บทที่ 3

วิธีดำเนินงานวิจัย

จากงานวิจัยที่กล่าวมา แสดงให้เห็นว่าวิธีการจับคู่โปรตีนที่ทำหน้าที่เหมือนกันยังมีเครื่องมือที่มีประสิทธิภาพไม่เพียงพอ ส่วนใหญ่ใช้วิธีการจับคู่ลำดับกรดอะมิโนในรูปแบบ 1 มิติ แต่การจับคู่กรดอะมิโนด้วยกลุ่มที่ไม่ชอบน้ำในรูปแบบ 2 มิติซึ่งเป็นวิธีการจับคู่กลุ่มที่ไม่ชอบน้ำยังไม่ได้รับความนิยมจากกลุ่มผู้วิจัยมากนัก เนื่องจากวิธีการจับคู่ยังคงต้องใช้ความรู้และความชำนาญจากผู้วิจัยเองในการจับคู่ แต่อย่างไรก็ตามยังคงมีกลุ่มผู้วิจัยที่ใช้วิธีการจับคู่กลุ่มที่ไม่ชอบน้ำแบบ 2 มิติอย่างต่อเนื่อง ดังนั้นเพื่อให้เป็นประโยชน์ต่องานวิจัยในอนาคต จึงควรมีเครื่องมือจับคู่กลุ่มที่ไม่ชอบน้ำอย่างอัตโนมัติแบบ 2 มิติ เพื่อใช้เป็นเครื่องมือนำไปสู่งานวิจัยที่มีคุณภาพ

จากแนวคิดข้างต้นการพัฒนาเครื่องมือสามารถแบ่งออกได้เป็น 4 ส่วน เริ่มจากการรับข้อมูลนำเข้าเป็นลำดับกรดอะมิโนและเพื่อให้สามารถศึกษากลุ่มที่ไม่ชอบน้ำได้อย่างละเอียด ในส่วนแรกจึงต้องสร้างเครื่องมือแสดงผลแผนภูมิ 2 มิติ ในส่วนที่ 2 จากลำดับข้อมูลนำเข้าเป็นลำดับกรดอะมิโนต้องแปลงข้อมูลให้อยู่ในรูปแบบที่สามารถใช้จับคู่กลุ่มได้ ต่อมาในส่วนที่ 3 การสร้างเครื่องมือจับคู่กลุ่มที่ไม่ชอบน้ำแบบ 2 มิติโดยอัตโนมัติ และในส่วนสุดท้ายเป็นการทดสอบความถูกต้องของเครื่องมือจับคู่กลุ่มที่เสนอขึ้นใหม่

3.1 การสร้างเครื่องมือแสดงผลแผนภูมิ 2 มิติ

ในงานวิจัยนี้ได้สร้างเครื่องมือที่ใช้ในการแสดงผลตามหลักของการวิเคราะห์กลุ่มที่ไม่ชอบน้ำ [13] โดยพัฒนาเครื่องมือการแสดงผลด้วยภาษา C#.NET เครื่องมือนี้สามารถแสดงผลกรดอะมิโนในรูปแบบแผนภูมิ 2 มิติ นอกจากนั้นเครื่องมือยังสามารถลากเส้นคั่นระหว่างกลุ่มเพื่อใช้สำหรับการแสดงผลการจับคู่กลุ่มที่ไม่ชอบน้ำ กำหนดสีของกรดอะมิโนตามการกำหนดสีมาตรฐานของ Lemesle-Varloot [32] ดังแสดงในตารางที่ 3.1

ตารางที่ 3.1 ตารางแสดงการแบ่งกลุ่มของอะมิโน และสีที่ถูกกำหนดไว้ในแต่ละกลุ่ม

กลุ่มของกรดอะมิโน	สัญลักษณ์และสีที่ได้กำหนดไว้
กลุ่มไม่ชอบน้ำ	ระดับแรง (Strong) คือ V I L F สีเขียว ระดับปานกลาง (Medium) คือ W M Y ใช้สีเขียว ระดับเบา (Mimetic) คือ A C ใช้สีดำ
กลุ่มชอบน้ำ	D E N Q ใช้สีแดง H K R ใช้สีน้ำเงิน
กลุ่มพิเศษ	P ใช้ ★ G ใช้ ♦ T ใช้ □ S ใช้ ⊡

สำหรับกรดอะมิโนในกลุ่มพิเศษ (P,G,T,S) ยังคงใช้เป็นตัวอักษรเช่นเดิม นอกจากนั้นเครื่องมือแสดงผลแผนภูมิ 2 มิติมีการคำนวณตำแหน่งในแกน y ของลำดับกรดอะมิโนแต่ละตัว (*ID*) ซึ่งสามารถคำนวณได้จากสมการของ Lemesle-Varloot [41] สมการที่ 3.1

