Abstract

The lysine succinylation is found as an important post-translational modification where succinyle 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

S-nitrosylation sites in proteins by incorporating position specific amino acid propensity into
modifications: the chemistry of proteome diversifications. Angewandte Chemie International
succinylation sites in proteins by incorporating sequence-coupling effects into pseudo
components and optimizing imbalanced training dataset. Analytical Biochemistry, 497, 48-56.
succinylation sites using an iterative semi-supervised learning technique. Journal of Theoretical
Biology, 374, 60-65.
sites in proteins with PseAAC and ensemble random forest approach. Journal of Theoretical
Biology, 394, 223-230.
tool for the prediction of protein succinylation sites by exploiting the amino acid patterns and
online prediction tool via enhanced characteristic strategy. Bioinformatics, 31(23), 3748-3750.
International Journal of Pattern Recognition and Artificial Intelligence, 23(04), 687-719.
predicting the interactions between drug compounds and target proteins in cellular networking
via benchmark dataset optimization approach. Journal of Biomolecular Structure and Dynamics,
33(10), 2221-2233.
vector machines. In Proceedings of the International Joint Conference on Artificial Intelligence,
pp. 55-60.
sites prediction in proteins using support vector machine with resolving data imbalanced issue.
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