Proteins and nucleic acids are the two most important classes of molecules in living organisms. DNA, or deoxyribonucleic acid, contains the genetic blueprint of the organism and stores the information necessary for cellular function. RNA, or ribonucleic acid, is chemically similar to DNA but performs more diverse and dynamic functions. It primarily transmits information from DNA to the cellular machinery responsible for protein synthesis, but many RNA molecules also serve regulatory, catalytic, and sensory roles. RNA can control gene activity, participate in chemical reactions, and influence numerous cellular processes. Most of these RNA functions are carried out in cooperation with proteins.

Proteins are built from 20 types of amino acids. Their sequence determines how each protein molecule folds into a three-dimensional structure, which in turn defines its function. RNA molecules are composed of four basic ribonucleotides A, C, G, and U, which can form base pairs and other interactions, enabling RNA to adopt specific spatial structures necessary for cellular function. The nucleotide sequence determines RNA structure in a way analogous to how the amino acid sequence determines protein structure.

RNA–protein interactions play a vital role in regulating biological processes. They are involved in gene regulation, protein biosynthesis, stress responses, and defense against viruses. Despite their fundamental importance, scientists still do not fully understand the mechanisms by which proteins recognize specific RNA sequences and structures, nor do they possess methods to predictably design new RNA and protein systems. This represents one of the most important challenges in modern molecular biology and biotechnology.

In recent years, there has been a breakthrough in protein structure prediction thanks to artificial intelligence methods such as AlphaFold. It has become possible not only to accurately predict protein folding but also to design new molecules with defined properties. Unfortunately, comparable tools are not yet available for RNA or RNA–protein complexes. There is a lack of methods for designing RNA and protein molecules that interact with high precision and selectivity.

Our team has been developing computational methods for modeling RNA and RNA–protein interactions for many years. We created the SimRNA program for RNA 3D structure prediction and its extended version SimRNA/SimRNP for modeling RNA–protein complexes. We also developed DesiRNA, a method for designing RNA sequences with defined structures, and ClaPNAC, a tool for analyzing contacts between amino acids and RNA bases in known complexes.

In this project, we will integrate our existing tools with new methods to develop a computational platform for designing RNA and protein molecules that interact with high specificity and stability. We will create two computational methods, which can operate independently or be used in combination. The first will use neural networks to predict RNA–protein sequence pairs that fit user-defined spatial structures. The second will extend the functionality of our SimRNA/SimRNP modeling package to enable simultaneous modeling of RNA and protein structure and dynamics, as well as optimization of their sequences with respect to mutual fit, taking into account conformational changes resulting from their interactions. The computational predictions will be experimentally validated using biochemical and biophysical methods, and for the most promising RNA–protein complexes, we will attempt to determine their 3D structures using crystallography and cryo-electron microscopy. The results will be used to further improve the algorithms and increase the accuracy of the design process.

By combining artificial intelligence methods with biomacromolecular structure modeling, we will create universal tools for designing RNA–protein complexes. This will enable the creation of new molecular systems and their variants, a better understanding of biological processes, and the development of new technologies useful in medicine and biotechnology.