

## บทที่ 4

### การทดลองและผลการทดลอง

ในบทนี้จะกล่าวถึงวิธีการทดลองและผลการทดลองของงานวิจัยเรื่อง การหาโปรตีนที่ทำหน้าที่เหมือนกันโดยใช้เครื่องมือจับคู่กลุ่มที่ไม่ชอบน้ำแบบ 2 มิติโดยอัตโนมัติ ซึ่งเป็นการทดลองหาอัตราความถูกต้องของการจับคู่กลุ่มที่ไม่ชอบน้ำ และการจำแนกกลุ่มหน้าที่ของโปรตีนด้วยเครื่องมือที่นำเสนอ โดยข้อมูลลำดับกรดอะมิโนที่ใช้ในการทดลองจำนวน 3 ชุดข้อมูลมาจาก 3 แหล่งคือฐานข้อมูล PIR SISYPHUS และ HOMSTRAD

#### 4.1 วิธีการเตรียมข้อมูลสำหรับการทดสอบ

ชุดข้อมูลที่ใช้เพื่อทดสอบความถูกต้องของเครื่องมือชุดนี้จะต้องมีชุดข้อมูลอยู่ด้วยกัน 3 ชุดข้อมูลหลักๆ คือ ชุดข้อมูลสำหรับทดลองเพื่อหาค่าหักลบช่องว่างที่เหมาะสม ชุดข้อมูลเพื่อทดสอบความถูกต้องของเครื่องมือจับคู่กลุ่มที่ไม่ชอบน้ำแบบ 2 มิติโดยอัตโนมัติ และชุดข้อมูลเพื่อการทดลองเพิ่มขนาดข้อมูลเปรียบเทียบ การกำหนดรายละเอียดของชุดข้อมูล และการเตรียมข้อมูล มีวิธีการดังต่อไปนี้

##### 4.1.1 การเตรียมข้อมูลสำหรับทดลองเพื่อหาค่าหักลบช่องว่าง

ชุดข้อมูลที่ใช้สำหรับการทดลองเพื่อหาค่าหักลบช่องว่างประกอบไปด้วยชุดข้อมูลสอนและชุดข้อมูลทดสอบจากฐานข้อมูล HOMSTRAD โดยมีการแบ่งกลุ่มโปรตีนไว้ทั้งหมด 1032 หน้าที่ของโปรตีน และทำการสุ่มเลือกข้อมูลใส่ไว้ในชุดข้อมูลเปรียบเทียบ และชุดข้อมูลทดสอบ โดยที่ชุดข้อมูลเปรียบเทียบจะมีจำนวนสายลำดับกรดอะมิโนต่อกลุ่มประมาณ 1-4 สายลำดับกรดอะมิโน และในชุดข้อมูลทดสอบจะมีสายลำดับกรดอะมิโน 1 สายลำดับกรดอะมิโนต่อกลุ่มหน้าที่ของโปรตีน รายละเอียดของจำนวนข้อมูลและจำนวนกลุ่มที่นำมาทำการทดลองแสดงไว้ในตารางที่ 4.1 สำหรับรายละเอียดของชนิดโปรตีนแสดงในภาคผนวก ก.2

ตารางที่ 4.1 ชุดข้อมูลสำหรับการทดลองเพื่อหาค่าหักลบช่องว่างที่เหมาะสม

ประเภทชุดข้อมูล	จำนวนข้อมูล	จำนวนกลุ่ม
ชุดข้อมูลเปรียบเทียบ	3448	1032
ชุดข้อมูลทดสอบ	1032	1032

#### 4.1.2 ชุดข้อมูลเพื่อทดสอบความถูกต้องของเครื่องมือ

ชุดข้อมูลที่ใช้สำหรับทดสอบความถูกต้องใช้ชุดข้อมูลจาก SISYPHUS เนื่องจากว่าเป็นชุดข้อมูลที่มีการจับคู่กลุ่มที่ไม่ชอบน้ำโดยผู้เชี่ยวชาญ ซึ่งใช้สำหรับการเปรียบเทียบความถูกต้อง รายละเอียดของชุดข้อมูลแสดงในภาคผนวก ก.1 ชุดข้อมูลนี้มีการจับคู่รวมทั้งหมด 8,474 คู่

#### 4.1.3 ชุดข้อมูลเพื่อทดลองเพิ่มขนาดข้อมูลสอน

ชุดข้อมูลนี้เตรียมมาจากชุดข้อมูล HOMSTRAD ซึ่งได้มีการจัดแบ่งชุดข้อมูลสอนออกเป็นขนาดต่างๆ ดังแสดงในตารางที่ 4.2

ตารางที่ 4.2 ชุดข้อมูลสำหรับทดลองเพิ่มขนาดข้อมูลสอน

ลำดับที่	จำนวนข้อมูล เปรียบเทียบ	จำนวนข้อมูล ทดสอบ	จำนวนกลุ่ม
1	97	30	13
2	3448	1032	1032
3	44813	1032	1032
4	86678	1032	1032
5	187991	1032	1032
6	336827	1032	1032

#### 4.2 วิธีการทดลอง

งานวิจัยนี้ได้แบ่งการทดสอบออกเป็น 4 การทดลอง ได้แก่ การทดลองเพื่อหาค่าหักลบช่องว่าง การทดลองเพื่อเปรียบเทียบความถูกต้องของเครื่องมือจับคู่กลุ่มที่ไม่ชอบน้ำแบบ 2 มิติโดยอัตโนมัติ การทดลองเพื่อเปรียบเทียบความถูกต้องของการเลือกลำดับที่เหมือนที่สุด และการทดลองเพิ่มขนาดข้อมูลสอนให้มีขนาดใหญ่ขึ้น

#### 4.2.1 การทดลองเพื่อหาค่าหักลบช่องว่าง

การทดลองเพื่อหาค่าหักลบช่องว่างสำหรับการใช้ตารางให้คะแนน BLOSUM62 และตารางให้คะแนนใหม่ ค่าหักลบช่องว่างซึ่งมีอยู่ด้วยกัน 2 ชนิดคือ Affine Gap Penalty และค่าหักลบช่องว่างใหม่ โดยในการทดลองนี้ได้ทำการเปลี่ยนค่าหักลบช่องว่าง ดังแสดงในตารางที่ 4.3

