บทที่ 3
วัสดุอุปกรณ์และวิธีการ (Material and Methods)

วัสดุอุปกรณ์ (Materials)

Computer Hardware

- Pentium II 300
- Main memory: 64 MB
- Hard disk: 6.4 GB
- VGA Card: 8 MB

Computer Software

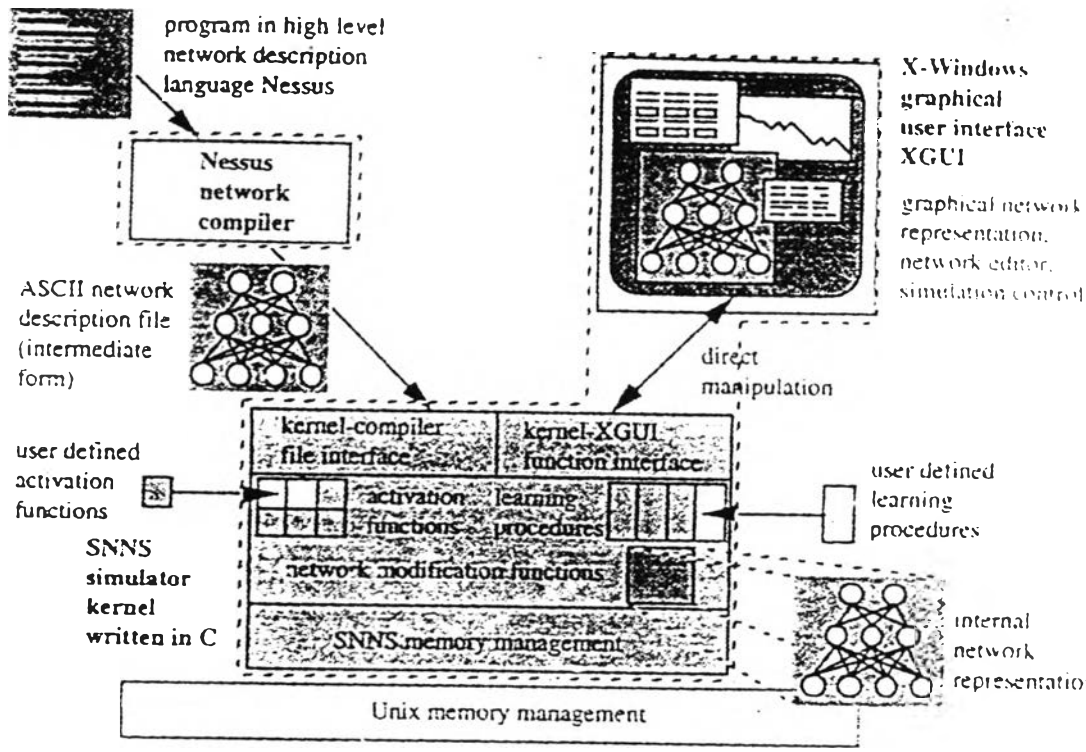
Linux

เป็นระบบปฏิบัติการที่มีการเปิดเผย source code และเป็นของฟรี ที่มีการพัฒนาระบบให้เทียบเท่าระบบ UNIX มีความสมบูรณ์ มีโปรแกรมให้ใช้มากมาย เป็นระบบที่มีความเสถียรและมีความคงตัวเป็นเวลาหลายทศวรรษ และเหมาะสำหรับงานด้านวิจัย เนื่องจากมีโปรแกรมให้ใช้ครบทุกด้าน ทั้ง word processing, databases, presentation, spred sheets, และอื่น ๆ โดยโปรแกรมเหล่านี้มีทั้งที่เสียค่าใช้จ่ายและไม่เสียค่าใช่จ่าย

SNNS (Stuttgart Neural Network Simulator)

เป็นโปรแกรมที่ใช้ในการจำลองลักษณะการทำงานของ neural network ประกอบไปด้วยส่วน kernel และ ส่วนแสดงผลทางจอภาพ (GUI) โดยที่ SNNS นั้นสามารถใช้งานได้บนระบบคอมพิวเตอร์หลายๆ แบบ SNNS พัฒนาโดย The Institue for Parallel and Distributed High Performance System at Stattgart University (Andreas, *et al.*,1987)

โปรแกรม SNNS ประกอบด้วยส่วนหลักๆ ดังภาพข้างล่างนี้ โดยแบ่งออกเป็น simulator kernel, graphical user interface, batch simulator version snnsbat



รูปที่ 3.1 รูปแสดง องค์กอบของ SNNS (Andreas, *et al.*, 1987.)