$$ID(i) = (i \times NT) \bmod NR \tag{3.1}$$

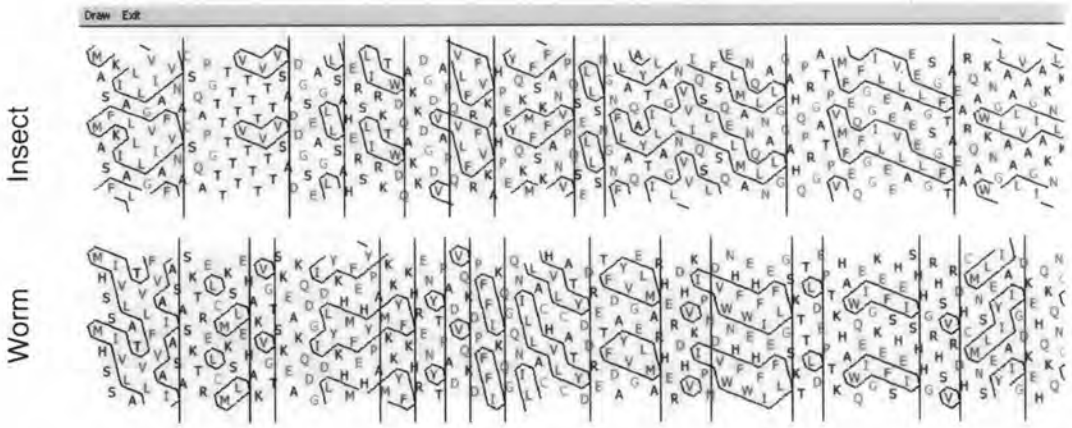
การวางตำแหน่งแกน y ของกรดอะมิโนจะสอดคล้องกันการหมุนเป็นเกลียวแบบแอลฟา ซึ่งในหนึ่งรอบเกลียวจะใช้กรดอะมิโนจำนวน 3.6 ตัว ดังนั้นเมื่อมีการหมุนเกียวจนครบ 5 รอบ (*NT*) จะทำให้ต้องใช้กรดอะมิโนทั้งหมดจำนวน 18 ตัว (*NR*) สามารถคำนวณและวางตำแหน่งของกรดอะมิโนได้ดังแสดงในรูปที่ 3.1

จากนั้นสร้างเส้นล้อมรอบกลุ่มที่ไม่ชอบน้ำเพื่อแสดงเป็นขอบเขตของกลุ่มที่อยู่ติดกัน วิธีการแบ่งกลุ่มสามารถศึกษาได้จาก การแทนข้อมูล ในหัวข้อถัดไป นอกจากนั้นเครื่องมือแสดงผลยังสามารถลากเส้นตรงคั่นระหว่างกลุ่ม เพื่อแบ่งแต่ละกลุ่มออกจากกัน เครื่องมือแสดงผลแผนภูมิ 2 มิติสามารถแสดงผลได้ดังตัวอย่างในรูปที่ 3.2

ลำดับกรดอะมิโนที่แสดงในเครื่องมือเป็นลำดับอะมิโนของฮีโมโกลบิน สายบนคือลำดับกรดอะมิโนของ *Daphnia magna* ซึ่งเป็นสิ่งมีชีวิตจำพวกของแมลง และสายล่างคือลำดับกรดอะมิโนของ *Pseudoterranova decipiens* (ลำดับกรดอะมิโนแสดงในภาคผนวก ก) ซึ่งเป็นสิ่งมีชีวิตจำพวกหนอน นอกจากนี้เครื่องมือวาดแผนภูมิจะสร้างเส้นล้อมรอบกรดอะมิโนที่ไม่ชอบน้ำ (V, I, L, F, W, M, Y) ให้เห็นชัดเจน

i	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
ID																															
0	0																		0												
1												1																		1	
2					2																		2								
3																3															
4									4																		4				
5		5																		5											
6												6																		6	
7						7																	7								
8																	8														
9										9																		9			
10			10																		10										
11														11																	
12							12																		12						
13																			13											
14											14																		14		
15				15																		15									
16																	16														
17									17																			17			

รูปที่ 3.1 รูปแสดงการจัดเรียงกรดอะมิโนโดยใช้ *NT* เท่ากับ 5 และ *NR* เท่ากับ 18 ตัวเลขที่แสดงอยู่ภายในรูปแสดงถึงตำแหน่งของ ID ที่กรดอะมิโนลำดับที่ *i*



รูปที่ 3.2 แผนภูมิ 2 มิติของกรดอะมิโนที่สร้างขึ้นจากเครื่องมือที่ใช้สำหรับแสดงผล

3.2 การแทนข้อมูล

วิธีการวิเคราะห์กลุ่มที่ไม่ชอบน้ำเป็นการวิเคราะห์กลุ่มของกรดอะมิโนที่เกิดจากกรดอะมิโนที่ไม่ชอบน้ำเพียงอย่างเดียว จึงทำให้เหลือกรดอะมิโนที่สนใจเพียง 7 ชนิดรวมกับกรดอะมิโนที่ทำให้เกิดการแบ่งกลุ่มก็คือ โพรลีน (P) ซึ่งตรงกับผลการวิจัยของ G. Von Heijine [42] ที่ศึกษาค่า Hydrophobicity สัมพัทธ์โดยเทียบกับไกลซีนได้มาจากการคำนวณค่าจากตัวทำละลายที่ต่างกันจากผิวสัมผัสของกรดอะมิโนแต่ละตัว และใช้สำหรับการแบ่งกลุ่มกรดอะมิโนตามลักษณะความชอบน้ำ