ตารางที่ 4.3 การเปลี่ยนค่าหักลบช่องว่างในการทดลอง

ตารางให้คะแนน	ชนิดของค่าหักลบช่องว่าง		ค่าหักลบที่ทำการทดสอบ		
			ค่าเริ่มต้น	ค่าสิ้นสุด	อัตราการเพิ่มค่า
BLOSUM62	Affine Gap Penalty	เปิด	-10	0	1
		ขยาย	-5	0	1
BLOSUM62	ค่าหักลบช่องว่างใหม่	ค่าเริ่มต้น	-10	0	1

การทดลองนี้ใช้การเปรียบเทียบโดยการเลือกลำดับกรดอะมิโนที่เหมือนที่สุดว่าอยู่ในกลุ่มเดียวกับลำดับกรดอะมิโนที่นำมาเปรียบเทียบหรือไม่ คิดเปอร์เซ็นต์ความถูกต้องแล้วนำมาเปรียบเทียบกัน จากการทดลองนี้จะได้ค่าหักลบที่ดีที่สุดสำหรับแต่ละตารางให้คะแนนและแต่ละชนิดของค่าหักลบช่องว่าง เพื่อใช้ในการทดลองขั้นต่อไป

#### 4.2.2 การทดลองเพื่อเปรียบเทียบความถูกต้องของเครื่องมือจับคู่กลุ่มที่ไม่ชอบน้ำแบบ 2 มิติโดยอัตโนมัติ

การทดสอบนี้เป็นการทดสอบความถูกต้องของการจับคู่กลุ่มที่ไม่ชอบน้ำแบบ 2 มิติโดยอัตโนมัติ โดยใช้ชุดข้อมูลจากการจับคู่กลุ่มที่ไม่ชอบน้ำโดยผู้เชี่ยวชาญ ซึ่งการทดสอบนี้เป็นการทดสอบระหว่างเครื่องมือจับคู่กลุ่มที่ไม่ชอบน้ำแบบ 2 มิติโดยอัตโนมัติ กับการจับคู่กลุ่มโดยผู้เชี่ยวชาญ การทดลองนี้ใช้ตารางการให้คะแนน BLOSUM62 และตารางให้คะแนนใหม่ และค่าหักลบช่องว่างทั้ง 2 ชนิดที่ได้จากข้อ 4.2.1 ใช้ชุดข้อมูลของ SISYPHUS ในการทดสอบความถูกต้อง โดยวิธีการทดลองมีลำดับขั้นตอนดังนี้

1. หาค่าความเหมือนของลำดับที่ไม่ชอบน้ำ (Hydrophobic score) จากชุดข้อมูล
2. นำลำดับกรดอะมิโนไปหาคะแนนโดยใช้เครื่องมือแบบอัตโนมัติ
3. นำคะแนนที่ได้มาหาความแตกต่างระหว่างกัน และเปรียบเทียบโดยใช้ค่าเฉลี่ย ค่าน้อยสุด ค่ามากที่สุด และค่าความแปรปรวน
4. สร้างกราฟเปรียบเทียบคะแนนระหว่างคะแนนจากผู้เชี่ยวชาญ และคะแนนจากเครื่องมืออัตโนมัติ

การหาความแตกต่างคำนวณได้จากสมการ 4.1

$$Difference(a, b) = AHCAmIdentity(sq_a, sq_b) - HumanIdentity(sq_a, sq_b) \quad (4.1)$$

โดยที่ *AHCAmIdentity* คือค่าความเหมือนของลำดับที่ไม่ชอบน้ำจากเครื่องมือจับคู่กลุ่มที่ไม่ชอบน้ำแบบ 2 มิติโดยอัตโนมัติ และ *HumanIdentity* คือความเหมือนของลำดับที่ไม่ชอบน้ำจากการจับคู่โดยผู้เชี่ยวชาญ  $sq_a$  และ  $sq_b$  คือลำดับกรดอะมิโนที่ใช้ในการจับคู่กลุ่มที่ไม่ชอบน้ำ เมื่อเปรียบเทียบจนครบจะมีความแตกต่างทั้งหมดจำนวน  $N$  ข้อมูล จากนั้นหาค่าเฉลี่ยความแตกต่าง ซึ่งสามารถคำนวณได้จากสมการ 4.2

$$AverageDifference = \frac{\sum_1^N |Difference(i)|}{N} \quad (4.2)$$

#### 4.2.3 การทดลองเพิ่มขนาดชุดข้อมูลสอนให้มีขนาดใหญ่ขึ้น

สำหรับการทดลองเพิ่มขนาดชุดข้อมูลสอน ผู้วิจัยได้ใช้ชุดข้อมูลสอนจากฐานข้อมูล HOMSTRAD ที่ได้เตรียมไว้ตามขนาดชุดข้อมูลสอนขนาดต่างๆ ตามข้อที่ 4.1.3 ทำการทดลองด้วยวิธีการเลือกลำดับที่เหมือนมากที่สุด 1 ลำดับแล้วตรวจสอบว่าลำดับที่ได้มาจากกลุ่มโปรตีนเดียวกันกับลำดับกรดอะมิโนที่นำมาเปรียบเทียบหรือไม่ โดยใช้ตารางคะแนน BLOSUM62 และ Affine Gap Penalty

#### 4.2.4 การทดลองเพื่อเปรียบเทียบความถูกต้องของการเลือกลำดับที่เหมือนกันที่สุด

การทดสอบนี้เป็นการทดสอบเพื่อแสดงถึงประสิทธิภาพในการจับคู่กลุ่มที่ไม่ชอบน้ำแบบ 2 มิติโดยอัตโนมัติ โดยใช้ชุดข้อมูลจากฐานข้อมูล HOMSTRAD และคำนวณออกมาเป็นเปอร์เซ็นต์ความถูกต้อง ซึ่งเปรียบเทียบความถูกต้องกับเครื่องมือจับคู่ที่ใช้วิธีกำหนดการพลวัตแบบ 1 มิติ โดยมีรายละเอียดการทดลองดังตารางที่ 4.4

