Short stems in RNA secondary structure

Polina Tikhonova

National Research University - Higher School of Economics, Moscow, Russia,
`tikhonova.polly@mail.ru`

Eugene Baulin

Institute of Mathematical Problems of Biology RAS - the Branch of Keldysh Institute of Applied Mathematics of Russian Academy of Sciences, Pushchino, Moscow region, Russia, `baulin@lpm.org.ru`

Mikhail Roytberg

Institute of Mathematical Problems of Biology RAS - the Branch of Keldysh Institute of Applied Mathematics of Russian Academy of Sciences, Pushchino, Moscow region, Russia, Moscow Institute of Physics and Technology (State University), Dolgoprudny, Moscow region, Russia, `mroytberg@lpm.org.ru`

RNA along with DNA and proteins is one of the most important classes of biopolymers. Aside from the three types of RNA molecules (mRNA, tRNA, and rRNA) involved in protein synthesis there are plenty of noncoding RNA molecules with a variety of functions (see for example [1]). The function of RNA is in correspondence with its structure and that is why we need to fully understand principles of RNA structure formation. One of the least investigated class of RNA structures comprises structures containing so called pseudoknots (a.k.a. pseudoknotted structures).

In this work we study the role of short helical regions (short stems) which are thought to be energetically unfavorable and hence unstable. The work is based on analysis of all RNA-containing structures from Protein Data Bank [2]. Annotations of stems and pseudoknots were obtained from URS database [3]. In particular, we analyze the 2D-environment of short stems and refine the classification of pseudoknots appearing in RNA 3D-structures assuming that short stems are not taken into account. The assumption allows us significantly simplify the classification that may result in improvement of algorithms predicted pseudoknotted RNA secondary structures.

According to the empirical thermodynamic parameters [4] the formation of a helix starts to decrease free energy of the structure only after the formation of the third stack of base pairs. With respect to this rule we believe that short stems, i.e. stems of length 2 or 3, may be unstable and should not contribute significantly to the overall free energy of folded RNA molecule. We compared stem length distributions separately for all structures and pseudoknotted structures. Short stems of length 2 have been found to be overrepresented in pseudoknotted structures (32% vs. 25% in all structures) although short stems of length 3 have not supported this trend. For indirect evidence of instability of short stems we compared stem length distributions separately for each type of respective loop. As we expected the majority of short stems were found near internal loops and bulges where they can be strengthened by longer adjacent stems.

Then we considered the role of short stems in forming of pseudoknots. The overall number of considered pseudoknots is 5233 (the data is highly redundant). The overall number of unique pseudoknot signatures is 16, see definition of pseudoknot signature in [5]. For each pseudoknot short stems have been excluded and its signatures have been updated via two stage reduction process. First the short stems are removed and then the signature is rebuilded. If needed, interchain stems have been excluded from interchain pseudoknots prior to short stems removal. After the procedure almost all pseudoknots have been reduced to the two simplest types – either H-knot (signature abAB) or kissing loops (signature abAcBC). The only exception is a pseudoknot from Group II Intron from Oceanobacillus iheyensis with signature abcdCefAFDEB. This case will be further examined in more detail.

This work was supported by grant 16-04-01640 A from RFBR.

1. Nie, Lei, Hsing-Ju Wu, Jung-Mao Hsu, Shih-Shin Chang, Adam M. LaBaff, Chia-Wei Li, Yan Wang, Jennifer L. Hsu, and Mien-Chie Hung. "Long non-coding RNAs: versatile master regulators of gene expression and crucial players in cancer." *American journal of translational research* **4**, no. 2 (2012): 127.

2. Berman, Helen M., John Westbrook, Zukang Feng, Gary Gilliland, Talapady N. Bhat, Helge Weissig, Ilya N. Shindyalov, and Philip E. Bourne. "The protein data bank." *Nucleic acids research* **28**, no. 1 (2000): 235-242.

3. Baulin, Eugene, Victor Yacovlev, Denis Khachko, Sergei Spirin, and Mikhail Roytberg. "URS DataBase: universe of RNA structures and their motifs." *Database* 2016 (2016): baw085.

4. Tan, Zhijie, Wenbing Zhang, Yazhou Shi, and Fenghua Wang. "RNA folding: structure

prediction, folding kinetics and ion electrostatics." *Advance in Structural Bioinformatics*, pp. 143-183. Springer Netherlands, 2015.

5. Bon M, Vernizzi G, Orland H, Zee A. Topological classification of RNA structures. *Journal of molecular biology*. 2008 Jun 13;**379**(4):900-11.