The evolution of the RNA molecular alphabet.

A long-standing question in evolutionary biology is the evolution of molecular alphabets. Why is the natural alphabet composed of only 4 monomers? Empirical evidence indicates on one hand, that it is possible to synthesize ribozymes with only 2 or 3 types of nucleotides and, on the other hand, that artificial forms of nucleotides can be used to enlarge the molecular alphabet. The goal of this project is to study evolutionary properties of RNA sequence and structure space as a function of the composition and size of the monomer alphabet. We use computer simulations to address the following questions: Are there peculiarities in the alphabet composition of RNA compared to random energy functions generated by random alphabets of size n? Does the natural RNA alphabet possess any evolutionary advantage?

The origin of non-coding RNA in bacterial genomes.

Recently, a variety of non-coding RNA families have been discovered to play important roles on metabolism and information processing in the cell. Although, biochemical and high-throughput genomic studies reveal their ubiquity across prokaryotes and eukaryotes, there is no evidence as to what extent the origin of non-coding RNAs is the result of non-adaptive processes. In this project, we first analyze the taxonomy of non-coding RNAs. Second, we study the underlying RNA structural repertoire accessible to bacterial genomes. Third, we analyze the genomic constraints and evolutionary scenarios that would favor non-coding RNA to arise by non-adaptive means. Fourth, we consider particular examples of bacterial non-coding RNAs characterized empirically.

Protein structural innovations in the *Escherichia* lineage.

Despite the enormous amount of sequence and structure information accumulated over the last decades, we do not understand, neither the evolutionary mechanisms, nor their relative contribution to the origin and evolution of macromolecules. In this project, we explore the role and contribution of mutational mechanisms to the evolution and innovation of protein structures. We study the *Escherichia* lineage, a set of 40 closely related genomes, which provide a high-resolution account on the evolution of proteins at relatively short time scales.