ตารางที่ 4.4 รายละเอียดการทดลองความถูกต้องของการเลือกลำดับที่เหมือนกันที่สุด

วิธีการเปรียบเทียบ	ชุดข้อมูล	ตารางคะแนน	ค่าหักลบช่องว่าง	
			เปิด	ขยาย
แบบ 1 มิติ	ชุดที่เลือกจากข้อ 4.2.3	BLOSUM62	-11	-1
แบบ 2 มิติที่นำเสนอ	ชุดที่เลือกจากข้อ 4.2.3	BLOSUM62	Affine Gap Penalty	
	ชุดที่เลือกจากข้อ 4.2.3	BLOSUM62	ค่าหักลบช่องว่างใหม่	

การคิดค่าความถูกต้องจะคิดจากลำดับที่เลือกได้เป็นลำดับแรก โดยดูจากค่าความเหมือนของลำดับที่ไม่ชอบน้ำที่สูงที่สุด จากนั้นดูว่าเป็นโปรตีนที่มาจากกลุ่มโปรตีนเดียวกับลำดับกรดอะมิโนที่นำมาเปรียบเทียบหรือไม่ ถ้ามาจากกลุ่มโปรตีนเดียวกันจะถือว่า ถูกต้อง หลังจากได้ผลรวมจากการเปรียบเทียบทั้งหมด หาเปอร์เซ็นต์ความถูกต้อง และรายงานผลการทดลอง

#### 4.3 ผลการทดลอง

จากการทำการทดลองดังกล่าวข้างต้น ได้ผลการทดลองแยกตามหัวข้อที่ทำการทดลองได้ดังต่อไปนี้

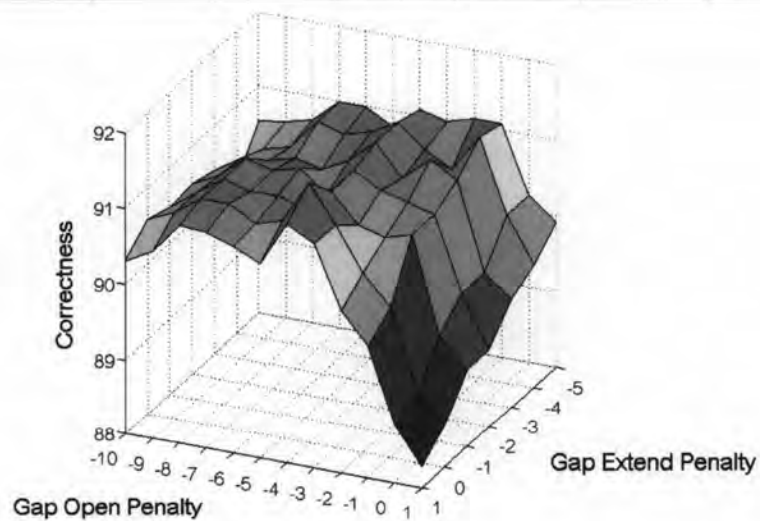
ตารางที่ 4.5 แสดงเปอร์เซ็นต์ความถูกต้องที่ค่าหักลบช่องว่างที่แตกต่างกัน โดยแสดงเป็นเปอร์เซ็นต์ความถูกต้องของการเลือกลำดับที่เหมือนกันที่สุด โดยมีการเปลี่ยนแปลงค่าหักลบเมื่อเปิดช่องว่างตั้งแต่ -10 ถึง 0 และค่าหักลบเมื่อขยายช่องว่างตั้งแต่ -5 ถึง 0

ผลการทดลองที่ได้ตามตารางที่ 4.5 แสดงให้เห็นว่าเปอร์เซ็นต์ความถูกต้องของค่าหักลบเมื่อเปิดช่องว่างที่ -4 และค่าหักลบเมื่อขยายช่องว่างที่ 0 มีค่าความถูกต้องมากที่สุดได้เท่ากับ 91.47 เปอร์เซ็นต์ จากผลการทดลองนี้จึงเลือกค่าหักลบช่องว่างทั้ง 2 ชนิดตามผลการทดลองนี้เพื่อใช้สำหรับการทดลองต่อไป

### 4.3.1 ผลการทดลองหาค่าหักลบช่องว่างที่เหมาะสม

ตารางที่ 4.5 ผลการทดลองเปรียบเทียบความถูกต้องของการเลือกลำดับกรดอะมิโนที่เหมือนที่สุด จากข้อมูลเปรียบเทียบ เพื่อหา Affine Gap Penalty ที่เหมาะสม

		ค่าหักลบเมื่อขยายช่องว่าง						
		1	0	-1	-2	-3	-4	-5
ค่าหักลบเมื่อเปิดช่องว่าง	1	88.28	88.57	89.05	89.05	89.44	89.63	89.92
	0	88.77	88.95	89.44	89.73	89.73	90.21	90.21
	-1	89.78	90.02	90.70	90.70	90.89	91.18	91.09
	-2	90.17	90.41	90.60	90.79	91.09	90.70	91.09
	-3	90.99	90.97	90.70	90.89	90.79	90.89	90.99
	-4	91.18	91.47	91.09	91.09	91.18	91.18	91.09
	-5	90.60	90.89	90.79	90.70	90.89	90.99	90.79
	-6	90.76	90.79	90.89	90.99	90.70	90.89	90.99
	-7	90.88	90.99	90.89	90.89	90.79	90.99	90.99
	-8	90.90	90.99	90.99	90.99	90.69	90.59	90.68
	-9	90.50	90.74	90.74	90.70	90.65	90.50	90.59
-10	90.31	90.58	90.48	90.52	90.42	90.00	90.54	



