

การหาโปรตีนที่ทำหน้าที่เหมือนกันโดยใช้การจัดเรียงกลุ่มที่ไม่ชอบน้ำแบบ 2 มิติโดยอัตโนมัติ

นายภิสิต์ วรรณสุต

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

สาขาวิชาวิทยาศาสตร์คอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2551

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

DETECTING PROTEIN HOMOLOGY USING AUTOMATIC 2-D HYDROPHOBIC  
CLUSTER ALIGNMENT

Mr. Phisit Kannasut

A Thesis Submitted in Partial Fulfillment of the Requirements  
for the Degree of Master of Science Program in Computer Science

Department of Computer Engineering

Faculty of Engineering

Chulalongkorn University

Academic Year 2008

Copyright of Chulalongkorn University

511554

หัวข้อวิทยานิพนธ์

การทำโปรตีนที่ทำหน้าที่เหมือนกันโดยใช้การจัดเรียง  
กลุ่มที่ไม่ชอบน้ำแบบ 2 มิติโดยอัตโนมัติ

โดย

นายภิสสิทธิ์ กรรณสูต

สาขาวิชา

วิทยาศาสตร์คอมพิวเตอร์

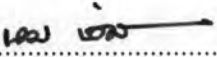
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก

ดร. โชติรัตน์ รัตนามัทธนะ

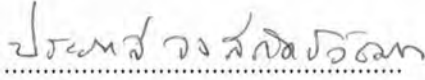
อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม

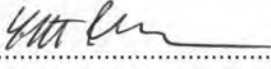
ผู้ช่วยศาสตราจารย์ ดร. รัฐ พิชญางกูร


คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้รับวิทยานิพนธ์ฉบับนี้  
เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรบัณฑิต

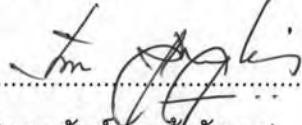
  
..... คณบดีคณะวิศวกรรมศาสตร์  
(รองศาสตราจารย์ ดร.บุญสม เลิศหิรัญวงศ์)

คณะกรรมการสอบวิทยานิพนธ์

  
..... ประธานกรรมการ  
(ศาสตราจารย์ ดร.ประภาส จงสถิตย์วัฒนา)

  
..... อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก  
(ดร.โชติรัตน์ รัตนามัทธนะ)

  
..... อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม  
(ผู้ช่วยศาสตราจารย์ ดร.รัฐ พิชญางกูร)

  
..... กรรมการภายนอกมหาวิทยาลัย  
(ดร.อนันต์ จงแก้ววัฒนา)

  
..... กรรมการ  
(ผู้ช่วยศาสตราจารย์ ดร.โปรตปราน บุญยพุกกณะ)



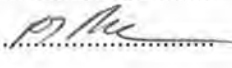
ภิลิทธิ กรรมรสุต : การหาโปรตีนที่ำหน้าทีเหมือนกันโดยใช้การจ้ดเรียงกลุ่มทีไม่ชอบน้ำแบบ 2 มิติโดยอ้ตโนม้ติ. (DETECTING PROTEIN HOMOLOGY USING AUTOMATIC 2-D HYDROPHOBIC CLUSTER ALIGNMENT) อ.ที่ปริกษาวิทยานิพนธ์หลัก : ดร. โชติร้ตน์ ร้ตนามห้ทชนะ, อ.ที่ปริกษาวิทยานิพนธ์ร่วม : ผศ.ดร. ร้ฐ พิษญางกูร, 110 หน้า.

