Warsaw, 03/07/2023

Abstract

The subject of the research thesis is the development of methods for analyzing functional mechanisms of (bio)molecular systems.

The first part concerns protein structure networks defined based on a set of nodes (amino acids) connected by edges (contacts), which are assigned an attribute, contact length. In an effort to identify networks that provide an optimal representation of protein structures, the focus was on finding the suitable contact cutoff distance, i.e. the length above which a contact is considered broken. Using an approach based on the information theory, with the aim to determine representation characterized by the highest possible information content per contact, it was shown that the cutoff distance ensuring the fulfillment of this criterion for a wide class of proteins is the same, regardless of the type of protein secondary structure and size. Specifically for the purpose of the analyses, a mean correlation measure was introduced and used, which, when applied to Gaussian signals, is equivalent to a measure of mutual information that quantifies the amount of redundancy in the signal.

In the second part of the study, a contact network model, obtained based on molecular dynamics trajectory of the chignolin protein, was used in the framework of Granger causality analysis to determine the temporal relationships in the process of reversible protein folding into β-hairpin structure. A number of descriptors were introduced to simplify the interpretation of the Granger coefficients matrix. The analysis made it possible to objectively indicate which of the two main mechanisms of chignolin folding, widely debated in the literature, is captured by the considered simulation.

In the last section, a numerically efficient formalism based on quaternion algebra was proposed, allowing the application of Granger causality analysis to temporal signals composed of three-dimensional vectors. Such a formalism allows convenient application of the causality analysis to vector physical quantities, that are commonly found in biomolecular simulations. Its applicability was demonstrated based on simulations of a model system.