โปรแกรมวิเคราะห์ฮอมอโลยีของโปรตีนโดยใช้แผนภูมิ 2 มิติของกรดอะมิโน

นาย วิทูร วิริยพิพัฒน์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาวิทยาการคณนา ภาควิชาคณิตศาสตร์
บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2542

ISBN 974-333-158-1

PROTEIN HOMOLOGY ANALYSIS USING 2-DIMENSIONAL AMINO ACID PATTERN PROGRAM

Mr. Witoon Wiriyapiat

A Thesis Submitted in Partial Fulfillment of the Requirements

for the Degree of Master of Science in Computational Science

Department of Mathematics

Graduate School

Chulalongkorn University

Academic year 1999

ISBN 974-333-158-1

หัวข้อวิทยานิพนธ์	โปรแกรมวิเคราะห์ฮอมอโลยีของโปรตีนโดยใช้แผนภูมิ 2 มิติของกรดอะมิโน
โดย	นายวิทูร วิริยพิพัฒน์
ภาควิชา	คณิตศาสตร์
อาจารย์ที่ปรึกษา	อ.ดร.รัฐ พิชญางกูร
อาจารย์ที่ปรึกษาร่วม	รองศาสตราจารย์.ดร.เดวิด รูฟโฟโล

บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

.....คณบดีบัณฑิตวิทยาลัย

(รองศาสตราจารย์ ดร.สุชาดา กีระนันทน์)

คณะกรรมการสอบวิทยานิพนธ์

.....ประธานกรรมการ

(ศาสตราจารย์ ดร.ชิดชนก เหลือสินทรัพย์)

.....อาจารย์ที่ปรึกษา

(อ.ดร.รัฐ พิชญางกูร)

.....อาจารย์ที่ปรึกษาร่วม

(รองศาสตราจารย์ ดร.เดวิด รูฟโฟโล)

.....กรรมการ

(อ.ดร.เลอสรร ธนสุกาญจน์)

วิทูร วิริยพิพัฒน์ : โปรแกรมวิเคราะห์ฮอมอโลยีของโปรตีนโดยใช้แผนภูมิ 2 มิติของกรดอะมิโน (Protein Homology Analysis Using 2-Dimensional Amino Acid Pattern Program) อ. ที่ปรึกษา อ.ดร.รัฐ พิชญางกูร อ. ที่ปรึกษาร่วม รองศาสตราจารย์ ดร.เดวิด รูฟโฟโล , 101 หน้า ISBN 974-333-158-1.

ความก้าวหน้าทางพันธุวิศวกรรม และชีวเคมี ทำให้ทราบถึงลำดับของกรดอะมิโนของโปรตีนมากมาย การทดสอบฮอมอโลยีของโปรตีนจากลำดับของกรดอะมิโนยังมีข้อจำกัดอยู่มาก วิธีการที่นิยมใช้และเป็นมาตรฐานในปัจจุบันคือ วิธี alignment ลำดับของกรดอะมิโน แต่วิธีการนี้ยังประสบปัญหาอันเกิดจาก การเว้นช่วง (gap) และการมี sequence identity ต่ำ ซึ่งทำให้ไม่สามารถทดสอบความคล้ายกันของลำดับของกรดอะมิโนได้

การวิเคราะห์ลำดับของกรดอะมิโนโดยปรับข้อมูลให้อยู่ในรูปแผนภูมิ 2 มิติเป็นเทคนิคที่มีผู้นำมาใช้อย่างได้ผลในการทำนายโครงสร้างทุติยภูมิและความคล้ายของโปรตีน โครงการนี้ได้นำแนวคิดนี้มาผนวกกับข้อมูลทางเคมี และชีวเคมีของกรดอะมิโนแต่ละตัว สร้างแผนภูมิ 2 มิติใหม่เพื่อใช้ในการเปรียบเทียบฮอมอโลยีหรือความคล้ายของโปรตีน ด้วยแนวคิดที่ว่าโปรตีนที่ฮอมอโลกัสกันจะให้แผนภูมิ 2 มิติที่เหมือนกันหรือคล้ายกัน แล้วใช้ neural networks ช่วยในการพิจารณาความเหมือนกันหรือคล้ายกันของแผนภูมิ 2 มิติใหม่ที่สร้างขึ้น พบว่าเมื่อนำโปรตีนที่มี sequence identity ต่างๆกัน ตั้งแต่ร้อยละ 35 ถึง ร้อยละ 97 มาสร้างแผนภูมิ 2 มิติ ผลที่ได้คือ โปรตีนที่ฮอมอโลกัสกันก็จะให้แผนภูมิ 2 มิติที่เหมือนหรือคล้ายกัน และได้สร้างโครงข่ายประสาทเทียมพิจารณาความเหมือนหรือคล้ายกันของแผนภูมิ 2 มิติขนาดเล็ก ที่เกิดจากลำดับของกรดอะมิโน 18 ตัว จากโปรตีน 5 กลุ่ม พบว่าโครงข่ายประสาทเทียมสามารถพิจารณาเปรียบเทียบได้ถูกต้องประมาณร้อยละ 86 ถึง ร้อยละ 92

ภาควิชา คณิตศาสตร์ ลายมือชื่อนิสิต .....