การแทนข้อมูลนี้มีวัตถุประสงค์ เพื่อเปลี่ยนข้อมูลแผนภูมิให้อยู่ในรูปแบบที่ง่ายต่อการเปรียบเทียบกลุ่ม ควรมีคุณสมบัติดังต่อไปนี้

1. สามารถเก็บและแสดงรูปแบบของกลุ่มที่ไม่ชอบน้ำได้
2. สามารถตัดข้อมูลที่ไม่จำเป็นออกเช่น ส่วนที่ใช้เชื่อมต่อระหว่างกลุ่มที่ไม่ชอบน้ำ
3. สามารถนำไปใช้สำหรับการจับคู่กลุ่มที่ไม่ชอบน้ำแบบอัตโนมัติ

ตามคุณสมบัติที่กล่าวมาข้างต้นสามารถเปลี่ยนลำดับกรดอะมิโนให้เป็นรูปแบบการแทนข้อมูลได้โดยมีวิธีการทั้งหมด 2 ขั้นตอนตามลำดับดังนี้

1. สำหรับกรดอะมิโนที่ไม่ชอบน้ำทั้ง 7 ชนิดให้คงสัญลักษณ์กรดอะมิโน และเปลี่ยนกรดอะมิโนที่เหลือทั้งหมดเป็นเลข 0 ตามรูปที่ 3.3 (ข)
2. แบ่งกลุ่มออกเป็นกลุ่มย่อย โดยแต่ละกลุ่มสามารถแสดงรูปแบบแผนภูมิแบบ 2 มิติ โดยใช้หลักเกณฑ์ในการแบ่งกลุ่ม 2 ข้อ [43] ดังนี้
    - 2.1. ถ้ามีกรดอะมิโนโพรลีนอยู่ที่ใดก็ตาม จะทำการแบ่งข้อมูลออกเป็น 2 กลุ่มในตำแหน่งโพรลีน (Proline, P)
    - 2.2. ถ้ามี 0 อยู่ติดกันมากกว่า 4 ตัว จะทำการแบ่งข้อมูลออกเป็น 2 กลุ่ม

ก)	MASFKIALLLGVIAFVNACSQAPGTTTTTVTTTVTTVSADDGSEAGLLS
ข)	M00F0I0LLL0VI0FV0000000000000V000V00V000000000LL0
ค)	M00F0I0LLL0VI0FV V000V00V LL

รูปที่ 3.3 การแบ่งกลุ่มลำดับกรดอะมิโนตามกฎเกณฑ์ข้างต้น

การแทนข้อมูลเช่นนี้สามารถนำไปใช้ในการจับคู่กลุ่มที่ไม่ชอบน้ำต่อไปได้ โดยนำแต่ละกลุ่มในรูปที่ 3.3 (ค) มาเชื่อมต่อกันโดยใช้สัญลักษณ์ "," จะได้การแทนข้อมูล ซึ่งจะได้การแทนข้อมูลที่มีลักษณะดังนี้ "M00F0I0LLL0VI0FV,V000V00V,LL" การแทนข้อมูลสามารถเปลี่ยนไปอยู่ในรูปแบบของแผนภูมิ 2 มิติได้ดังตารางที่ 3.2

ตารางที่ 3.2 ตัวอย่างการแทนข้อมูลจากกลุ่มที่ไม่ชอบน้ำ

ตัวอย่างกลุ่ม	แผนภูมิ 2 มิติ	การแทนข้อมูล
กลุ่มที่ 1		M00F0I0LLL0VI0FV
กลุ่มที่ 2		V000V00V

3.3 การสร้างเครื่องมือจับคู่กลุ่มที่ไม่ชอบน้ำแบบ 2 มิติ

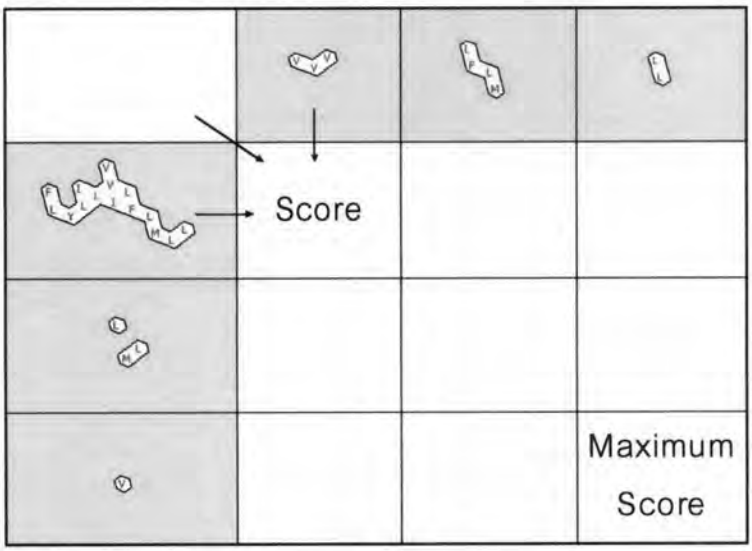
เครื่องมือจับคู่กลุ่มที่ไม่ชอบน้ำแบบ 2 มิติใช้การแทนข้อมูลที่ได้จากข้อที่ 3.2 มาจับคู่กลุ่มที่มีลักษณะโครงสร้างเหมือนกันให้เข้าคู่กัน และเพื่อให้การจับคู่มีประสิทธิภาพ ดังนั้นจึงเลือกใช้หลักการกำหนดการพลวัตเข้ามาใช้จับคู่เพื่อเปรียบเทียบแต่ละกลุ่ม และคะแนนความเหมือนของแต่ละกลุ่มจะทำโดยใช้การจับคู่แบบสายอักขระเพื่อหาตำแหน่งที่ให้คะแนนสูงที่สุด และให้คะแนนโดยดูจากตารางคะแนน นอกจากนั้นยังหักลบค่าช่องว่างเข้าไปด้วย เพื่อให้การจับคู่มีลักษณะดียิ่งขึ้น ซึ่งจะอธิบายวิธีทำงานในแต่ละหัวข้ออย่างละเอียดต่อไป