ในชวงระยะเวลาทีผ่านมา ข้อมูลทางชีววิทยาได้มีปริมาณเพิ่มขึ้นอย่างมากมาย ไม่ว่าจะเป็ข้อมูลลำดับนิวคลีโอไทด์ หรือข้อมูลลำดับกรดอะมิโน นักวิทยาศาสตร์สามารถใช้ข้อมูลเหล่านี้เพื่อการศึกษาหน้าทีการทำงานของโปรตีนนั้นๆ จากทีผ่านมาการศึกษาโปรตีนยังคงใช้เครื่องมือจากห้องปฏิบัติการทางการเคมีหรือห้องปฏิบัติการทางชีววิทยา ซึ่งต้องใช้เวลาและงบประมาณสูงมาก ต่อมานักวิจัยได้เริ่มใช้วิธีระบุน้ำทีจากลำดับกรดอะมิโนแทน แต่ทว่ายังไม่ีเครื่องมือทีมีความถูกต้องสูงและรวดเร็ว และยังคงเป็งานวิจัยทีทำทายอย่างมาก เครื่องมือทีใช้รูปแบบการจับคู่ลำดับกรดอะมิโนแบบ 1 มิติมีข้อจ้ก้ตอยู่มากและไม่สามารรถเปรียบเทียบลำดับกรดอะมิโนจากคู่ทีมีค่าความเหมือนต่ำได้

การวิเคราะห์กลุ่มทีไม่ชอบน้ำเป็การวิเคราะห์โครงสร้างระดับทุติยภูมิของลำดับกรดอะมิโน ได้พัฒนารวมเข้ากับเครื่องมือทีใช้เปรียบเทียบลำดับ วิธีการนี้ช่วยเพิ่มประสิทธิภาพในการเปรียบเทียบได้มาก แต่ในปัจจุบันนี้เครื่องมือนี้ยังใช้การจับคู่กลุ่มโดยใช้ความรู้และความชำนาญของผู้เชี่ยวชาญในการเปรียบเทียบ สำหรับงานวิจัยขึ้นนี้ได้นำวิธีทีใช้เปรียบเทียบความเหมือนด้วยวิธีจับคู่กลุ่มทีไม่ชอบน้ำมาพัฒนาต่อให้เป็ระบบอ้ตโนม้ติ และทดสอบการระบุน้ำทีของโปรตีนกับชุดข้อมูลทีรวบรวมจากฐานข้อมูลต่างๆ เช่น PIR SISYPHUS และ HOMSTRAD เพื่อแสดงถึง ความถูกต้อง และประสิทธิภาพของเครื่องมือใหม่นี้

งานวิจัยนี้ได้เสนอเครื่องมือแสดงแผนภูมิ 2 มิติของการวิเคราะห์กลุ่มทีไม่ชอบน้ำ และเครื่องมือจับคู่กลุ่มทีไม่ชอบน้ำแบบ 2 มิติโดยอ้ตโนม้ติทีมีประสิทธิภาพ โดยการทำการทดสอบกับชุดข้อมูลจากฐานข้อมูล PIR ฐานข้อมูล SISYPHUS และฐานข้อมูล HOMSTRAD และแสดงให้เห็นว่าเครื่องมือทีใช้จับคู่กลุ่มทีไม่ชอบน้ำแบบ 2 มิติโดยอ้ตโนม้ติทีมีประสิทธิภาพในการเลือกลำดับกรดอะมิโนจากฐานข้อมูลได้ดีกว่าเครื่องมือเปรียบเทียบแบบ 1 มิติเดิม

ภาควิชา.....วิศวกกรรมคอมพิวเตอร์.....  
สาขาวิชา.....วิทยวศาสตร้คอมพิวเตอร์.....  
ปีการศึกษา.....2551.....

ลายมือชื่อนิสิติ .....  .....  
ลายมือชือ อ.ที่ปริกษาวิทยานิพนธ์หลัก .....  .....  
ลายมือชือ อ.ที่ปริกษาวิทยานิพนธ์ร่วม .....  .....

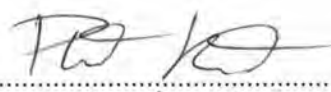


## 4970501521 : MAJOR COMPUTER SCIENCE

KEY WORD: BIOINFORMATICS / HYDROPHOBIC CLUSTER ANALYSIS / PROTEIN / ALIGNMENT TOOL

PHISIT KANNASUT : DETECTING PROTEIN HOMOLOGY USING AUTOMATIC 2-D HYDROPHOBIC CLUSTER ALIGNMENT, THESIS PRINCIPAL ADVISOR : CHOTIRAT RATANAMAHAHATANA, Ph.D., THESIS CO-ADVISOR : RATH PICHYANGKURA, 110 pp.

