元智大學

資訊工程學系

碩士論文

硫端棕櫚醯化位置於肝組織具有受質專一性之辨識調查

Investigation and identification of S-palmitoylation sites with substrate specificity in the liver tissues.

研究生： Ho Thi Trang

指導教授： Dr. Tzu-Ya Weng

共同指導教授: Dr. Tzong-Yi Lee

中華民國 一百零五年 一月

硫端棕櫚酰化位置於肝組織具有受質專一性之辨識調查
Investigation and identification of S-palmitoylation sites with substrate specificity in the liver tissues.

研 究 生： 胡氏妝　　Student: HO THI TRANG

指導教授: 翁資雅博士　　Advisor: Dr. TZU-YA WENG

共同指導教授: 李宗夷博士　　Joint advisor: Dr. TZONG-YI LEE

元 智 大 學

資 訊 工 程 學 系

碩 士 論 文

A Thesis
Submitted to Department of Computer Science & Engineering
Yuan Ze University
in Partial Fulfillment of the Requirements
for the Degree of Master
in
Computer Science and Engineering
January 2016
Chungli, Taiwan, Republic of China.

中華民國 一百零五年一月

硫端棕櫚醯化位置於肝組織具有受質專一性之辨識調查

學生： 胡氏敘　　指導教授： 翁資雅 博士
共同指導教授: 李宗英博士
元智大學資訊工程學系

摘要

硫端棕櫚醯化是經由酶的催化下飽和的 16 個碳的棕櫚酸鹽通過硫酯鍵共價修飾到蛋白質半胱氨酸的鹼基上硫鍵共價連接，是轉譯後修飾的形式之一。它是動態、可逆的，在蛋白質-蛋白質交互作用與蛋白質運輸過程中息息相關，也涉及到生物過成的多樣性...等等，是非常重要的蛋白質轉譯後修飾之一。而現今對於硫端棕櫚醯化位置受質專一性的辨識還沒有找到一個有效的方法去辨認。因此運用資訊技術找尋有效的計算方法成為了十分迫切的需求，希望能以資訊的方法快速辨別硫端棕櫚醯化位置。本研究於 UniProtKB 資料庫以及兩篇文獻(Forrester MT et al; Yang W et al)中蒐集了肝組織中硫端棕櫚醯化位置的實驗驗證資料，總共整理出 710 個位置資料。並提出了一種邏迴統計的方法來辨識肝組織中硫端棕櫚醯化位置受質之專一性。並以五倍交叉驗證之形式使用 SVM 分類器，進行模擬測試及評估效能，結果表示所開發的方法的確有效的提升敏感性，特異性和準確性。另外使用獨立測試資料對硫端棕櫚醯化位置測試之結果進行評估。最後，本研究將此新穎的方法研究結果，呈現於免費的網站，並可幫助生物學家分析硫端棕櫚醯化位置。(網址: http://csb.cse.yzu.edu.tw/MDDPalm/)

關鍵字：後轉譯修飾作用, 最大依賴分解度.

Investigation and identification of S-palmitoylation sites with substrate specificity in the liver tissues.

Student : Ho Thi Trang

Advisor : Dr. Tzu-Ya Weng
Joint advisor: Dr. Tzong-Yi Lee

Department of Computer Science and Engineering
Yuan Ze University

Abstract

S-palmitoylation, the covalent attachment of 16-carbon palmitic acids to a cysteine residue via a thioester linkage, is an important reversible lipid modification (PTM) that regulates protein trafficking, protein-protein interaction. It also related to diversity of physiological and biological process, etc… However, the substrate specificity of cysteine S-palmitoylation remains unknown. Thus, finding the effective computational method to predict S-palmitoylation sites is urgent demand in bioinformatics. Based on total of 710 experimentally verified S-palmitoylation sites in the liver tissues were collected from UniProtKB, Forrester MT et al and Yang W et al papers. This study presents a recursively statistical method to identify conserved substrate motifs for S-palmitoylation in the liver tissues. Statistical significance Support vector machine (SVM) was applied to construct predictive model learned from verified substrate motifs. The evaluation of five-fold cross-validation indicated that the model trained with identified motifs were effective in identification of S-palmitoylation sites with an enhanced sensitivity, specificity, and accuracy. It also provided a promising performance in an independent testing set. The correct identification of previous report S-palmitoylation sites of mouse protein demonstrated the effectiveness of the proposed method and it indicated that the proposed method could be a practicable ways of conducting primary analyses of proteins S-palmitoylation in the liver tissues. Finally, the constructed models have been implemented as a web-based system freely available at http://csb.cse.yzu.edu.tw/MDDPalm/ for identifying uncharacterized S-palmitoylation sites on the protein sequences.

Keywords: S-palmitoylation, Support vector machine, maximal dependence decomposition.

Acknowledgement

Firstly, I would like to express my thanks to Yuan Ze University for the full sholarship which has illuminated my long-lasting dream of obtaining a Master Degree in Computer Science and Engineering.

Second, I am thankful to my thesis advisors, Professor Tzong-Yi Lee, Professor Tzu-Ya Weng who suggested an interesting research topic to me and helped to keep it on the right track.

Finally, my sincere thanks to my family, my bear friend, my labmates, all the YZU teachers and friends that I have ever met during the two years of study in Taiwan. This thesis could be finished on time is largely due to their support.

Contents

List of Tables

List of Figures