[BIOINFO 2023]

CAFA-Powered Insights: Predicting Protein Function in the Post-Genomic Era

Kim Rubiga[1,*]

[1] *Department of Life Science, Handong Global University, Pohang 37554, Korea*
**Corresponding author: kimruby86@handong.ac.kr*

Proteins are complex biological molecules composed of chains of amino acids. The three-dimensional structure of proteins is crucial for their function. Understanding protein function is essential for drug development, disease diagnosis, and therapy design. In this research, I would utilize computational algorithms that aim to predict Gene Ontology (GO) terms for protein sequences based on their amino acid composition. The Gene Ontology (GO) provides a framework for describing protein functions and their relationships. By leveraging a training set of annotated proteins with known GO terms, the algorithm employs machine learning to uncover relationships between amino acid sequences and protein functions. Therefore, by learning the relationship between amino acid sequences and their functions, the model can predict GO terms for unannotated proteins. The model incorporates various features, such as physicochemical properties and evolutionary conservation. In order to improve the accuracy of protein function prediction, I plan to utilize the Critical Assessment of Protein Function Annotation (CAFA) challenge as a benchmark. Through the combination of machine learning and the Gene Ontology Framework, this study aims to enhance our understanding of proteins and their complex functions, ultimately offering significant benefits to biomedical applications through the utilization of robust computational tools.