Functional motif-targeted RNA structure modeling

Marta Szachniuk

Poznan University of Technology

The subject of our research project is immersed in the world of bioinformatics, a cutting-edge field that blends biology with computer science to unravel the mysteries of life. Bioinformatics provides scientists with the tools to create theoretical models of biological phenomena and develop computational methods for their simulation and analysis. These models and methods may be used, among other things, to explain life processes at the cellular level or design molecular constructs for biotechnology or biomedicine. In bioinformatics – as in other areas of science – various specializations have developed. For example, structural bioinformatics, to which this project is related, focuses on the study of proteins, DNA and RNA, their structures, structural features, characteristic motifs and the relationship between structure and function. The latter molecule is featured in our research. RNA is an extremely versatile and powerful biopolymer that has been harnessed for use in therapeutics, diagnostics, and biological systems for *in vivo* information processing such as gene regulatory networks or synthetic biology circuits. It can perform computations, store data, and control biological processes within the cellular environment. The function of RNA is closely related to its structure. Particular functional importance is attributed to structural motifs being the subject of many experimental studies. They are not just repetitive strings of genetic code. Some well-known motifs – like triplexes, quadruplexes, etc. – are complex three-dimensional shapes that may be composed of multiple strands held together by both canonical and non-canonical, near- and long-range interactions. The distinguishing feature of others, such as hairpins or internal loops, on the other hand, is the topology of the secondary structure that determines the mutual arrangement of single- and double-stranded fragments of RNA molecule. Understanding structural patterns will lead to major advances in medicine and biotechnology, especially, if we can design molecules containing specific functional motifs. However, predicting and modeling structures with expert-designated intricate shapes is a challenging task, and - so far - not even the best AI-based systems like AlphaFold 3 can handle it. Our project steps in just at this point by addressing the challenge of motif-centric modeling and analysis of functional RNA molecules.

In the project, we will develop new biologically interpretable models for inferring functional motifs in RNA molecules. Based on these models, we will formulate computational problems regarding motif search, identification, classification, clustering, similarity assessment, evaluation, quality assessment, and modeling. We will study the computational complexity of these problems to select appropriate algorithmic techniques and hardware resources, and estimate the probability of obtaining optimal solutions in a reasonable amount of time. We will design efficient computational algorithms and AI models that address the defined problems. We will implement the algorithms in the form of user- and environment-friendly bioinformatics tools, and make their codes available in open repositories like GitHub. We will evaluate the created tools in terms of their accuracy, practical application, and environmental sustainability. We will conduct the first computational experiments using the developed tools to model and analyze exemplary RNA molecules containing motifs representing different structural and functional classes. We will publish the obtained results in open access scientific articles.

In pursuing the above goals, we will rely on the team's experience in modeling and analyzing nucleic acid structures. Our research group has developed more than 20 different computational analytical tools and databases for structural bioinformatics. It is also internationally recognized for its expert modeling of three-dimensional RNA structures with high precision and quality - in RNA-Puzzles and CASP competitions the team has obtained some of the best results for years. The latter are due primarily to the excellent cooperation of the members of our multidisciplinary lab and the ability to combine biological and computational experimentation. These qualities will ensure the smooth implementation of the present project. As a result of the project, practitioners will be able to quickly, accurately and reliably analyze, model and refine RNAs with functional motifs, facilitating advances in structural science, nanotechnology and RNA-oriented biomedicine (diagnostics and therapeutics). The project also responds to the community's growing demand for expert-based tools to evaluate and process models of molecular structures generated by deep learning systems, among others, and extract new knowledge from them.