บทที่ 2

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

งานวิจัยเกี่ยวกับเครื่องมือที่ใช้จับคู่กลุ่มที่ไม่ชอบน้ำแบบ 2 มิติโดยอัตโนมัติ มีการใช้ทฤษฎีและงานวิจัยต่างๆ ซึ่งเป็นพื้นฐานต่อการทำการวิจัย โดยมีรายละเอียดดังต่อไปนี้

2.1 โปรตีน (Protein)

โปรตีนเป็นสารชีวโมเลกุลที่พบในสิ่งมีชีวิตและมีความสำคัญมาก เนื่องจากโปรตีนเป็นส่วนประกอบหลักในสิ่งมีชีวิตเกือบทุกชนิด ซึ่งสามารถพบได้ทั่วไป เช่น ในโครงสร้างเนื้อเยื่อ อวัยวะ ผิวหนัง เล็บ ขน ผม รวมถึงสารที่อยู่ภายในเซลล์ จากที่ทราบกันแล้วว่าโปรตีนประกอบด้วยหน่วยย่อยของกรดอะมิโนที่แตกต่างกันเชื่อมต่อกันเป็นสายยาวด้วยพันธะเปปไทด์ (Peptide Bond) [17] โดยหน้าที่การทำงานของโปรตีนจะขึ้นกับความแตกต่างของกรดอะมิโน เช่น ฮีโมโกลบินซึ่งเป็นโปรตีนชนิดหนึ่งที่อยู่ในเซลล์เม็ดเลือดแดงทำหน้าที่ส่งออกซิเจนไปเลี้ยงส่วนต่างๆ ของร่างกาย มีจำนวนกรดอะมิโนประมาณ 570 ตัว

เราสามารถแบ่งกลุ่มโปรตีนตามโครงสร้างได้ดังนี้ [17]

2.1.1 โปรตีนกลุ่มโกลบิน (Globin Family)

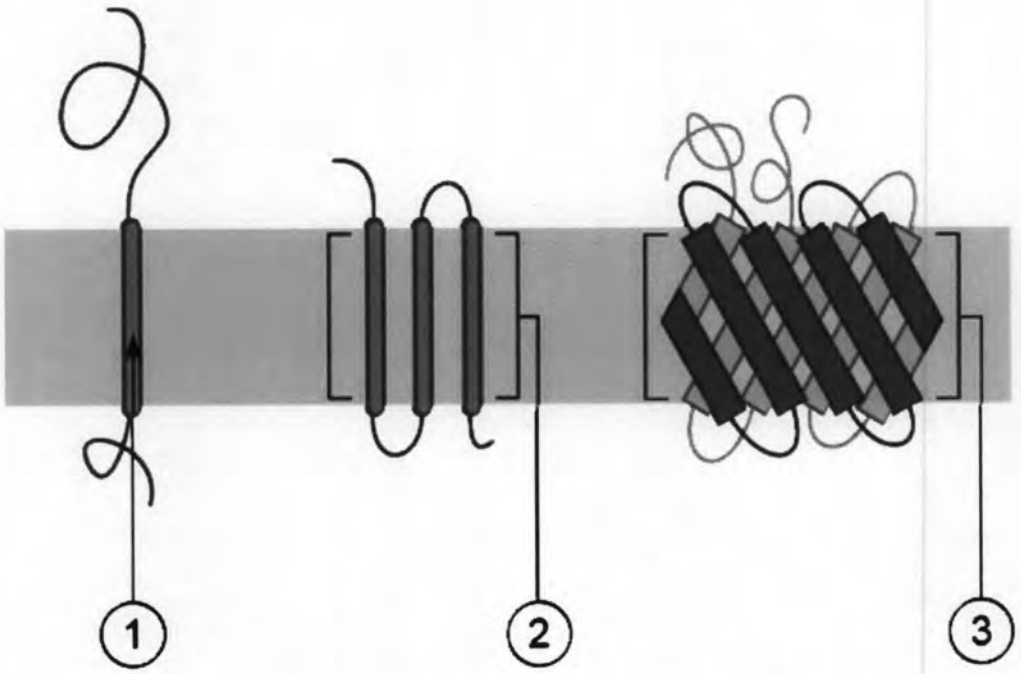
โปรตีนกลุ่มนี้มีลักษณะโครงสร้างที่รวมกลุ่มกันเป็นก้อน พบอยู่ทั่วไปอยู่ในสภาพแวดล้อมที่ชอบน้ำ (Hydrophilic) เช่น ฮีโมโกลบิน (Hemoglobin) หรือ มีโอโกลบิน (Myoglobin) ฮีโมโกลบิน เป็นโปรตีนที่ทำหน้าที่ขนส่งออกซิเจนไปเลี้ยงเซลล์ต่างๆ ในร่างกายและอยู่ภายนอกเซลล์ มิโอโกลบิน เป็นโปรตีนที่ทำหน้าที่รับออกซิเจนจากฮีโมโกลบินเพื่อเก็บสะสมไว้ภายในเซลล์

2.1.2 **โปรตีนที่รวมตัวกับเยื่อหุ้มเซลล์** (Integral Membrane Proteins)

โปรตีนชนิดนี้เป็นโปรตีนที่อยู่ติดกับเยื่อหุ้มเซลล์ และสามารถสกัดแยกจากเยื่อหุ้มเซลล์ได้โดยใช้น้ำยาทำความสะอาด (Detergent) สารละลายไม่มีขั้ว (Non-Polar Solvents) หรือสารที่ทำให้ผิดไปจากธรรมชาติ (Denaturing Agents) โปรตีนกลุ่มนี้มีอยู่ด้วยกัน 2 ชนิด คือ

2.1.2.1 **โปรตีนที่อยู่ภายในเยื่อหุ้มเซลล์** (Transmembrane Protein)

โปรตีนที่อยู่ภายในเยื่อหุ้มเซลล์จะมีส่วนที่ไม่ชอบน้ำหันออกมาด้านนอกของก้อนโปรตีน เพื่อให้โปรตีนสามารถอยู่ภายในเยื่อหุ้มเซลล์ที่มีสภาพที่ไม่ชอบน้ำได้ ดังรูปที่ 2.1

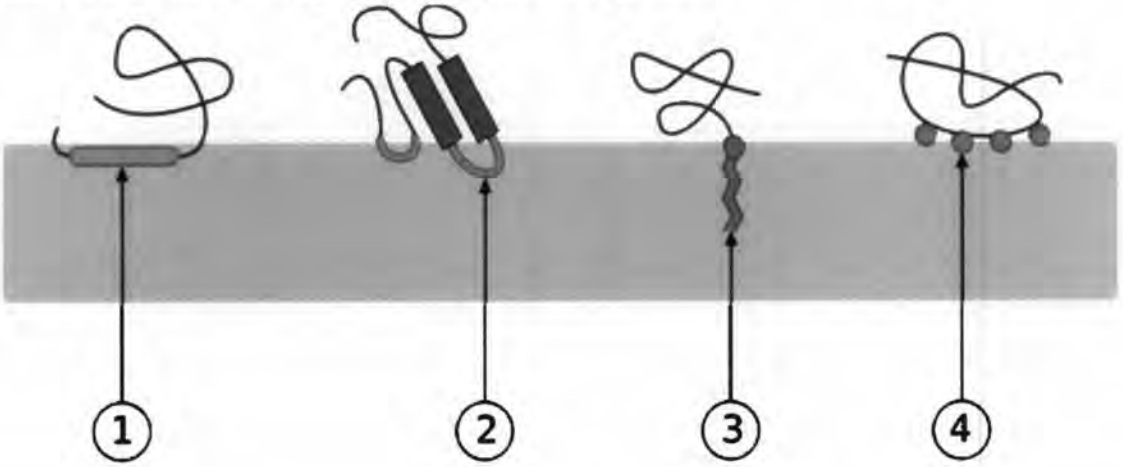


รูปที่ 2.1 โครงสร้างและตำแหน่งของโปรตีนที่อยู่ภายในเยื่อหุ้มเซลล์
(ที่มา : Foobar [18])