ส่วนประกอบเหล่านี้ใช้ในการสร้างระบบโครงข่ายประสาทของ SNNS โดยที่ simulator kernel ทำหน้าที่ประมวลผลข้อมูล ส่งผ่านผลลัพธ์ที่ได้ นำเสนอในรูป graphic โดย graphical user interface ที่มีชื่อว่า xgui ในขณะเดียวกัน สามารถใช้ graphical user interface ในการสร้างประมวลผลและแสดงผลลัพธ์ของระบบโครงข่าย-ประสาทได้ และมีคำแนะนำการใช้งานให้กับผู้ใช้งานด้วย ส่วนนี้จึงเหมาะสำหรับผู้ใช้งานเบื้องต้น ส่วน batch simulator version snnsbat ใช้สำหรับประมวลผลแบบ batch สามารถกำหนดเวลาสำหรับประมวลผลได้ ส่วน network compiler snns2c ใช้สำหรับแปลความหมายของ network (ศึกษาได้จากคู่มือการใช้ SNNS หรือที่ http://durandal.mass.ubordeaux2.fr/~corsini/SNNS_Manual/node1.html)

โปรแกรม SNNS นี้ยังมีความสามารถหลายอย่างเช่น เราสามารถหยุดการทำงานได้ โดยนำผลลัพธ์นั้นไปใช้ประมวลผลได้อีกในครั้งถัดไป โปรแกรม SNNS ทำงานได้บนหลายระบบปฏิบัติการ ดังตารางด้านล่าง

ชนิดของเครื่องคอมพิวเตอร์ (Machine type)	**ระบบปฏิบัติการ** (Operating System)
SUN SparcSt. ELC,IPC	SunOS 4.1.2,4.1.3,5.3,5.4
SUN SparcSt. 2	SunOS 4.1.2
SUN SparcSt. 5,10,20	SunOS 4.1.3,5.3,5.4
DECstation 3100,5000	Uitrix V4.2
DEC Alpha AXP 3000	OSF V2.1
IBM-PC 80486, Pentium family	Linux
IBM RS 6000/320,3220H, 530H	AIX V3.1, AIX V32.2
HP 9000/720,730	HP-UX 8.07
SGI Indigo 2	IRIX 4.0.5,5.3

ตารางที่ 3.1 **ตารางแสดงเครื่องคอมพิวเตอร์ และระบบปฏิบัติการที่ใช้** SNNS (Andreas, *et al.*, 1987)

โปรแกรม SNNS พัฒนามาจากโปรแกรมภาษา C (ANSI-C) โดยที่ส่วน simulator ได้ผ่านการทดสอบจากระบบปฏิบัติการดังตารางข้างต้นแล้ว ส่วน grapnical user interface นั้นทดสอบผ่าน x-windows manager เช่น twm,tvtwm, olwm,fvwm บนมาตรฐาน X11R5 และ X11R6 ถ้าไม่สามารถใช้ graphical user interface ได้ สามารถทำงานแบบตัวอักษร (text mode) แทนได้

Visual C++ เป็นเครื่องมือที่ใช้ในการสร้าง โปรแกรม 32 บิท สำหรับระบบปฏิบัติการ Windows 95 และ Windows NT ซึ่งมีรูปแบบการติดต่อกับผู้ใช้หรือ user interface ที่ง่ายและสะดวก นอกจากนี้ยังสามารถใช้ในการโปรแกรมได้หลายๆชนิดไม่ว่าจะเป็นโปรแกรมทั่วไป โปรแกรมจัดการฐานข้อมูล หรือโปรแกรมสร้างระบบเครือข่ายเป็นต้น

Protein Database ข้อมูลของลำดับของโปรตีนได้มาจาก NCBI และ PBD

URL => http://www.ncbi.nlm.nih.gov/

URL => http://www.rcsb.org/pdb/

Methods

ในการสร้างแผนภูมิ 2 มิติจะใช้รูปแบบของ 2-D α-Helical pattern (Schiffer and Edmunson,1967.) ซึ่งประสบความสำเร็จในการทำนายโครงสร้างทุติยภูมิ (secondary structure) ของโปรตีน (Woodcock, Protein Engineering. , 1992.) แล้วมีการรวมกรดอะมิโนเป็น cluster โดยใช้สมบัติทางเคมีของกรดอะมิโนแต่ละตัวเป็นตัวจำแนก จากนั้นก็จะใช้โครงข่ายประสาทเทียมเป็นตัวช่วยในการทดสอบความเหมือนหรือต่างกันของแผนภูมิเล็กๆ ที่สร้างขึ้น