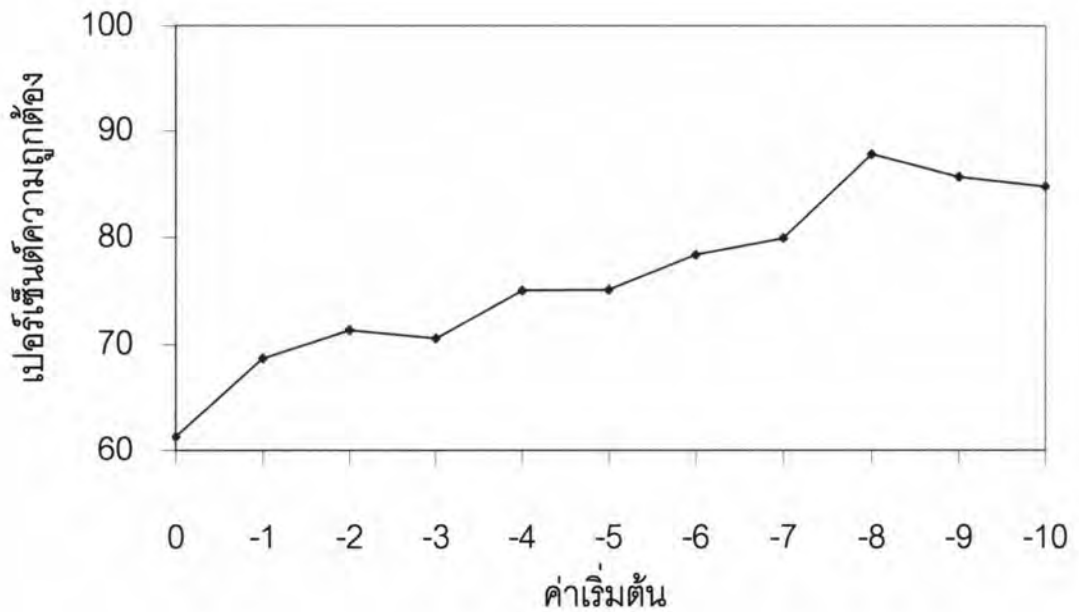
รูปที่ 4.1 กราฟแสดงความถูกต้องเมื่อเปลี่ยนแปลงค่าหักลบเมื่อเปิดช่องว่าง และค่าหักลบเมื่อขยายช่องว่างของ Affine Gap Penalty

จากการทดลองข้างต้นแสดงให้เห็นว่าค่าหักลบเชิงสัมพรรคที่ทำให้ได้ความถูกต้องมากที่สุดคือ ค่าหักลบเมื่อเปิดช่องว่างที่ -4 และค่าหักลบเมื่อขยายช่องว่างที่ 0 ซึ่งทำให้ได้ความถูกต้องมากที่สุดที่ 94.38 เปอร์เซ็นต์

Affine Gap Penalty ที่ได้มีค่าไม่เท่ากับค่าที่ตั้งเป็นมาตรฐานไว้สำหรับโปรแกรม BLAST ซึ่งเป็นที่นิยมให้กันมาก เนื่องจากโปรแกรม BLAST ใช้ค่าหักลบสำหรับหักลบตำแหน่งของช่องว่างตามจำนวนลำดับกรดอะมิโนที่หายไป เช่น MLS-D-SAFVY จะมีขนาดช่องว่างเท่ากับ 2 กรดอะมิโน แต่ค่าหักลบสำหรับเครื่องมือใหม่จะเป็นค่าหักลบสำหรับการเกิดช่องว่างภายในกลุ่มที่ไม่ชอบน้ำที่เป็นโครงสร้างของโปรตีน ดังนั้นทำให้วิธีการใช้ค่าหักลบไม่เหมือนกันกับของโปรแกรม BLAST ค่าที่ได้ออกมาจึงต่างจากของโปรแกรม BLAST ที่ใช้ค่าหักลบเมื่อเปิดช่องว่างที่ -11 และค่าหักลบเมื่อขยายช่องว่างที่ -1

ตารางที่ 4.6 ผลการทดลองเปรียบเทียบความถูกต้องของการเลือกลำดับกรดอะมิโนที่เหมือนที่สุด จากข้อมูลเปรียบเทียบ หาค่าหักลบช่องว่างใหม่ที่เหมาะสม

ค่าเริ่มต้น	ความถูกต้อง
0	61.32
-1	68.70
-2	71.32
-3	70.56
-4	74.97
-5	75.03
-6	78.34
-7	79.93
-8	87.81
-9	85.70
-10	84.80



รูปที่ 4.2 กราฟแสดงความถูกต้องเมื่อเปลี่ยนแปลงค่าเริ่มต้นของค่าหักลบช่องว่างใหม่

จากผลการทดลองข้างต้นแสดงให้เห็นว่าค่าหักลบช่องว่างใหม่ที่ทำให้ได้ความถูกต้องมากที่สุดอยู่ที่ -8 ให้ความถูกต้องของการเลือกลำดับสูงที่สุดที่ 87.81 เปอร์เซ็นต์ ซึ่งน้อยกว่าวิธีการหักลบช่องว่างแบบเดิม

#### 4.3.2 ผลการทดลองเปรียบเทียบความถูกต้องของเครื่องมือจับคู่กลุ่มที่ไม่ชอบน้ำแบบ 2 มิติโดยอัตโนมัติ

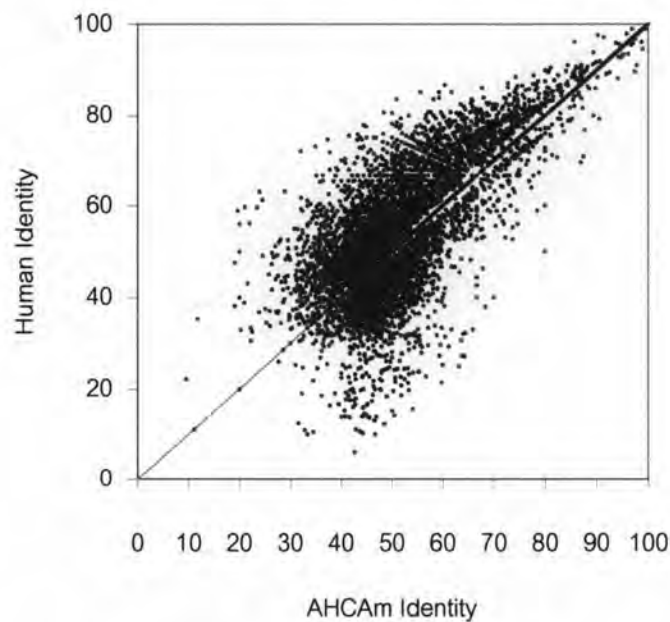
ตารางที่ 4.7 ผลการทดลองเปรียบเทียบค่าความเหมือนของลำดับที่ไม่ชอบน้ำระหว่างเครื่องมือจับคู่กลุ่มที่ไม่ชอบน้ำแบบ 2 มิติโดยอัตโนมัติ กับการจับคู่กลุ่มโดยผู้เชี่ยวชาญ