จากรูปที่ 2.1 (1,2) คือการอยู่ภายในเยื้อหุ้มเซลล์ของโปรตีนที่มีโครงสร้างแบบเกลียวแอลฟา โดยจะมีแรงที่กระทำกันกับบริเวณที่ไม่ชอบน้ำ นอกจากนี้ยังมีโปรตีนที่มีโครงสร้างแบบบีตาชีทรวมตัวกันเป็นรูปเป็นรูปถัง (3) ซึ่งภายนอกถังจะเป็นกรดอะมิโนที่ไม่ชอบน้ำ แต่สำหรับภายในถังจะเป้นกรดอะมิโนอื่นๆ

2.1.2.2 **โปรตีนที่เกาะอยู่กับเยื่อหุ้มเซลล์** (Peripheral Membrane Protein)

โปรตีนชนิดนี้จะใช้เพียงส่วนหนึ่งของก้อนโปรตีนเพื่อยึดเกาะกับเยื่อหุ้มเซลล์ไว้ โดยที่ส่วนทำงานจะยังคงอยู่ภายนอกเยื่อหุ้มเซลล์ ดังรูปที่ 2.2



รูปที่ 2.2 โครงสร้างและตำแหน่งของโปรตีนที่เกาะติดกับเยื่อหุ้มเซลล์
(ที่มา : Foobar [19])

จากรูปที่ 2.2 (1) คือโปรตีนที่มีโครงสร้างแบบ Amphipathic α-Helic เป็นลักษณะแบบเกลียวแอลฟา ซึ่งสามารถแบ่งเกลียวออกเป็น 2 ส่วนตามความยาวโดยมีส่วนหนึ่งเป็นส่วนที่ไม่ชอบน้ำ และอีกส่วนเป็นส่วนที่ชอบน้ำ และใช้สำหรับยึดเกาะกับเยื่อหุ้มเซลล์ ในรูปแบบที่ (2) คือโปรตีนที่มีกรดอะมิโนที่ไม่ชอบน้ำอยู่บริเวณลูป และใช้ส่วนลูปเพื่อยึกเกาะกับเยื่อหุ้มเซลล์ นอกจากนี้ยังมีโปรตีนที่ไม่ได้ใช้สภาพที่ไม่ชอบน้ำสำหรับยึดเกาะเยื่อหุ้มเซลล์ 2 ชนิด คือ โปรตีนที่ใช้การเชื่อมพันธะของกรดอะมิโนเข้ากับโครงสร้างเยื่อหุ้มเซลล์ (3) และชนิดสุดท้ายคือ โปรตีนที่ใช้การสร้างพันธะทางไฟฟ้ากับเยื่อหุ้มเซลล์ (4)

2.1.3 **โปรตีนเส้นใย** (Fibrous Proteins)

โปรตีนเส้นใยเป็นโปรตีนที่มีลักษณะเป็นท่อนเหมือนกับสายไฟ โปรตีนทำหน้าที่เป็นส่วนยึดเกาะระหว่างโครงสร้าง (Inert Structural Protein) หรือโปรตีนเก็บสะสม (Storage Protein) โครงสร้างของโปรตีนเกิดขึ้นจากลำดับกรดอะมิโนที่เรียงตัวกันเป็นเกลียวแอลฟา โดยแต่ละเกลียวจะต้องมีจำนวนกรดอะมิโนที่จำกัดและคงที่ ซึ่งทำให้เกิดโครงสร้างดังรูปที่ 2.3 เป็นโครงสร้างของคอลลาเจนที่เกิดจากเกลียวแอลฟา 3 เส้น ประกอบเข้าด้วยกัน

รูปที่ 2.3 แบบจำลองโครงสร้างของคอลลาเจน
(ที่มา : Rulez [20])

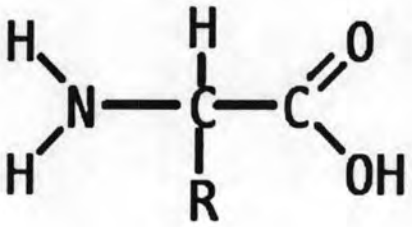
2.2 วิวัฒนาการของสิ่งมีชีวิต

โปรตีนแต่ละชนิดจะมีลำดับการเรียงตัวของลำดับกรดอะมิโนที่แตกต่างกัน โดยที่เซลล์จะเก็บข้อมูลลำดับของกรดอะมิโนที่ใช้สำหรับการสร้างโปรตีนไว้บนลำดับนิวคลีโอไทด์ที่อยู่ในใจกลางของเซลล์ที่เรียกว่านิวเคลียส เมื่อเซลล์ต้องการแบ่งตัวเพื่อการสืบพันธุ์จะมีกระบวนการไขว้เปลี่ยน (Crossing Over) ระหว่างโครโมโซมในนิวเคลียส ทำให้เกิดการเปลี่ยนแปลงลำดับนิวคลีโอไทด์ ซึ่งลำดับกรดอะมิโนจะเกิดการเปลี่ยนแปลงด้วย ถ้าการเปลี่ยนแปลงที่เกิดขึ้นส่งผลกระทบทำให้หน้าที่การทำงานของโปรตีนเสียสภาพไปอาจทำให้เซลล์ไม่สามารถอยู่รอดได้ ในทางกลับกัน ถ้าการเปลี่ยนแปลงทำให้กรดอะมิโนที่เปลี่ยนไปยังคงสภาพโปรตีนไว้ได้เช่นเดิม ทำให้เซลล์สามารถอยู่รอดต่อไปได้ การเปลี่ยนแปลงลำดับกรดอะมิโนอาจทำให้เกิดลักษณะที่ไม่ปรากฏในรุ่นพ่อแม่ และลักษณะใหม่ที่เกิดขึ้นนี้ก็สามารถเปลี่ยนแปลงได้อีกในรุ่นต่อไป ลักษณะเช่นนี้เรียกว่า วิวัฒนาการของสิ่งมีชีวิต

จากที่กล่าวข้างต้น กรดอะมิโนเป็นสิ่งที่ทำให้เกิดหน้าที่การทำงานที่แตกต่างกัน ดังนั้น หัวข้อต่อไปจะกล่าวถึงลักษณะของกรดอะมิโนที่ทำให้เกิดความแตกต่างเหล่านี้

2.3 กรดอะมิโน (Amino Acid)

กรดอะมิโนมีอยู่ด้วยกันทั้งหมด 20 ชนิด ซึ่งประกอบด้วยธาตุหลัก 4 ชนิด คือ คาร์บอน (C) ไฮโดรเจน (H) ออกซิเจน (O) และไนโตรเจน (N) นอกจากนี้อาจมีธาตุอื่นรวมอยู่ด้วย เช่น ฟอสฟอรัส (P) หรือกำมะถัน (S) ซึ่งมีโครงสร้างทั่วไปดังรูปที่ 2.4



