Abstract

Prediction of protein secondary structure is an important step on the way to spell out its three dimensional structure and its function. This paper describes a new technique for prediction of secondary structure of protein based on contemporary machine learning methodology and data mining approach. More than one method has been developed to predict the protein secondary
structure from the amino acids sequence; these methods show that we can achieve accuracy up to 80%. The work in this research is consists of three parts. In the first part, the secondary structure of each amino acid is predict alone with naive bays classifier, this method is based on amino acid preferences for different secondary structure. In the second part, an evolutionary algorithm to ameliorate this prediction is used; this method is based on physicochemical properties of protein regions. In the last part, a fragments bank which contains the protein fragments frequently detected in the Protein Data Bank (PDB) was developed; this method is based on the sequence alignment of protein but with a reduced database. The results of this research shows that the proposed method is improved the best know predictive accuracy by 4.5%, and attain 85% accuracy with different datasets.

Reference

- Donald Voet Judith-G Voet Biochimie. 2e édition. De boeck

Index Terms

Computer Science Pattern Recognition
Enhancing Accuracy for Protein Prediction Secondary Structure by a New Hybrid Method

Key words
Protein secondary structure prediction

Genetic algorithm
K nearest neighbor
Data mining
Amino acids
Hybrid method
Supervised learning