Biological databases in the past decade have tremendously grown in size. However, effective retrieval of these data is still a great challenge. In particular, we need a high-quality tool to measure similarity among protein sequences within the database. Current techniques in protein homology testing involve a 1-dimensional alignment of Nucleotide or Amino acid sequencing. Due to its various constraints and low sequence identity values, Hydrophobic Cluster Alignment has increasingly been used to predict the structure and functionality of protein. However, this Hydrophobic Cluster Alignment still needs to be done manually and solely depends on experience and expertise of a researcher.

This work introduces a 2-D visualization tool for amino acid, 2-D Automatic hydrophobic cluster alignment tool, and proposes a new protein representation that could be used effectively in the 2-D automatic Hydrophobic Cluster Alignment software. The 2-D alignment was performed on the amino acid sequences from the PIR database, the SISYPHUS database, and HOMSTRAD database. The results have demonstrated that the 2-D automatic hydrophobic cluster alignment is more accurate than the existing 1-D sequence alignment is.

Department : ....Computer Engineering....	Student's signature : 
Field of study : ....Computer Science.....	Principal advisor's signature : 
Academic year : ....2008.....	Co-advisor's signature : 

## กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปได้ด้วยดี เนื่องด้วยความช่วยเหลือ และ คำแนะนำจากอาจารย์ที่ปรึกษา อาจารย์ ดร. โชติรัตน์ รัตนามัทธนะ พร้อมด้วยอาจารย์ที่ปรึกษา ร่วม ผู้ช่วยศาสตราจารย์ ดร. รัฐ พิชญางกูร ที่กรุณาให้คำปรึกษาและตลอดจนให้กำลังใจในการทำวิทยานิพนธ์ฉบับนี้อย่างดียิ่ง ผู้วิจัยขอกราบขอบพระคุณไว้เป็นอย่างสูง ณ โอกาสนี้ กราบ ขอบคุณศาสตราจารย์ ดร.ประภาส จงสถิตยวัฒนา ประธานการสอบวิทยานิพนธ์ ผู้ช่วย ศาสตราจารย์ ดร.โปรดปราน บุญยทุกณะ และ ดร.อนันต์ จงแก้ววัฒนา ที่กรุณาให้คำแนะนำ ข้อคิดเห็น และรวมถึงแนวทางการทำวิทยานิพนธ์ฉบับนี้

ขอกราบขอบคุณคณาจารย์ภาควิชาคอมพิวเตอร์ทุกท่าน ที่ให้วิชาความรู้แก่ ผู้วิจัยอย่างไม่รู้จักเหน็ดเหนื่อย อีกทั้งคอยเป็นกำลังใจให้เสมอมา

ขอขอบคุณ คุณสมฤดี ชุณหโรจน์ฤทธิ์ ที่คอยให้กำลังใจเสมอมา และขอบคุณ เพื่อนๆ พี่ๆ น้องๆ ในห้องปฏิบัติการชั้น 18 ทุกคน ที่คอยให้เสียงหัวเราะ เป็นกำลังใจ และรวมถึง ให้คำแนะนำในการทำวิทยานิพนธ์ฉบับนี้ ให้สำเร็จลุล่วงไปได้ด้วยดี

สุดท้ายนี้ต้องขอขอบพระคุณคุณพ่อและคุณแม่เป็นอย่างสูงที่คอยสนับสนุน การศึกษา คอยให้ความรัก และคอยเป็นกำลังใจ ทำให้งานวิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปได้ ด้วยดี