สาขาวิชา วิทยาการคณนา ลายมือชื่ออาจารย์ที่ปรึกษา .....

ปีการศึกษา 2540 ลายมือชื่ออาจารย์ที่ปรึกษาร่วม .....

4072388523 : Major Computational Sci

Key word : Protein / Homology / Amino Acid / Neural Network /

WITOON WIRIYAPIPAT : PROTEIN HOMOLOGY ANALYSIS USING 2-DIMENSIONAL AMINO ACID PATTERN PROGRAM. THESIS ADVISOR : DR. RATH PICHYANGKURA, Ph.D., THESIS COADVISOR : ASSOC. PROF. DAVID RUFFOLO, Ph.D. 101 pp. ISBN 974-333-158-1

Presently, an enormous number of protein sequences has been generated by advanced techniques in molecular biology and genetic engineering. However, protein homology searching by comparing amino acid sequences is complicated. The standard and most popular method is linear alignment. This method has many limitations, for example, gaps present in the protein sequence and low sequence identity could interfere with the protein alignment.

A two-dimensional representation of proteins has been successfully used in classification of secondary structure as well as similarity of proteins. This research uses this method to generate 2-dimensional patterns and incorporate data on the chemical/biochemical characteristics of each amino acid are incorporated. Computer neural networks were used to identify the similarity of the amino acid patterns. When proteins with a sequence identity of 35%-97% were used to generate the 2-dimensional patterns, similar patterns of amino acids were found. Neural networks were built to identify the similarity category of a 18-amino acid window from 5 groups of homologous proteins. The neural networks we built can identify the similarity category to an accuracy of approximately 86-92%.

ภาควิชา คณิตศาสตร์ ลายมือชื่อนิสิต .....

สาขาวิชา วิทยาการคณนา ลายมือชื่ออาจารย์ที่ปรึกษา .....

ปีการศึกษา 2540 ลายมือชื่ออาจารย์ที่ปรึกษาร่วม .....

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปได้ด้วยความช่วยเหลืออย่างดียิ่งของอ.ดร.รัฐ พิชญางกูร อาจารย์ที่ปรึกษาวิทยานิพนธ์และ รองศาสตราจารย์ ดร. เดวิด รูฟโฟโล อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม ซึ่งท่านได้ให้คำแนะนำและข้อคิดเห็นต่างๆ ในการทำวิจัยมาด้วยดีตลอด

เนื่องจากงานวิจัยนี้ ได้รับทุน Graduate Consortium Scholarship Computational Science & Engineering Program (GREC01-41-010) จากสำนักงานวิทยาศาสตร์และเทคโนโลยีแห่งชาติ จึงขอกราบขอบพระคุณสำนักงานวิทยาศาสตร์และเทคโนโลยีแห่งชาติมา ณ ที่นี้ด้วย

ขอขอบพระคุณที่ท่านคณะกรรมการสอบวิทยานิพนธ์ ศาสตราจารย์.ดร.ชิดชนก เหลือสินทรัพย์ และ อ.ดร.เลอสรร ธนสุกาญจน์ ที่กรุณาให้คำแนะนำ และแก้ไขวิทยานิพนธ์ฉบับนี้จนเสร็จสมบูรณ์

และขอขอบคุณนายวรวัฒน์ วรศิลป์ และนายจุมพล ตั้นประยูร ผู้ให้คำแนะนำในการเขียนโปรแกรม Visual C++ และ การใช้ โปรแกรม SNNS และพี่อนๆที่ให้กำลังใจ

ท้ายนี้ ผู้วิจัยใคร่ขอขอบพระคุณ บิดา-มารดา ซึ่งสนับสนุนในด้านการเงิน และให้กำลังใจแก่ผู้วิจัย เสมอมาจนสำเร็จการศึกษา

สารบัญ

หน้าที่

สารบัญตาราง

สารบัญภาพ

หน้าที่

หน้าที่

หน้าที่