Development a Predictor of Aggregation-protein using Supporting Vector Machine

Carlos Moreira, Prof. Dr. Luis Paulo Barbour Scott

Universidade Federal do ABC

Protein has important tasks for the body, such as the catalysis of chemical reactions, transport, and recognition and signal transmission. Although proteins could acquire a huge amount of conformations, they tend to a preferred conformation of lower energy known as native structure. However, there is a class of proteins that although it can be soluble in certain tissues it can be found as insoluble aggregates known as amyloid. According to several researches this kind of conformation is linked with diseases like: Alzheimer, Diabetes-II, and Parkinson. This proteins class coexists in two extremely different stable conformations: native and amyloid forms, the last one consist mainly of β sheets. This work aims to develop a Predictor of Protein Aggregation concerning some features as: sequences of amino acids, physicochemical characteristics and trend to aggregation. In predictor will have a module considerer energy frustration degree and secondary structures. For this purpose we investigated some tools described in literature like Aggrescan, Zyggregator, Pafig, Tango. We have been analyzing their algorithms and models adopted and the circumstances to use them. we are selecting and mapping some features of physicochemical characteristics, propensity of aggregation and statistical aspects of amino-acids. A local database is being created to store proteins amyloid, and other attributes. To set up the data base we are using PostgreSQL tool. For the learning machine method we are using WEKA system/ Support Vector Machine (SVM) at first. We are testing and making adjustments to reach higher performance for prediction. We are using Java language to program into the Linux environment. In the end the result will can be compared with others predictors already mentioned in order to compare performance and accuracy.