	Affine Gap Penalty	ค่าหักลบช่องว่างใหม่
ค่าเฉลี่ยความแตกต่าง	7.98	7.45
ค่าความแปรปรวน	6.75	6.69
ค่าความแตกต่างที่น้อยที่สุด	0.00	0.00
ค่าความแตกต่างที่มากที่สุด	44.00	45.36



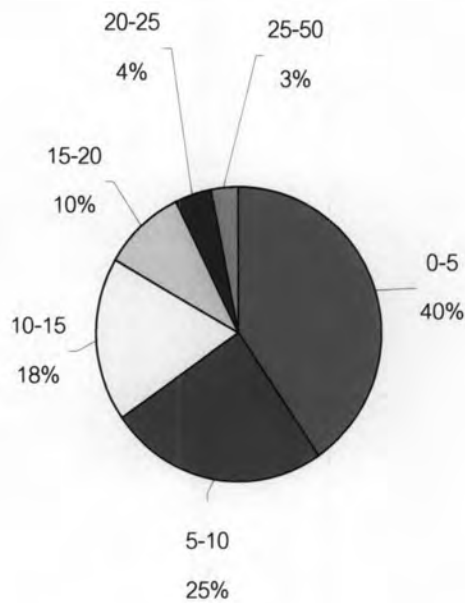
จากผลการทดลอง แสดงให้เห็นว่าเครื่องมือจับคู่กลุ่มที่ไม่ชอบน้ำแบบ 2 มิติโดยอัตโนมัติเมื่อใช้ตารางคะแนน BLOSUM62 และ Affine Gap Penalty พบว่า มีค่าความแตกต่างเฉลี่ยอยู่ที่ 7.98 เปอร์เซ็นต์ ค่าความแปรปรวนอยู่ที่ 6.75 เปอร์เซ็นต์ ความแตกต่างที่มากที่สุดอยู่ที่ 44.00 เปอร์เซ็นต์ และมีการจับคู่ที่สามารถจับคู่ได้เหมือนกับผู้เชี่ยวชาญ นอกจากนี้เมื่อเปลี่ยนเป็นค่าหักลบช่องว่างใหม่ พบว่า มีค่าความแตกต่างเฉลี่ยอยู่ที่ 7.45 เปอร์เซ็นต์ ค่าความแปรปรวนอยู่ที่ 6.69 เปอร์เซ็นต์ ความแตกต่างที่มากที่สุดอยู่ที่ 45.36 เปอร์เซ็นต์ และมีการจับคู่ที่สามารถจับคู่ได้เหมือนกับผู้เชี่ยวชาญ

จากผลการทดลองผู้วิจัยสร้างกราฟระหว่างค่าความเหมือนของลำดับที่ไม่ชอบน้ำจากเครื่องมือจับคู่กลุ่มโดยอัตโนมัติ กับการจับคู่โดยผู้เชี่ยวชาญ ได้ดังรูปที่ 4.3



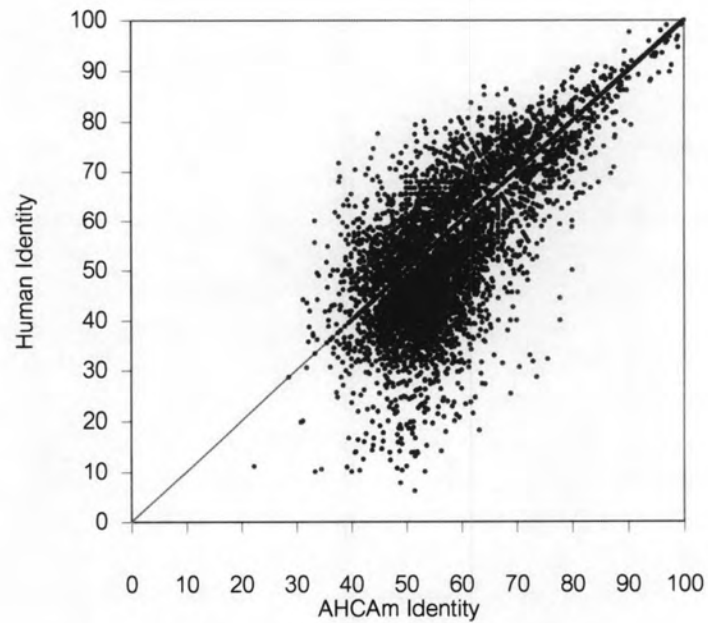
รูปที่ 4.3 การเปรียบเทียบค่าความเหมือนของลำดับที่ไม่ชอบน้ำระหว่างเครื่องมือจับคู่โดยอัตโนมัติกับการจับคู่โดยผู้เชี่ยวชาญโดยใช้ตาราง BLOSUM62 และ Affine Gap Penalty

จากรูปที่ 4.3 เป็นการวาดกราฟระหว่างค่าความเหมือนของลำดับที่ไม่ชอบน้ำจากผู้เชี่ยวชาญกับค่าความเหมือนของลำดับจากเครื่องมือจับคู่กลุ่มที่ไม่ชอบน้ำแบบ 2 มิติโดยอัตโนมัติ และเส้นกลางแสดงตำแหน่งที่ค่าความเหมือนของลำดับที่ไม่ชอบน้ำที่เท่ากัน ซึ่งถ้าจุดทุกจุดที่เกิดขึ้นอยู่บนเส้นกลาง หมายความว่า เครื่องมือใหม่มีความสามารถในการจับคู่กลุ่มได้เหมือนกับผู้เชี่ยวชาญ นอกจากนี้ผู้วิจัยได้ทดลองนับจำนวนความถี่ของความแตกต่างที่เกิดขึ้นแสดงดังรูปที่ 4.4



