2'-OH, if it is in a specific conformation. The result of the experiment is the modification intensity of each nucleotide (so called reactivity), that correlates with the nucleotide state in secondary structure. Combination of reactivities with Zuker's algorithm significantly increases prediction quality, thus SHAPE-data has high information content about secondary structure [2]. However, modification intensity depends on many RNA structural features besides nucleotide state. In other words, through experiment we get indirect information on secondary structure in current experimental conditions.

Here we propose a model for signal detection from secondary and non-secondary components of RNA structure from SHAPE experiment. We used 25 small RNAs with known secondary structure and experimental data. We designed probabilistic data interpretation based on reactivity distribution for possible nucleotide states (paired, unpaired, involved in non-canonical interaction) and secondary structure conditional probability with given reactivities. Secondary structure ensemble analysis revealed the fact that new function correlates better with the proximity to annotated structure than free energy for the majority of small RNAs. While RNA molecules can form a variety of structures, the result mentioned above indicates existence of a dominant structure in SHAPE assessed conditions. Therefore, it is possible to predict secondary structure having only the results of SHAPE experiments. To accomplish this, Nussinov algorithm modification was developed. It uses dynamic programming approach, as addition of a nucleotide to smaller optimal structure has weight according to its reactivity. In the case of the good reactivity separation for different states (simulated data) such an algorithm predicts more than 95% of pairing events. However, experimental data usage critically decreases quality.

1. Zuker M. (1994) Prediction of RNA secondary structure by energy minimization, Methods Mol Biol, 25:267-94.
2. Weeks KM et al. (2013) Principles for understanding the accuracy of SHAPE-directed RNA structure modeling, Biochemistry, 52(4):588-95.