

A method for predicting the active sites of protein tyrosine phosphatases  
based on protein surface patches

黃裕翔, 劉志俊, 張慧玫

Bioinformatics

Computer Science and Informatics

wmchang@chu.edu.tw

Abstract

Proteins are main macro-molecules to operate the physical activities and catalytic functions in organisms. Proteins interacting with other proteins regulate the flows and various responses of biological signal transduction pathways. To ensure timing-switch of protein activities in signaling, the phosphorylation and de-phosphorylation reactions of the proteins involved are the most common and critical regulation. The main focus of the present study is to predict the active site of protein family such as protein tyrosine phosphatases. We start off the training set using a PTP subclass-DSPs. DSPs are named after their ability to regulate the substrate activity by simultaneously removing the phosphate group from phosphate-modified tyrosine and serine/threonine residues in the substrate proteins. We first collected the conserved "CXXXXXR" (X as arbitrary amino acid) active site sequences, and recorded the spatial coordinates of the cysteine (C) and arginine (R) from the active sites plus the aspartate (D) from catalytic sites of structure-solved DSPs, and those of the phosphates (P) from substrates. The distances between enzymatic C, R, and D are estimated as active site parameters to predict the possible presence of DSP active site cavity from unknown proteins. To prevent some possible skew coordinates obtained from DSPs structure without their substrates, the distances from enzymatic C, R, and D to substrate P are estimated as interaction parameters. The  $\alpha$ -ball protein surface modeling was applied to screen unknown proteins for the possible presence of active site topology before entering DSP-parameter comparison. We were able to predict the active sites of 16 structure-solved DSPs as well as 23 PTPs with unknown structure.

Keyword : protein tyrosine phosphatases, dual-specificity phosphatases, active site,  $\alpha$ -ball protein surface structure modeling