รูปที่ 4.4 ความถี่ของความแตกต่างที่เกิดขึ้นจากการใช้ตาราง BLOSUM62 และ Affine Gap Penalty

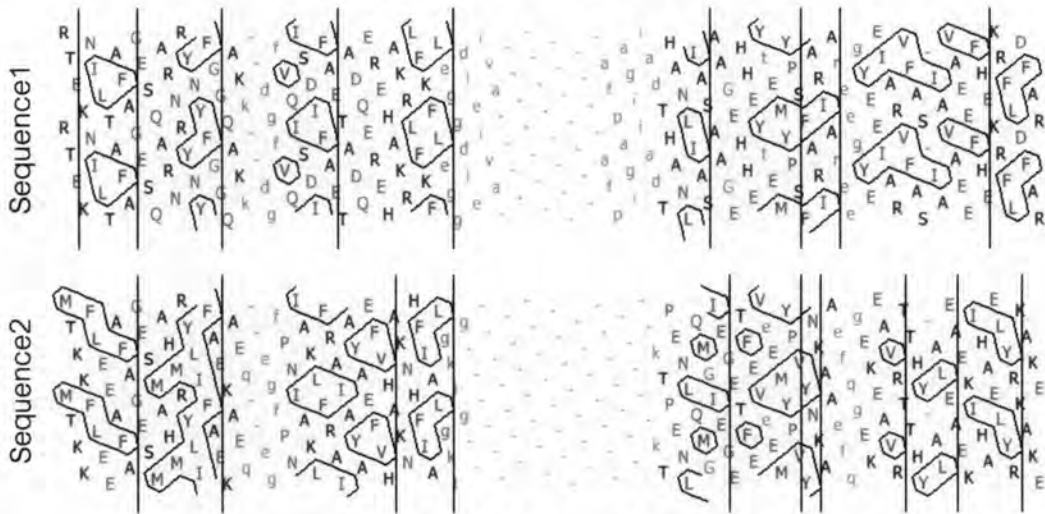
จากรูปที่ 4.4 จะเห็นได้ว่าความแตกต่างไม่เกิน 15 เปอร์เซ็นต์มีจำนวนมากถึง 83 เปอร์เซ็นต์ของการเปรียบเทียบทั้งหมด และสำหรับความแตกต่างที่อยู่ระหว่าง 0-5 เปอร์เซ็นต์ มีจำนวนมากถึง 40 เปอร์เซ็นต์ของการเปรียบเทียบทั้งหมด จากผลการทดลองนี้แสดงให้เห็นว่า เครื่องมือจับคู่กลุ่มที่ไม่ชอบน้ำมีประสิทธิภาพใกล้เคียงกับการจับคู่กลุ่มโดยผู้เชี่ยวชาญมาก



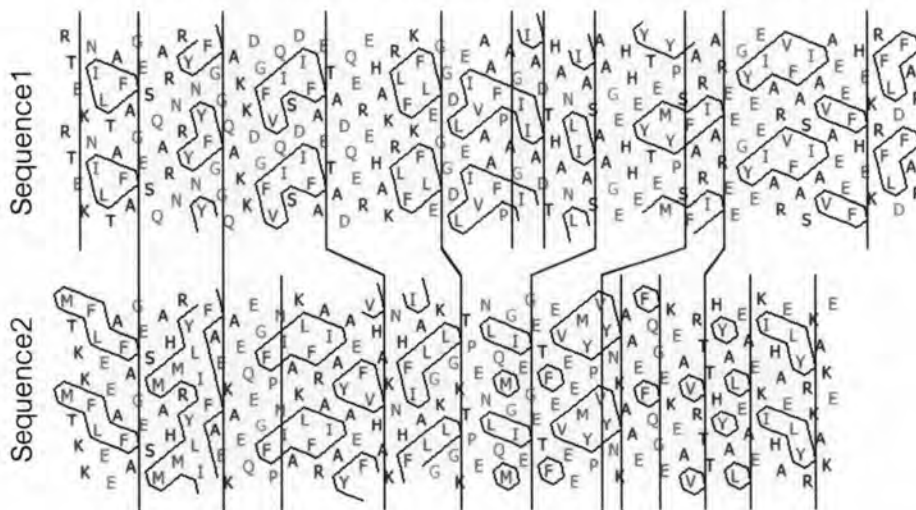
รูปที่ 4.5 การเปรียบเทียบค่าความเหมือนของลำดับที่ไม่ชอบน้ำระหว่างเครื่องมือจับคู่โดยอัตโนมัติกับการจับคู่โดยผู้เชี่ยวชาญโดยใช้ตาราง BLOSUM62 และค่าหักลบช่องว่างใหม่

นอกจากนี้วิเคราะห์ผลของการจับคู่ โดยเปรียบเทียบลักษณะของการจับคู่ทั้งจากผู้เชี่ยวชาญและจากเครื่องมือจับคู่กลุ่มโดยอัตโนมัติมาวาดเป็นแผนภูมิ 2 มิติเพื่อศึกษา ลักษณะการจับคู่กลุ่มที่เกิดขึ้น แสดงดังรูปที่ 4.6