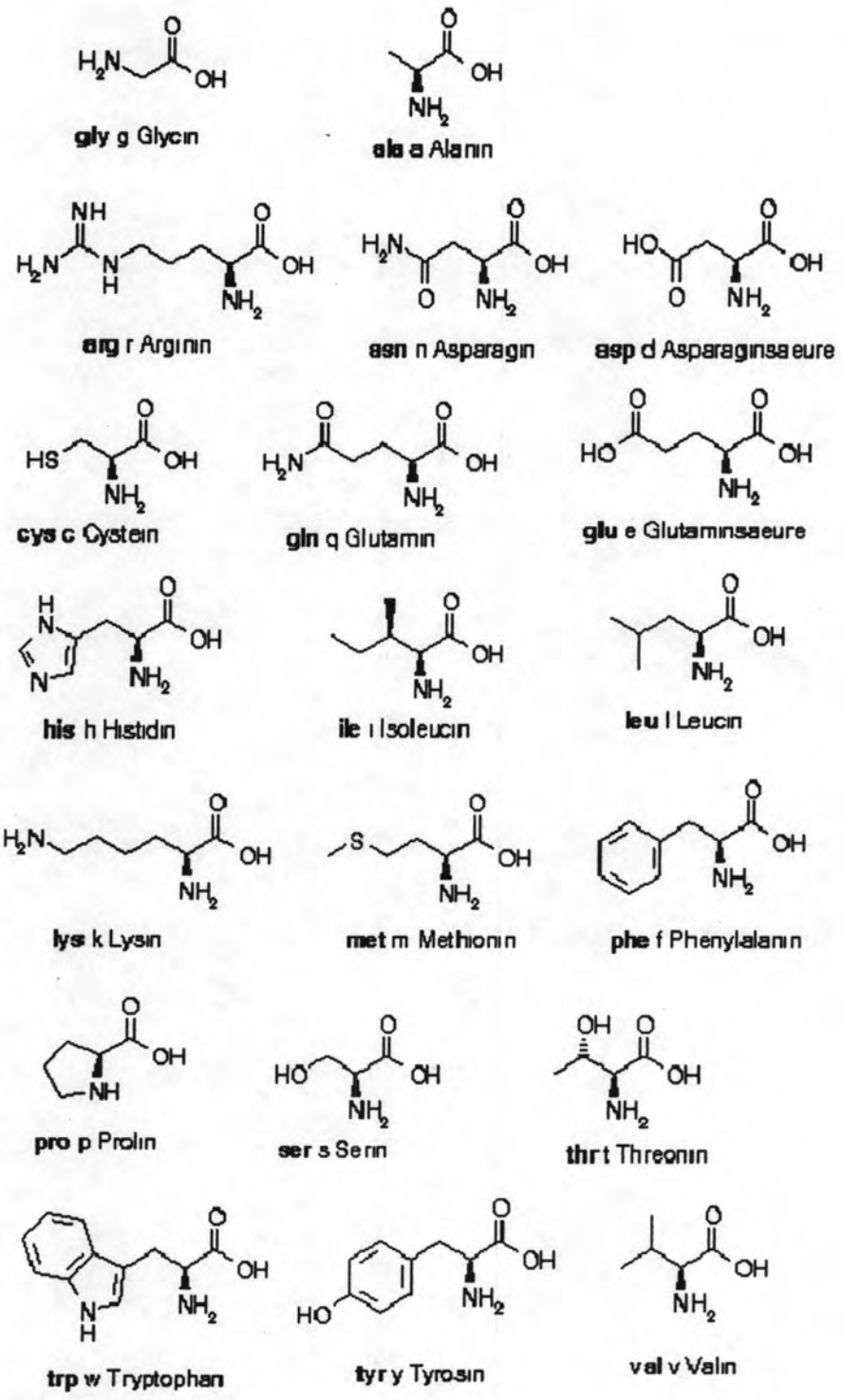
รูปที่ 2.4 โครงสร้างของกรดอะมิโน หมู่ทางขวามือคือ หมู่คาร์บอกซิล (Carboxyl Group, -COOH) หมู่ทางซ้ายมือคือ หมู่อะมิโน (Amino Group, $-NH_2$) และ R คือหมู่อัลคีน (Alkyl Group) เป็นหมู่ข้างเคียงของกรดอะมิโน

(ที่มา : Liebecq [21])

หมู่ที่ทำให้คุณสมบัติของกรดอะมิโนแตกต่างกันคือหมู่อัลคีน ซึ่งทำให้เกิดคุณสมบัติที่ชอบน้ำ (Hydrophilic) และคุณสมบัติที่ไม่ชอบน้ำ (Hydrophobic) โดยลักษณะโครงสร้างของกรดอะมิโนและความแตกต่างกันของหมู่แอลคีนได้แสดงไว้ดังรูปที่ 2.5 และการแบ่งกลุ่มกรดอะมิโนตามคุณสมบัติความชอบน้ำสามารถแบ่งได้ตามตารางที่ 2.1

รูปที่ 2.5 แสดงให้เห็นว่า โพรลีน (Proline) เป็นกรดอะมิโนที่มีโครงสร้างหมู่อัลคีนมีการเชื่อมต่อกับหมู่อะมิโนของตัวเองทำให้เกิดเป็นวงแหวน มีลักษณะเป็น 3 เหลี่ยมแบบราบทำให้เกิดมุมที่มีลักษณะเป็นมุมหัก (Dihedral Angle) ขนาด 75 องศา และไม่สามารถหมุนรอบตัวเอง หรือเปลี่ยนแปลงขนาดของมุมได้ ดังนั้นตรงตำแหน่งที่มีโพรลีนอยู่จะทำให้เกิดการหักงอแบบ U-turn

คุณสมบัติความชอบน้ำของกรดอะมิโนทั้งหมด 20 ชนิด แสดงดังตารางที่ 2.1 สามารถแบ่งกลุ่มกรดอะมิโนตามคุณสมบัติความชอบน้ำได้เป็น 3 กลุ่ม ได้แก่ กรดอะมิโนที่ไม่ชอบน้ำ กรดอะมิโนที่ชอบน้ำ และกรดอะมิโนที่มีคุณสมบัติเป็นกลาง งานวิจัยนี้สนใจเฉพาะกรดอะมิโนที่มีความสำคัญสำหรับการจัดโครงสร้างของโปรตีนคือกรดอะมิโนที่ไม่ชอบน้ำซึ่งมีอยู่เพียง 7 ชนิด คือ เวลีน (Valine) ไอโซลิวซีน (Isoleucine) ลิวซีน (Leucine) ฟีนิลอะลานีน (Phenylalanine) ทริปโตเฟน (Tryptophan) เมไธโอนีน (Methionine) และ ไทโรซีน (Tyrosine)



รูปที่ 2.5 โครงสร้างของกรดอะมิโนทั้ง 20 ชนิด

(ที่มา : Heinau V. และ Kirste B. [22]

ตารางที่ 2.1 การแบ่งกลุ่มกรดอะมิโนแต่ละชนิดตามคุณสมบัติความชอบน้ำ (ที่มา : Lesser และคณะ [23])

คุณสมบัติความชอบน้ำ	ชนิดกรดอะมิโน	ตัวย่อ
ไม่ชอบน้ำ (Hydrophobic)	เวลีน (Valine)	V
	ไอโซลิวซีน (Isoleucine)	I
	ลิวซีน (Leucine)	L
	ฟีนิลอะลานีน (Phenylalanine)	F
	ทริปโตเฟน (Tryptophan)	W
	เมไธโอนีน (Methionine)	M
	ไทโรซีน (Tyrosine)	Y
ชอบน้ำ (Hydrophilic)	กรดแอสปาร์ติก (Aspartic acid)	D
	กรดกลูตามิก (Glutamic acid)	E
	แอสพาราจีน (Asparagine)	N
	กลูตามีน (Glutamine)	Q
	อาร์จินีน (Arginine)	R
	ไกลซีน (Glycine)	G
	ไลซีน (lysine)	K
	โพรลีน (Proline)	P
	ซีรีน (Serine)	S
	ธรีโอนีน (Threonine)	T
เป็นกลาง (Neutral)	ฮีสติดีน (Histidine)	H
	อะลานีน (Alanine)	A
	ซีสเตอีน (Cysteine)	C

2.4 **โครงสร้างของโปรตีน** (Protein Structure)

โปรตีนมีโครงสร้างเป็นแบบ 3 มิติ โดยโปรตีนส่วนมากจะมีโครงสร้างที่ซับซ้อน ดังนั้นจึงได้มีการแบ่งระดับของโครงสร้างของโปรตีนออกเป็น 4 ระดับ คือ

2.4.1 **โครงสร้างระดับปฐมภูมิ** (Primary Structure)

โครงสร้างระดับปฐมภูมิคือระดับที่มีลักษณะเป็นลำดับกรดอะมิโน ซึ่งไม่มีความสัมพันธ์ต่อหน้าที่การทำงานของโปรตีนโดยตรง แต่ในระดับนี้เองที่เก็บข้อมูลต่างๆของโปรตีนแต่ละชนิดไว้ในลำดับการเรียงตัวของกรดอะมิโน

2.4.2 **โครงสร้างระดับทุติยภูมิ** (Secondary Structure)

