

# ANALYSIS AND CLASSIFICATION OF NONSTANDARD RNA MOTIFS

Baulin E.F.<sup>1,2\*</sup>, Spirin S.A.<sup>3</sup>, Roytberg M.A.<sup>1,2,4</sup>

<sup>1</sup>*Institute of Mathematical Problems in Biology, Pushchino, Russia;*

<sup>2</sup>*Higher School of Economics, Moscow, Russia;*

<sup>3</sup>*Belozersky Institute of Moscow State University, Moscow, Russia;*

<sup>4</sup>*Moscow Institute of Physics and Technology, Moscow Region, Russia*

*e-mail: baulin@lpm.org.ru*

*\*Corresponding author*

**Key words:** RNA structure, database, nonstandard motif, pseudoknot, multiplet

*Motivation and Aim:* Investigation of RNA and its spatial structure is one of the main areas of modern molecular biology, both in theoretical (understanding of intracellular processes, the theory of evolution), and in practical parts (drug discovery). Development of research is constrained by the fact that there is no universal annotated database of spatial structures of RNA. One of the reasons for that is the lack of a common classification of RNA structure elements, which would allow one to describe both classical secondary structures and non-standard structures. In our previous work [1] we have proposed the universal classification of RNA structure elements; the database of spatial structures of RNA based on this classification was developed [2]. In this work we study structures formed by interactions between RNA helices (*multiplets*), and some other non-standard RNA structure motifs, e.g. pseudoknots.

*Methods and Algorithms:* RNA-containing structures were selected from Protein Data Bank [3] and then base-pairs were annotated using new program DSSR [4]. These two types of data were used as input of original program package to mine new data and create the database [2]. The database architecture is based on the original classification of RNA secondary structure elements.

*Results:* Classification of multipliers is based on their graph representation; the vertices correspond to helices, the edges reflect H-bonds between helices. Most common multipliers are formed by two helices (12772 cases). The complete list of multiplet types that occur in PDB RNA structures augmented with all occurrences of multipliers was created. The list of all types of multipliers includes 75 types, the most complicated multiplet contain 12 helices (see structure with PDB code 3UZM); a helix may be connected with up to 5 other helices (see structures 3UZM, 1FCW and 3KC4).

*Conclusion:* We have demonstrated usefulness of the created database and proposed classification of non-standard RNA structures.

*Availability:* The beta-version of the database is available at <http://server2.lpm.org.ru/~baulin/home.html>.

## References:

1. Баулин Е. Ф., Астахова Т. В., Ройтберг М. А. // *Математическая биология и биоинформатика*. – 2012. – Т. 7. – №. 2. – С. 567-571.
2. E.F. Baulin, M.A. Roytberg. // *MCCMB 2013: Proceedings*. 2013.
3. Berman H. M. et al. // *Nucleic acids research*. – 2000. – Т. 28. – №. 1. – С. 235-242.
4. Lu X. J., Olson W. K., Bussemaker H. J. // *Nucleic acids research*. – 2010. – Т. 38. – №. 14. – С. 4868-4876.