

## บทที่ 2 ปรีทัศน์วรรณกรรม

### ความรู้เบื้องต้นเกี่ยวกับโปรตีน (Introduction Protein)

โปรตีน (protein) เป็นสารชีวโมเลกุลที่พบในสิ่งมีชีวิต และมีความสำคัญต่อสิ่งมีชีวิตมาก เพราะโปรตีนเป็นส่วนประกอบในส่วนต่างๆของสิ่งมีชีวิต ไม่ว่าจะเป็นโครงสร้าง เนื้อเยื่อ อวัยวะต่างๆ ตลอดจน ผิวหนัง กล้ามเนื้อ ขน ผม เล็บ ตลอดจนสารต่างๆที่อยู่ภายใน เซลล์ เช่น ฮอรโมน เอนไซม์ แอนติบอดี ฮีโมโกลบิน และอื่นๆอีกมาก นอกจากนี้โปรตีนยังมีบทบาทในการสร้างและซ่อมแซมเซลล์ต่างๆในร่างกายอีกด้วย

โปรตีนหรือโพลีเปปไทด์จะ ประกอบด้วยหน่วยย่อยที่เรียกว่ากรดอะมิโนซึ่งมีอยู่ 20 ชนิด ต่อกันด้วยพันธะเปปไทด์ ขนาดของโปรตีนจะต่างกันไปขึ้นกับจำนวนของกรดอะมิโน เช่น Cytochrome c ในมนุษย์ ประกอบด้วยกรดอะมิโน 104 อะมิโน ฮีโมโกลบินในมนุษย์ประกอบด้วยกรดอะมิโน 574 อะมิโน ส่วนโครงสร้างของโปรตีนมีลักษณะเป็นสามมิติ โดยโครงสร้างหรือรูปร่างของโปรตีนจะถูกกำหนดโดยลำดับของกรดอะมิโนในโปรตีน

ส่วนใหญ่เราจะรู้จักโปรตีนในรูปของสารอาหาร แหล่งอาหารโปรตีนที่สำคัญของมนุษย์ได้แก่เนื้อสัตว์และผลิตภัณฑ์จากสัตว์ รongมาได้แก่พวกถั่วต่างๆ เมื่ออาหารเหล่านี้เข้าสู่ร่างกายจะถูกย่อยจนกระทั่งได้กรดอะมิโนแล้วจึงนำไปใช้ ร่างกายของมนุษย์เราไม่สามารถสะสมโปรตีนไว้ได้มากมาย และไม่สามารถสังเคราะห์กรดอะมิโนที่จำเป็นต่อร่างกายได้ จึงทำให้มนุษย์เรามีความจำเป็นต้องบริโภคโปรตีนอย่างสม่ำเสมอ

### ประเภทของโปรตีน

เราสามารถแบ่งโปรตีนตามหน้าที่ของโปรตีนได้ดังนี้

#### 1. โปรตีนที่ทำหน้าที่เป็นองค์ประกอบของเอนไซม์ (Enzymes)

เอนไซม์ทำหน้าที่เป็นเป็นตัวเร่ง (catalyst) ในปฏิกิริยาต่างๆในเซลล์ โดยปฏิกิริยาเหล่านี้จะเกิดได้เมื่อมีเอนไซม์ที่จำเพาะเจาะจงไปจับกับซับสเตรต

## 2. โปรตีนที่ทำหน้าที่เป็นขนส่งสารอาหารและออกซิเจน (Transport Proteins)

โปรตีนหลายตัวทำหน้าที่เป็นตัวขนส่งโมเลกุลต่างๆในเซลล์เช่น ฮีโมโกลบินในเลือดทำหน้าที่ขนส่งออกซิเจนและคาร์บอนไดออกไซด์จากปอดไปยังส่วนต่างๆในเซลล์ในร่างกาย หรือ ไลโปโปรตีนที่อยู่ในน้ำเลือดทำหน้าที่ขนส่งไขมันจากผนังลำไส้เล็กไปสู่ตับ

## 3. โปรตีนที่ทำหน้าที่เป็นสารอาหารและคลังอาหาร (Nutrient and Storage Proteins)

เมล็ดพืชต่างๆจะมีการสะสมโปรตีนไว้เลี้ยงตัวอ่อน เช่น ข้าว ข้าวโพด ข้าวสาลี เคซีน (casein) ในน้ำนม หรือ โอวัลบูมิน (ovalbumin) ในไข่ขาว เป็นต้น

## 4. โปรตีนที่ทำหน้าที่ในการเคลื่อนไหว (Contractile or Motile Proteins)

การเคลื่อนไหวและการเปลี่ยนแปลงรูปร่างในสิ่งมีชีวิตมีผลมาจากโปรตีนบางชนิด เช่นการเคลื่อนไหวในสิ่งมีชีวิตชั้นสูงเป็นผลมาจากโปรตีน 2 ชนิดคือแอกติน (actin) และไมโอซิน หรือโปรตีนไดเอนินในไมโครทิวบูล (microtubules) มีส่วนช่วยในการเคลื่อนไหวในซิเลีย (cilia) หรือแฟลเจลลา (flagella) ในสิ่งมีชีวิตชั้นต่ำ

## 5. โปรตีนที่ทำหน้าที่เป็นโครงสร้าง (Structural Proteins)

โปรตีนชนิดนี้เป็นองค์ประกอบหลักในร่างกาย เช่น คอลลาเจน (collagen) ในกล้ามเนื้อเกี่ยวพัน เอ็น และกระดูก หรือเคราติน (keratin) ในเส้นผมเล็บ ขน โปรตีนจะทำหน้าที่โดยตรงในการในการสร้างความเจริญเติบโตให้กับร่างกาย

## 6. โปรตีนที่ทำหน้าที่เป็นภูมิคุ้มกันโรค (Defensive Proteins)

แอนติบอดี (antibody) เป็นโปรตีนชนิดหนึ่งในร่างกายที่ทำหน้าที่ป้องกันและทำลายสิ่งแปลกปลอมที่เข้าสู่ร่างกาย ดังนั้นคนที่รับประทานอาหารประเภทโปรตีนไม่เพียงพอ จะมีภูมิต้านทานต่ำและเกิดโรคภัยไข้เจ็บได้ง่าย

## 7. โปรตีนที่ทำหน้าที่เป็นฮอร์โมน (Regulatory Proteins)

ฮอร์โมนทำหน้าที่ควบคุมอัตราเมตาบอลิซึมเช่น อินซูลิน (insulin) ทำหน้าที่ควบคุมระดับน้ำตาลในกระแสเลือด โกรทฮอร์โมน (growth Hormone) ทำหน้าที่ควบคุมการเจริญเติบโต เป็นต้น

## 8. โปรตีนที่ทำหน้าที่อื่น ๆ (Other Protein)

โปรตีนบางชนิดไม่สามารถจำแนกได้ว่าอยู่ในประเภทไหน เช่น โปรตีน monellin ในพืชแถบแอฟริกา ทำมีรสหวานอย่างรุนแรง หรือโปรตีนในเลือดของปลาที่ทำให้เลือดไม่แข็งตัวเมื่ออุณหภูมิภายนอกต่ำมากๆ เป็นต้น

## กรดอะมิโนและพันธะเปปไทด์ (Amino acid and Peptide bond)

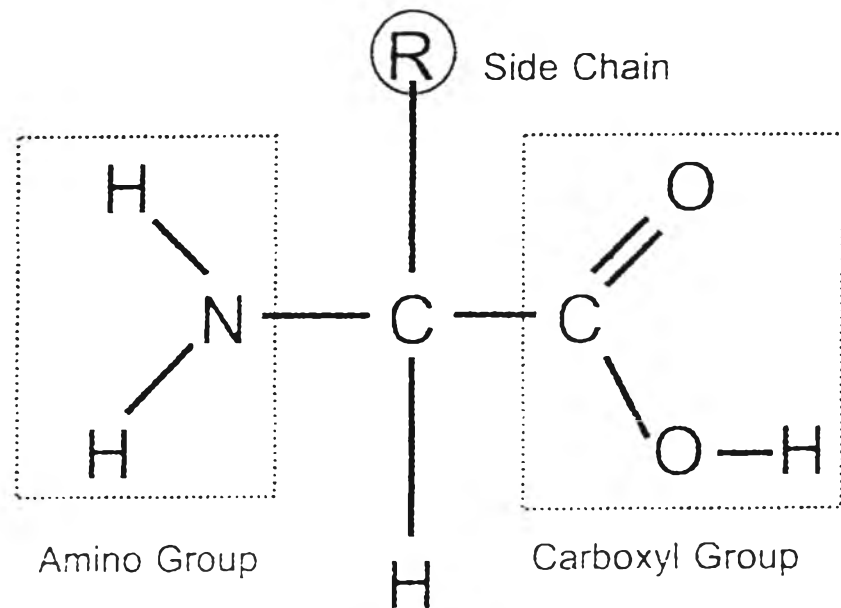
โปรตีนโดยทั่วไปจะประกอบด้วยกรดอะมิโนเพียงอย่างเดียวต่อกันด้วยพันธะเปปไทด์ แต่ก็มีโปรตีนบางชนิดมีองค์ประกอบของสารอื่นๆรวมอยู่ด้วย เช่น ฮีโมโกลบิน ในเม็ดเลือดประกอบด้วยอะมิโนและธาตุเหล็กซึ่งอยู่ตรงกลางของโมเลกุล