3.3.1 ขั้นตอนวิธีการทำงานของเครื่องมือจับคู่กลุ่มที่ไม่ชอบน้ำ

เครื่องมือจับคู่กลุ่มที่ไม่ชอบน้ำสามารถวิเคราะห์ได้ดังนี้

1. ข้อมูลเข้า คือ ลำดับกรดอะมิโน 2 สายลำดับ
2. การประมวลผล คือ นำลำดับกรดอะมิโนคำนวณหาตำแหน่งที่ให้คะแนนมากที่สุดด้วยวิธีกำหนดการพลวัต
3. ข้อมูลออก คือ ค่าคะแนนสะสมที่มากที่สุด และค่าความเหมือนของลำดับ

จากการวิเคราะห์ดังที่กล่าวมาสามารถรูปการทำงานของเครื่องมือได้ดังรูปที่ 3.4 และสามารถแสดงเป็นรหัสเทียมได้ดังรูปที่ 3.5



รูปที่ 3.4 การจับคู่กลุ่มที่ไม่ชอบน้ำโดยอาศัยวิธีกำหนดการพลวัต

ฟังก์ชันหลัก รับข้อมูลเป็นลำดับกรดอะมิโน 2 สายลำดับ (aminoSQ1,aminoSQ2)
เริ่มการทำงาน
เปลี่ยนลำดับกรดอะมิโนให้อยู่ในรูปแบบการแทนข้อมูลทั้ง 2 สายลำดับ (representation1 และ representation2)
กำหนดค่าสำหรับแถวขอบของวิธีกำหนดการพลวัตให้กับ ตารางเก็บตำแหน่งสุดท้ายของทั้ง 2 ลำดับ (endResidue1 และ endResidue2) ตารางเก็บคะแนนสะสม (scoreMatrix) และตารางเก็บความเหมือนของลำดับ (identityMatrix)
วนซ้ำในทุกกลุ่มที่ไม่ชอบน้ำใน representation1
วนซ้ำในทุกกลุ่มที่ไม่ชอบน้ำใน representation2
คำนวณหาคำแนนที่ดีที่สุดด้วยวิธีกำหนดการพลวัตในฟังก์ชัน HCAAlign และเก็บค่าที่คำนวณได้ในตาราง endResidue1 endResidue2 scoreMatrix identityMatrix และตารางเก็บเส้นทางการเดิน (k)
จบการวนซ้ำ
จบการวนซ้ำ
คำนวณค่าความเหมือนของลำดับจากตำแหน่งสุดท้ายของตาราง identityMatrix
คืนค่า คะแนนสูงสุดจากตำแหน่งสุดท้ายของตาราง scoreMatrix และคืนค่าความเหมือน
สิ้นสุดการทำงาน

รูปที่ 3.5 ขั้นตอนวิธีการทำงานของฟังก์ชันหลักในเครื่องมือที่ใช้จับคู่กลุ่มที่ไม่ชอบน้ำ