โครงสร้างระดับทุติยภูมิเป็นโครงสร้างที่เกิดจากกรดอะมิโนมีแรงกระทำระหว่างกันด้วยพันธะไฮโดรเจนทำให้เกิดการขดตัวของลำดับกรดอะมิโน เกิดเป็นโครงสร้าง 2 รูปแบบคือ แบบเกลียวแอลฟา (Alpha-Helix) และแบบบีตาชีท (Beta-Sheet) ดังแสดงในรูปที่ 2.6 โครงสร้างที่เกิดขึ้นมีความสัมพันธ์กับหน้าที่การทำงานของโปรตีนด้วย

รูปที่ 2.6 โครงสร้างระดับตติยภูมิ แสดงให้เห็นถึงลักษณะที่ม้วนเป็นเกลียวอยู่ด้านนอก เรียกว่าแบบเกลียวแอลฟา (Alpha-Helix) และลักษณะแผ่นที่พาดผ่านด้านใน เรียกว่าแบบบีตาชีท (Beta-Sheet)
(ที่มา : Whitford [17])

การเกิดโครงสร้างในระดับทุติยภูมิเกี่ยวข้องกับชนิดและตำแหน่งของกรดอะมิโนที่อยู่ในลำดับกรดอะมิโน กรดอะมิโนที่มีความสำคัญกับการเกิดโครงสร้างคือกรดอะมิโนที่ไม่ชอบน้ำทั้ง 7 ชนิดได้แก่ เวลีน (V) ไอโซลิวซีน (I) ลิวซีน (L) ฟีนิลอะลานีน (F) ทริปโตเฟน (W)

เมไธโอนีน (M) และไทโรซีน (Y) การเกิดโครงสร้างแบบเกลียวแอลฟาเกิดจากกรดอะมิโนที่ไม่ชอบน้ำอยู่ในตำแหน่งที่สามารถบิดเป็นเกลียวและเกิดพันธะไฮโดรเจนระหว่างออกซิเจน (O) ในหมู่ C=O ของกรดอะมิโนโมเลกุลหนึ่งกับไฮโดรเจน (H) ของหมู่ N-H ของอะมิโนอีกโมเลกุลหนึ่งที่อยู่ถัดไป 4 ตำแหน่ง และเกิดลักษณะเช่นนี้ต่อเนื่องกันจนกลายเป็นลักษณะแบบเกลียวแอลฟา โครงสร้างอีกแบบหนึ่งคือโครงสร้างแบบบีตาชีท ซึ่งเกิดจากกรดอะมิโนที่ไม่ชอบน้ำที่อยู่ในลำดับกรดอะมิโนแบบเรียงขนานกันสองสาย และมีพันธะไฮโดรเจนระหว่างกรดอะมิโนจากสายที่ 1 กับกรดอะมิโนจากสายที่ 2 โดยการเกิดพันธะจะเป็นแบบสลับตัวเว้นตัว เมื่อโปรตีนเกิดโครงสร้างในระดับทุติยภูมิ กลุ่มของโปรตีนที่มีการจัดตัวขึ้นจะเสมือนมีการแบ่งส่วนของกรดอะมิโนที่ชอบน้ำและกรดอะมิโนที่ไม่ชอบน้ำออกจากกัน เมื่อโปรตีนมีเกิดโครงสร้างในระดับตติยภูมิ จะหันส่วนต่างๆ ที่มีลักษณะเดียวกันเข้าหากัน เช่นหันส่วนที่ชอบน้ำเข้าด้วยกัน หรือหันส่วนที่ไม่ชอบน้ำเข้าด้วยกัน ตามแต่ลักษณะของสภาพแวดล้อมภายนอก

2.4.3 **โครงสร้างระดับตติยภูมิ** (Tertiary Structure)

โครงสร้างระดับตติยภูมิเป็นโครงสร้างที่เกิดจากกรดอะมิโนมีแรงกระทำระหว่างกันกับสภาพแวดล้อมภายนอก กล่าวคือ สภาพที่ชอบน้ำ (Hydrophilic) และ สภาพที่ไม่ชอบน้ำ (Hydrophobic) จากการกระทำระหว่างกันดังกล่าวนี้เองทำให้ลำดับกรดอะมิโนเกิดการจัดตัวรวมเข้าเป็นกลุ่มเป็นก้อนที่มีขนาดใหญ่ขึ้น

2.4.4 **โครงสร้างระดับจตุรภูมิ** (Quaternary Structure)

โครงสร้างระดับสุดท้ายคือโครงสร้างระดับจตุรภูมิเป็นโครงสร้างที่เกิดจากการรวมตัวกันของกลุ่มโปรตีนหลายกลุ่มก้อนเข้าด้วยกัน เพื่อให้ได้โปรตีนที่มีความอยู่ตัวและสามารถทำงานได้

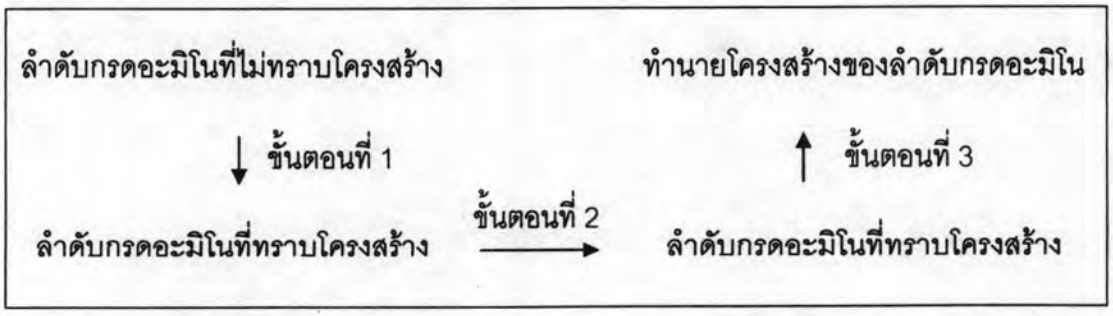
แม้ว่าความแตกต่างกันของลำดับกรดอะมิโนทำให้การเกิดโครงสร้างของโปรตีนนั้นมีหน้าที่การทำงานที่แตกต่างกันไป แต่โปรตีนที่ทำหน้าที่เหมือนกันจะมีโครงสร้างที่คล้ายกัน ดังนั้นจึงเป็นที่มาของการใช้โครงสร้างของโปรตีนในการศึกษาความเหมือนกันของโปรตีน แต่สำหรับการจำลองโครงสร้างในระดับ 3 มิติ เช่น โครงสร้างตติยภูมิและโครงสร้างจตุรภูมิซึ่งเป็นโครงสร้างที่มีความซับซ้อน และยังไม่สามารถสร้างแบบจำลองโครงสร้างที่ถูกต้องได้อย่างสมบูรณ์ ซึ่งอาจมีความผิดพลาดขึ้นในการวางตำแหน่งของโครงสร้าง 3 มิติ ประกอบกับคอมพิวเตอร์ในปัจจุบันยังมีความสามารถไม่เพียงพอกับการเปรียบเทียบหาลำดับกรดอะมิโนที่เหมือนที่สุดโดยใช้

โครงสร้าง 3 มิติ ซึ่งต้องมีการเปรียบเทียบข้อมูลจำนวนมาก ผู้วิจัยจึงสนใจโครงสร้างในระดับทุติยภูมิที่มีข้อมูลของโครงสร้างโปรตีนเพียงพอสำหรับการหาความเหมือนของหน้าที่การทำงานของโปรตีน ในโครงสร้างระดับทุติยภูมินี้เองที่กรดอะมิโนที่ไม่ชอบน้ำมีความสำคัญในการจัดโครงสร้างและสามารถเชื่อมโยงไปถึงหน้าที่การทำงานของโปรตีนได้ ซึ่งจะกล่าวถึงในหัวข้อการศึกษาโปรตีนต่อไป

