Proteins are very large molecules that are responsible for nearly every task of cellular life, including cell shape and inner organisation, product manufacture and waste cleanup, and routine maintenance. Determination of a protein chemical structure, *i.e.* the sequence of its amino acid is relatively straightforward task and can be performed on a scale of whole genomes. To the contrary, experimental determination of a protein three-dimensional structure is a very laborious and time consuming task. Experimental methods are too expensive to apply them for all proteins. Therefore, the number of known protein sequences exceeds the number of known structures by two orders of magnitude and that gap is steadily widening. This stimulates the development of computational approaches.

A protein molecule is a relatively flexible chain of atoms connected with bonds. In an extended conformation, chemical groups can nearly freely rotate around these bonds. These rotations introduce structural changes to a protein chain and may eventually lead to one of immense number of compact conformation. Only one of them is biologically relevant and active in biological conditions. Computer simulations of that process require generating compact protein states that obey rules of physics and is a computationally demanding task. Energy of each conformation is assessed by a force field function, which depends on coordinates of all N atoms in a system and is typically computed at cost $\mathcal{O}(N \ln N)$. For large molecular system containing tens of thousands of atoms this may be a serious obstacle. Even more severe problem results from the shape of the energy function which features numerous local minima with very steep gradients. The search for global minimum is therefore a very difficult task.

A very promising solution to that problem is the use of coarse-grained model of a protein molecule. In that approach a group of atoms is replaced by a single center of interaction (pseudo-atom). In a SURPASS model, recently proposed by the PI of this project, a single united atom is used to represent whole amino acid residue, which results in almost twentyfold reduction in the number of atoms N. The model defines also a statistical energy function, which greatly reduces energy barriers and facilitates easier search through the conformational space. Thanks to its definition, SURPASS model dramatically reduces computational cost; coarse grained conformations however are quite far from physically relevant all-atom structures. Coarse grained simulation therefore must be followed by reconstruction of atomistic details, structural refinement and optimisation.

In this project, the PI proposes a novel multiscale algorithm for modelling protein structures. The method will combine SURPASS approach with ROSETTA modelling toolkit into a single protocol. The project will require development of a few algorithms related to coarse grained and multiscale modelling. SURPASS model will be further developed towards modelling very large proteins and their complexes. Also, novel method to combine SURPASS with ROSETTA will be devised. An important element of this project is to utilise experimental data to guide conformational sampling and to detect the native (biologically relevant) conformation.