จากรหัสเทียมของฟังก์ชันหลักเริ่มจากการเปลี่ยนข้อมูลให้อยู่ในรูปแบบการแทนข้อมูลพร้อมทั้งแบ่งกลุ่มแต่ละกลุ่มออกจากกันและคั่นแต่ละกลุ่มด้วย "," จากนั้นนำแต่ละกลุ่มที่ได้ไปหาคะแนนที่มากที่สุดด้วยวิธีกำหนดการพลวัตตามฟังก์ชัน HCAAlign ดังรูปที่ 3.6 ทั้งนี้การคำนวณคะแนนมีการเพิ่มค่าหักลบช่องว่างเข้าไปด้วย (อธิบายในหัวข้อ 3.3.2) และในระหว่างการดำเนินไปของโปรแกรมมีการเก็บตำแหน่งสุดท้ายของกรดอะมิโนซึ่งถูกใช้ไปในการจับคู่ เพื่อสำหรับการคำนวณคะแนน อีกทั้งมีการเก็บเส้นทางการเดินไว้ด้วย

```
ฟังก์ชัน HCAAlign รับข้อมูลเป็น (Cluster1 Cluster2 endResidue1 endResidue2
scoreMatrix i j และ Gap)
  เริ่มการทำงาน โดยหลักกำหนดการพลวัต
    //นำคะแนนจากด้านบนมาคำนวณต่อ
        หาตำแหน่งสุดท้ายที่ถูกใช้ไปจาก endResidue2[i][j-1]
        หาคะแนนที่ได้จาก scoreMatrix[i][j-1]
        คำนวณคะแนนระหว่างกลุ่มจากฟังก์ชัน Alignment โดยใช้ Cluster1 Cluster2
            endResidue ในการคำนวณ คืนค่าเป็นคะแนนและตำแหน่งสุดท้ายที่ใช้ไป
        เก็บคะแนนที่คำนวณค่าหักลบช่องว่างแล้วได้ไว้ใน Score[1]
    //นำคะแนนจากด้านซ้ายมาคำนวณต่อ
        หาตำแหน่งสุดท้ายที่ถูกใช้ไปจาก endResidue1[i-1][j]
        หาคะแนนที่ได้จาก scoreMatrix[i-1][j]
        คำนวณคะแนนระหว่างกลุ่มจากฟังก์ชัน Alignment โดยใช้ Cluster1 Cluster2
            endResidue ในการคำนวณ คืนค่าเป็นคะแนนและตำแหน่งสุดท้ายที่ใช้ไป
        เก็บคะแนนที่คำนวณค่าหักลบช่องว่างแล้วได้ไว้ใน Score[2]
    //นำคะแนนจากแนวทแยงมาคำนวณต่อ
        ตำแหน่งสุดท้ายของทั้ง 2 กลุ่มไม่ได้ถูกใช้
        หาคะแนนที่ได้จาก scoreMatrix[i-1][j-1]
        คำนวณคะแนนระหว่างกลุ่มจากฟังก์ชัน Alignment โดยใช้ Cluster1 Cluster2
            endResidue ในการคำนวน คืนค่าเป็นคะแนนและตำแหน่งสุดท้ายที่ใช้ไป
        เก็บคะแนนที่คำนวณได้ไว้ใน Score[3]
    เปรียบเทียบคะแนนที่ได้ทั้งหมดเพื่อหาคะแนนที่มากที่สุด
    คืนค่า คะแนนที่ได้ ตำแหน่งสุดท้าย และเส้นทาง
  สิ้นสุดการทำงาน
```

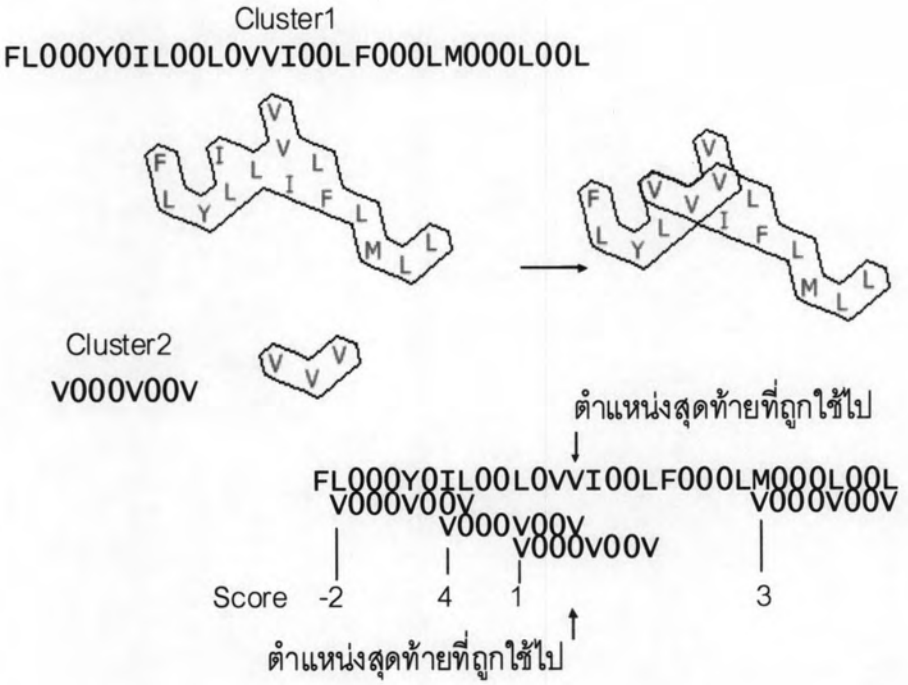
รูปที่ 3.6 ขั้นตอนวิธีการทำงานของฟังก์ชัน HCAAlign

```
ฟังก์ชัน Alignment รับข้อมูลเป็น Cluster1 Cluster2 endResidue
เริ่มการทำงาน
กำหนดตัวแปรเก็บคะแนนที่คำนวณได้ tmpScore = 0
ถ้าตำแหน่งสุดท้ายมีความยาวน้อยกว่าความยาวของ Cluster1 ให้เริ่มการคำนวณ
  กำหนดตำแหน่งเริ่มต้นจากตำแหน่งสุดท้ายที่ถูกใช้ไปให้กับ Cluster1 และ Cluster2
  เปรียบเทียบตำแหน่งที่สามารถเป็นไปได้ทั้งหมดระหว่าง Cluster1 และ Cluster2
      คำนวณคะแนนจากตารางให้คะแนนไว้ใน newScore
      คำนวณตำแหน่งสุดท้ายที่ใช้ไปเก็บค่าไว้ใน newEndResidue
      ถ้าคะแนนใหม่ (newScore) มีค่ามากกว่าคะแนนที่เก็บไว้ (tmpScore) ให้ทำ
          เปลี่ยนคะแนนใน tmpScore เป็น newScore
          เปลี่ยนตำแหน่งสุดท้ายใน tmpEndResidue = newEndResidue
      ถ้าไม่ใช่ ให้ตรวจสอบว่าคะแนนใหม่ (newScore) เท่ากับ (tmpScore) ให้ทำ
          เปลี่ยนตำแหน่งสุดท้ายใน tmpEndResidue = newEndResidue สำหรับ
Cluster2 เท่านั้น
      จบ
  จบการวนซ้ำ
จบ
คืนค่า tmpScore และ tmpEndResidue
สิ้นสุดการทำงาน
```