2.5 การศึกษาโปรตีน

โปรตีนเป็นสารชีวโมเลกุลที่มีความสำคัญต่อชีวิต ดังนั้นจึงมีการศึกษาโปรตีนกันอย่างกว้างขวาง ไม่ว่าจะเป็นการศึกษาการทำงานของโปรตีน ความไวต่อปฏิกิริยาของเอนไซม์ สภาพแวดล้อมที่เหมาะกับการทำงานของเอนไซม์ ทั้งนี้ในการศึกษาหน้าที่การทำงานของโปรตีนโดยทั่วไป จะนำโปรตีนมาศึกษาจนทราบหน้าที่การทำงานและคุณสมบัติก่อน แล้วจึงนำไปหาลำดับกรดอะมิโน แต่ในปัจจุบันมีความก้าวหน้าทางด้านชีวโมเลกุลเป็นอย่างมากจึงสามารถทราบลำดับของกรดอะมิโนได้ง่ายขึ้น ประกอบกับการมีฐานข้อมูลที่เติบโตขึ้นอย่างรวดเร็วและมีลำดับของกรดอะมิโนจากสิ่งมีชีวิตมากมาย ทำให้ผู้วิจัยหันมาสนใจศึกษาโปรตีนโดยการศึกษาจากลำดับกรดอะมิโน โดยมีขั้นตอนในการศึกษา 3 ขั้นตอน ดังแสดงในรูปที่ 2.7

1. นำลำดับกรดอะมิโนใหม่ไปเปรียบเทียบหาลำดับกรดอะมิโนที่ทราบโครงสร้างที่เก็บไว้ในฐานข้อมูล ด้วยเครื่องมือที่ใช้เปรียบเทียบ
2. จากการเปรียบเทียบในขั้นตอนที่ 1 จะได้ลำดับกรดอะมิโนที่ทราบโครงสร้าง ซึ่งจะเป็นลำดับกรดอะมิโนที่มีค่าความเหมือนมากที่สุด
3. ใช้ลำดับกรดอะมิโนจากขั้นตอนที่ 2 เพื่อทำนายโครงสร้างของลำดับกรดอะมิโนใหม่



รูปที่ 2.7 ขั้นตอนการศึกษาโครงสร้างและหน้าที่การทำงานของโปรตีน
(ที่มา : ปรับปรุงจาก Ruddle [24])

2.6 การศึกษาโปรตีนที่มีหน้าที่เหมือนกัน

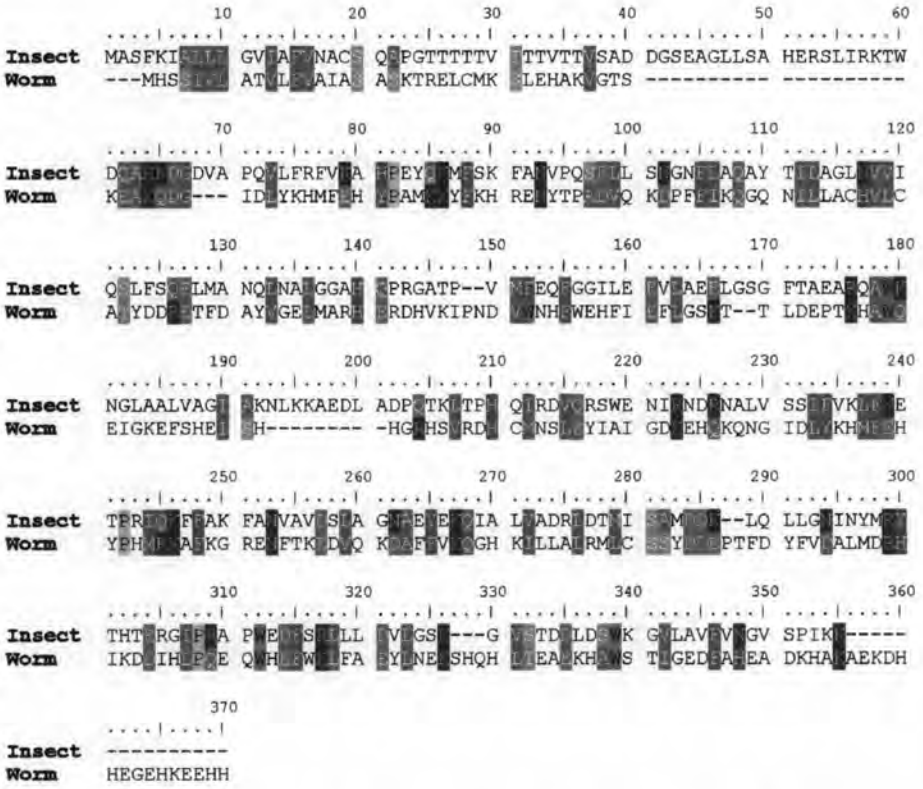
โปรตีนที่มีหน้าที่การทำงานเหมือนกัน (Protein Homology) คือโปรตีนที่ทำหน้าที่เหมือนกันแม้ว่าจะอยู่ในสิ่งมีชีวิตที่ต่างชนิดกัน เช่น ฮีโมโกลบินในสิ่งมีชีวิตใดก็ตามก็จะมีหน้าที่ขนส่งออกซิเจน ถึงแม้ว่าฮีโมโกลบินในมนุษย์กับฮีโมโกลบินในหนอนตัวกลมอาจไม่มีกรดอะมิโนที่เหมือนกันเลย ฮีโมโกลบินก็ยังคงทำหน้าที่ขนส่งออกซิเจน ซึ่งเป็นผลจากการวิวัฒนาการทางธรรมชาติที่ทำให้เกิดการเปลี่ยนแปลงลำดับกรดอะมิโน อย่างไรก็ตาม การหาโปรตีนที่มีหน้าที่เหมือนกันต้องอาศัยเครื่องมือที่ใช้จับคู่ลำดับกรดอะมิโน

2.7 เครื่องมือจับคู่ (Alignment Tools)

ในงานวิจัยด้านชีววิทยามีการนำเครื่องมือจับคู่มาประยุกต์ใช้เพื่อเปรียบเทียบความเหมือนของลำดับ ไม่ว่าจะเป็นลำดับนิวคลีโอไทด์หรือลำดับกรดอะมิโน โดยอาศัยวิธีการจับคู่ที่หลากหลาย เช่น วิธีกำหนดการพลวัต (Dynamic Programming Methods) [25] BLAST [26] แบบจำลองฮิดเดนมาร์คอฟ (Hidden Markov Models) [27] หรือข่ายงานประสาทเทียม (Artificial Neural Network) [28] เป็นต้น

2.7.1 วิธีกำหนดการพลวัต (Dynamic Programming Methods)

วิธีกำหนดการพลวัตเป็นวิธีการที่มีใช้กันมานาน หลักการคือการสร้างเมทริกซ์ที่เป็นตัวแทนของลำดับกรดอะมิโนแต่ละตัว โดยนำลำดับกรดอะมิโนแต่ละสายมาเขียนตามแถวและตามหลักของเมทริกซ์ จากนั้นเริ่มหาตัวที่เหมือนกันไปจนได้ลักษณะของการจับคู่ที่ดีที่สุดจากตารางเมทริกซ์ ในท้ายที่สุดเครื่องมือจับคู่จะสามารถติดตามหาเส้นทางที่ทำให้เกิดคะแนนสูงสุดและสร้างผลการจับคู่ของลำดับกรดอะมิโน ดังรูปที่ 2.8 ที่เป็นผลจากการใช้วิธีกำหนดการพลวัตในการจับคู่ลำดับกรดอะมิโน 2 สายในแบบ 1 มิติ



