Complex structures that comprise cellular machinery: proteins, nucleic acids, and lipids consist of molecules and macromolecules. The forces, which are mostly responsible for the formation of these structures and their motions which are, in turn, responsible for the functioning of a living cell are know since Coulomb times; each atom of a molecule carries a net electric charge and the attraction and repulsion of the charges result in the formation of a given three-dimensional structure. We do not know, though, how the elemental inter-atomic forces combine to put in order the elements of the sophisticated puzzle termed the dynamic structure of a protein. This phenomenon can be compared to the performance of a piece of music: although the musical piece consists of a sequence of sounds with varying frequency, amplitude, and dynamics, its perception is much more than that of a sum of single sounds.

Proteins are the molecules of life. They are polymers of amino acids connected with each other in a given order, termed the sequence. Sequence direction matters as does reading a word forward or backward. Aminoacid sequence determines protein structure and the structure determines, in turn, its properties, for example whether a protein with a slightly modified (mutated) sequence will be cancerogenic. Experimental methods for protein-structure determination require a lot of time and labor, hence the great importance of theoretical methods for protein-structure prediction. Recently, the methods based on artificial intelligence, in particular AlphaFold developed by Deep Mind, enabled us to predict even very complicated protein structures. However, this success did not move us any closer to learning the merging of elementary interactions results in structure. An attempt to answer this question is the most important goal of the proposed project. It should also be noted that AlphaFold is a very advanced method of recognition of the “meaning” (structure) of a “text” (sequence), which relies upon a database of the already existing patterns, thus relying on database completeness. Moreover, it does not predict protein dynamics or the fuzzy structures of the so-called plastic proteins, which adapt their structures to the binding site. Thus, this method, albeit extremely efficient, is neither errorless nor universal. The latter feature could only be attributed to a method which will be based on learning and understanding the rules of the interplay of interactions to form a complex structure. Statistical physics is the right basis to accomplish this task.

Since many years ago, we have been developing the coarse-grained UNRES model of proteins, which is based on statistical physics. Each amino-acid residue is represented by two objects: the peptide group, which accounts for the formation of hydrogen bonds, responsible for the formation of regular helical and beta-sheet structure, and a sidechain responsible for the specificity of the interactions involving a given aminoacid residue. This simplified representation accelerates the calculations by orders of magnitude, enabling us to simulate the elements of a living cell. At the same time, its statistical-physics foundations result in its high accuracy and reliability.

The scale-consistent approach to coarse-graining developed by our group, which was applied in the derivation of the UNRES model, is based on incorporation of atomic-structure details of the coarse-grained sites in their interaction patterns. As a result, these sites are not spherical (in contrast to most of the coarse-grained models) but have the axial symmetry (as e.g., a rolling pin) and, second, correlation interactions appear, which encompass extensive sections of a chain (e.g., a helix fragment) and their motions. These interactions can be compared to sequences of notes of a musical scale and other organized sequences of sounds. On the other hand, the correlated motion can be compared to turning a key in the lock, which could be described by looking at every atom of the key but it is simpler to consider the rotation of the whole key as a body. In our recent work, we derived the formulas describing the long-range domino-effect-type correlations in protein sequences: a perturbation of residue position at one end is carried along the sequence. We demonstrated that these correlation interactions are reflected in protein structures and, in particular, probably determine the orientation of the portions of protein chain preceding and following regular helix and beta-sheet structures.

In the proposed project, we intend to introduce the correlations mentioned above into the energy function of the UNRES model and to investigate other potentially significant correlation contributions. In particular, we will consider those which involve the interactions between the peptide groups and between the side chains. Introduction of the long-range correlations should result in a major improvement of the ability of UNRES to predict protein structures and dynamics; their absence results in a significant decrease of the accuracy of the predicted structure with chain length even for the so far best-performing AlphaFold. We also intend to investigate whether these correlations enable us to describe allosteric communication (propagation of protein-structure perturbation at one site to another site), which is crucial in the action of most of the receptors. No consistent description of allosteric communication has been developed so far and its practical significance is best reflected in the appearance of an increasing number of the so-called allosteric drugs. We will also apply the long-range correlations in the description of the action of the very efficient molecular rotators (e.g., those moving bacterial flagella), which is important, e.g., in the design of artificial molecular motors. The success of the project will thus contribute both to the development of a very effective method of modeling protein structure and dynamics and to a uniform description of the very important aspects of their dynamics. In the future, its results can be applied in drug design and nanotechnology.