กรดอะมิโนที่พบในธรรมชาติมีอยู่ 20 ชนิดด้วยกัน การเรียงลำดับของกรดอะมิโนทั้ง 20 สามารถทำให้เกิดโปรตีนชนิดต่างๆขึ้นมาได้มากมาย และทำหน้าที่ต่างกันไป โดยลำดับของกรดอะมิโนที่ใช้ในการสร้างโปรตีนจะถูกควบคุมโดยยีนหรือ DNA

ในธรรมชาติยังมีกรดอะมิโนอีกมากมายหลายชนิดนอกเหนือจาก 20 ชนิดที่พบในโปรตีน เช่น กรดเจนโคลิค (djenkolic acid) ในลูกเนียงซึ่งเป็นสารพิษ เมื่อรับประทานจำนวนมากๆจะเกิดอาการคล้ายคลึงกับอาการของโรคนี้

## โครงสร้างของกรดอะมิโน

กรดอะมิโนประกอบด้วยธาตุหลัก 4 ชนิดคือ คาร์บอน (C), ไฮโดรเจน (H), ออกซิเจน (O) และไนโตรเจน (N) และอาจมีธาตุอื่นๆ เช่น ฟอสฟอรัส (P), กำมะถัน (S), เป็นองค์ประกอบอยู่ด้วย ธาตุเหล่านี้จะจัดเรียงตัวเข้าด้วยกันโดยมีสูตรโครงสร้างทั่วไปดังต่อไปนี้



รูปที่ 2.1 รูปแสดงโครงสร้างของกรดอะมิโน

COOH คือหมู่คาร์บอกซิล (carboxyl group)

NH<sub>2</sub> คือหมู่อะมิโน (amino group)

R คือ side chain

โดยกรดอะมิโนแต่ละตัวก็จะมีสมบัติและ โครงสร้างต่างกันไปดังตารางที่ 2.1

ชื่อ	สูตรโครงสร้าง	Mass (daltons)	Volume (Å <sup>3</sup> )	Accessible surface area (Å <sup>2</sup> )
Alanine (Ala,A)	$  \begin{array}{c}  \text{COO}^- \\    \\  ^+\text{H}_3\text{N}-\text{C}-\text{H} \\    \\  \text{CH}_3  \end{array}  $	71.08	88.6	115
Arginine (Arg,R)	$  \begin{array}{c}  \text{COO}^- \\    \\  ^+\text{H}_3\text{N}-\text{C}-\text{H} \\    \\  \text{CH}_2 \\    \\  \text{CH}_2 \\    \\  \text{CH}_2 \\    \\  \text{NH} \\    \\  \text{C}=\text{NH}_2^+ \\    \\  \text{NH}_2  \end{array}  $	156.20	173.4	225
Asparagine (Asn,N)	$  \begin{array}{c}  \text{COO}^- \\    \\  ^+\text{H}_3\text{N}-\text{C}-\text{H} \\    \\  \text{CH}_2 \\    \\  \text{C} \\  / \quad \backslash \\  \text{H}_2\text{N} \quad \text{O}  \end{array}  $	114.11	117.7	160
Aspartic Acid (Asp,D)	$  \begin{array}{c}  \text{COO}^- \\    \\  ^+\text{H}_3\text{N}-\text{C}-\text{H} \\    \\  \text{CH}_2 \\    \\  \text{COO}^-  \end{array}  $	115.09	111.1	150
Cysteine (Cys,C)	$  \begin{array}{c}  \text{COO}^- \\    \\  ^+\text{H}_3\text{N}-\text{C}-\text{H} \\    \\  \text{CH}_2 \\    \\  \text{SH}  \end{array}  $	103.14	108.5	135
Glutamic Acid (Gln,Q)	$  \begin{array}{c}  \text{COO}^- \\    \\  ^+\text{H}_3\text{N}-\text{C}-\text{H} \\    \\  \text{CH}_2 \\    \\  \text{CH}_2 \\    \\  \text{COO}^-  \end{array}  $	128.14	143.9	180

Glutamine (Glu,E)	$  \begin{array}{c}  \text{COO}^- \\    \\  ^+\text{H}_3\text{N}-\text{C}-\text{H} \\    \\  \text{CH}_2 \\    \\  \text{CH}_2 \\    \\  \text{C} \\  / \quad \backslash \\  \text{H}_2\text{N} \quad \text{O}  \end{array}  $	129.12	138.4	190
Glycine (Gly,G)	$  \begin{array}{c}  \text{COO}^- \\    \\  ^+\text{H}_3\text{N}-\text{C}-\text{H} \\    \\  \text{H}  \end{array}  $	57.06	60.1	75
Histidine (His,H)	$  \begin{array}{c}  \text{COO}^- \\    \\  ^+\text{H}_3\text{N}-\text{C}-\text{H} \\    \\  \text{CH}_2 \\    \\  \text{C} \\  / \quad \backslash \\  \text{C}=\text{N}^+ \quad \text{CH} \\    \quad \backslash \\  \text{HC} \quad \text{N} \\    \\  \text{H}  \end{array}  $	137.15	153.2	195
Isoleucine (He,I)	$  \begin{array}{c}  \text{COO}^- \\    \\  ^+\text{H}_3\text{N}-\text{C}-\text{H} \\    \\  \text{H}-\text{C}-\text{CH}_3 \\    \\  \text{CH}_2 \\    \\  \text{CH}_3  \end{array}  $	113.17	166.7	175
Leucine (Leu,L)	$  \begin{array}{c}  \text{COO}^- \\    \\  ^+\text{H}_3\text{N}-\text{C}-\text{H} \\    \\  \text{CH}_2 \\    \\  \text{CH} \\  / \quad \backslash \\  \text{H}_3\text{C} \quad \text{CH}_3  \end{array}  $	113.17	166.7	170
Lysine (Lys,k)	$  \begin{array}{c}  \text{COO}^- \\    \\  ^+\text{H}_3\text{N}-\text{C}-\text{H} \\    \\  \text{CH}_2 \\    \\  \text{CH}_2 \\    \\  \text{CH}_2 \\    \\  \text{CH}_2 \\    \\  \text{NH}_3^+  \end{array}  $	128.18	168.6	200

