A New Scheme to Predict Kinase-Specific Phosphorylation Sites on Protein Three-Dimensional Structures

蘇敏綱, 黃凱堯, 董其樺, 李宗夷
Bioinformatics
Computer Science and Informatics
chihua.tung@chu.edu.tw

Abstract

Due to the high-throughput of mass spectrometry-based phosphoproteomics experiment, the desire to annotate the catalytic kinases for in vivo phosphorylation sites has motivated. Many researches are undertaken to develop a computational method for the identification of kinase-specific phosphorylation sites using linear amino acid sequences. With an increasing interest in the structural environment of protein phosphorylation sites, herein, a new scheme has been developed for identifying kinase-specific phosphorylation sites on protein three-dimensional (3D) structures. For a large-scale investigation on 3D structures, all of the experimental phosphorylation sites are mapped to the protein entries of Protein Data Bank by sequence identity. In this work, a support vector machine (SVM) is applied to generate the predictive model learned from the information of spatial amino acid composition and structural alphabet. After the cross-validation evaluation, most of the kinase-specific models trained with the consideration of structural information outperform the models considering only the sequence information. Moreover, the independent testing set which is not included in training set has demonstrated that the proposed method could provide a stable performance. This study has demonstrated that the consideration of spatial context could improve the predictive performance compared to the model only considering the local sequence motifs.

Keyword：Phosphorylation, protein kinase, three-dimensional structure, structural alphabet, spatial amino acid composition.