## สารบัญ

หน้า

บทคัดย่อภาษาไทย .....	ง
บทคัดย่อภาษาอังกฤษ .....	จ
กิตติกรรมประกาศ.....	ฉ
สารบัญ .....	ช
สารบัญภาพ.....	ญ
สารบัญตาราง.....	ฎ
บทที่ 1 บทนำ.....	1
1.1 ที่มาและความสำคัญของปัญหา.....	1
1.2 วัตถุประสงค์ของการวิจัย .....	2
1.3 ขอบเขตของการวิจัย .....	3
1.4 ขั้นตอนของการวิจัย .....	3
1.5 ประโยชน์ที่ได้รับ.....	3
1.6 โครงสร้างของวิทยานิพนธ์.....	4
1.7 ผลงานตีพิมพ์จากวิทยานิพนธ์.....	4
บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง .....	5
2.1 โปรตีน (Protein) .....	5
2.1.1 โปรตีนกลุ่มโกลบิน (Globin Family).....	5
2.1.2 โปรตีนที่รวมตัวกับเยื่อหุ้มเซลล์ (Integral Membrane Proteins) .....	6
2.1.3 โปรตีนเส้นใย (Fibrous Proteins).....	7
2.2 วิวัฒนาการของสิ่งมีชีวิต .....	8
2.3 กรดอะมิโน (Amino Acid).....	8
2.4 โครงสร้างของโปรตีน (Protein Structure).....	12
2.4.1 โครงสร้างระดับปฐมภูมิ (Primary Structure).....	12
2.4.2 โครงสร้างระดับทุติยภูมิ (Secondary Structure) .....	12
2.4.3 โครงสร้างระดับตติยภูมิ (Tertiary Structure) .....	13
2.4.4 โครงสร้างระดับจตุรภูมิ (Quaternary Structure).....	13
2.5 การศึกษาโปรตีน .....	14

2.6	การศึกษาโปรตีนที่มีหน้าที่เหมือนกัน .....	15
2.7	เครื่องมือจับคู่ (Alignment Tools) .....	15
2.7.1	วิธีกำหนดการพลวัต (Dynamic Programming Methods) .....	15
2.7.2	วิธีการประมาณการจับคู่อักขระ (Approximate string matching) .....	17
2.7.3	แบบจำลองฮิดเดนมาร์คอฟ (Hidden Markov Model) .....	18
2.7.4	การวิเคราะห์กลุ่มที่ไม่ชอบน้ำ (Hydrophobic Cluster Analysis) .....	18
2.8	Local Alignment และ Global Alignment.....	19
2.9	ตารางให้คะแนน BLOSUM (BLOSUM Substitution Matrix) .....	20
2.10	ค่าหักลบช่องว่าง (Gap Penalty).....	22
2.10.1	ค่าหักลบช่องว่างแบบคงที่ (Constant Gap Penalty) .....	22
2.10.2	ค่าหักลบช่องว่างแบบเส้นตรง (Linear Gap Penalty).....	22
2.10.3	Affine Gap Penalty.....	22
บทที่ 3	วิธีดำเนินงานวิจัย .....	24
3.1	การสร้างเครื่องมือแสดงผลแผนภูมิ 2 มิติ .....	24
3.2	การแทนข้อมูล.....	27
3.3	การสร้างเครื่องมือจับคู่กลุ่มที่ไม่ชอบน้ำแบบ 2 มิติ .....	28
3.3.1	ขั้นตอนวิธีการทำงานของเครื่องมือจับคู่กลุ่มที่ไม่ชอบน้ำ.....	28
3.3.2	การคิดคะแนนการจับคู่กลุ่ม.....	33
3.3.3	ค่าหักลบช่องว่าง.....	34
3.3.4	การคิดค่าความเหมือนของลำดับที่ไม่ชอบน้ำ.....	35
บทที่ 4	การทดลองและผลการทดลอง .....	36
4.1	วิธีการเตรียมข้อมูลสำหรับการทดสอบ.....	36
4.1.1	การเตรียมข้อมูลสำหรับทดลองเพื่อหาค่าหักลบช่องว่าง.....	36
4.1.2	ชุดข้อมูลเพื่อทดสอบความถูกต้องของเครื่องมือ .....	37
4.1.3	ชุดข้อมูลเพื่อทดลองเพิ่มขนาดข้อมูลสอน .....	37
4.2	วิธีการทดลอง.....	37
4.2.1	การทดลองเพื่อหาค่าหักลบช่องว่าง .....	38



