Many functional RNAs have been recently discovered, including catalytic ribozymes, ligand-binding riboswitches, and long non-coding RNAs, all playing key roles in various aspects of gene expression regulation. Advances in RNA research are transforming modern biology and medicine, exemplified by the development of mRNA vaccines and RNA-guided genome editing systems. Functional RNAs often adopt complex structures that define their function, e.g., by binding to proteins or small molecules. Similarly, DNA can form various non-canonical structures involved in maintaining genetic stability and linked to human disorders, such as muscular dystrophy and Alzheimer's disease.

Nucleic acid 3D structure is organized hierarchically and can be seen as a set of building blocks, 3D modules. The recurrent modules that retain their structural features across various molecules are called structural motifs and often act as functional units. For instance, ribosomal RNAs use A-minor motifs (planar base triples) to recognize cognate transfer RNAs. Catalytic centers of many ribozymes serve as binding sites for magnesium ions required for their activation. Telomeric DNA G-quadruplexes (planar guanosine tetrads) are known to play a role in cancer and are an active target in drug discovery. Identifying and characterizing these recurrent motifs provide critical insights into the mechanisms underlying the folding and functioning of structured nucleic acids.

Yet, our knowledge of RNA and DNA structural motifs remains sporadic. **No common resource or software tool provides comprehensive coverage of even the most frequent and well-described motifs.** Thus, A-minor motifs and kink-turn motifs, both discovered in 2001, still do not have their articles on Wikipedia. **Consequently, key discoveries are sometimes under-recognized or inconsistently referenced, causing important insights to be missed.** For example, researchers who showed in 2015 that disruption of the kink-turn motif in a spliceosomal RNA leads to neurodevelopmental disorders did not use the term “kink-turn”. Furthermore, although many motifs are known to form only in the presence of specific ions or ligands, this information is not consistently applied to resolve ambiguous small molecule sites in experimental 3D structures. **This gap likely arises from existing computational methods being limited by the annotation of pairwise interactions only. Recently, we developed a novel algorithm ARTEM, the first to identify nucleic acid structural motifs and binding sites of arbitrary size, interaction network, and backbone connectivity.**

The main goal of the proposed project is to develop ARTEMANN, a software tool for identifying recurrent structural motifs in experimental nucleic acid 3D structures, and apply it to annotate motifs and metal ion-, ligand-, and protein-binding sites in known structures, thereby revealing so far hidden structural and functional similarities and potential evolutionary implications. With the new data, we will build an improved classification of nucleic acid 3D structures and their modules, and enrich the description of their characteristic features and variations.

The new method will be based on our ARTEM superposition algorithm that can identify similarities between arbitrary nucleic acid structures with no a priori knowledge of their residue-residue matching. ARTEMANN will be equipped with motif-specific parameter configurations tailored for a rich set of currently known recurrent 3D modules and binding sites, which will be compiled into a manually curated reference library.

The expected results of this project will be twofold. The curated reference library of motifs along with methodological advances of the proposed method and its software implementation have the potential to be of interest to thousands of researchers in the biology of nucleic acids, as it provides easily accessible annotations currently lacking in RNA and DNA structural studies. Moreover, the comprehensive analysis of structured RNAs and DNAs using the proposed method will elucidate new structural and functional features of these molecules that can lead to crucial insights into their roles in vital cellular processes.