

A non-redundant set of long-range RNA tertiary interactions and their classification

Bohdan D.^{1*}, Baulin E.²

¹ *Moscow Institute of Physics and Technology (National Research University), Dolgoprudny, Russia*

² *International Institute of Molecular and Cell Biology in Warsaw, Warsaw, Poland*

* *bogdan.d@phystech.edu*

Key words: RNA structure, non-coding RNA, long-range interactions, tertiary motifs

Motivation and Aim: It is now widely known that non-coding RNA molecules are a key player of molecular machinery in living cells. Non-coding RNA functions are largely determined by their 3D structure, which is of modular architecture and is made of building blocks of various sizes called tertiary motifs. In the context of RNA secondary structure, all the motifs can be divided into two groups – local motifs and long-range motifs. Local motifs are relatively well described and are subject to a natural comprehensive classification based on five standard secondary structure elements: a hairpin, a bulge, an internal loop, a multiple junction, and a stem. In turn, long-range motifs are formed by interactions between distant secondary structure elements and, at the moment, lack a systematic description. Thus, the purpose of the work is to search for long-range motifs and determine their systemic characteristics.

Methods and Algorithms: In this work, we compiled a non-redundant set of 89 PDB entries with resolution better than 4.0 Å. The set included RNA 3D structure representatives of 97 non-coding RNA families from Rfam. The share of ribosomal RNA residues made up only 58 %, unlike up to 82 % in conventional non-redundant lists of RNA structures. Using the original description of RNA secondary structure elements, we detected 17272 nucleotide residue pairs forming long-range interactions which we defined as two nucleotides on a distance closer than 6 Å that belong to loops and stems distant from each other. Such a simple definition allowed us to annotate all possible direct and mediated interactions without having to set any additional thresholds to define particular interaction types. Then we intersected our nucleotide pairs with the results of several automatic RNA tertiary motifs and interactions annotation tools. The annotations were improved by superposition and clustering procedures and used to build a comprehensive classification.

Results: The result of our work is a dataset of nucleotide pairs forming long-range interactions along with a comprehensive annotation of the interaction types and the structural context of the interacting residues. The dataset is highly heterogeneous, e.g. the adenine(loop)-cytosine(stem) and adenine(loop)-guanine(stem) pairs that correspond to the most common type of long-range RNA tertiary motifs, the A-minor motif, compose only 9 and 12 % of the dataset respectively.

Conclusion: The produced classification and the dataset improves our understanding of the diversity of long-range RNA tertiary interactions and will be a helpful contribution to the field of non-coding RNA 3D structure modeling.