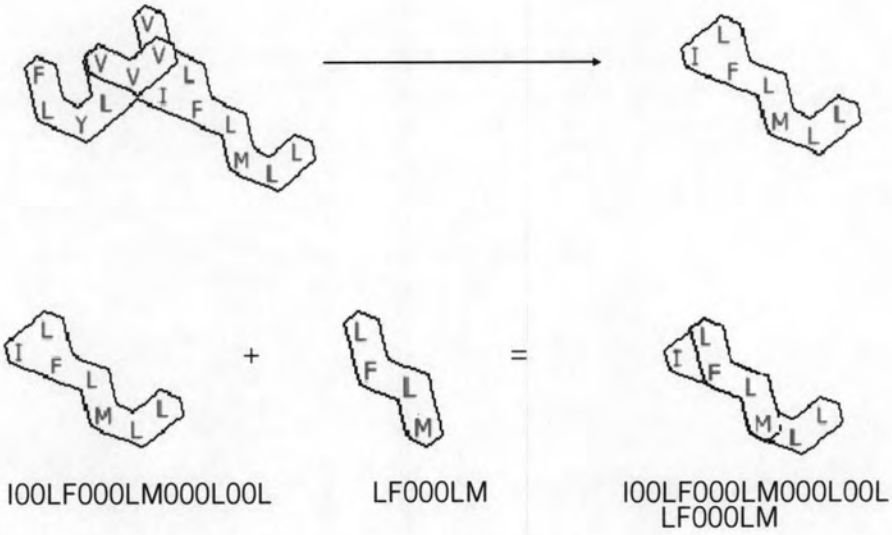
รูปที่ 3.7 ขั้นตอนวิธีการทำงานของฟังก์ชัน Alignment

จากขั้นตอนวิธีการ Alignment จะนำลำดับกรดอะมิโนทั้ง 2 กลุ่มมาเปรียบเทียบและหาตำแหน่งที่ได้คะแนนมากที่สุด โดยวิธีการเปรียบเทียบคือการนำรูปแบบการแทนข้อมูลเข้ามาเปรียบเทียบอักขระแบบไม่เว้นช่องว่าง ซึ่งจะเลื่อนการเปรียบเทียบทั้งกลุ่มจนครบทุกตำแหน่งที่เป็นไปได้ ดังแสดงในรูปที่ 3.8 จะเห็นได้ว่าในตำแหน่งที่ทำให้คะแนนสูงสุด เป็นตำแหน่งที่ทำให้ได้คะแนนถึง 4 คะแนน นั่นคือตำแหน่งที่แสดงในรูป มีการใช้กรดอะมิโนของ Cluster1 ไป 15 ตัว และใช้กรดอะมิโนใน Cluster2 ไป 8 ตัว และหลังจากนั้นไม่มีคะแนนใดที่เกิดขึ้นทำให้ได้คะแนนเพิ่มมากขึ้น โดยที่คะแนนที่ได้มากที่สุดเท่ากับ 4 คะแนน

ในครั้งต่อไปถ้ามีการใช้ Cluster1 อีกจำนวนกรดอะมิโนที่นำมาเปรียบเทียบก็จะถูกตัดไปจนถึงตำแหน่งที่ 15 แล้วจึงนำตำแหน่งที่เหลือมาเปรียบเทียบอีกครั้งหนึ่ง ดังรูปที่ 3.9



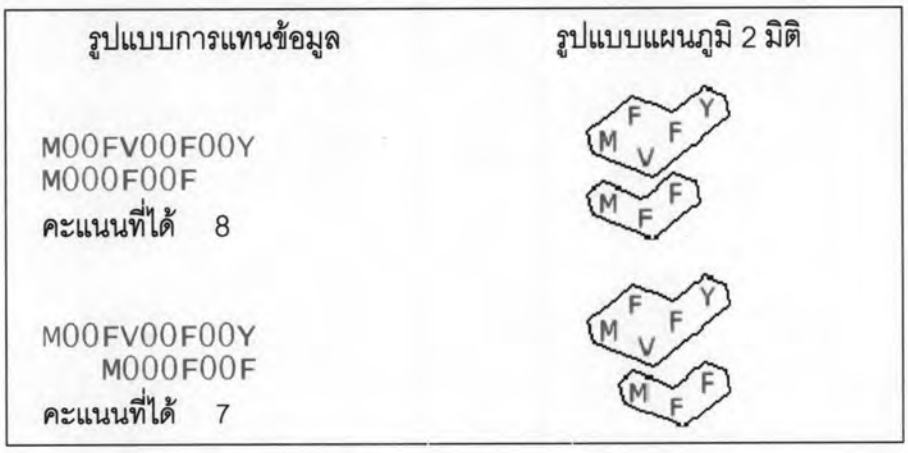
รูปที่ 3.8 ตัวอย่างการเปรียบเทียบจากกลุ่มที่ไม่มีกรดอะมิโนที่ถูกใช้