ดังนั้นในงานวิจัยครั้งนี้จะแบ่งงานออกเป็น 2 ส่วนคือ

1. สร้างแแผนภูมิ 2 มิติของกรดอะมิโน

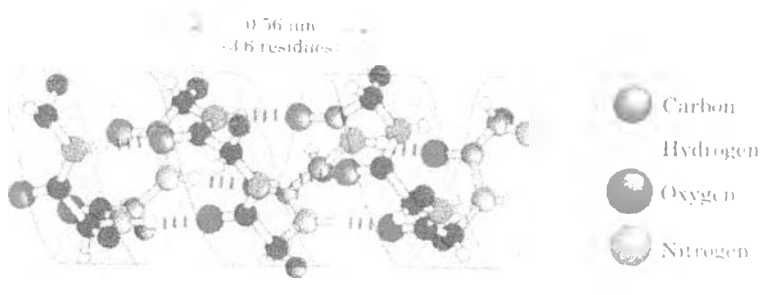
2. สร้างโครงข่ายประสาทเทียม

สร้างแแผนภูมิ 2 มิติของกรดอะมิโน

แบ่งเป็นขั้นตอนดังนี้

1. การสร้างแผนภูมิโดยการเปลี่ยนลำดับของกรดอะมิโน ให้อยู่ในรูปของ 2-D α-helical pattern โดย

1.1 นำลำดับของกรดอะมิโน มาพันเป็นรูปบันไดเวียนเรียบ (helix smoothed) รอบทรงกระบอก โดยแต่ละรอบจะประกอบด้วย 3.6 อะมิโน (Henrissat *et al.*,1988.)
ดังนั้น กรดอะมิโนตัวแรกที่จะกลับมาที่ตำแหน่งเดิม จะเป็นกรดอะมิโนตัวที่ 19 นั้นคือจะประกอบด้วยกรดอะมิโน จำนวน 18 อะมิโนต่อคาบ(period) หรือ 5 รอบ ต่อคาบนั่นเอง

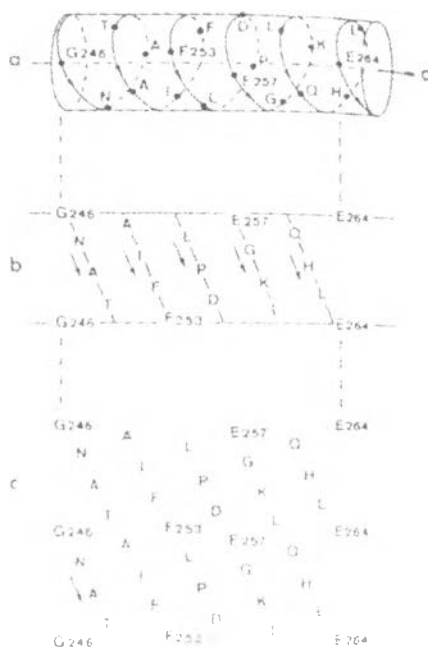


รูปที่ 3.2 **รูปแสดงอัลฟาฮีลิกที่มีขนาดกรดอะมิโน** 3.6 **ตัวต่อรอบ** (Cox., Lenhinger. and Nelson., 1993.)

1.2 ตัดและคลี่ทรงกระบอกออกเป็นแผ่น (b)

1.3 นำรูปที่ได้จากข้อ 2. สองรูปมาต่อกัน (c)

ดังรูป 3.3



รูปที่ 3.3 **รูปแสดงการสร้าง** 2-D α-Helical pattern (Bissery V., Gadorriaud C., Henrissat B., Lemesie L., Morgat V. and Mornon. JP. 1990.)

I 19861143

ขั้นตอนวิธีการ(Algorithm)

จะทำการเปลี่ยนลำดับของกรดอะมิโนให้อยู่ในรูปของ2-D α-Helical pattern นั้นคือเปลี่ยนจาก ข้อมูลที่เป็นลำดับให้กลายเป็นคู่ลำดับ

นั่นคือจาก i => (i , ID[i]) เช่น 1 => (1,ID[1])

