

Genes specify the kinds of proteins that are made by cells, but DNA is not directly involved in protein synthesis. The key players in all steps of gene expression are RNA molecules. Their crucial functions for living organisms strongly depend on the ability to form elaborate, three-dimensional (3D) structures. Indeed, entities such as ribozymes are complex RNA machines that perform chemical tasks in cells. There are many such fascinating species in the zoo of RNAs. The native 3D structures of few hundreds of them have been captured by experimental techniques (e.g. X-ray crystallography, NMR, Cryo-EM) with the highest resolutions and stored in databases (e.g. Protein Data Bank, PDB). These structures have been used for building the knowledge-based statistical potentials of SimRNA model recently introduced by the Bujnicki Group. This method performs computational simulations of RNA folding and tertiary structure prediction. Using only sequence information, SimRNA can successfully fold small and medium-sized RNAs into their native-like structures. Despite this and many other mile-stone achievements of the SimRNA approach, it still has certain limitations. The program usually correctly predicts which bases are paired, but the exact type of pairing is predicted less frequently. Why? Although the nucleotide sequence plays a major role in the formation of biomolecular RNA structures, the neglect of non-sequence specific interactions in the current version of SimRNA does not allow for a complete exploration of the conformational space available in RNA folding. Non-canonical interactions, especially of rare-type like Hoogsteen pairs in G-quadruplexes, are not scored as highly probable and therefore are more difficult to capture. G-quadruplexes are non-canonical four-stranded secondary structures, assembled from repeated stretches of guanine-rich sequences. The quadruplex structure is further stabilized by the presence of a cation, especially potassium, which sits in a central channel between each pair of tetrads. Reports have indicated the wide occurrence of RNA G-quadruplexes in non-coding RNAs. It is now clear that G-quadruplexes play an important role in translational regulation and are also considered as potential targets for anti-cancer treatment. Despite the growing interest in RNA G-quadruplexes, only 20 experimentally resolved structures of them have been deposited so far in the Nucleic Acid Database. Hence, discovering the folding mechanism and being able to predict it, is becoming *the Holy Grail* aimed by the state-of-the-art bioinformatics techniques. The difficulty in predicting the folding of an intramolecular RNA G-quadruplex comes from the fact that the most stable conformation is determined by numerous energetic contributions, including: (i) stacking interactions, (ii) hydrogen bonding, (iii) cation coordination and solvent interactions. The occurrence order of interactions and their relative importance in the course of RNA folding is a central question that remains unsolved. In this project we aim is to explore fundamental limitations of the SimRNA simulation methodology and develop a self-consistent computational framework to overcome these problems. By expanding the functionalities of the SimRNA software, the proposed research will deliver a general method for predicting tertiary structures created by non-canonical base pairing and/or stabilized by non-sequence specific interactions, in particular cation coordination. The improved version of SimRNA will hopefully shed light on the mechanisms of G-quadruplexes formation and will therefore unlock valuable information for advancing therapeutics research.