รูปที่ 3.9 ตัวอย่างการเปรียบเทียบจากกลุ่มที่มีกรดอะมิโนที่ถูกใช้

3.3.2 การคิดคะแนนการจับคู่กลุ่ม

การคิดคะแนนระหว่างกลุ่มจะใช้วิธีการเปรียบเทียบกลุ่มลำดับอักษร (String Matching) โดยใช้การเลื่อนพร้อมกันทั้งกลุ่ม จะไม่มีการเว้นช่องว่างตรงกลาง ทั้งนี้เนื่องจากภายในโครงสร้างทุติยภูมิของลำดับกรดอะมิโนมีข้อจำกัดในด้านตำแหน่งกรดอะมิโนสำหรับการสร้างโครงสร้างอยู่ด้วย ด้วยเหตุนี้การคำนวณคะแนนจึงไม่อนุญาตให้มีการเว้นช่องว่างได้ ดังรูปที่ 3.10 และใช้ตาราง BLOSUM62 ซึ่งเป็นตารางคะแนนที่นิยมใช้โดยทั่วไปและยังเป็นตารางที่คำนวณมาจากการเปรียบเทียบลำดับที่มีค่าความเหมือนปานกลาง แต่ผู้วิจัยได้มีการปรับปรุงตาราง BLOSUM62 สำหรับคะแนนที่จะต้องให้ระหว่างกรดอะมิโนที่ไม่ชอบน้ำกับกรดอะมิโนอื่นที่ไม่ได้อยู่ในตารางโดยใช้การเฉลี่ยคะแนนที่เหลือทั้งหมด เพื่อให้สามารถนำมาใช้กับเครื่องมือชิ้นนี้ได้อย่างเหมาะสม ดังรูปที่ 3.11



รูปที่ 3.10 การเปรียบเทียบกลุ่มโดยใช้การแทนข้อมูล

	M	I	L	V	F	Y	W	Other
M	**5**	1	2	1	0	-1	-1	-2
I		**4**	2	3	0	-1	-3	-3
L			**4**	1	0	-1	-2	-2
V				**4**	-1	-1	-3	-2
F					**6**	3	1	-2
Y						**7**	2	-2
W							**11**	-3

รูปที่ 3.11 ตารางคะแนน BLOSUM62 ที่ปรับปรุงจากเดิม

จากรูปที่ 3.10 เป็นรูปแสดงการเปรียบเทียบกลุ่มที่ไม่ชอบน้ำ 2 กลุ่ม กลุ่มแรกมีการแทนข้อมูลเป็น "M00FV00F00Y" และกลุ่มที่ 2 มีการแทนข้อมูลเป็น "M000F00F" ตัวอักษรสีสีเขียวแสดงกรดอะมิโนที่ไม่ชอบน้ำ และตัวหนาแสดงถึงตำแหน่งที่มีเป็นกรดอะมิโนที่ไม่ชอบน้ำเหมือนกัน สำหรับการคำนวณหาค่าความเหมือนของลำดับที่ไม่ชอบน้ำระหว่าง 2 กลุ่ม พบว่ามีตำแหน่งที่น่าจะเป็นคะแนนสูงสุดอยู่ 2 ตำแหน่ง คือตำแหน่งดังรูปแถวบนและแถวล่าง สำหรับแถวบนได้คะแนนเท่ากับ 8 และแถวล่างได้คะแนนเท่ากับ 7 ดังนั้นคะแนนสูงสุดจะได้ตามรูปแบบของแถวบน คือได้คะแนนเท่ากับ 8 และสำหรับแถวบนยังเหลือกรดอะมิโนที่ไม่ชอบน้ำเหลืออยู่อีก 1 ตัว ทั้งนี้ถ้ามีการนำกลุ่มถัดไปข้ามมาจับคู่ด้วยก็จะต้องมีการคิดค่าหักลบเมื่อเปิดช่องว่างด้วย

