Application of Markov state model in studies on protein folding pathways

Mateusz Pikora

Proteins are linear polymers, usually of biological origins. They consist of amino-acid residues linked by peptide bonds. They have complicated, multi-level, three-dimensional structure, which determines their specialized functions. Studying this structure is an important issue in molecular biology and biochemistry. In living organisms proteins fulfill many different roles, most importantly enzymatic, structural, mechanical and signaling.

Molecular dynamics is a technique enabling simulation of atoms in a given system using Newtonian mechanics. Replica exchange molecular dynamics is a subtype of this method, where several copies (replicas) of given system are simulated simultaneously in different temperatures. Temperatures are exchanged between replicas at defined time intervals, which helps in escaping local energy minima. The outcome of such simulation is a time series of structures representing evolution of studied molecular system, with information about each structure energy and temperature. One method used to analyze this data is clustering. It is a process of dividing structures into groups by defined criteria, such as structural similarity. Different groups represent different system states and transitions between them can be analyzed with Markov Chains. Markov Chain is a mathematical, stochastic model of behavior of given system.

The aim of this doctoral thesis was to investigate the folding mechanism, i.e. the path molecules follow to accomodate a tertiary structure, for several selected proteins. I wanted to verify the following research hypothesis: Proteins do not have one dominant folding pathway understood as an unambiguous and repetitive sequence of events leading from a fully extended structure to a native conformation. Moreover, the native structure is not perfectly stable and a protein can unfold into various, partially folded structures. This creates a network of partially stable states, which I wanted to show.

My strategy to achieve this goal was to use the clustering analysis of molecular structures obtained from simulations, and to build a Markov model using obtained clusters and subsequent preparation of transition graphs between obtained groups of structures. To reach it I used a modified Nearest Neighbor Algorithm. The technical means to achieve my goal was a computer program, which I designed and written using C programming language with OpenMP library for parallelization. In cooperation with the Chemistry Faculty of UG I carried out replica exchange molecular dynamics simulations of chosen proteins and analyzed their results with mentioned program. I succesfully verified my research hypothesis for several selected proteins in two different force fields.