4.2.2	การทดลองเพื่อเปรียบเทียบความถูกต้องของเครื่องมือจับคู่กลุ่มที่ไม่ชอบน้ำแบบ 2 มิติโดยอัตโนมัติ.....	38
4.2.3	การทดลองเพิ่มขนาดชุดข้อมูลสอนให้มีขนาดใหญ่ขึ้น.....	39
4.2.4	การทดลองเพื่อเปรียบเทียบความถูกต้องของการเลือกลำดับที่เหมือนที่สุด.....	40
4.3	ผลการทดลอง .....	40
4.3.1	ผลการทดลองหาค่าหักลบช่องว่างที่เหมาะสม.....	41
4.3.2	ผลการทดลองเปรียบเทียบความถูกต้องของเครื่องมือจับคู่กลุ่มที่ไม่ชอบน้ำแบบ 2 มิติโดยอัตโนมัติ.....	43
4.3.3	ผลการทดลองเพิ่มขนาดชุดข้อมูลสอนให้มีขนาดใหญ่ขึ้น .....	49
4.3.4	ผลการทดลองเปรียบเทียบความถูกต้องของการเลือกลำดับที่เหมือนที่สุด .....	50
บทที่ 5	สรุปผลการวิจัยและข้อเสนอแนะ.....	52
5.1	สรุปผลการวิจัย .....	52
5.2	ข้อเสนอแนะ.....	53
	รายการอ้างอิง.....	55
	ภาคผนวก.....	60
	ภาคผนวก ก ข้อมูลที่ใช้ในการทดลอง .....	61
	ภาคผนวก ข ผลงานตีพิมพ์.....	79
	ประวัติผู้เขียนวิทยานิพนธ์.....	110

