

Abstract

Project Code : MRG5580188
Project Title : Sequence-based interaction prediction for mouse PDZ domains and peptide ligands
Investigator : Assoc. Prof. Dr. Songyot Nakariyakul
Department of Electrical and Computer Engineering,
Thammasat University
E-mail Address : nsongyot@engr.tu.ac.th
Project Period : 2 years (2 July 2012 – 1 October 2014)

Abstract:

The PDZ domain is one of the most ubiquitous protein domains that is involved in coordinating signaling complex formation and protein networking by reversibly interacting with multiple binding partners. It has been linked to many devastating diseases such as avian influenza, Fraser syndrome, Usher syndrome and Dejerine-Sottas neuropathy. Understanding the selectivity of PDZ domains can help elucidate how defects in PDZ proteins and their binding partners lead to human diseases. Since experimental methods to determine the interaction specificity of the PDZ domains are expensive and labor intensive, an accurate computational method is tremendously needed. Our developed support vector machine-based predictor using dipeptide composition is shown to qualitatively predict PDZ domain-peptide interaction with a high accuracy rate. Furthermore, since most of the dipeptide compositions are redundant and irrelevant, we propose a new hybrid feature selection technique to select only a subset of these compositions for interaction prediction. The experimental results show that only approximately 25% of dipeptide features are needed and that our method improves the prediction results significantly. The selected dipeptide features are also analyzed and shown to play important roles in specificity patterns of PDZ domains. Our method is based only on primary sequence information, and it can be used for the research of drug target and drug design in identifying PDZ domain-ligand interactions.

Keywords : Dipeptide composition; Feature selection; PDZ domain-peptide interaction; Protein interaction; Protein sequence encoding