RNA triplexes within the known RNA 3D structures

Anastasia Medvedeva

Moscow Institute of Physics and Technology (State University), Dolgoprudny, Moscow region, Russia, `anastasia.medvedeva@phystech.edu`

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`

Investigation of RNA spatial structure and tertiary motifs is an important problem of structural bioinformatics. Various biological functions of RNA molecules are in correspondence with their structures that are stabilized by various secondary and tertiary interactions between nucleotides. RNA triplexes are one of the most important RNA tertiary motifs [1]; they are usually formed by canonical Watson-Crick base-pairs and noncanonical base-pairs among three nucleotides, two of which are usually involved in double helical regions. RNA triplexes are often found in RNA pseudoknots, for instance forming A-minor motifs within H-knots. Taking into account the importance of RNA triplexes it is necessary to get better understanding of their nature, in particular to classify them by their features with respect to both geometry and secondary structure environment.

Klosterman et al. [2] have proposed a classification of internal loops of RNA secondary structure containing three-dimensional interactions. More specifically internal loops are classified by the number and features of RNA triplexes within the loops. There are nine different classes: internal loops with dinucleotide platforms involved in triplexes (two classes), internal loops with triplexes formed by one looped nucleotide and a base-pair from an adjacent stem (six classes) and internal loops with multiple triplexes (one class). Note that in [2] only 47 loops were considered

Almakarem et al. [3] provide a classification of RNA triplexes according to involved base-pairs. Authors selected 108 families of RNA triplexes, each family is defined by types

of two involved base-pairs according to Leontis-Westhof classification of base-pairs. All families are grouped into 18 superfamilies which are determined by 18 types from Leontis-Westhof classification of base-pairs. The overall number of RNA triplex instances is about 3930.

In our work we have analyzed all RNA triplexes found in RNA-containing entries from Protein Data Bank [4]. The overall number of triplex instances is about 180000 that is significantly greater than in cited works. The data have been prepared using our database of RNA structures URSDB [5].

For each triplex we have extracted the following characteristics: the bases of the three nucleotides, types of involved base-pairs according to the Leontis-Westhof classification, description of secondary structure elements containing the three nucleotides, features of the triplex within the elements and indication of the presence of pseudoknotted regions around the triplex. It allows us to make a classification of triplexes that generalizes both above classifications. The collected data also form a background for further machine learning classification.

We plan to integrate the classification into the web-interface of URSDB available at http://server3.lpm.org.ru/urs/.

1. Devi G, Zhou Y, Zhong Z, Toh DF, Chen G. RNA triplexes: from structural principles to biological and biotech applications. *Wiley Interdisciplinary Reviews: RNA*. 2015 Jan 1;**6**(1):111-28.   -

2. Klosterman PS, Hendrix DK, Tamura M, Holbrook SR, Brenner SE. Three dimensional motifs from the SCOR, structural classification of RNA database: extruded strands, base triples, tetraloops and U turns. *Nucleic acids research*. 2004 Apr 15;**32**(8):2342-52.

3. Almakarem AS, Petrov AI, Stombaugh J, Zirbel CL, Leontis NB. Comprehensive survey and geometric classification of base triples in RNA structures. *Nucleic acids research*. 2011 Nov 3:gkr810.

4. 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.

5. Baulin E, Yacovlev V, Khachko D, Spirin S, Roytberg M. URS DataBase: universe of RNA structures and their motifs. *Database.* 2016 Jan 1;2016:baw085.