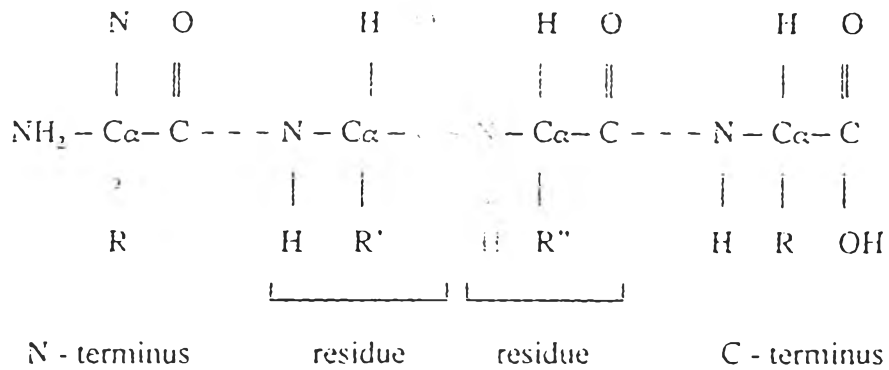
Methionine (Met,M)	$  \begin{array}{c}  \text{COO}^- \\    \\  ^+\text{H}_3\text{N}-\text{C}-\text{H} \\    \\  \text{CH}_2 \\    \\  \text{CH}_2 \\    \\  \text{S} \\    \\  \text{CH}_3  \end{array}  $	131.21	162.9	185
Phenylalanine (Phe,F)	$  \begin{array}{c}  \text{COO}^- \\    \\  ^+\text{H}_3\text{N}-\text{C}-\text{H} \\    \\  \text{CH}_2 \\    \\  \text{C}_6\text{H}_5  \end{array}  $	147.18	189.9	210
Proline (Pro,P)	$  \begin{array}{c}  \text{COO}^- \\    \\  ^+\text{H}_2\text{N}-\text{C}-\text{H} \\    \quad   \\  \text{H}_2\text{C} \quad \text{CH}_2 \\  \backslash \quad / \\  \text{C} \\  / \quad \backslash \\  \text{H}_2 \quad \text{H}_2  \end{array}  $	97.12	122.7	145
Serine (Ser,S)	$  \begin{array}{c}  \text{COO}^- \\    \\  ^+\text{H}_3\text{N}-\text{C}-\text{H} \\    \\  \text{H}-\text{C}-\text{OH} \\    \\  \text{H}  \end{array}  $	87.08	89.0	115
Threonine (Thr,T)	$  \begin{array}{c}  \text{COO}^- \\    \\  ^+\text{H}_3\text{N}-\text{C}-\text{H} \\    \\  \text{H}-\text{C}-\text{OH} \\    \\  \text{CH}_3  \end{array}  $	101.11	116.1	140
Tryptophan (Trp,W)	$  \begin{array}{c}  \text{COO}^- \\    \\  ^+\text{H}_3\text{N}-\text{C}-\text{H} \\    \\  \text{CH}_2 \\    \\  \text{C} \\    \quad   \\  \text{HC} \quad \text{N} \\  \backslash \quad / \\  \text{C}_6\text{H}_4  \end{array}  $	186.21	227.8	255
Tyrosine (Try,Y)	$  \begin{array}{c}  \text{COO}^- \\    \\  ^+\text{H}_3\text{N}-\text{C}-\text{H} \\    \\  \text{CH}_2 \\    \\  \text{C}_6\text{H}_4 \\    \\  \text{OH}  \end{array}  $	163.18	193.6	230

Valine (Val,V)	$  \begin{array}{c}  \text{COO}^- \\    \\  \text{}^+\text{H}_3\text{N}-\text{C}-\text{H} \\    \\  \text{CH} \\  / \quad \backslash \\  \text{H}_3\text{C} \quad \text{CH}_3  \end{array}  $	99.14	140.0	155
-------------------	---	-------	-------	-----

ตารางที่ 2.1 ตารางแสดง สูตรโครงสร้าง น้ำหนักโมเลกุล ปริมาตรและพื้นที่ผิวของกรดอะมิโนแต่ละชนิด (Creighton T.E., 1984.)

**พันธะเปปไทด์**

พันธะที่ยึดระหว่างกรดอะมิโนแต่ละโมเลกุลเข้าด้วยกันนั้นเราเรียกว่าพันธะเปปไทด์ (peptide bond) ซึ่งจัดเป็นพันธะโควาเลนต์ (covalent bond) ซึ่งเป็นพันธะที่แข็งแรง ไม่แตกออกง่ายๆ



**รูปที่ 2.2 รูปแสดงพันธะเปปไทด์**

รูปร่างของโปรตีนมีความสัมพันธ์ต่อหน้าที่ แต่ถ้าเราปรับสภาพแวดล้อมเช่น อุณหภูมิ pH จะให้รูปร่างของโปรตีนเปลี่ยนไปจากเดิม และเสียสมบัติในการทำหน้าที่ของโปรตีนเหล่านั้นด้วย

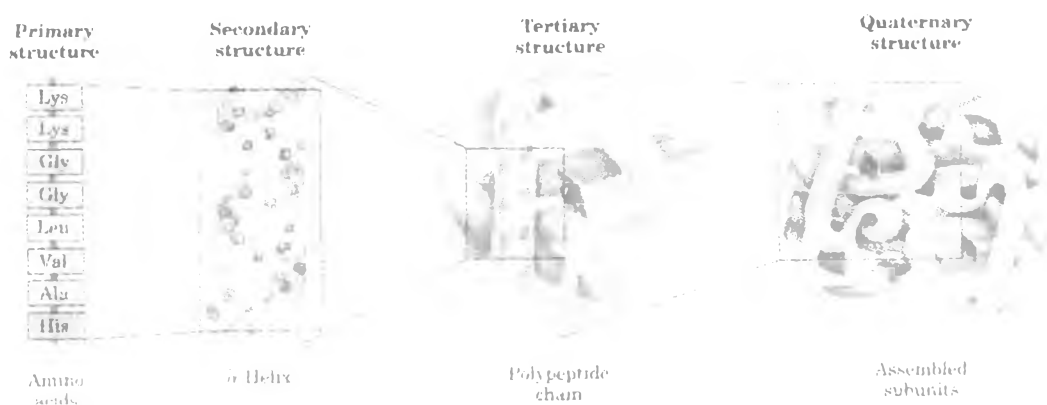


## โครงสร้างของโปรตีน (Protein Structure)

โปรตีนมีโครงสร้างแบบสามมิติ โดยโปรตีนส่วนใหญ่จะมีโครงสร้างซับซ้อน ดังนั้นจึงได้มีการแบ่งระดับโครงสร้างของโปรตีนออกเป็น 4 ระดับคือ

1. โครงสร้างแบบปฐมภูมิ (primary structure)
2. โครงสร้างแบบทุติยภูมิ (secondary structure)
3. โครงสร้างแบบตติยภูมิ (tertiary structure)
4. โครงสร้างแบบจตุรภูมิ (quaternary structure)

ดังรูป



รูปที่ 2.3 รูปแสดงระดับโครงสร้างของโปรตีน (Cox., Lenhinger. and Nelson., 1993.)

### 1. โครงสร้างแบบปฐมภูมิ (Primary Structure)

เป็นการพิจารณาโครงสร้างของโปรตีนในระดับแรกคือ การเรียงลำดับของกรดอะมิโนในโปรตีน พบว่าจากการเรียงลำดับของกรดอะมิโนซึ่งมีอยู่ด้วยกัน 20 ชนิดนี้ สามารถทำให้เกิดโปรตีนที่ต่างกันชนิดกันมากมาย

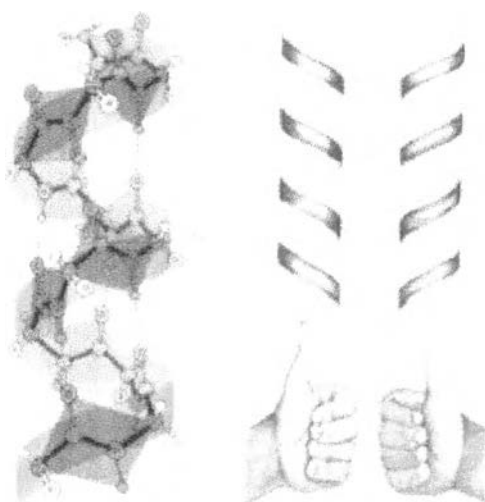
## 2. โครงสร้างแบบทุติยภูมิ(Secondary Structure)

จากการที่สายโพลีเปปไทด์มีแบคโบนเป็นแกน และมี side chain ยื่นออกมาจากแนวแกนหมุนข้างเมื่อมีปฏิสัมพันธ์ระหว่างสายโพลีเปปไทด์เดียวกันหรือระหว่างโพลีเปปไทด์ต่างสายกันจะทำให้สายโพลีเปปไทด์ ขดเป็นเกลียวหรือแผ่นพับ

โครงสร้างทุติยภูมิที่พบบ่อยมี 2 แบบคือ

### 2.1 แบบเกลียวแอลฟา ( $\alpha$ -helix)

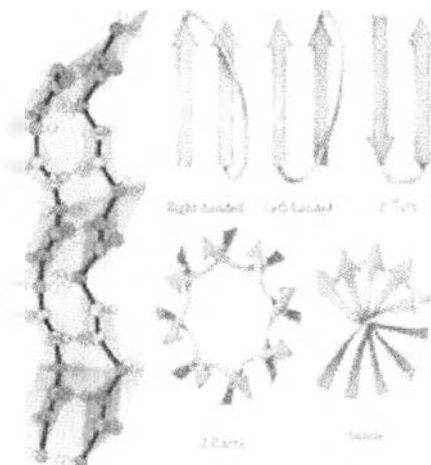
สายโพลีเปปไทด์บิดเป็นเกลียวได้โดยการเกิดพันธะไฮโดรเจนระหว่าง O ในหมู่ C=O ของกรดอะมิโนโมเลกุลที่หนึ่งกับ H ในหมู่ N-H ของกรดอะมิโนซึ่งอยู่ถัดไป 4 ตำแหน่ง โดยเกิดพันธะเช่นนี้ตลอดสายโพลีเปปไทด์ ทำให้เกิดเป็นเกลียวแอลฟา ซึ่งจะวนซ้ายหรือขวาก็ได้



รูปที่ 2.4 รูปแสดงโครงสร้างโปรตีนแบบเกลียวแอลฟา (Cox., Lenhinger. and Nelson., 1993.)

