Proteins are macromolecules, which form the base of the life on the Earth. They constitute cell building stuff, function as enzymes, including those which control the genetic information, relay signals from the outside to the inside of the cell and between tissues or organs, and are a key part of the immune system. It has long been accepted that each protein has a unique structures necessary to perform its function, called the *native structure*. However, the functions of proteins are inseparably associated with their flexibility, which enables them to adapt to variable environment. In particular, the *intrinsically disordered proteins* (IDPs) and proteins with large *intrinsically disordered regions* (IDRs), in other words *plastic proteins*, are encoded by more than 40% of the human genome. Their flexibility enables them to perform different functions, because they can associate with various receptors. On the other hand, the undesirable feature of IDPs is their propensity towards association into oligomers and fibrils, which can be deadly for a cell and, thus, cause the amyloidosis and other conformational diseases.

Due to the flexibility, the experimental investigation of the structure of the IDPs is difficult, because not just one structure but all major structures, which coexist in solution need to be found. The nuclear magnetic resonance (NMR) technique, which is usually applied to determine protein structure in solution enables us to determine average properties (e.g., the distances between selected atoms), which can be compared to taking a picture of a cat at long exposure time. Finding out individual structures from average observables is difficult and needs to be accomplished by using computer modeling. Typically, molecular dynamics (MD) simulations with restraints from experiment imposed on a single structure are carried out. However, usually, no single structure satisfies all restraints. Alternatively, unrestrained MD simulations are carried out and, after they are finished, the conformations, which together reproduce the experimental averages are selected. However, the simulated conformations do not necessarily contain those which, jointly conform with the experimentally measured quantities. Therefore, **the first goal of the proposed project is to develop the Ensemble Oriented Data Assisted Molecular Dynamics (EODAMD) approach to determine the dynamic structure of the IDPs and other flexible proteins.**

The two groups of applicants possess complementary expertise. The Chinese PI is an established scientist in investigating proteins by NMR and other experimental techniques. He has been successful in determining the structure of multistate and flexible proteins. The Polish PI is an established scientist in theoretical modeling of protein structure and dynamics. The proposed approach has two features. First, MD will be used to run multiple trajectories in the so-called replica-exchange mode and the average observables will be computed from all trajectories, thus mimicking the situation in solution. Second, to run calculations faster (typically 1000 times faster compared to standard molecular dynamics packages), the UNited RESidue (UNRES) model of polypeptide chains (www.unres.pl) developed in the group of the Polish PI will be used. Apart from the results of NMR measurements, the results of chemical cross-link (CXMS) and small-angle X-ray measurements (SAXS) will be used. These measurements are highly complementary in studying the IDPs: NMR provides the information of local chain geometry and limited information of interatomic distances, CXMS provides information of the contacts between selected amino-acid residues, while SAXS provided the information of the average shape of the molecule. **The integration of these measurements into the UNRES package is the second goal of the proposed project.**

The third goal of the proposed project is to apply the EODAMD approach to three representative IDPs with the aim to understand how IDP conformation is modulated by environmental cues and by self-crowding and thus to establish a relationship between IDP structure and the propensity of coacervation. The joint effort of the Polish and the Chinese groups will result in the development of a novel approach for the study of dynamic ensemble structures of highly flexible proteins. The successful execution of this project will lead to a clearer picture of the initial assembly process of IDPs and contribute to our understanding of the origin of conformational diseases.