

พีระ ลีวลม 2551: การพัฒนาตัวแทนสายโปรตีนสำหรับการทำนายประเภท
ฟังก์ชันเอนไซม์ ปริญญาวิทยาศาสตรดุษฎีบัณฑิต (วิศวกรรมคอมพิวเตอร์) สาขา
วิศวกรรมคอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์ ปรชานกรรมการที่ปรึกษา:
รองศาสตราจารย์กฤษณะ ไวยมัย, Doctorat d'Universite 102 หน้า

วิทยานิพนธ์นี้มุ่งพัฒนาตัวแทนสายโปรตีนประเภทโมทีฟสำหรับการทำนายประเภท
ฟังก์ชันเอนไซม์ โดยนำเสนอตัวแทนสายโปรตีนชนิดใหม่ที่เรียกว่า รีแอกทีฟโมทีฟ (reactive
motif) ซึ่งมีความสัมพันธ์โดยตรงกับการเกิดฟังก์ชันเอนไซม์ในบริเวณจับและบริเวณเร่ง โดย
ปัญหาวิจัยที่สำคัญก็คือ การขาดแคลนข้อมูล (มีเพียงประมาณ 3.34% เท่านั้นจากจำนวนเอนไซม์
ทั้งหมด) ในการนำมาสร้างรีแอกทีฟโมทีฟ ดังนั้นในวิทยานิพนธ์นี้จึงนำเสนอวิธีการด้านสถิติ
ร่วมกับการใช้ความรู้พื้นหลังด้านชีวเคมีในการสร้างและพัฒนาคุณภาพของรีแอกทีฟโมทีฟ
สำหรับวิธีการทางสถิติใช้ในการขยายกลุ่มข้อมูลบริเวณจับและบริเวณเร่งให้ได้ปริมาณที่
เหมาะสมและมีคุณภาพ ส่วนการใช้ความรู้พื้นหลังเช่น แผนภาพของ Taylor และ BLOSUM62
ถูกนำมาใช้สำหรับเพิ่มคุณภาพองค์ประกอบย่อยของรีแอกทีฟโมทีฟ โดยใช้คอนเซ็ปต์ของ การ
ควบคุมการกลายพันธุ์ (mutation control) ในการพัฒนากลุ่มแทนที่กรดอะมิโนเป็น กลุ่มแทนที่
กรดอะมิโนที่สมบูรณ์ โดยคอนเซ็ปต์การควบคุมการกลายพันธุ์ดังกล่าวถูกอธิบายและจัดให้อยู่ใน
ระเบียบแบบแผนของทฤษฎีคอนเซ็ปต์เลขทิส ทั้งนี้เครื่องมือวัดประเภท sensitivity precision
specificity และ coverage ถูกนำมาใช้วัดคุณภาพของรีแอกทีฟโมทีฟที่ได้จากงานวิจัย ผลการวัด
พบว่ารีแอกทีฟโมทีฟมีคุณภาพในเชิง sensitivity และ coverage ดีกว่าโมทีฟ PROSITE ที่พัฒนา
โดยผู้เชี่ยวชาญ ในขณะที่ค่าคุณภาพที่เหลือมีค่าใกล้เคียงกัน สำหรับเครื่องมือวัดประเภท
accuracy ที่ได้จากการทำนายประเภทฟังก์ชันเอนไซม์ ใช้วัดคุณภาพของรีแอกทีฟโมทีฟเมื่อ
นำมาใช้เป็นคุณลักษณะเด่นในการเรียนรู้ด้วยขั้นตอนวิธี C4.5 สำหรับทำนายประเภทฟังก์ชัน
เอนไซม์ ผลวิจัยพบว่าการใช้รีแอกทีฟโมทีฟดังกล่าวให้ค่า accuracy ถึง 72% เมื่อเปรียบเทียบกับ
67% ที่ใช้ PROSITE

พีระ ลีวลม

ลายมือชื่อนิติ

ลายมือชื่อประธานกรรมการ

3 / 6 / 2551

Peera Liewlom 2008: Development of Protein Sequence Representation for Enzyme Function Classification. Doctor of Engineering (Computer Engineering), Major Field: Computer Engineering, Department of Computer Engineering. Thesis Advisor: Associate Professor Kitsana Waiyamai, Doctorat d'Universite ;102 pages.

This thesis is concentrated in developing a motif-based protein sequence representation for enzyme function classification. The main objective of this thesis is to propose a new protein sequence representation called reactive motifs. Reactive motifs are motifs that are related directly to enzyme functions and are generated from binding and catalytic sites. Main challenge is a lack of data (only 3.34% data available of all enzymes) at binding and catalytic sites to generate reactive motifs. Therefore, a method that combines statistics and bio-chemistry background knowledge is proposed to generate and improve reactive motifs. Statistics are used to extend binding and catalytic sites data. Bio-chemistry background knowledge such as Taylor – Venn’s diagram and BLOSUM62 are used to improve the quality of the reactive motif elements. The concept of mutation control is introduced, which uses amino acid substitution groups to generate maximal amino acid substitution groups. Mutation control operations are described and formalized using concept lattice. Sensitivity, precision, specificity, and coverage measures are used to assess the quality of the discovered reactive motifs. Experimental results show that the discovered reactive motifs provide better quality in terms of sensitivity and coverage compared to PROSITE expert-based motifs. To assess the accuracy in enzyme function classification, the reactive motifs are used as input to the C4.5 learning algorithm. Experiments using reactive motifs as feature predict enzyme function with 72% accuracy compared with 67% accuracy using PROSITE motifs.



Student’s signature



Thesis Advisor’s signature

3 / 6 / 2008