## 2.2 แบบแผ่นพับ ( $\beta$ - sheet)

ประกอบด้วยโพลีเปปไทด์หลายสายเรียงขนานกัน และยึดกันด้วยพันธะไฮโดรเจน โปรตีนชนิดนี้พบในเส้นไหม โครงสร้างลักษณะนี้ทำให้เส้นไหมมีโครงสร้างแข็งแรง ทนต่อแรงดึงได้ดี



รูปที่ 2.5 รูปแสดงโครงสร้างโปรตีนแบบแผ่นพับ (Cox., Lenhinger. and Nelson., 1993.)

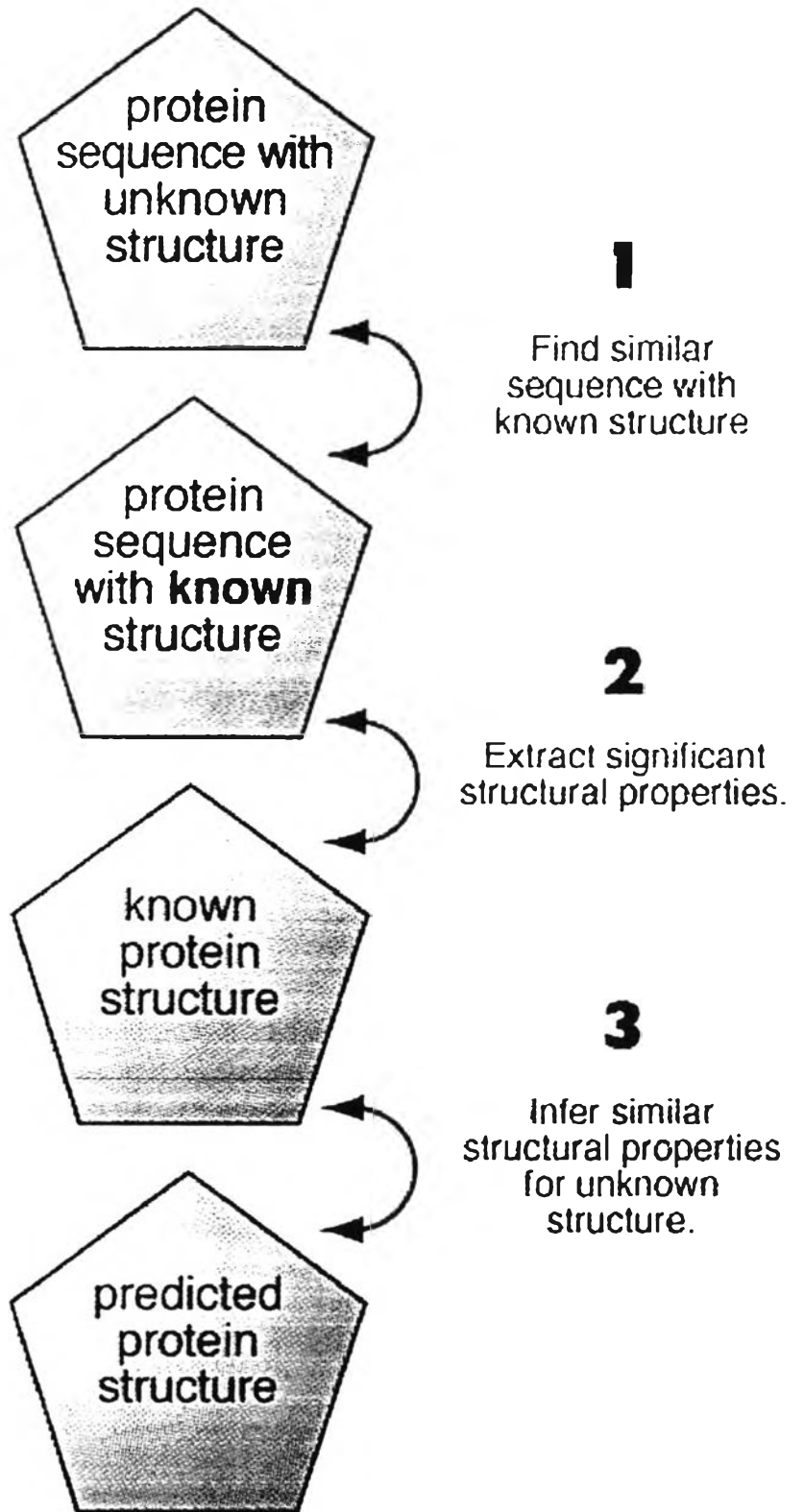
## 3. โครงสร้างแบบตติยภูมิ (Tertiary structure)

เป็นโครงสร้างที่เกิดจากการม้วนตัว หรือขดไปมาอย่างเป็นระเบียบของโครงสร้างทุติยภูมิ เนื่องจากมีพันธะที่เกิดจากหมู่ side chain ของกรดอะมิโนในสายโพลีเปปไทด์เป็นตัวยึดเหนี่ยว จึงทำให้โปรตีนในโครงสร้างระดับนี้ดูเสมือนเป็นรูปทรงกลม ซึ่งเรียกโปรตีนที่มีลักษณะเช่นนี้ว่า กลอบูลาร์โปรตีน (globular protein) ตัวอย่างโปรตีนที่มีโครงสร้างระดับนี้คือ ไมโอโกลบิน

## 4. โครงสร้างแบบจตุรภูมิ (Quaternary Structure)

โปรตีนบางชนิดประกอบด้วยสายโพลีเปปไทด์มากกว่า 1 สาย สายโพลีเปปไทด์รวมกันอยู่เป็นกลุ่มได้พันธะต่างๆ เช่นเดียวกับในโครงสร้างตติยภูมิ โพลีเปปไทด์แต่ละสายเป็นองค์ประกอบในโครงสร้างเรียกว่า หน่วยย่อย (subunit) โดยจะเป็นชนิดเดียวกันหรือต่างชนิดกันก็ได้ โปรตีนที่มีโครงสร้างระดับนี้ได้แก่ ฮีโมโกลบิน

เนื่องจากโปรตีนเป็นสารชีวโมเลกุลที่มีความสำคัญต่อสิ่งมีชีวิต ดังนั้นจึงมีการศึกษาโปรตีนกันอย่างกว้างขวาง โดยทั่วไปการศึกษาโปรตีน จะทำโดยการนำโปรตีนมาศึกษาจนทราบสมบัติและหน้าที่ แล้วจึงนำโปรตีนตัวนั้นไปหาลำดับของกรดอะมิโน แต่เนื่องจากปัจจุบันความก้าวหน้าทางวิศวกรรมพันธุศาสตร์และชีวโมเลกุล ทำให้สามารถทราบถึงลำดับของกรดอะมิโนจากการแปลรหัสบนดีเอ็นเอ และประกอบกับได้มีการเริ่มโครงการอ่านรหัสบนสารพันธุกรรมของมนุษย์ (Human Genome Project) และในสิ่งมีชีวิตอีกหลายชนิด ทำให้ฐานข้อมูลของลำดับอะมิโนของโปรตีนต่างๆมีการขยายตัวอย่างรวดเร็ว ทำให้นักวิทยาศาสตร์สนใจศึกษาโปรตีนโดยศึกษาจากลำดับของกรดอะมิโนมากขึ้น ซึ่งโครงการนี้นับเป็นส่วนหนึ่งของการศึกษาโปรตีนโดยศึกษาจากลำดับของกรดอะมิโนนั่นเอง



รูปที่ 2.6 รูปแสดงการศึกษาโปรตีน (Smith W., 1998)