รูปที่ 2.8 ผลของการจับคู่ลำดับกรดอะมิโนด้วยวิธีกำหนดการพลวัตแบบ 1 มิติ

จากรูปที่ 2.8 สายบนเป็นลำดับกรดอะมิโนของฮีโมโกลบินที่อยู่ใน *Daphnia magna* ซึ่งเป็นสิ่งมีชีวิตในกลุ่มของแมลง และสายล่างเป็นลำดับกรดอะมิโนของฮีโมโกลบินที่อยู่ใน *Pseudoterranova decipiens* ซึ่งเป็นสิ่งมีชีวิตจำพวกหนอน และสัญลักษณ์ "-" แสดงช่องว่าง (Gap) จากผลการจับคู่กรดอะมิโนได้แสดงตำแหน่งกรดอะมิโนที่คล้ายกันโดยเน้นเป็นแถบสีไว้ จะเห็นได้ว่า ลำดับกรดอะมิโนทั้งสองลำดับมีความแตกต่างกันมาก และในกรดอะมิโนลำดับที่ 271-280 มีการจับคู่กรดอะมิโนที่ไม่ถูกต้อง ซึ่งที่ถูกต้องควรเลื่อนลำดับกรดอะมิโนสายล่างไปทางซ้ายอีก 2 ตำแหน่ง ซึ่งจะทำให้โครงสร้างที่เหมือนกันเข้าคู่กัน (กรดอะมิโนที่ไม่ชอบน้ำคู่กับกรดอะมิโนที่ไม่ชอบน้ำ) และนอกจากนั้น ผลที่ได้จากการจับคู่ทั้งที่เป็นลำดับกรดอะมิโมของฮีโมโกลบินเหมือนกัน ซึ่งควรมีค่าความเหมือนของลำดับที่สูง แต่ผลที่ได้กลับมีค่าความเหมือนของลำดับ (Sequence Identity) เพียง 19 เปอร์เซ็นต์ ดังนั้นการจับคู่ลำดับกรดอะมิโนด้วยวิธีกำหนดการพลวัตแบบ 1 มิติ จึงใช้ไม่ได้ผลกับการหาความเหมือนของโปรตีนทำหน้าที่เดียวกันแต่อยู่ในสิ่งมีชีวิตที่แตกต่างกัน

2.7.2 วิธีการประมาณการจับคู่อักขระ (Approximate string matching)

งานวิจัยที่เกี่ยวข้องกับวิธีการนี้ ส่วนใหญ่เป็นงานวิจัยเกี่ยวกับการเพิ่มประสิทธิภาพให้กับโปรแกรม BLAST (Basic Local Alignment Search Tool) [26] ซึ่งเป็นโปรแกรมที่นิยมใช้กันมาก โปรแกรมใช้วิธีการทางสถิติและความรู้ในการค้นหาข้อมูลในฐานข้อมูล โดยมีการพัฒนามาตั้งแต่ปี 1980 การเปรียบเทียบลำดับกรดอะมิโนกับชิ้นส่วนกรดอะมิโนขนาดเล็กหลายชิ้นก่อน หลังจากนั้นนำความเหมือนที่ได้จากชิ้นส่วนกรดอะมิโนขนาดเล็กมาวิเคราะห์ทางสถิติ ตัวอย่างชิ้นส่วนกรดอะมิโนขนาดเล็กแสดงดังตารางที่ 2.2 จากนั้นใช้ชิ้นส่วนกรดอะมิโนขนาดเล็กเลือกลำดับกรดอะมิโนที่จะนำมาเปรียบเทียบต่อ สุดท้ายจะได้ลำดับกรดอะมิโนที่มีค่าความเหมือนมากที่สุดเรียงตามลำดับจากมากที่สุดจำนวน 100 ลำดับ โดยรวมแล้วเครื่องมือ BLAST เป็นเครื่องมือที่มีประสิทธิภาพในการเปรียบเทียบหาลำดับกรดอะมิโนที่มาจากสิ่งมีชีวิตชนิดเดียวกันและมีลำดับกรดอะมิโนที่คล้ายกันมาก แต่มีข้อจำกัดค่อนข้างมากสำหรับโปรตีนที่มีหน้าที่การทำงานเหมือนกันแต่อยู่ในสิ่งมีชีวิตที่แตกต่างกัน

ตารางที่ 2.2 ตัวอย่างชิ้นส่วนกรดอะมิโนขนาดเล็กที่นำมาเปรียบเทียบภายใน BLAST

PIR Code of query sequence	Superfamily searched
MYMQW	Globin
KVMSTI	Immunoglobulin
OKBOG	Protein kinase
ITHU	Serpin
KYBOA	Serine protease
CCHU	Cytocrome c
FECF	Feredoxin

2.7.3 แบบจำลองฮิดเดนมาร์คอฟ (Hidden Markov Model)

แบบจำลองฮิดเดนมาร์คอฟเป็นวิธีการใช้การคำนวณทางสถิติเพื่อเลือกจับคู่ข้อมูลที่ให้คะแนนมากที่สุด ซึ่งต่อมามีกลุ่มผู้วิจัยได้ประยุกต์ใช้แบบจำลองฮิดเดนมาร์คอฟกับเครื่องมือที่ใช้จับคู่ความเหมือนอีกไม่น้อย เช่น SAMT02 [29] และ HMMer [30] แต่ทว่าเครื่องมือที่ใช้แบบจำลองฮิดเดนมาร์คอฟเป็นการใช้เพื่อดึงข้อมูลบางอย่างที่อยู่ในลำดับกรดอะมิโน เช่น โปรแกรม HMMer เป็นโปรแกรมที่วิเคราะห์หาโปรตีนโดเมน (Protein Domain) ซึ่งเป็นส่วนหนึ่งของลำดับกรดอะมิโนที่สามารถทำงานได้ โดยอาศัยรูปแบบที่เก็บอยู่ในฐานข้อมูล และโปรแกรมจะเปรียบเทียบความเหมือนของโปรตีนโดเมนจากทั้งสองลำดับ ข้อเสียของวิธีการนี้คือ โปรแกรมต้องวิเคราะห์หาโปรตีนโดเมนก่อน แล้วจึงนำมาเปรียบเทียบกัน ซึ่งในบางครั้งลำดับกรดอะมิโนอาจมีเพียงหนึ่งโดเมนเท่านั้น เมื่อนำไปเปรียบเทียบอาจเกิดความผิดพลาดได้ นอกจากนี้การวิเคราะห์หาโปรตีนโดเมนสามารถระบุได้เฉพาะโปรตีนโดเมนที่มีอยู่ในฐานข้อมูลเท่านั้น เนื่องจากไม่มีข้อมูลรูปแบบโปรตีนใหม่อยู่ในฐานข้อมูล จึงทำให้ไม่สามารถระบุโปรตีนโดเมนชนิดใหม่ได้

2.7.4 การวิเคราะห์กลุ่มที่ไม่ชอบน้ำ (Hydrophobic Cluster Analysis)

การวิเคราะห์กลุ่มที่ไม่ชอบน้ำ (HCA) เป็นเครื่องที่มีประสิทธิภาพ เมื่อใช้สำหรับการศึกษาโครงสร้างระดับทุติยภูมิของโปรตีน [31] การวิเคราะห์สามารถทำได้โดยการสร้างแผนภูมิ 2 มิติของลำดับกรดอะมิโน โดยมีการเรียงลำดับกรดอะมิโนขึ้นใหม่ให้คล้ายกับโครงสร้างระดับทุติยภูมิแบบเกลียวแอลฟา ซึ่งวิธีนี้จะทำให้เห็นรูปแบบโครงสร้างระดับทุติยภูมิที่ซ่อนอยู่ในลำดับกรดอะมิโนได้ง่ายขึ้น ในปัจจุบันโปรแกรมที่ใช้การวิเคราะห์กลุ่มที่ไม่ชอบน้ำมีอยู่ไม่มาก ส่วนใหญ่แล้วเป็นโปรแกรมสำหรับสร้างแผนภูมิ 2 มิติเพียงอย่างเดียว แต่มีบางโปรแกรมที่อำนวยความสะดวกในการแสดงผล 2 มิติและมีเครื่องมือให้ผู้เชี่ยวชาญจับคู่กลุ่มที่เหมือนกันได้ เช่น โปรแกรม MANSEK และ SUNHCA [32]

MANSEK และ SUNHCA เป็นโปรแกรมสำหรับการสร้างแผนภูมิ 2 มิติจากลำดับกรดอะมิโน ทำงานบนระบบ Vax และ Sun ซึ่งได้มีการพัฒนาอย่างต่อเนื่อง เพื่อให้สามารถใช้งานได้ดียิ่งขึ้น มีความสามารถในการย่อ ขยายแผนภูมิ และยังสามารถสร้างแผนภูมิให้อยู่ในรูปของเอกสารที่ต้องการพิมพ์ได้ นอกจากนั้น โปรแกรมมีการรวมเครื่องมือสำหรับจับคู่กลุ่มอะมิโนที่ไม่ชอบน้ำได้โดยให้ผู้เชี่ยวชาญจับคู่เอง ซึ่งต้องใช้ความรู้และความชำนาญเป็นอย่างมาก ดังนั้นวิธีการจับคู่กลุ่มอะมิโนที่ไม่ชอบน้ำจึงยังไม่ได้รับความนิยม ยังมีงานวิจัยอีกไม่น้อยที่พยายามใช้