3.3.3 ค่าหักลบช่องว่าง

ในงานวิจัยใช้วิธีคิดค่าหักลบช่องว่าง 2 วิธีคือ

3.3.3.1 Affine Gap Penalty

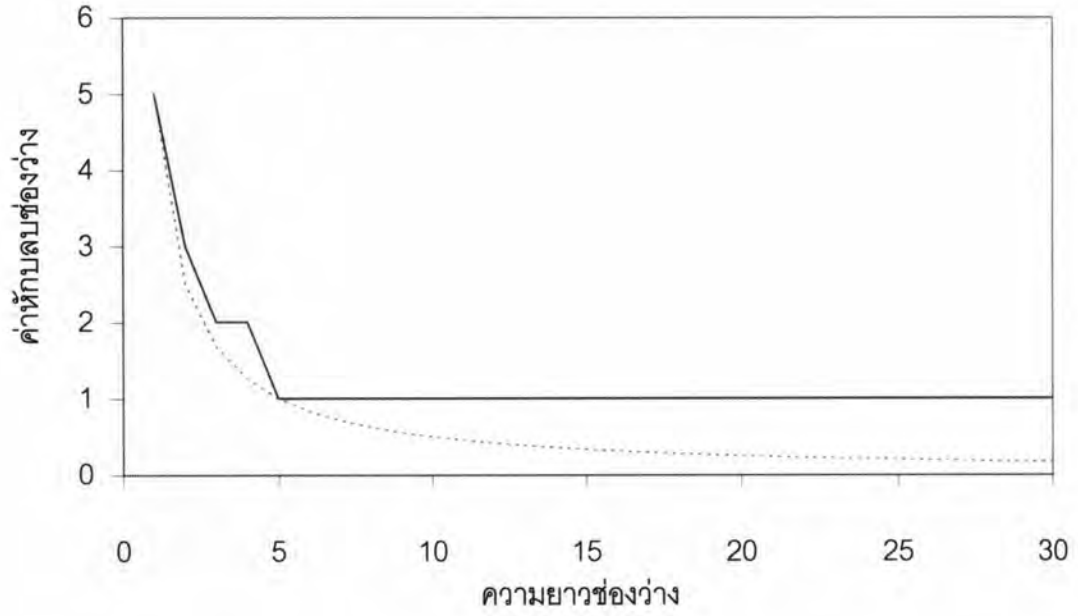
ค่าการหักลบนี้จะมีอยู่ด้วยกัน 2 ชนิด คือ การหักลบเมื่อเปิดช่องว่าง (Gap Open Penalty) และค่าหักลบเมื่อขยายช่องว่าง (Gap Extension Penalty) ซึ่งค่าหักลบจะต้องมีการปรับค่าให้เข้ากับตาราง BLOSUM62 และให้ความถูกต้องของการเลือกลำดับที่สูงด้วยเมื่อใช้ฐานข้อมูลเป็นกรดอะมิโนกลุ่มใดๆ ทั้งนี้การทดลองได้แสดงในหัวข้อ 4.2.1

3.3.3.2 ค่าหักลบช่องว่างใหม่

ผู้วิจัยได้คิดค่าหักลบช่องว่างใหม่ จากการศึกษาความถี่ของขนาดช่องว่างที่เกิดขึ้นภายในกลุ่มที่ไม่ชอบน้ำ และผู้วิจัยได้เสนอฟังก์ชันการคำนวณค่าหักลบช่องว่างใหม่เป็นไปดังสมการ 3.2

$$g(i) = \frac{u}{l(i)} \tag{3.2}$$

ขณะที่ $g(i)$ คือค่าหักลบช่องว่างที่ตำแหน่ง i u คือค่าหักลบช่องว่างเริ่มต้น และ $l(i)$ คือขนาดช่องว่างที่เกิดขึ้น ณ ตำแหน่งที่ i เมื่อนำสมการไปสร้างกราฟจะได้ดังรูปที่ 3.12



รูปที่ 3.12 ค่าหักลบช่องว่างแบบใหม่ที่นำเสนอโดยมีค่าเริ่มต้นที่ 5

จากรูปที่ 3.12 แสดงค่าหักลบเมื่อมีค่าเริ่มต้นที่ -5 และแต่ละตำแหน่งที่เพิ่มขึ้นจะมีค่าหักลบช่องว่างมีค่าลดลงเรื่อยๆ เส้นประสีเขียวแสดงค่าหักลบช่องว่างจากสมการที่ 3.2 สำหรับเส้นสีน้ำเงินคือค่าหักลบช่องว่างที่ปรับคะแนนและนำมาใช้ในเครื่องมือจับคู่กลุ่มที่ไม่ชอบน้ำแบบ 2 มิติโดยอัตโนมัติ โดยการใช้ค่าปัดขึ้นเพื่อให้เป็นจำนวนเต็มและค่าน้อยสุดที่หักลบเมื่อมีการขยายช่องว่างเท่ากับ -1

3.3.4 การคิดค่าความเหมือนของลำดับที่ไม่ชอบน้ำ

เครื่องมือนี้จะต้องคิดค่าความเหมือนของลำดับของกลุ่มที่ไม่ชอบน้ำ (HCA Homology Score) สามารคำนวณได้จากสมการ 3.3

$$HydrophobicScore = \frac{2 \times Correct}{HydroAA_1 + HydroAA_2} \qquad (3.3)$$

จากสมการข้างต้น *Correct* คือ จำนวนกรดอะมิโนไม่ชอบน้ำอยู่ที่ตำแหน่งเดียวกัน $HydroAA_1$ คือจำนวนกรดอะมิโนไม่ชอบน้ำในลำดับกรดอะมิโนสายที่ 1 และ $HydroAA_2$ คือ จำนวนกรดอะมิโนไม่ชอบน้ำในลำดับกรดอะมิโนสายที่ 2

I 44681054