## ฮอมอโลยีของโปรตีน (Homology of protein)

### ฮอมอโลยีและความคล้าย (Homology and similarity)

โปรตีนที่ฮอมอโลกัสกัน (Homologous protein) คือโปรตีนที่มีความสัมพันธ์ทางวิวัฒนาการใกล้ชิดกัน และโดยส่วนมากโปรตีนพวกนี้จะทำหน้าที่เดียวกันแม้ว่าจะอยู่ในสิ่งมีชีวิตที่ต่างกัน เช่น ฮีโมโกลบินก็จะทำหน้าที่ขนส่งออกซิเจนเหมือนกันไม่ว่าจะอยู่ในสิ่งมีชีวิตที่มีกระดูกสันหลังชนิดใดก็ตาม

โดยทั่วไปจะพบว่าโปรตีนที่มีลำดับของกรดอะมิโนคล้ายกันก็มักจะเป็น โปรตีนที่ฮอมอโลกัสกัน แต่จะพบว่าโปรตีนที่ฮอมอโลกัสกันไม่จำเป็นต้องมีลำดับคล้ายกันอาจเป็นผลมาจากความต่างกันในส่วนของ variable residues หรือ อาจเป็นเพราะ ช่องว่าง (Gap) ทำให้การวิเคราะห์ฮอมอโลกัส ของโปรตีนโดยพิจารณาความเหมือนกันของลำดับของกรดอะมิโนจึงใช้ไม่ได้ผล

### ความสัมพันธ์ระหว่าง โครงสร้างสามมิติ และ ฮอโมโลยี

โปรตีนที่ฮอมอโลกัสกันคือโปรตีนที่มีต้นกำเนิดเดียวกันแต่ทำหน้าที่เดียวกัน และเราพบว่าโปรตีนที่ทำหน้าที่เดียวกันก็มักมีโครงสร้าง 3 มิติ (three-dimensional structure) คล้ายๆกัน ดังนั้นหากเรานำโปรตีน A ที่ไม่ทราบโครงสร้าง 3 มิติ ไปทดสอบฮอมอโลยีกับโปรตีน B ซึ่งทราบโครงสร้าง 3 มิติแล้ว ถ้าโปรตีน A ฮอมอโลยีกับโปรตีน B แสดงว่าโปรตีน A น่าจะมีโครงสร้าง 3 มิติคล้ายกับโปรตีน B นั่นคือการทดสอบฮอมอโลยีของโปรตีนเป็นเครื่องมืออย่างหนึ่งที่จะช่วยในการพิจารณาโครงสร้าง 3 มิติของโปรตีน

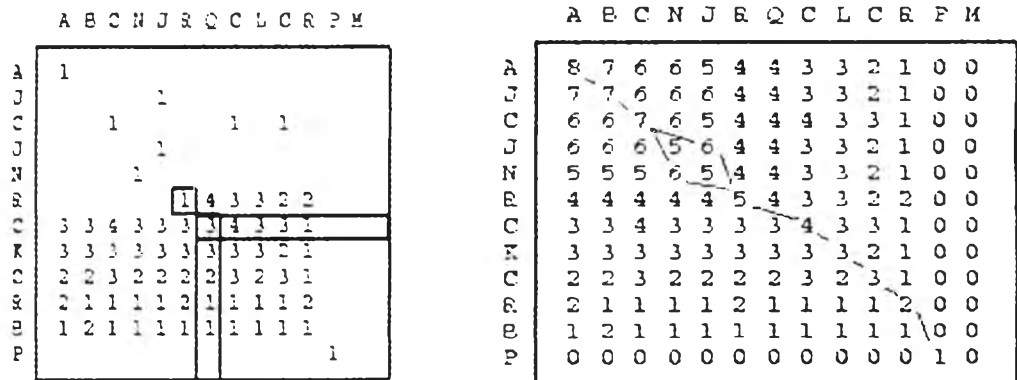
### เครื่องมือในการทดสอบความคล้ายกันของโปรตีน

มีวิธีการมากมายในการที่จะทดสอบความคล้ายกันของโปรตีนไม่ว่าจะเป็น Dynamic programming methods ของ (Needleman., et al., 1969), BLAST (Altschul et al., 1990), Hidden markov models (Kevin., et al ,1999) และอื่นๆอีกมากมายแต่ในที่นี้จะกล่าวถึง 2 วิธีคือ

1. Dynamic Programming methods ของ Needleman

วิธีการนี้แม้ว่าจะเป็นวิธีที่มีมานาน และมีความถูกต้องน้อยแต่ในปัจจุบันก็ยังใช้กันอยู่ เนื่องจากเป็นวิธีการที่ง่าย

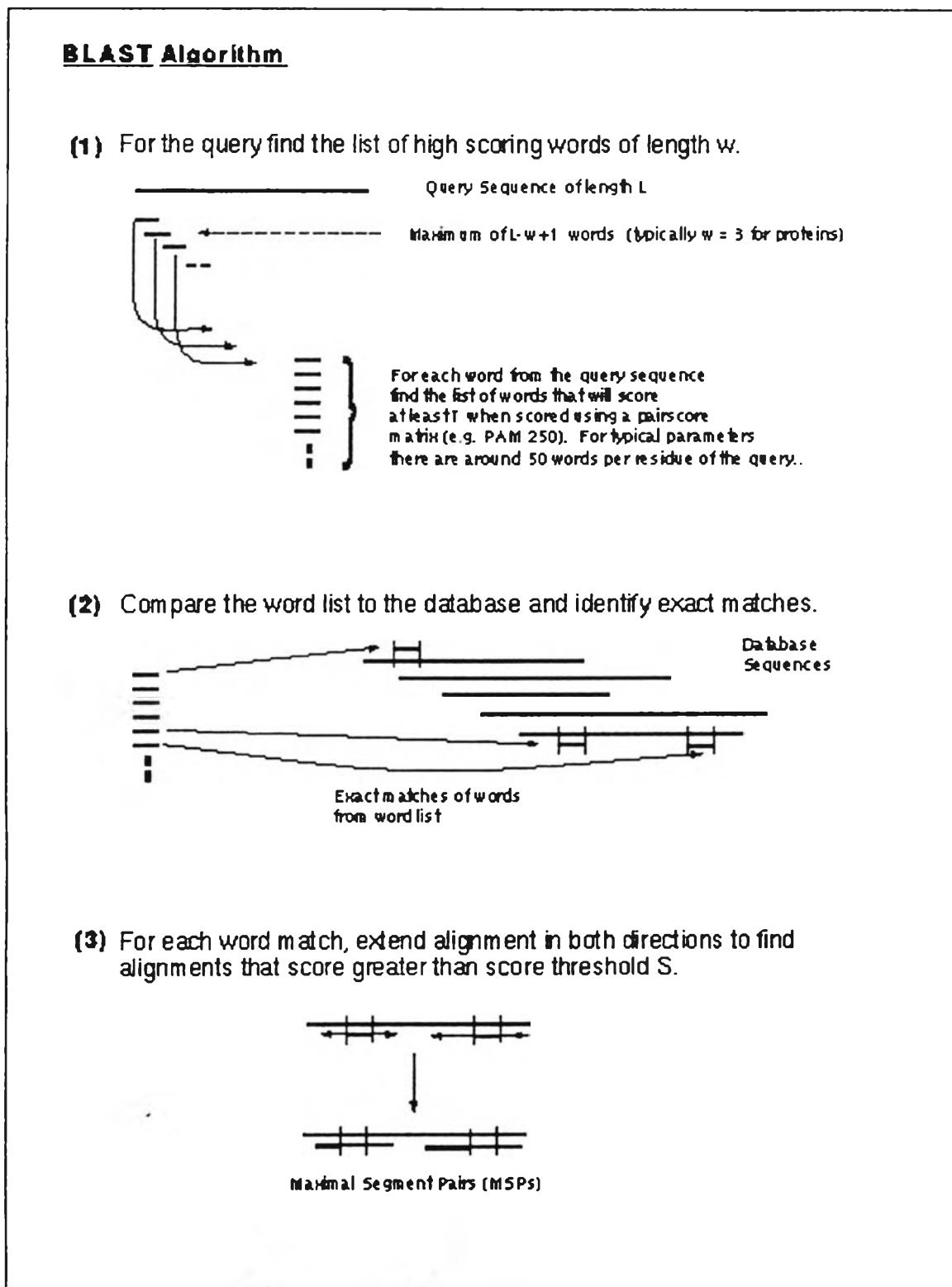
หลักการคือเริ่มจากการ สร้างเมตริกซ์ที่เป็นตัวแทนของลำดับของกรดอะมิโนขึ้นมาโดยนำลำดับของกรดอะมิโนมาโดยนำลำดับของกรดอะมิโนมาเขียนตามแถวและหลักของเมตริกซ์ แล้วกำหนดและปรับแต่งค่าให้ตำแหน่งที่กรดอะมิโน แล้วก็เริ่มจากมุมล่างด้านขวา ดูไปตามแถวและหลักว่ามีค่าตัวเลขตัวไหนมากที่สุด จากนั้นก็สร้างทางเดินโดยเลือกจากตัวที่มีค่ามากที่สุดแล้วก็ทำการบวกันในแถวหรือหลักถัดไปจนครบ ที่มุมบนด้านซ้ายดูไปทางตำแหน่งที่อยู่ที่อยู่ทางขวา และต่ำกว่าว่าตัวเลขที่มีค่ามากที่สุดให้เลือกตัวนั้น ทำไปจนหมดเมตริกซ์ แล้วจะได้ว่าเส้นทางที่ได้จะเป็นเส้นทางที่แสดงตำแหน่งของอะมิโนที่ตรงกัน



รูปที่ 2.7 รูปแสดง Dynamic programming methods ของ Needleman(Term H., 1999.)

