Feature exploration and modelling of quadruplex structures

Marta Szachniuk

Institute of Computing Science, Poznan University of Technology

The theme of the project is embedded in bioinformatics, which is one of the youngest areas of computational sciences. Bioinformatics helps to formulate theoretical models of phenomena that are the subject of research in life sciences and develops computational methods that allow computer simulation of these phenomena and their analysis. It also supports experimental research in biosciences by providing tools for their processing.

This project focuses on quadruplexes, specific quadruple-helix motifs that form in molecules of nucleic acids (DNA and RNA). Their existence in cells of living organisms has been confirmed relatively recently. The recent decade has seen the intensification of experimental research on these structural motifs. Their great potential in biomedicine has been confirmed, among others, in medical diagnostics and therapy of some diseases, including cancer. The possible functions of quadruplexes in cellular processes are currently being analyzed with great commitment. The functions of molecules are closely related to their structure. The analysis of molecule function and structure is, therefore, complementary research, which to a large extent is performed *in silico*. In the case of quadruplexes, however, there are not many computer methods that support the study of these structural motifs. So far, about 10 algorithms have been developed, which on the basis of nucleotide sequence predict potential places of their creation and two databases, which collect selective information about these motifs [1]. Recently, it has also become possible to find quadruplexes in 3D structures determined experimentally. It is offered by two programs, DSSR [2] and RNApdbee [3]. This second program was developed by our team and published in 2018. The analysis of quadruplexes and their tetrads, which we performed using RNApdbee, allowed us to observe topological patterns in the structures of these motifs. This resulted in defining a new ONZ classification for tetrads and quadruplexes based on the topological features of their structures [4].

Continuing the research on quadruplexes, we plan to develop computational methods that will enable a comprehensive analysis of their structure at all levels, the primary, secondary and tertiary. Within the framework of the project, we will create a tool for determining various parameters describing the properties of the quadruplex structure. We will analyze all publicly available nucleic acid structures containing quadruplexes in terms of these parameters' values. We will conduct statistical analyses and cluster structures on the basis of a selected subset of structural features. We will examine the properties of structures for different classes of quadruplexes in the ONZ nomenclature. We also plan to investigate the relationship between the architecture of the quadruplex, its potential function and the vicinity of the quadruplex in the nucleic acid molecule. In this research, we will use neural networks and machine learning methods. We will create a database that will collect comprehensive information on tetrad and quadruplex structures and will allow us to effectively search for structural data. We also plan to develop a method for modelling nucleic acid structures containing quadruplex motifs. It will be based on models obtained from machine learning and on the computational engine of the RNAComposer system [5], which we have developed to predict the 3D RNA structure.

Bioinformatic tools developed by us will be made available within RNApolis [6] - a computational platform dedicated to structural bioinformatics of nucleic acids. They will significantly contribute to the development of structure-functional relationship analysis for nucleic acids. They will influence the development of new methods of modelling and designing nucleic acid molecules with quadruplexes. It will be of great importance for designing new ways of diagnostics and therapy focused on DNA and RNA or generating artificial nucleic acid molecules for the needs of nanotechnology.

[1] E. P. Lombardi, A. Londono-Vallejo, A guide to computational methods for G-quadruplex prediction, *Nucleic Acids Research*, gkz1097, 2019.

[2] X.-J. Lu, H. J. Bussemaker, W. K. Olson, DSSR: an integrated software tool for dissecting the spatial structure of RNA, *Nucleic Acids Research* 43, 2015, 142.

[3] T. Zok, M. Antczak, M. Zurkowski, M. Popenda, J. Blazewicz, R.W. Adamiak, M. Szachniuk, RNApdbee 2.0: multifunctional tool for RNA structure annotation, *Nucleic Acids Research* 46(W1), 2018, W30-W35.

[4] M. Popenda, J. Miskiewicz, J. Sarzynska, T. Zok, M. Szachniuk, Topology-based classification of tetrads and quadruplex structures, *Bioinformatics*, 2019, in press (epub: 7.10.2019, doi:10.1093/bioinformatics/btz738).

[5] Popenda, M, Szachniuk, M, Antczak, M, Purzycka, KJ, Lukasiak, P, Bartol, N, Blazewicz, J, Adamiak, RW, Automated 3D structure composition for large RNAs. *Nucleic Acids Research* 40, 2012, e112.

[6] M. Szachniuk, RNApolis: computational platform for RNA structure analysis, *Foundations of Computing and Decision Sciences* 44(2), 2019, 241-257.