a) ค่าความเหมือนของลำดับที่ไม่ชอบน้ำจากผู้เชี่ยวชาญ เท่ากับ 47.37

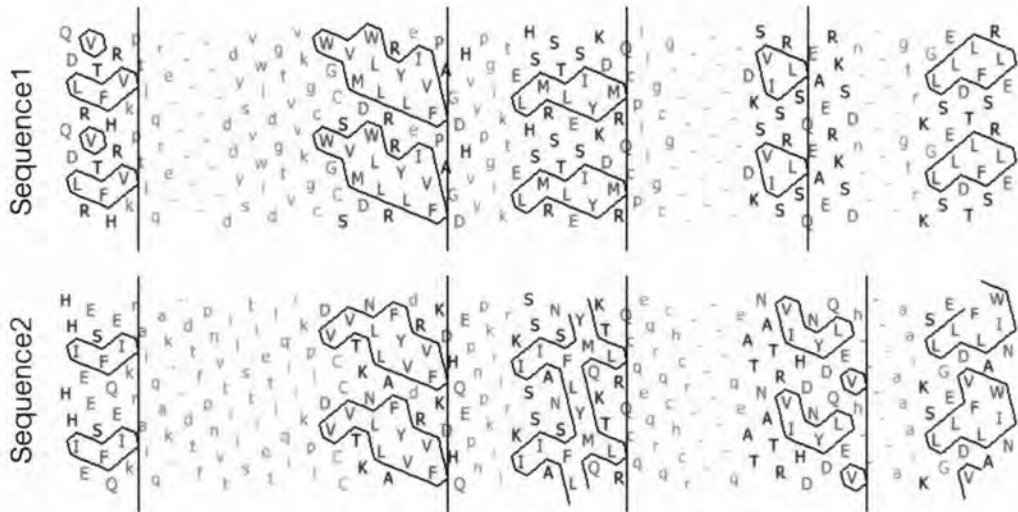


b) ค่าความเหมือนของลำดับที่ไม่ชอบน้ำจากเครื่องมือจับคู่โดยอัตโนมัติ เท่ากับ 63.16

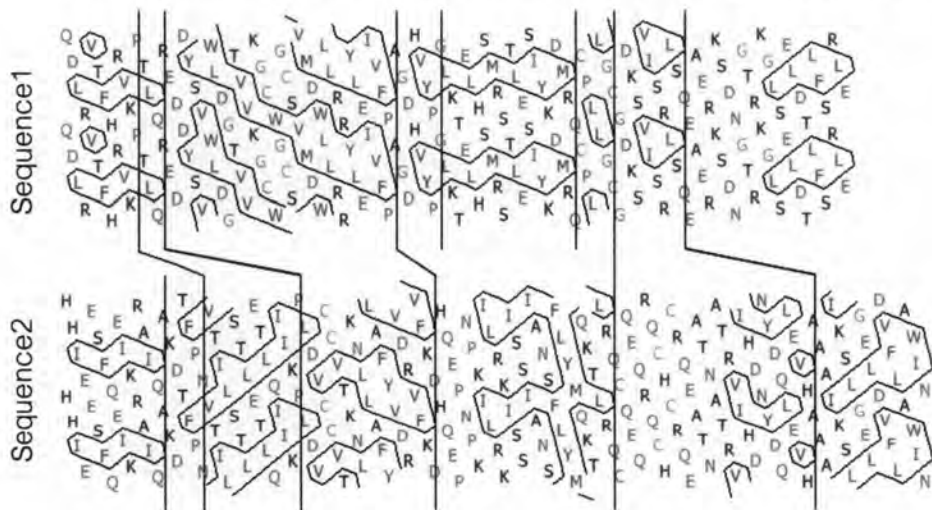


รูปที่ 4.6 การจับคู่กลุ่มที่ไม่ชอบน้ำของโปรตีน Rubrerythrin จากผู้เชี่ยวชาญ และจากเครื่องมือจับคู่โดยอัตโนมัติ

a) ค่าความเหมือนของลำดับที่ไม่ชอบน้ำจากผู้เชี่ยวชาญ เท่ากับ 63.53



b) ค่าความเหมือนของลำดับที่ไม่ชอบน้ำจากเครื่องมือจับคู่โดยอัตโนมัติ เท่ากับ 54.12



รูปที่ 4.7 การจับคู่กลุ่มที่ไม่ชอบน้ำของโปรตีน Interleukin จากผู้เชี่ยวชาญ และจากเครื่องมือจับคู่โดยอัตโนมัติ

จากผลการทดลองตามรูปที่ 4.6 แสดงการจับคู่กลุ่มที่ไม่ชอบน้ำของโปรตีน Rubrerythrin สำหรับ Sequence1 คือลำดับกรดอะมิโนของสิ่งมีชีวิตจำพวกแบคทีเรีย *Desulfovibrio vulgaris* และ Sequence2 คือลำดับกรดอะมิโนของสิ่งมีชีวิตจำพวกแบคทีเรียโบราณ (Archaea หรือ Eubacteria) *Pyrococcus furiosus* (a) คือการจับคู่กลุ่มโดยผู้เชี่ยวชาญ และ (b) คือการจับคู่กลุ่มด้วยเครื่องมือจับคู่กลุ่มโดยอัตโนมัติ พบว่าการจับคู่กลุ่มโดยผู้เชี่ยวชาญ จะมีการจับคู่กลุ่มไม่ครบทุกรูปแบบของกลุ่มที่ไม่ชอบน้ำที่เกิดขึ้น ตัวอักษรตัวพิมพ์เล็กเป็นบริเวณที่ผู้เชี่ยวชาญไม่ใช้ในการจับคู่ ในขณะที่เครื่องมือจับคู่กลุ่มที่ไม่ชอบน้ำมีการจับคู่ทุกรูปแบบกลุ่มที่ไม่ชอบน้ำที่เกิดขึ้นซึ่งมีความละเอียดมากกว่า จึงทำให้เกิดความแตกต่างขึ้น และทำให้การแบ่งกลุ่มที่ไม่ชอบน้ำไม่เหมือนกันอีกด้วย

