Force fields are models that use classical Newtonian mechanics laws to describe molecular behavior at the smallest, atomistic scale. Despite the fundamentally quantum nature of atomistic phenomena, such an approximation enables the study of molecular processes thanks to the significantly larger spatial and temporal scale of phenomena within the reach of classical models. Over many years, the computational biophysics community has developed models of proteins, membrane systems, and DNA to the point where they exhibit predictive power, i.e., they also prove effective in new and never-tested situations.

However, simulations of RNA molecules remain an exception to this rule, as RNA's exceptional flexibility and ability to form metastable, alternative structures cause the computational models of even the simplest structural motifs to exhibit a lack of thermodynamic stability. Moreover, years of developing models based on RNA molecules with defined structure have led to the situation where leading models do not reflect the properties of molecular systems lacking a single structure: in simulations, RNA chains known to be more flexible have an excessive tendency to adopt a compact, static form, mainly due to the overestimated affinity between nucleobases. Despite many voices emphasizing the importance of RNA as promising molecular targets for innovative therapies, these problems prevent the reliable exploration of RNA interactions with potential drugs, as well as the study of numerous molecular mechanisms based on the RNA component, such as splicing or translation.

Within the framework of the project, we propose an innovative solution to this state of affairs: using advanced quantum chemical and machine learning approaches, we will build classical corrections to existing force fields, compatible with existing methodology and with practically identical computational efficiency. For our re-training workflows, we selected the interaction terms responsible for the known problems of existing force fields: the tendency toward aggregation, intramolecular tensions, and the dynamic influence of the solvent. After training appropriate models, we will thoroughly validate them using recently acquired experimental datasets obtained by nuclear magnetic resonance method, as well as selected cases well-known to the RNA community where existing models yield behavior that does not match reference data. Finally, we will release an optimized protocol for performing parametrization of any nucleic acid, as well as use it to conduct independent validation on DNA and selected examples of modified nucleic acids, so-called XNAs.

Thanks to the combination of speed and accuracy, a predictive approach to interactions involving RNA molecules will be possible, both with drug-like molecules and with proteins or other nucleic acids, both natural and designed by humans. This, in turn, will open up new paths for biomedical research in the area of rational drug design, interpretation of experimental data, as well as discovering the functions of an ever-increasing number of classes and motifs of RNA in humans and human pathogens.