RNA and DNA are the only carriers of genetic information for all living forms on Earth and scrutinizing the relation between their structure/dynamics and functions is the key to understanding biochemistry, finding targets for therapies and developing new treatments. Currently, one of the most challenging questions related to the origins of life on Earth is how these primary genetic information carriers could undergo self-replication in the complete absence of enzymes. Furthermore, it is unclear how RNA and DNA were selected as the primary informational polymers for biology from the rich pool of prebiotically formed nucleotides, which also constituted alternative sugars and various nucleobase analogs. While recent experimental advances in prebiotic chemistry shed some light on possible mechanisms of nonenzymatic RNA self-replication, the progress in this field has been stalled by lack of understanding of the structure and dynamics of self-replicating systems and very fragmentary knowledge about the underlying molecular mechanisms.

With this project, I aim to apply accurate methods of computational chemistry in order to address these challenges and understand the origins of RNA and DNA on Earth. Here, computational chemistry will serve as a molecular microscope and allow to understand the structure of such self-replicating biomolecules or describe the mechanisms and energetics of the underlying chemical reactions. In particular, large-scale molecular dynamics simulations will allow us to characterize the dynamical behavior of self-replicating RNA fragments, whereas quantum chemical calculations will offer the description of key bond breaking and forming events leading do the elongation of the copied nucleic acid strand. Energetic and structural differences observed for different self-replicating polymers will allow us to identify the most efficient variants of the self-replication process and directly support current experimental efforts with mechanistic explanations and predictions of reactivity. To push the current methodological boundaries, I also aim to train machine learning (ML) potentials for nucleic acids based on accurate quantum chemical calculations. Current machine learning techniques enable very efficient fitting of multidimensional functions such as potential energy surfaces of large molecules. As a result, one can obtain a highly accurate empirical potential which can reproduce the physical and chemical properties of a molecular system at fractional costs when compared to standard methods of quantum chemistry. For this purpose, I will use the most recent neural network models (e.g. PhysNet or SpookyNet) to account for the key electrostatic, van der Walls and long-range interactions shaping the structures of nucleic acids. Such an ML potential will then be applied to the most challenging fragments of canonical and alternative nucleic acids which could undergo nonenzymatic self-replication.

In summary, this project will allow us to understand why RNA was selected from the rich pool of different nucleic acid building blocks and how it started copying itself to initiate the proliferation of the first protocells. The results of this study will also allow to better target current searches for biosignatures which is one of the main missions of the recently launched James Webb Space Telescope. Finally, the trained ML potential aimed at a broad spectrum of nucleic acids could also become an accurate and cost-effective tool for the discovery and design of RNA therapeutics.