วิธีการจับคู่กลุ่มที่ไม่ชอบน้ำ ตัวอย่างเช่น Laget และคณะ [33] ได้ใช้วิธีการจับคู่กลุ่มที่ไม่ชอบน้ำเพื่อดูโครงสร้างของโปรตีนที่มีหน้าที่การทำงานเหมือนกันในสิ่งมีชีวิตที่ต่างกัน พบว่าโปรตีนที่นำมาจับคู่กลุ่มที่ไม่ชอบน้ำมีโครงสร้างคล้ายกัน Henrissat และคณะ [31] ได้ใช้การจับคู่กลุ่มที่ไม่ชอบน้ำในแผนภูมิ 2 มิติ มาวิเคราะห์หาโครงสร้างที่เป็นลำดับกรดอะมิโนในส่วนคาร์บอนจากข้าวสาลี (*Triticum aestivum*) ซึ่งเป็นลำดับกรดอะมิโนที่อยู่ทางปลายสุดที่เป็นหมู่คาร์บอกซิล พบว่าโปรตีนที่นำมาเปรียบเทียบนั้นมีส่วนของโครงสร้างเหมือนกัน Henrissat และคณะสรุปว่า

1. ส่วนที่เปลี่ยนแปลง และส่วนที่อนุรักษ์ สามารถอยู่ที่ตำแหน่งใดของลำดับกรดอะมิโนก็ได้
2. ส่วนที่อนุรักษ์สามารถบอกถึงความเป็นมาของบรรพบุรุษของสิ่งมีชีวิตได้
3. คุณสมบัติที่เป็นเอกลักษณ์ของโปรตีนแต่ละชนิด มีความสัมพันธ์กับส่วนที่เปลี่ยนแปลงในลำดับกรดอะมิโน

โดยที่ส่วนที่เปลี่ยนแปลงคือบริเวณที่กรดอะมิโนมีความแตกต่างกันในสิ่งมีชีวิตแต่ละสายพันธุ์ และส่วนที่อนุรักษ์คือบริเวณที่กรดอะมิโนไม่มีการเปลี่ยนแปลงหรือเปลี่ยนแปลงน้อยมากในสิ่งมีชีวิตแต่ละสายพันธุ์

นอกจากนี้ยังมีงานวิจัยอีกมากที่ใช้การวิเคราะห์กลุ่มที่ไม่ชอบน้ำเพื่อหาความสัมพันธ์ทางโครงสร้างของโปรตีน [14, 31, 34-38] แต่ทั้งหมดนี้นักวิจัยยังคงใช้เครื่องมือที่ใช้จับคู่กลุ่มที่ไม่ชอบน้ำเพื่อใช้สำหรับเปรียบเทียบความเหมือนของลำดับกรดอะมิโน ซึ่งยังไม่เป็นระบบอัตโนมัติ

2.8 Local Alignment และ Global Alignment

เป็นวิธีการที่ใช้เปรียบเทียบความคล้ายกัน โดย Global Alignment จะเป็นวิธีการเปรียบเทียบความเหมือนของลำดับทั้งหมดของทั้ง 2 ลำดับกรดอะมิโน สำหรับเทคนิคนี้เป็นที่นิยมใช้กันอย่างแพร่หลายอยู่ในขั้นตอนวิธีการของ Needleman-Wunsch [25] เป็นวิธีกำหนดการพลวัต แต่สำหรับ Local Alignment เป็นการเปรียบเทียบแยกเป็นส่วนย่อยได้ โดยจะเปรียบเทียบเป็นท่อนสั้นๆ ใช้สำหรับเปรียบเทียบลำดับที่ไม่เหมือนกันมาก สามารถใช้หาลำดับที่คงไว้ (Motifs Sequence) ซึ่งอยู่ในลำดับขนาดใหญ่ ส่วนใหญ่ใช้อยู่ในขั้นตอนวิธีการ Smith-Waterman [39] เป็นวิธีกำหนดการพลวัตเช่นกัน นอกจากนี้ยังมีอีกวิธีการที่ผสมผสานวิธีการทั้ง 2 แบบเข้าด้วยกัน เรียกว่า Glocal วิธีการนี้มีประสิทธิภาพในการเปรียบเทียบหาลำดับที่มีการทับซ้อนกัน การเปรียบเทียบแบบ Global Alignment และ Local Alignment แสดงในรูปที่ 2.9

```
Global FTFTALILLAVAV
       F--TAL-LLA-AV

Local  FTFTALILL-AVAV
       --FTAL-LLAAV--
```

รูปที่ 2.9 Global Alignment vs. Local Alignment

จากรูปที่ 2.9 วิธี Global Alignment ที่แสดงดังรูปบนจะมีการจับคู่โดยจะไม่มีช่องว่างที่จุดเริ่มต้นและจุดสิ้นสุด ในขณะที่ Local Alignment จะยอกให้มีช่องว่างเกิดขึ้นที่จุดเริ่มต้น หรือจุดสิ้นสุดได้

2.9 ตารางให้คะแนน BLOSUM (BLOSUM Substitution Matrix)

ตารางให้คะแนน BLOSUM เป็นตารางที่คำนวณมาจากการคำนวณความน่าจะเป็น (Likelihood Method) ใช้การประมาณความน่าจะเป็นที่เกิดการจับคู่ระหว่างกรดอะมิโนแต่ละคู่ ลำดับกรดอะมิโนที่ใช้ สำหรับชุดข้อมูลที่ใช้คำนวณความน่าจะเป็นมีอยู่ด้วยกันหลายชุดข้อมูล เช่น Protein Block และ Structural Block ซึ่งแต่ละชุดข้อมูลจะมีวิธีการเปรียบเทียบที่แตกต่างกันไป เราสามารถคำนวณตารางการให้คะแนนได้จากสมการ 2.1

$$S(a,b) = \frac{1}{\lambda}\log(\frac{P_{ab}}{f_a f_b}) \qquad (2.1)$$

จากสมการที่ 2.1 $S(a,b)$ คือคะแนนที่คำนวณได้ของกรดอะมิโน a และ b P_{ab} คือความน่าจะเป็นของการพบกรดอะมิโน a อยู่ตำแหน่งเดียวกับ b ซึ่งได้จากการเก็บข้อมูลจากชุดข้อมูล f_a และ f_b คือความน่าจะเป็นของการพบกรดอะมิโน a หรือ b ในลำดับกรดอะมิโนทั้งหมด ส่วน λ คืออัตราส่วนย่อขยาย (scaling factor)

ดังนั้นเมื่อใช้ชุดข้อมูล Protein Block ที่มีความเหมือนของลำดับสูงสุด (Identity cutoff) ที่ 62 เปอร์เซ็นต์ เราจะได้ตารางคะแนน BLOSUM62 ดังแสดงในรูปที่ 2.10

R	-1	5																			
N	-2	0	6																		
D	-2	-2	1	6																	
C	0	-3	-3	-3	9																
Q	-1	1	0	0	-3	5															
E	-1	0	0	2	-4	2	5														
G	0	-2	0	-1	-3	-2	-2	6													
H	-2	0	1	-1	-3	0	0	-2	8												
I	-1	-3	-3	-3	-1	-3	-3	-4	-3	4											
L	-1	-2	-3	-4	-1	-2	-3	-4	-3	2	4										
K	-1	2	0	-1	-3	1	1	-2	-1	-3	-2	5									
M	-1	-1	-2	-3	-1	0	-2	-3	-2	1	2	-1	5								
F	-2	-3	-3	-3	-2	-3	-3	-3	-1	0	0	-3	0	6							
P	-1	-2	-2	-1	-3	-1	-1	-2	-2	-3	-3	-1	-2	-4	7						
S	1	-1	1	0	-1	0	0	0	-1	-2	-2	0	-1	-2	-1	4					
T	0	-1	0	-1	-1	-1	-1	-2	-2	-1	-1	-1	-1	-2	-1	1	5				
W	-3	-3	-4	-4	-2	-2	-3	-2	-2	-3	-2	-3	-1	1	-4	-3	-2	11			
Y	-2	-2	-2	-3	-2	-1	-2	-3	2	-1	-1	-2	-1	3	-3	-2	-2	2	7		
V	0	-3	-3	-3	-1	-2	-2	-3	-3	3	1	-2	1	-1	-2	-2	0	-3	-1	4	
X	0	-1	-1	-1	-2	-1	-1	-1	-1	-1	-1	-1	-1	-1	-2	0	0	-2	-1	-1	-1
	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V	X

