Proteins are essential components of all living organisms; the name comes from Greek πρωτεοσ, which means "of origin". Proteins participate in almost all processes within living cells. The functioning of all organisms is largely dependent on the fact that each of its constituent protein adopts a unique three dimensional structure during folding under physiological conditions. A protein molecule consists of long chains of amino acid residues. Different proteins have a different sequence of amino acids, which assume the well-defined spatial structure that, together with its mobility determines its function. Each protein molecule is thus a nanomachine. Knowledge of the structure and motion pattern is necessary to understand a function of a protein. Errors in protein sequences, termed *mutations*, cause errors in structure (misfolding or aggregation) and can seriously impair the functioning of the entire organism, resulting in serious diseases from cancer to amyloidoses (e.g., the Alzheimer disease).

At present, experimental three dimensional structures are available for only about 7% of proteins with known sequence. The gap between the number of known sequences and known structures is still growing as new protein sequences become available. The experimental techniques for structure determination, primarily X-ray crystallography and nuclear magnetic resonance require a lot of time, labor, and resources to determine a single structure. Moreover, the experimental techniques can provide only limited and fragmentary information about protein dynamics. Therefore, theoretical methods have become a common means to predict protein structure and to simulate the dynamic behavior of proteins. To date, the most successful methods are based on the similarity between the sequence of a target protein and those of the proteins whose structures are already known; these methods are termed knowledge-based method. The candidate structures are selected based on this similarity. This approach is justified because proteins with similar sequences are usually related by evolution and performed similar functions in the ancestral organisms; therefore, their structure should also be similar. This observation can be compared to the similarity of the same meaning of words that sound and are spelled similarly in related languages; e.g., '*house*' in English, '*Haus*' in German and '*hus*' in Danish or Swedish. However, these methods fail when the new sequence is too distantly related to those of known protein. Again, a linguistic analogy can be made: 'to spend' (money) in English and 'spenden' in German, the latter actually meaning 'to donate' (money). On the other hand, the physics-based methods, which are independent of the menu of protein structures known so far, are another approach to protein-structure prediction. In these methods, a mathematical model of physical interactions (such as, e.g., the attraction or repulsions between electric charges) termed a *force field* is constructed and the motion of a protein towards its final functional structure (termed the *native structure*) is simulated by using the mathematical description of motion. These methods are, however, more expensive and not so efficient as knowledge-based methods and, moreover, very sensitive to the inaccurate modeling of the interactions and motion.

The best way to handle the problem of the prediction of protein structure and motion is probably to combine the two approaches. Thus, parts of a protein whose sequence is sufficiently similar to that of the proteins with known structures can be modeled with knowledge-based methods and the weakly similar parts with the physics-based methods. In our previous project, we started to develop such an approach by furnishing knowledge-based information to the UNRES force field developed in our laboratory. In contrast to commonly used force fields such as, e.g., AMBER or CHARMM, UNRES uses a highly simplified representation of proteins, in which each amino-acid is modeled by only two bodies; yet the force field is rigorously physics-based. A simple representation enables us to carry out simulations about 10000 faster than with the all-atom representation, which makes it possible to tackle large proteins or even elements of cellular machinery. In the proposed project, we will improve the developed approach to achieve higher prediction accuracy and extend it towards simulations of dynamics behavior of proteins, as well as to the determination of protein structure from deficient experimental data. With these improvements and extensions, we expect to create a software package with which to study the structure and dynamics of subcellular systems (e.g., respiratory complexes, motor proteins, etc.) and we expect that this tool will find applications in biophysical and health sciences. This software package will be made available to the scientific community on the UNRES web page (www.unres.pl) and through PL-GRID infrastructure.