## สารบัญญภาพ

หน้า

รูปที่ 2.1	โครงสร้างและตำแหน่งของโปรตีนที่อยู่ภายในเยื่อหุ้มเซลล์.....	6
รูปที่ 2.2	โครงสร้างและตำแหน่งของโปรตีนที่เกาะติดกับเยื่อหุ้มเซลล์ .....	7
รูปที่ 2.3	แบบจำลองโครงสร้างของคอลลาเจน.....	8
รูปที่ 2.4	โครงสร้างของกรดอะมิโน หมู่ทางขวามือคือ หมู่คาร์บอกซิล (Carboxyl Group, -COOH) หมู่ทางซ้ายมือคือ หมู่อะมิโน (Amino Group, -NH <sub>2</sub> ) และ R คือ หมู่อัลคิล (Alkyl Group) เป็นหมู่ข้างเคียงของกรดอะมิโน.....	9
รูปที่ 2.5	โครงสร้างของกรดอะมิโนทั้ง 20 ชนิด .....	10
รูปที่ 2.6	โครงสร้างระดับตติยภูมิ แสดงให้เห็นถึงลักษณะที่ม้วนเป็นเกลียวอยู่ด้านนอก เรียกว่าแบบเกลียวแอลฟา (Alpha-Helix) และลักษณะแผ่นที่พาดผ่านด้านใน เรียกว่าแบบบีตาชีท (Beta-Sheet).....	12
รูปที่ 2.7	ขั้นตอนการศึกษาโครงสร้างและหน้าที่การทำงานของโปรตีน.....	14
รูปที่ 2.8	ผลของการจับคู่ลำดับกรดอะมิโนด้วยวิธีกำหนดการพลวัตแบบ 1 มิติ .....	16
รูปที่ 2.9	Global Alignment vs. Local Alignment .....	20
รูปที่ 2.10	ตารางคะแนน BLOSUM62 .....	21
รูปที่ 2.11	Affine Gap Penalty .....	23
รูปที่ 3.1	รูปแสดงการจัดเรียงกรดอะมิโนโดยใช้ NT เท่ากับ 5 และ NR เท่ากับ 18 ตัวเลขที่แสดงอยู่ภายในรูปแสดงถึงตำแหน่งของ ID ที่กรดอะมิโนลำดับที่ $i$ .....	26
รูปที่ 3.2	แผนภูมิ 2 มิติของกรดอะมิโนที่สร้างขึ้นจากเครื่องมือที่ใช้สำหรับแสดงผล .....	26
รูปที่ 3.3	การแบ่งกลุ่มลำดับกรดอะมิโนตามกฎเกณฑ์ข้างต้น.....	27
รูปที่ 3.4	การจับคู่กลุ่มที่ไม่ชอบน้ำโดยอาศัยวิธีกำหนดการพลวัต .....	29
รูปที่ 3.5	ขั้นตอนวิธีการทำงานของฟังก์ชันหลักในเครื่องมือที่ใช้จับคู่กลุ่มที่ไม่ชอบน้ำ .....	29
รูปที่ 3.6	ขั้นตอนวิธีการทำงานของฟังก์ชัน HCAAlign .....	30
รูปที่ 3.7	ขั้นตอนวิธีการทำงานของฟังก์ชัน Alignment.....	31
รูปที่ 3.8	ตัวอย่างการเปรียบเทียบจากกลุ่มที่ไม่มีกรดอะมิโนที่ถูกใช้ .....	32
รูปที่ 3.9	ตัวอย่างการเปรียบเทียบจากกลุ่มที่มีกรดอะมิโนที่ถูกใช้ .....	32
รูปที่ 3.10	การเปรียบเทียบกลุ่มโดยใช้การแทนข้อมูล .....	33
รูปที่ 3.11	ตารางคะแนน BLOSUM62 ที่ปรับปรุงจากเดิม .....	33

รูปที่ 3.12	ค่าหักลบช่องว่างแบบใหม่ที่น่าเสนอโดยมีค่าเริ่มต้นที่ 5 .....	35
รูปที่ 4.1	กราฟแสดงความถูกต้องเมื่อเปลี่ยนแปลงค่าหักลบเมื่อเปิดช่องว่าง และค่าหักลบ เมื่อขยายช่องว่างของ Affine Gap Penalty .....	41
รูปที่ 4.2	กราฟแสดงความถูกต้องเมื่อเปลี่ยนแปลงค่าเริ่มต้นของค่าหักลบช่องว่างใหม่ .....	43
รูปที่ 4.3	การเปรียบเทียบค่าความเหมือนของลำดับที่ไม่ชอบน้ำระหว่างเครื่องมือจับคู่โดย อัตโนมัติกับการจับคู่โดยผู้เชี่ยวชาญโดยใช้ตาราง BLOSUM62 และ Affine Gap Penalty .....	44
รูปที่ 4.4	ความถี่ของความแตกต่างที่เกิดขึ้นจากการใช้ตาราง BLOSUM62 และ Affine Gap Penalty .....	45
รูปที่ 4.5	การเปรียบเทียบค่าความเหมือนของลำดับที่ไม่ชอบน้ำระหว่างเครื่องมือจับคู่โดย อัตโนมัติกับการจับคู่โดยผู้เชี่ยวชาญโดยใช้ตาราง BLOSUM62 และค่าหักลบ ช่องว่างใหม่ .....	46
รูปที่ 4.6	การจับคู่กลุ่มที่ไม่ชอบน้ำของโปรตีน Rubrerythrin จากผู้เชี่ยวชาญ และจาก เครื่องมือจับคู่โดยอัตโนมัติ .....	47
รูปที่ 4.7	การจับคู่กลุ่มที่ไม่ชอบน้ำของโปรตีน Interleukin จากผู้เชี่ยวชาญ และจาก เครื่องมือจับคู่โดยอัตโนมัติ .....	48
รูปที่ 4.8	ความถูกต้องเมื่อเพิ่มขนาดชุดข้อมูลสอน .....	50
รูปที่ ก.1	การจับคู่โครงสร้างโดยผู้เชี่ยวชาญใน SISYPHUS แสดงแบบแผนภูมิ 1 มิติ .....	67
รูปที่ ก.2	การจับคู่โครงสร้างโดยผู้เชี่ยวชาญใน SISYPHUS แสดงแบบแผนภูมิ 2 มิติ .....	68
รูปที่ ก.3	การจับคู่โครงสร้างใน HOMSTRAD แสดงแบบแผนภูมิ 1 มิติ .....	76
รูปที่ ก.4	การจับคู่โครงสร้างใน HOMSTRAD แสดงแบบแผนภูมิ 2 มิติ .....	77