พบว่าวิธีการนี้ หากจำนวนของกรดอะมิโนต่างกันมากๆ หรือ มีช่องว่าง (gap) ขนาดใหญ่ หรือจำนวนมาก วิธีการนี้ก็ใช้ไม่ได้ผล

## 2. BLAST



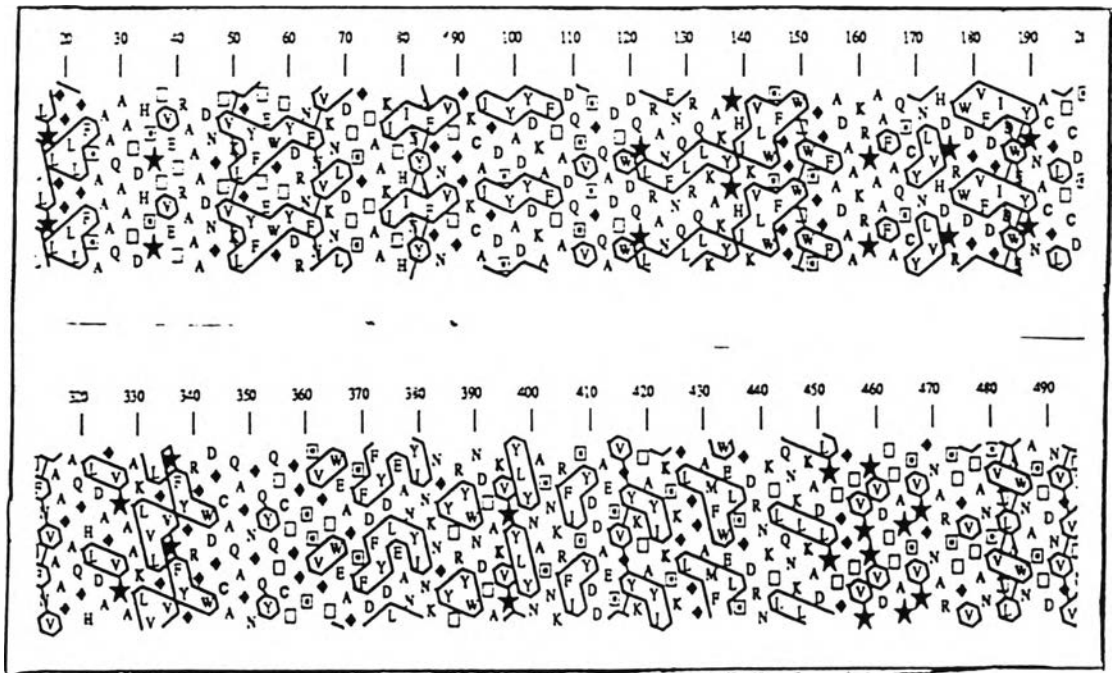
รูปที่ 2.8 รูปแสดงวิธีการ BLAST(Tern H., 1999.)



BLAST ย่อมาจาก Basic Local Alignment Search Tool เป็นการใช้วิธีการทางสถิติและความรู้ในการค้นหาข้อมูลในฐานข้อมูล โดยพัฒนามาตั้งแต่ปี 1980 ซึ่งเป็นที่นิยมในปัจจุบันเป็นอย่างมาก

#### Hydrophobic Cluster Analysis (HCA)

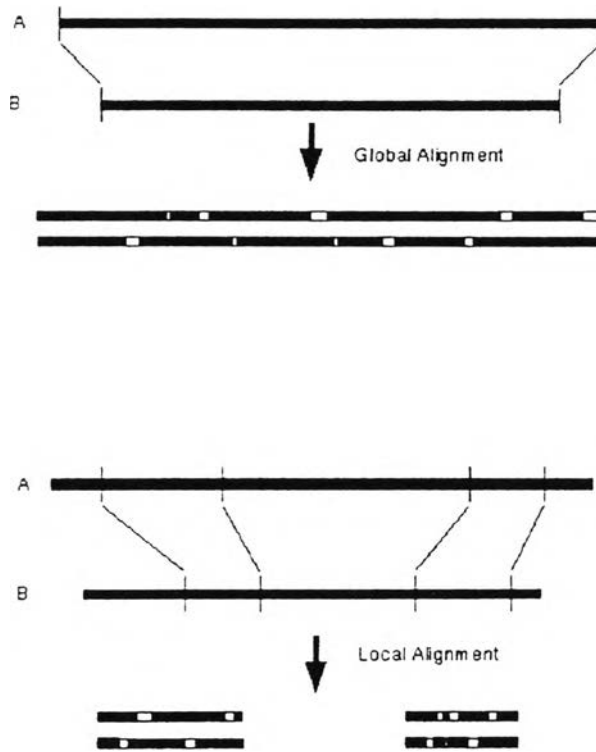
เป็นวิธีการที่สามารถใช้ได้ในการทำนายโครงสร้างทุติยภูมิและทดสอบความคล้ายกันของลำดับของกรดอะมิโนในวิธีการนี้ใช้หลักการเปลี่ยนลำดับของกรดอะมิโนให้อยู่ในรูปของรูปของแผนภูมิ 2 มิติ โดยเริ่มจากการนำมาทดสอบความคล้ายกันของลำดับของกรดอะมิโน แต่เนื่องจากเป็นวิธีการที่ต้องใช้ความชำนาญในการพิจารณาความเหมือนของแผนภูมิที่สร้างขึ้น จึงไม่นิยมใช้วิธีการนี้



รูปที่ 2.9 รูปแสดงแผนภูมิ 2 มิติจากวิธีการ HCA (Karlin and Altschul., 1987)

### Local VS Global Alignment

เป็นวิธีการทดสอบความคล้ายกันของกรดอะมิโน โดยglobal alignment จะเป็นการพิจารณาความเหมือนของลำดับของกรดอะมิโนตลอดทั้ง 2 โมเลกุลของโปรตีน แต่ Local alignment จะทำการแยกพิจารณาเป็นส่วนๆ โดยตัดโปรตีนออกเป็นช่องๆ



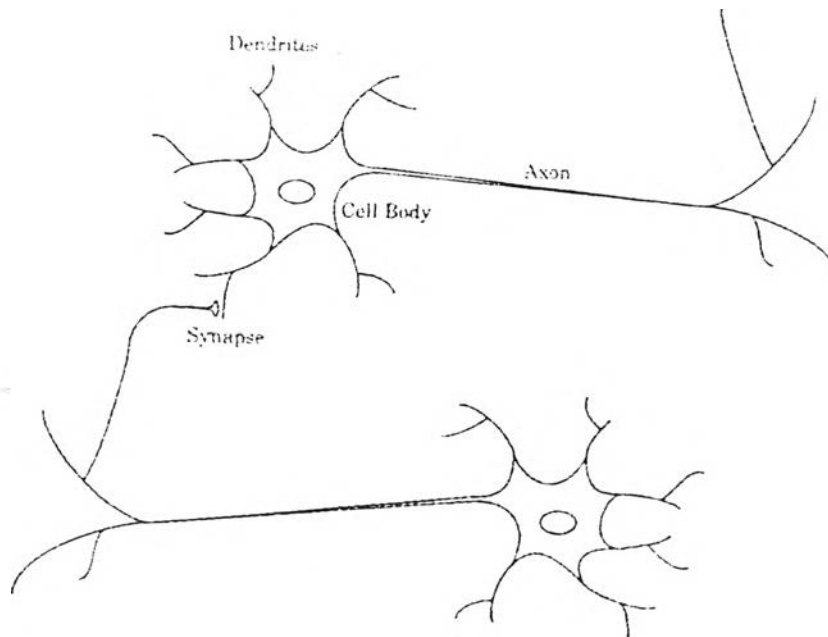
รูปที่ 2.10 รูปแสดง Local VS Global Alignment (Term H., 1999.)

