The basic rules of reading the genetic information stored in the DNA of modern organisms have been discovered over 50 years ago and called ‘the standard genetic code’. Today we know all the assignments of amino acids to triplets of nucleotides, called codons, and that is why we are able to translate a coding DNA sequence to the sequence of amino acids which build the encoded protein. However, the researchers do not agree on the issues of how did such mechanism work when life on Earth was just emerging, how did it evolve and why is it universal for almost every modern organism. There were many attempts to answer these questions and the most popular propositions are the stereo-chemical, co-evolution and adaptive hypotheses.

The first one postulates that there are some structural and physico-chemical affinities between amino acids and fragments of nucleic acids. However, such affinities have been proved only in the cases of 5 out of 20 amino acids and therefore we cannot explain the structure of the standard genetic code by using this hypothesis only.

The co-evolution hypothesis assumes that the codons in the ancestral and simpler version of the genetic code encoded only a few amino acids, but along with the evolution of cells newly synthesised amino acids took over some of the codons of their precursors on the biosynthetic pathways. This is a very feasible scenario for the evolution of the genetic code, but still it does not fully explain the particular codon-amino acid pairs present in the standard genetic code.

The third hypothesis postulates that the structure of the standard genetic code is optimal regarding the robustness to mutations and mistranslations which lead to amino acid replacements in the synthesised protein. Indeed, looking at the genetic code table we can notice that most of the amino acids are encoded by two or four codons which differ only in one nucleotide in the third codon position. It is also interesting that amino acids characterised by similar physico-chemical properties usually have the same nucleotide in the second codon position, thus even a mutation in the first position does not cause a replacement which would have very harmful effects on the functionality of the protein.

To check the validity of the adaptive hypothesis, many researchers did calculations comparing the robustness of the standard genetic code with the best optimized theoretical alternatives. However, finding the codes most robust to mutations is a challenge because many factors influence the probability of an amino acid replacement in the protein and all of them should be included in the model. Therefore, the main goal of this project is to investigate the impact of such factors as the probabilities of nucleotide substitutions, codon usage bias, different types of mutations like single-point mutations, insertions and deletions, and also various amino acid properties. By means of an evolutionary algorithm we will search the space of potential theoretical codes to find not only the best but also the worst optimized solutions, and then we will compare their features with the properties of the standard genetic code. It will enable us to place the standard genetic code in the global space of all alternatives more accurately than it was done so far by comparing the standard code with the best and random codes.

The results obtained in this kind of research are of great importance to the scientists involved in modifying the genetic code in living organisms and designing artificial biological systems. This field of synthetic biology is constantly developing and the techniques of changing the assignments of codons and introducing them to organisms are still refining. The knowledge of which assignments of codons to amino acid are beneficial to the organism and which could be changed to improve the desired characteristic is vital in this line of research. Such modifications can be used to produce peptides or proteins including unnatural amino acids with novel properties and construct organisms with new features.