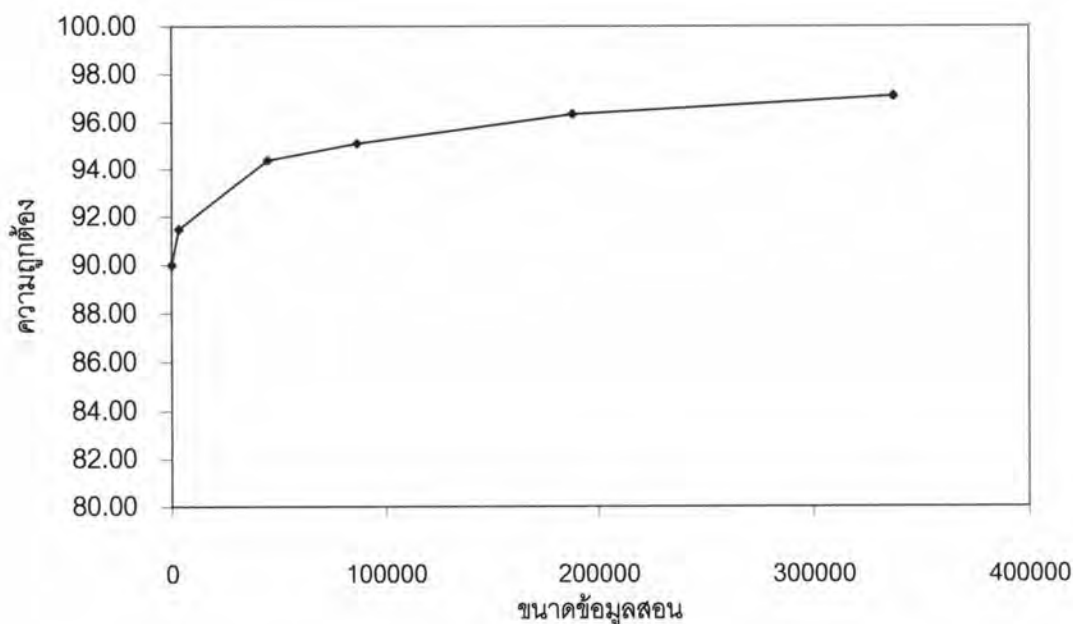
จากผลการทดลองตามรูปที่ 4.7 แสดงการจับคู่กลุ่มที่ไม่ชอบน้ำของโปรตีน Interleukin สำหรับ Sequence1 คือลำดับกรดอะมิโนของ Interleukin-10 ของ Human herpesvirus 5 และ Sequence2 คือลำดับกรดอะมิโนของ Interleukin-19 ของมนุษย์ (*Homo sapiens*) (a) คือการจับคู่กลุ่มโดยผู้เชี่ยวชาญ และ (b) คือการจับคู่กลุ่มด้วยเครื่องมือจับคู่กลุ่มโดยอัตโนมัติ พบว่าขณะที่เครื่องมือจับคู่กลุ่มที่ไม่ชอบน้ำมีการจับคู่กลุ่มใกล้เคียงกับการจับคู่กลุ่มโดยผู้เชี่ยวชาญ

#### 4.3.3 ผลการทดลองเพิ่มขนาดชุดข้อมูลสอนให้มีขนาดใหญ่ขึ้น

ตารางที่ 4.8 ผลการทดลองเพิ่มขนาดชุดข้อมูลสอนให้มีขนาดใหญ่ขึ้นตามลำดับ

ขนาดชุดข้อมูล เปรียบเทียบ	ขนาดชุดข้อมูลทดสอบ	จำนวนที่เลือกถูกต้อง	ความถูกต้อง
90	30	27	90.00%
3448	1032	944	91.47%
44813	1032	974	94.38%
86678	1032	982	95.15%
187991	1032	994	96.32%
336827	1032	1002	97.09%

จากตารางที่ 4.8 สามารถนำไปสร้างกราฟได้ดังรูปที่ 4.8



รูปที่ 4.8 ความถูกต้องเมื่อเพิ่มขนาดชุดข้อมูลสอน

จากรูปที่ 4.8 แสดงให้เห็นว่าการเพิ่มขนาดข้อมูลสอนในช่วงแรกจะสามารถเพิ่มความถูกต้องได้ แต่สำหรับการเพิ่มขนาดซึ่งมีขนาดใหญ่หลายๆ ความถูกต้องจะเพิ่มขึ้นเล็กน้อย ซึ่งความถูกต้องมากที่สุดที่สามารถทำได้คือ 97.09 เปอร์เซ็นต์ ดังนั้นผู้วิจัยจึงได้เลือกชุดข้อมูลที่ 3 ที่มีเปอร์เซ็นต์ความถูกต้องสูงถึง 94.38 เปอร์เซ็นต์ สำหรับใช้ในการทดลองในขั้นต่อไป

#### 4.3.4 ผลการทดลองเปรียบเทียบความถูกต้องของการเลือกลำดับที่เหมือนกันที่สุด

ตารางที่ 4.9 ผลการทดลองเปรียบเทียบความถูกต้องของการเลือกลำดับที่เหมือนกันที่สุดระหว่างวิธีการจับคู่แบบ 1 มิติ กับวิธีการจับคู่แบบ 2 มิติโดยอัตโนมัติ

วิธีการจับคู่	ตารางคะแนน	ค่าหักลบช่องว่าง	เปอร์เซ็นต์ความถูกต้อง
แบบ 1 มิติ	BLOSUM62	Affine Gap Penalty	61.82 %
แบบ 2 มิติโดยอัตโนมัติ	BLOSUM62	Affine Gap Penalty	94.38 %
	BLOSUM62	ค่าหักลบช่องว่างใหม่	90.67 %

จากตารางที่ 4.9 แสดงให้เห็นว่าวิธีการจับคู่แบบ 1 มิติโดยใช้ตารางคะแนน BLOSUM62 และ Affine Gap Penalty สามารถเลือกลำดับที่เหมือนกันที่สุดที่ถูกต้องได้เพียง 61.82 เปอร์เซ็นต์ ในขณะที่วิธีการจับคู่กลุ่มที่ไม่ชอบน้ำแบบ 2 มิติโดยอัตโนมัติที่ใช้ตารางคะแนน BLOSUM62 และ Affine Gap Penalty สามารถเลือกลำดับที่ถูกต้อง 94.38 เปอร์เซ็นต์ นอกจากนี้เมื่อใช้ค่าหักลบช่องว่างแบบใหม่จะให้ความถูกต้อง 90.67 เปอร์เซ็นต์ ดังนั้นวิธีการเลือกลำดับที่เหมือนกันมากที่สุดโดยการใช้วิธีการจับคู่กลุ่มแบบ 2 มิติโดยอัตโนมัติจึงเป็นวิธีที่ดีกว่าวิธีเดิม