## โครงข่ายประสาทเทียม (Neural Networks)

### ความรู้เกี่ยวกับระบบประสาท (Biological neural system)

สมองประกอบด้วยองค์ประกอบที่เชื่อมต่อกันเรียกว่า neuron (มีประมาณ  $10^{11}$  ตัว) มีส่วนประกอบ 3 อย่างคือ

- dendrite ทำหน้าที่รับ-ส่งไปยังตัวเซลล์
- cell body ทำหน้าที่แปรสัญญาณจาก dendrite
- axon ทำหน้าที่ส่งผลลัพธ์ ไปยัง neuron อื่น ๆ



รูปที่ 2.11 รูปแสดงระบบประสาท (Demuth H.B., Beale M. and Hagan M.T., 1995.)

synapse เป็นตำแหน่งที่ dendrite อยู่ใกล้กับ axon จนทำให้เกิดการส่งสัญญาณ โดยการแลกเปลี่ยนสารเคมี ทำให้เกิดการตอบสนองต่อสัญญาณต่างๆ ได้ และเมื่อเกิดไปนานแล้วจะทำให้เกิดการเรียนรู้และตอบสนองทันทีที่มีสัญญาณแบบที่เคยรับมาก่อน

โครงข่ายประสาทเทียมไม่ได้มีความซับซ้อนเหมือนกับสมอง แต่มี 2 อย่างที่เหมือนกันคือ

1. ลักษณะทางกายภาพของระบบประสาทของสิ่งมีชีวิต และ ระบบโครงข่ายประสาท เหมือนกัน แต่ ระบบโครงข่ายประสาท มีรูปแบบง่ายกว่า

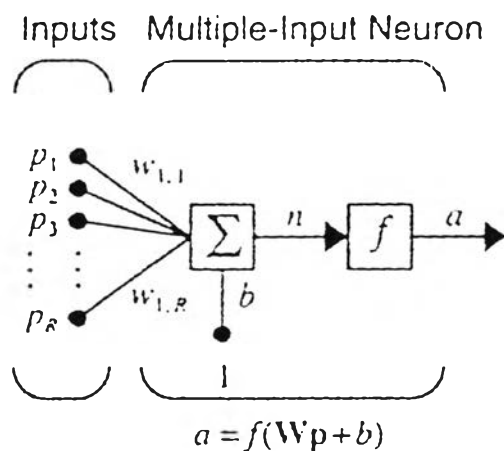
2. บน ระบบโครงข่ายประสาท มีการเชื่อมต่อแบบเดียวกับ synapse แต่ใช้อธิบายโดยใช้ function ของระบบโครงข่ายประสาท แทน

**แบบจำลองโครงข่ายประสาทเทียม (Neural Model)**

**ระบบโครงข่ายประสาทแบบหลาย input**

เป็น ระบบโครงข่ายประสาท ที่มี input มากกว่า 1 ตัว

- ให้  $R$  คือ Input ของ neuron
- $P_j$  คือ สมาชิกตัวที่  $j$
- $W_{1,j}$  คือ ค่าของสมาชิกตัวที่  $j$  ในครั้งที่  $1$
- $n$  คือ ผลรวมของผลคูณระหว่าง input ตัวที่  $P_j$



รูปที่ 2.12 รูปแสดงโครงข่ายประสาทเทียมแบบ Multiple-Input (Demuth H.B., Beale M. and Hagan M.T., 1995.)

$$n = w_{1,1}P_1 + w_{1,2}P_2 + \dots + w_{1,r}P_r + b$$

$$a = \text{ผลลัพธ์ของ neuron}$$

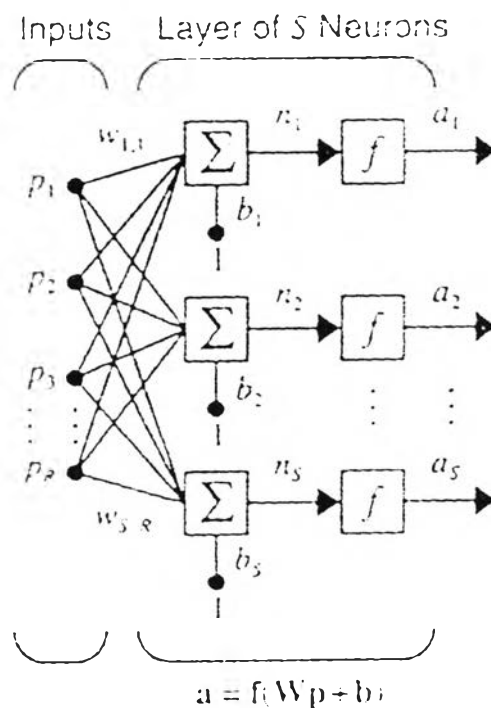
$$= f(Wp+b)$$

$$b = \text{bias}$$

### Network Architecture

บางครั้งการใช้ 1 neuron ในการตัดสินใจคำตอบ  $a$  ซึ่งอาจไม่เหมาะสมสำหรับ input หลาย ๆ ตัว เพราะคำตอบที่ได้อาจจะไม่ถูกต้อง จำเป็นต้องใช้หลาย ๆ neuron ในการสรุปคำตอบ เหมือนกับการลงมติในที่ประชุม ถ้าได้รับคะแนนมากกว่า ก็น่าจะมีความน่าเชื่อถือมากกว่าได้รับคะแนนน้อย

### Single Layer Network ของ S Neuron

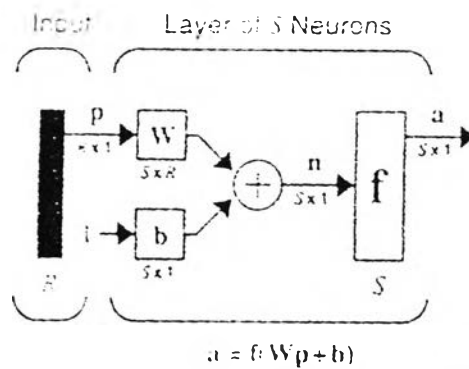


รูปที่ 2.13 รูปแสดง Layer ของโครงข่ายประสาทเทียมแบบ Single-Layer (Demuth H.B., Beale M. and Hagan M.T., 1995.)

$$a = f(Wp+b)$$

แต่ละ input  $p$  เชื่อมต่อกับแต่ละ neuron ด้วย weight matrix แต่ละ neuron มี bias  $b$  ของตัวเอง มี transfer function  $f$  และผลลัพธ์คือ  $a$ , สมาชิก input มี weight matrix ของ  $W$

$$W = \begin{bmatrix} W_{1,1} & W_{1,2} & \cdots & W_{1,R} \\ W_{2,1} & W_{2,2} & \cdots & W_{2,R} \\ \vdots & \vdots & \ddots & \vdots \\ W_{S,1} & W_{S,2} & \cdots & W_{S,R} \end{bmatrix}$$



รูปที่ 2.14 รูปแสดง Layer ของโครงข่ายประสาทเทียมแบบ Single-Layer ที่ได้มีการปรับแต่งแล้ว (Demuth H.B., Beale M. and Hagan M.T., 1995.)

