# Long-range stem-based RNA tertiary motifs

E.F. Baulin

*Educational Center of Mathematical Biology, Pushchino State Institute of Natural Sciences, Pushchino, Moscow region, Russia,* `baulin@lpm.org.ru`

A.V. Korinevskaya

*Faculty of Computer Science, National Research University "Higher School of Economics", Moscow, Russia,* `korinalice@gmail.com`

M.A. Roytberg

*Institute of Mathematical Problems of Biology, 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. There are two ways to categorize tertiary motifs: (1) by distance between its fragments: long-range motifs and motifs formed by tertiary interactions inside a secondary structure element (e.g. internal loop); (2) by secondary structure elements involved in a motif: loop-loop motifs, stem-loop motifs and stem-stem motifs (also called stem-based motifs). Hereinafter the word «stem» means two or more consecutive base-pairs of any type; «helix» is a synonym of «stem». To our knowledge long-range motifs are less studied compared to other ones [1]. The most known types of long-range motifs are A-minor, Ribose Zipper, Kissing Hairpins, Tetraloop-Receptor [2]. The ultimate goal of our work is to systematize long-range stem-based tertiary motifs. Here we present results related to stem-based motifs.

In our previous work [3] we have proposed a description of RNA tertiary motifs which are formed by stem-stem interactions. Each motif was described with the graph, the vertices correspond to stems, the edges reflect either H-bonds between stems or nucleotides shared by two stems). However it turned that this approach leads to too complicated graphs if a fragment of RNA belongs to two stems. In this work we present an improved approach to describe stem-based tertiary motifs. There are two key improvements compared to the previous work, namely, (1) graph representation of a motif was changed (now the vertices correspond to RNA chain fragments and the edges reflect various types of base-pairs between them) and (2) parallel helices formed by non-canonical base-pairs were included.

We used previously developed database URSDB [4] to conduct a search for stem-based motifs. The search revealed about 32000 stem-based motifs, having about 1800 motifs from non-redundant list of PDB structures [5]. As we expected the majority of such motifs are presented by triplets between two adjacent stems (~44%) and A-minor motifs (~28%), data are given for non-redundant database. In total we have revealed 20 types of graphs describing motifs that occur at least 10 times in the non-redundant database. The most complicated one contains 6 RNA fragments forming 7 interacting stems (see PDB 3J0W, chain B, fragments 2070-2074, 2077-2081, 2239-2245, 2435-2438, 2592-2595, 2598,2601).

The data on stem-based tertiary motifs have been integrated into URSDB database and are available via its web-interface (http://server3.lpm.org.ru/urs/) which provides the ability to search for such motifs and to analyze their individual cases.

This work was supported by grant 14-01-93106-NCNILa from RFBR.

1. Petrov, A. I., Zirbel, C. L., Leontis, N. B. (2013). Automated classification of RNA 3D motifs and the RNA 3D Motif Atlas. *RNA*, 19(10), 1327-1340.

2. Tamura, M., Hendrix, D. K., Klosterman, P. S., Schimmelman, N. R., Brenner, S. E., Holbrook, S. R. (2004). SCOR: Structural Classification of RNA, version 2.0. *Nucleic acids research*, 32(suppl 1), D182-D184.

3. Baulin, E. F., Roytberg, M. A. (2015) Stem Multiplets: a New Approach to the Description of Tertiary RNA Motifs. *Mathematical Biology & Bioinformatics*, 10(1), 54-59

4. Baulin, E. F., Roytberg, M. A. (2013) The database of RNA secondary structure elements. *Proceedings of Moscow Conference on Computational Molecular Biology (July 25-28, 2013, Moscow)*

5. Leontis, N. B., & Zirbel, C. L. (2012). Nonredundant 3D Structure Datasets for RNA Knowledge Extraction and Benchmarking. In N. Leontis & E. Westhof (Eds.), Vol. 27, pp. 281–298 (Springer Berlin Heidelberg).