

ศิริวรรณ แต้วิจิตร 2552: การจำแนกชั้นแฟมิลีและวิเคราะห์โครงสร้างของเอนไซม์
ด้วยวิธีการพิจารณาการติดกันของลำดับกรดอะมิโนและโพรไฟล์ฮิดเดินมาร์คอฟโมเดล
ปริญญานิพนธ์วิทยาศาสตรมหาบัณฑิต (วิศวกรรมคอมพิวเตอร์) สาขาวิศวกรรม
คอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์ อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก:
รองศาสตราจารย์กฤษณะ ไวยมัย, Ph.D. 111 หน้า

การค้นพบเอนไซม์ใหม่มีอัตราการค้นพบที่สูงขึ้นอย่างรวดเร็ว เพื่อลดเวลาและค่าใช้จ่าย
ในการจำแนกประเภทเอนไซม์ จึงมีการพัฒนางานวิจัยเพื่อสร้างระบบอัตโนมัติด้วยวิธีการต่างๆ
ได้แก่ (1) วิธีการเปรียบเทียบเอนไซม์ใหม่และเอนไซม์ต้นแบบที่ทราบฟังก์ชันแล้ว วิธีดังกล่าวให้
ความถูกต้องสูงเมื่อเอนไซม์ใหม่มีความเหมือนกับเอนไซม์ต้นแบบมากกว่า 60% (2) วิธีการ
พิจารณาส่วนย่อยของลำดับกรดอะมิโน ซึ่งสามารถบ่งบอกฟังก์ชันได้ (3) วิธีการสร้างฟีเจอร์เพื่อ
เป็นตัวแทนเอนไซม์ เช่น พิจารณาความถี่ในการเกิดกรดอะมิโน วิธีดังกล่าวให้ความถูกต้องได้ดี
แต่กระนั้นผลลัพธ์ที่ได้ไม่สามารถแสดงความสัมพันธ์เกี่ยวกับโครงสร้างของเอนไซม์ใหม่ได้

ในงานวิจัยนี้ทำการพัฒนาระบบจำแนกประเภทเอนไซม์ชั้นแฟมิลีที่ให้ความถูกต้องสูง
และให้ผลลัพธ์ที่สามารถนำไปวิเคราะห์โครงสร้างเอนไซม์ใหม่ ในรูปแบบการติดกันของ
โครงสร้างทุติยภูมิ โดยมีสมมติฐานที่ว่าส่วนย่อยของโครงสร้างเอนไซม์สามารถบ่งบอกหน้าที่
การทำงานของเอนไซม์ได้ สำหรับส่วนย่อยของโครงสร้างนั้นสืบค้นจากตารางเมตริกซ์คอนแทค
แมป ซึ่งเป็นการแปลงข้อมูลโครงสร้าง 3 มิติให้อยู่ในรูปตารางเมตริกซ์ที่เข้าใจง่าย และนำแต่ละ
ส่วนย่อยมาประกอบกันเป็นตัวแทนสายโปรตีนด้วยโพรไฟล์ฮิดเดินมาร์คอฟโมเดล ดังนั้นข้อมูล
ที่ใช้ในการสกัดฟีเจอร์จำเป็นต้องทราบโครงสร้าง 3 มิติแล้วเท่านั้น คิดเป็นสัดส่วนเพียง 11.21%
ของข้อมูลทั้งหมดในงานวิจัยนี้ และสร้างแบบจำลองด้วยซัพพอร์ตเวกเตอร์แมชชีน จากผลการ
ทดลองพบว่าฟีเจอร์จากตารางเมตริกซ์คอนแทคแมป มีความสัมพันธ์และมีนัยสำคัญเพียงพอใน
การจำแนกประเภทเอนไซม์ โดยแบบจำลองให้ค่าความถูกต้องถึง 73.71% เมื่อทดสอบด้วยวิธี
jackknife และเมื่อเพิ่มองค์ความรู้เกี่ยวกับการแทนที่ระหว่างกรดอะมิโน สามารถเพิ่มความถูกต้อง
ได้ถึง 75.87% เมื่อทดสอบด้วยวิธี 10-fold cross validation

Siriwon Taewijit 2009: Enzyme Subfamily Classification and Structure Analysis Using Residue Contact and Profile HMMs. Master of Engineering (Computer Engineering), Major Field: Computer Engineering, Department of Computer Engineering. Thesis Advisor: Associate Professor Kitsana Waiyamai, Ph.D.
111 pages.

The number of enzyme sequence grows exponentially over time. Identifying of class for a new found enzyme is usually performed by experiments in laboratory. Unfortunately, it is both time-consuming and costly. Large number of researches on enzyme classification have been carried out for decade such as (1) homology-based, this approach got high accuracy when similarity score between target and template is more than 60%. (2) subsequence-based, this approach aims to find useful portions that related to enzyme function. (3) feature-based, this approach tries to represent enzyme with the new form. Their results were able to generate good accuracy but they cannot exhibit relationships between enzyme and its structure.

In this research, we propose an enzyme subfamily classification method using residue contact and profile HMMs. Our hypothesis is based on the fact that enzyme function is directly related to its structure. First, each fragment from contact map matrix which is representative of 3D structure is extracted. Then, profile HMMs is used to build feature vector to represent each enzyme sequence. Notice that only enzymes with known PDB (11.21% of the training data) have been used to extract features. Finally, SVM was used to build a classifier. Our method yields high accuracy and provides structural description based on secondary structure contact on 3D domain. Empirical results show that our method yields 73.71% accuracy with jackknife test and the accuracy rises up to 75.87% when substitution group of amino acids is used for background knowledge. These suggest that our feature extraction from contact map and Profile HMMs are sufficiently significant for enzyme subfamily classification.