### McCulloch and Pitts Model (MCP)

เป็น model ง่าย ๆ ของการเลือกผลลัพธ์แบบทวิภาค เป็นการคำนวณผลรวมของค่าน้ำหนักของแต่ละหน่วย และผลลัพธ์มีเพียง 0 และ 1 เท่านั้น ผลลัพธ์พิจารณาจากฟังก์ชันต่อไปนี้

$$n_j(t + 1) = \theta(\sum_j w_{ij}n_j(t) - u_i)$$

และ  $\theta(x)$  เป็น step function โดย

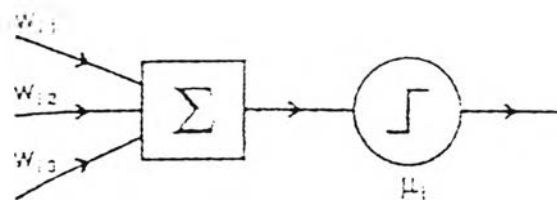
$\theta(x)$  จะเท่ากับ 1 ถ้า  $x \geq 0$

จะเท่ากับ 0 กรณีอื่น ๆ

ค่า  $w_{ij}$  เป็นค่าความยาวของ synapse ระหว่าง neuron j กับ neuron i คำนี้นี้เป็นบวก หรือ ลบ ก็ได้ขึ้นกับว่าเป็นภาวะถูกกระตุ้นหรือเป็นภาวะเฉื่อย และถ้าไม่มี synapse เลยระหว่าง neuron j กับ neuron i ค่า  $w_{ij}=0$

$u_i$  = เป็นค่าเข้าใกล้ของหน่วยที่ i คำนี้นักจะนำไปคำนวณในครั้งถัดไป

$n_j$  = เป็นสถานะ neuron i จะเป็น 1 หรือเป็น 0



รูปที่ 2.15 รูปแสดงแบบจำลองแบบ McCulloch and Pitts (Krongh A., Hertz J. and Palmer R.G., 1991)

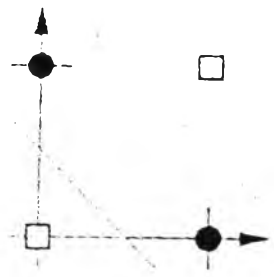
$$n_i = g(\sum_j w_{ij}n_{ij} - u_{ij})$$

$n_i$  เป็นค่าต่อเนื่องและบอกสถานะของหน่วยที่ i

$g(x)$  เป็นสมการ nonlinear อาจจะเป็น activation function หรือ gain function หรือ transfer function ก็ได้

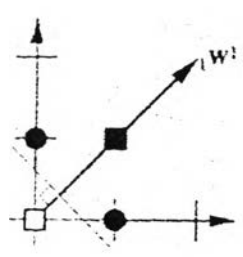
Pattern Classification

เป็นการแบ่งกลุ่มของ pattern ต่าง ๆ เช่นสำหรับ XOR แบ่งผลลัพธ์ได้เป็น 2 กลุ่มคือผลลัพธ์ที่เป็น 0 และผลลัพธ์ที่เป็น 1

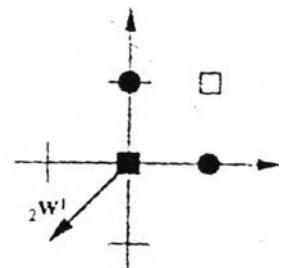


$p_1=[0 \ 0]^t$	,	$t_1=0$
$p_2=[0 \ 1]^t$	,	$t_2=0$
$p_3=[1 \ 0]^t$	,	$t_3=0$
$p_4=[1 \ 1]^t$	,	$t_4=0$

เราไม่สามารถใช้ single layer แบ่งกลุ่มของปัญหา XOR ได้ต้องใช้ two-layer network โดย layer แรกแบ่ง  $p_1$  และ layer ที่ 2 แบ่ง  $p_4$



Layer 1/Neuron 1



Layer 1/Neuron 2

รูปที่ 2.16 รูปแสดงการสร้างขอบเขตเพื่อจำแนกข้อมูลเป็นกลุ่ม (Demuth H.B., Beale M. and Hagan M.T., 1995.)

The Backpropagation

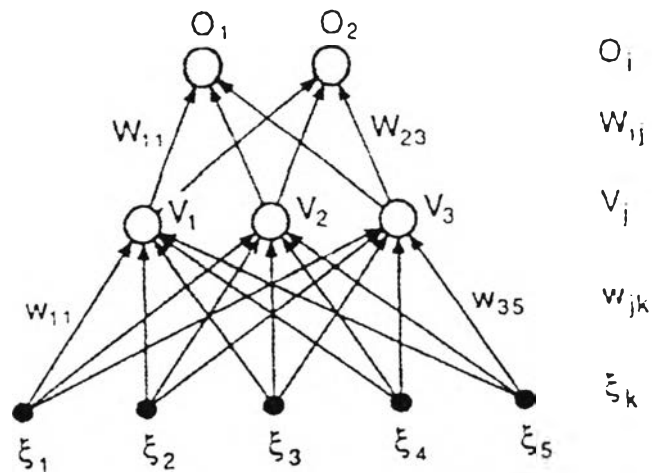
เป็นวิธีมาตรฐานสำหรับงานด้าน neuron network มีการคิดค้นโดย Bry และ Ho (1969), Werbos(1974), Parker(1985) และ Rumelhart et al(1986a,b) โดยมีวิธีการเสนอการเปลี่ยนค่าน้ำหนัก  $w_{pq}$  ในแต่ละชั้นของ feed-forward network ในการเริ่มต้นเรียนรู้ของชุดเรียนรู้แต่ละคู่ input และ output  $\{ \xi_i^u, \zeta_i^u \}$  วิธีการที่นิยมใช้วิธีหนึ่งคือ gradient descent



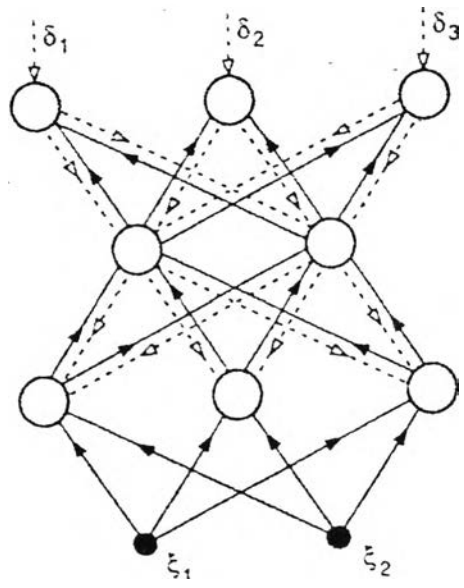
$$E[w] = \frac{1}{2} \sum_{iu} (\xi_i^u - O_i^u)^2 = \frac{1}{2} \sum_i (G_i^u - \sum_k w_{ik} \xi_k^u)^2$$

ให้ค่าความผิดพลาด  $E[w]$  เราสามารถปรับปรุงกลุ่มของ  $w_{ik}$  โดยลดค่าที่กำหนดในแต่ละมิติ  $w$  ลง โดยที่ gradient descent ของ  $E$  คือ

$$\begin{aligned} \Delta w_{ik} &= -\eta \frac{\partial E}{\partial w_{ik}} \\ &= \eta \sum_u (G_i^u - O_i^u) \xi_k^u \end{aligned}$$



รูปที่ 2.17 รูปแสดง Feedforward network (Krogh A., Hertz J. and Palmer R.G., 1991)



รูปที่ 2.18 รูปแสดง Black-propagation ใน 3-Layer network (Krongh A., Hertz J. and Palmer R.G., 1991)

$$\Delta w_{pq} = \eta \sum_{\text{pattern}} \delta_{\text{output}} \times V_{\text{input}}$$

**ขั้นตอน Backproagation**

1. เริ่มต้นโดยการสุ่มน้ำหนักน้อย ๆ
2. เลือก pattern  $\xi_k^u$  และผ่านเข้าไปใน input layer ( $m = 0$ ) จะได้

$$V_k^0 = \xi_k^u$$

3. propagate สัญญาณผ่านเข้าไปยัง network โดย

$$V_i^m = g(h_i^m) = g(\sum w_{ij}^m V_j^{m-1})$$

สำหรับแต่ละ  $i$  และ  $m$  จนกระทั่งได้ผลลัพธ์สุดท้าย  $V_i^m$

4. หาค่า delta สำหรับ output layer จาก

$$\delta_i^m = g'(h_i^m) [\xi_i^u - V_i^m]$$

เป็นการเปรียบเทียบค่าที่ได้กับค่าที่ได้จากการคำนวณ  $\xi_i^u$  สำหรับแต่ละชุดข้อมูล

5. คำนวณค่า delta สำหรับ preceding layers โดย propagating ค่าผิดพลาด

$$\delta_1^{m-1} = g'(h_1^{m-1}) \sum w_{1j}^m \delta_j^m$$

สำหรับ  $m = M, M-1, \dots, 2$  จนกระทั่งค่า delta ถูกคำนวณหมด

6. ใช้

$$\Delta w_{ij}^m = \eta \delta_i^m V_j^{m-1}$$

ในการปรับปรุงค่าเชื่อมต่อเป็น  $w_{ij}^{m+1} = w_{ij}^{old} + \Delta w_{ij}^m$

7. กลับไปทำขั้นที่ 2 และทำในชุดถัดไปเรื่อย ๆ