Multiple sequence alignments are at the core of modern computational biology and bioinformatics. They serve as a rich and unique source of evolutionary information for protein and genome sequences. Various approaches to aligning sequences exist to date that are used routinely by many thousands of researchers in life sciences. The efficient processing of multiple sequence alignments was one of the key features that led to the enormous success of the AlphaFold method in predicting 3D structures of proteins.

Recently, functional RNA molecules were discovered to play major roles in diverse cellular processes. A wide range of non-coding RNAs were discovered including catalytic ribozymes, ligand-binding riboswitches, long non-coding and intronic RNAs, all involved in various aspects of the regulation of gene expression. RNAs are responsible for the advent of mRNA vaccines and RNA-guided genome editing systems, significantly reshaping modern biology and medicine. Often, functional RNAs form complex 3D structures that define their functioning. Structural studies of such RNAs enable grasping the mechanisms underlying their folding.

Yet, the established sequence alignment methods perform relatively poorly when it comes to aligning sequences of structured RNA molecules. **Unlike the 20-letter amino acid alphabet, the 4-letter alphabet of nucleic acids usually does not capture the structural information sufficient to build the structurally correct sequence alignment. Therefore, an additional layer of RNA secondary structure information is commonly used to build sequence alignments of functional RNAs. The problem is that the only way to obtain the RNA secondary structure data with confidence is to have the correct sequence alignment in the first place.** Commonly, this vicious circle is broken down by multiple rounds of alternating sequence alignment and secondary structure prediction attempts facilitated by various kinds of auxiliary experimental data and manual curation by human experts.

The main goal of the proposed project is to develop a software tool encompassing new algorithms for RNA multiple sequence alignment, RNA secondary structure prediction, and structure-based RNA sequence comparison, and apply it for identifying conserved secondary structures, pseudoknots, and alternative folds, that remain hidden from existing approaches. The method will be applicable to a set of unaligned homologous sequences of structured RNAs. Moreover, the proposed approach will enable structure-based comparisons between RNA sequences and sequence alignments. This will be utilized to build an improved classification of non-coding RNA families along with a searchable database.

The new method will be based on the concept of approximate alignment of sequences and will be able to handle non-canonical and alternative structures efficiently.

The expected results of this project will be twofold. The methodological advances of the proposed algorithms and their software implementation have the potential to be of interest to thousands of researchers in RNA science, as it resolves the important bottleneck of many RNA-related pipelines. Moreover, the comprehensive comparative analysis of structured RNAs using the suggested method will elucidate new conserved structural elements of non-coding RNAs that can lead to crucial insights into their roles in vital cellular processes.