## สารบัญตาราง

หน้า

ตารางที่ 2.1	การแบ่งกลุ่มกรดอะมิโนแต่ละชนิดตามคุณสมบัติความชอบน้ำ.....	11
ตารางที่ 2.2	ตัวอย่างชิ้นส่วนกรดอะมิโนขนาดเล็กที่นำมาเปรียบเทียบภายใน BLAST .....	17
ตารางที่ 2.3	คุณสมบัติทางเคมีของกรดอะมิโน.....	21
ตารางที่ 3.1	ตารางแสดงการแบ่งกลุ่มของอะมิโน และสีที่ถูกกำหนดไว้ในแต่ละกลุ่ม .....	25
ตารางที่ 3.2	ตัวอย่างการแทนข้อมูลจากกลุ่มที่ไม่ชอบน้ำ.....	28
ตารางที่ 4.1	ชุดข้อมูลสำหรับการทดลองเพื่อหาค่าหักลบช่องว่างที่เหมาะสม .....	36
ตารางที่ 4.2	ชุดข้อมูลสำหรับทดลองเพิ่มขนาดข้อมูลสอน.....	37
ตารางที่ 4.3	การเปลี่ยนค่าหักลบช่องว่างในการทดลอง .....	38
ตารางที่ 4.4	รายละเอียดการทดลองความถูกต้องของการเลือกลำดับที่เหมือนที่สุด.....	40
ตารางที่ 4.5	ผลการทดลองเปรียบเทียบความถูกต้องของการเลือกลำดับกรดอะมิโนที่ เหมือนที่สุดจากข้อมูลเปรียบเทียบ เพื่อหา Affine Gap Penalty ที่เหมาะสม .....	41
ตารางที่ 4.6	ผลการทดลองเปรียบเทียบความถูกต้องของการเลือกลำดับกรดอะมิโนที่ เหมือนที่สุดจากข้อมูลเปรียบเทียบ หาค่าหักลบช่องว่างใหม่ที่เหมาะสม.....	42
ตารางที่ 4.7	ผลการทดลองเปรียบเทียบค่าความเหมือนของลำดับที่ไม่ชอบน้ำระหว่าง เครื่องมือจับคู่กลุ่มที่ไม่ชอบน้ำแบบ 2 มิติโดยอัตโนมัติ กับการจับคู่กลุ่มโดย ผู้เชี่ยวชาญ .....	43
ตารางที่ 4.8	ผลการทดลองเพิ่มขนาดชุดข้อมูลสอนให้มีขนาดใหญ่ขึ้นตามลำดับ.....	49
ตารางที่ 4.9	ผลการทดลองเปรียบเทียบความถูกต้องของการเลือกลำดับที่เหมือนที่สุด ระหว่างวิธีการจับคู่แบบ 1 มิติ กับการจับคู่แบบ 2 มิติโดยอัตโนมัติ .....	50
ตารางที่ ก.1	รายละเอียดของชุดข้อมูลจาก SISYPHUS .....	61
ตารางที่ ก.2	ตัวอย่างการแบ่งกลุ่มโปรตีนในฐานข้อมูล HOMSTRAD .....	70