predSucc-Site: Lysine Succinylation Sites Prediction in Proteins by using Support Vector Machine and Resolving Data Imbalance Issue

Abstract

The lysine succinylation is found as an important post-translational modification where succinyl group is added to a lysine (K) residue of a protein molecule. It plays major role not only in regulating the cellular processes but also associated with some diseases. As a result, it requires an easiest way to detect succinylation modification in proteins. However, since the experimental technologies are costly and time-consuming, so it is quite hard to detect the succinylation modification timely at low cost to face the explosive growth of protein sequences in postgenomic age. In this context, an accurate computational method for predicting succinylation sites is an urgent issue which can be useful for drug development. In this study, a novel computational tool termed predSucc-Site has been developed to predict protein succinylation sites by (1) incorporating the sequence-coupled information into the general pseudo amino acid composition, (2) balancing the effect of skewed training dataset by Different Error Costs (DEC) method, and (3) constructing a predictor using support vector machine as classifier. The experimental result shows that the predSucc-Site predictor achieves an average AUC (area under curve) score of 0.97 in predicting lysine succinylation sites. All of the experimental results
along with AUC of our system are found from the average of 5 complete runs of the 5-fold cross-validation and those results indicate significantly better performance of predSucc-Site than existing predictors. A user-friendly web server for the predSucc-Site is available at http://research.ru.ac.bd/predSucc-Site/

**References**

Analytical biochemistry, 525, 107-113.


prediction of sumoylation sites based on hydrophobic properties. PLoS One, 7(6), e39195.

Index Terms

Computer Science
Artificial Intelligence

Keywords

Lysine Succinylation Sites Prediction, Sequence-coupling Model, General PseAAC, Data Imbalance Issue, Support Vector Machine