รูปที่ 2.10 ตารางคะแนน BLOSUM62

โดยตารางคะแนนที่คำนวณได้จะสามารถบอกถึงความสัมพันธ์ของความเหมือนกันของโครงสร้าง ซึ่งแบ่งได้ตามคุณสมบัติทางเคมีของกรดอะมิโนดังตารางที่ 2.3

ตารางที่ 2.3 คุณสมบัติทางเคมีของกรดอะมิโน

อักษรย่อกรดอะมิโน	คุณสมบัติทางเคมี
G,A,V,L,I,M	Aliphatic
S,T,C	Hydroxyl, Sulfhydryl, Polar
N,Q	Amide side chains
F,W,Y	Aromatic
H,K,R	Basic
D,E	Acidic

จากตาราง BLOSUM62 จะเห็นว่าส่วนใหญ่กรดอะมิโนที่อยู่ในกลุ่มเดียวกันจะมีค่าคะแนนที่สูง และกรดอะมิโนที่อยู่คนละกลุ่มกันจะมีค่าคะแนนที่ต่ำลงไปเรื่อยๆ และสำหรับกรดอะมิโนทริปโตเฟน เป็นกรดอะมิโนที่หายาก และไม่ค่อยมีการเปลี่ยนแปลงเป็นลำดับกรดอะมิโนอื่น ดังนั้นคะแนนของทริปโตเฟน/ทริปโตเฟน จึงมีค่าสูงที่สุดในตาราง

2.10 ค่าหักลบช่องว่าง (Gap Penalty)

ค่าหักลบช่องว่างใช้ระหว่างการจับคู่ลำดับ เป็นคะแนนที่หักเมื่อลำดับกรดอะมิโนมีจำนวนเพิ่มขึ้น (Insertion) หรือมีจำนวนลดลง (Deletion) ทำให้เกิดช่องว่างขึ้น ซึ่งค่าที่หักลบแต่ละตำแหน่งจะขึ้นกับความยาวของช่องว่างที่เกิดขึ้น วิธีการคำนวณค่าหักลบช่องว่างมีอยู่ด้วยกันหลายวิธีเช่น

2.10.1 ค่าหักลบช่องว่างแบบคงที่ (Constant Gap Penalty)

เป็นค่าหักลบช่องว่างที่ง่ายที่สุด คือใช้ค่าหักลบค่าเดียวกันตลอดไม่ว่าขนาดช่องว่างจะกว้างเท่าใด ซึ่งไม่สนใจขนาดของช่องว่างที่เกิดขึ้น การคำนวณจะเป็นไปดังสมการ 2.2

$$g = d \tag{2.2}$$

ขณะที่ g คือค่าหักลบ และ d คือค่าหักลบที่กำหนด

2.10.2 ค่าหักลบช่องว่างแบบเส้นตรง (Linear Gap Penalty)

เป็นค่าหักลบที่ใช้ความยาวของช่องว่างเข้ามาคำนวณด้วย จะทำให้ค่าหักลบจะเพิ่มขึ้นเมื่อมีความยาวเพิ่มขึ้นเรื่อยๆ ดังแสดงในสมการที่ 2.3

$$g = dl \tag{2.3}$$

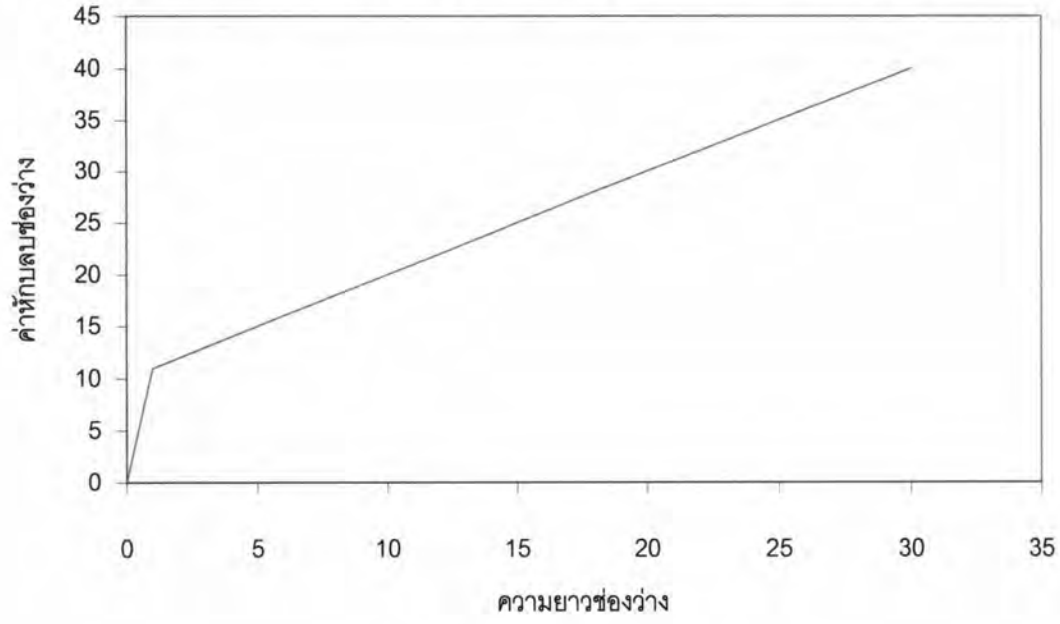
ขณะที่ g คือค่าหักลบ d คือค่าหักลบที่กำหนดต่อความยาวที่เพิ่มขึ้น และ l คือความยาวของช่องว่าง

2.10.3 Affine Gap Penalty

Affine Gap Penalty ได้รับการเสนอขึ้นโดย Nalin และคณะ [40] เพื่อใช้สำหรับการคำนวณการเปรียบเทียบ ซึ่งคำนวณมาจากการอัตราการเกิดช่องว่างที่เกิดขึ้น จากฐานข้อมูลการจับคู่โดยใช้โครงสร้าง ซึ่งจะได้ค่าหักลบช่องว่างดังสมการ 2.4

$$g = u + vl \tag{2.4}$$

ขณะที่ g คือค่าหักลบช่องว่าง u คือค่าหักลบเมื่อเปิดช่องว่าง v คือค่าหักลบเมื่อขยายช่องว่าง และ l คือความยาวช่องว่าง จากสมการที่ 2.4 แสดงเป็นกราฟได้ดังรูปที่ 2.11



รูปที่ 2.11 Affine Gap Penalty

งานวิจัยนี้ได้เสนอเครื่องมือจับคู่กลุ่มที่ไม่ชอบน้ำแบบ 2 มิติโดยอัตโนมัติ ซึ่งใช้การเปรียบเทียบรูปแบบกลุ่มที่ไม่ชอบน้ำที่เกิดขึ้น เพื่อเพิ่มประสิทธิภาพในการทำวิจัยให้มากขึ้น และผู้วิจัยได้นำเสนอค่าหักลบช่องว่างใหม่ที่น่าจะมีประสิทธิภาพมากกว่าวิธีการเดิมที่ใช้กันอยู่ในปัจจุบัน เพื่อใช้กับเครื่องมือจับคู่กลุ่มที่ไม่ชอบน้ำแบบ 2 มิติโดยอัตโนมัติ ซึ่งใช้วิธีการหักลบช่องว่างไม่เหมือนแบบเดิม ซึ่งค่าหักลบช่องว่างแบบเดิมจะคิดเป็นจำนวนช่องว่างต่อกรดอะมิโน โดยวิธีการดำเนินงานวิจัย และการทดสอบกับชุดข้อมูลจะกล่าวในบทที่ 3 ต่อไป