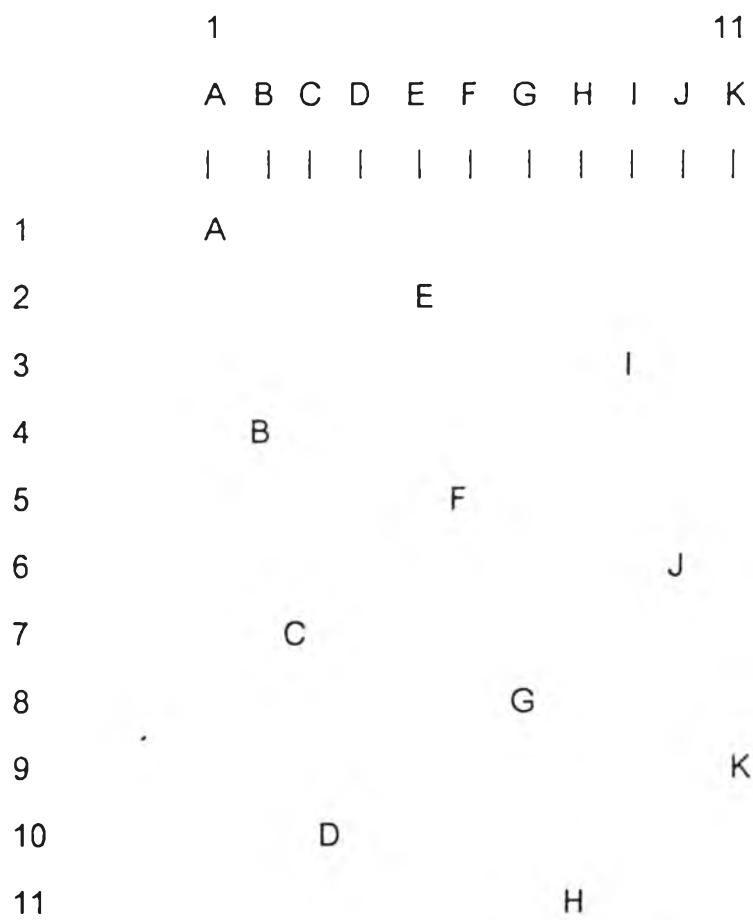
$$ID[i] = (I \times NT) \bmod NR$$

เมื่อ NT แทน จำนวนรอบใน 1 period

NR แทน จำนวนกรดอะมิโนใน 1 period

i แทน ลำดับของกรดอะมิโนตัวที่ I เมื่อ I = 0,1,2,3...

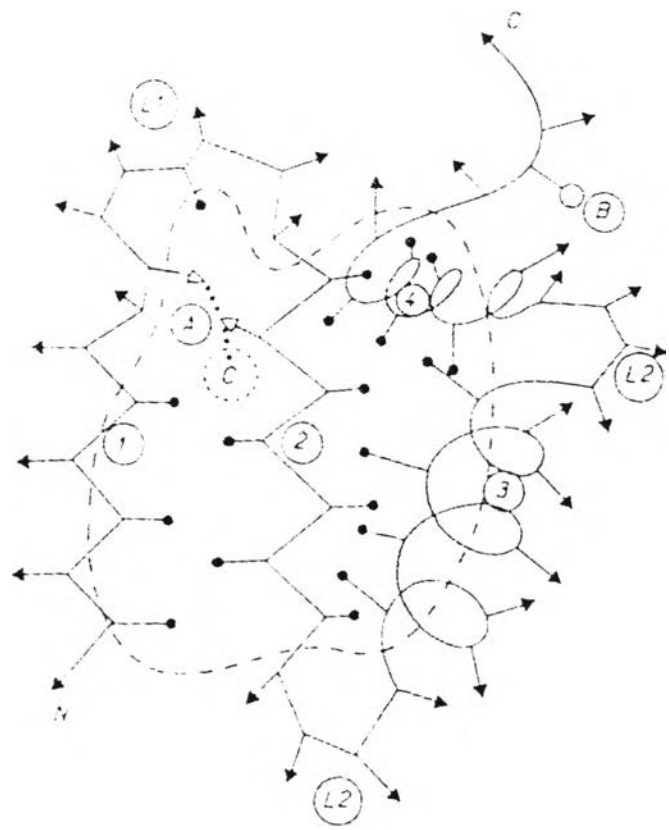
จะได้ดังรูป (NT = 3 , NR = 11)



โดยในโครงการนี้จะใช้ NT = 5 และ NR = 18

2. **การรวม** cluster

จากการสังเกตแผนภูมิ 2 มิติของกรดอะมิโนพบว่า กรดอะมิโนชนิด hydrophobic มักจะร่วมอยู่กันเป็นกลุ่ม ซึ่งมีลักษณะคล้ายกับการจัดเรียงตัวของโปรตีนในธรรมชาตินั้นคือ โปรตีนจะพยายามรักษาโครงสร้าง 3 มิติเอาไว้เพื่อที่จะทำให้สามารถทำหน้าที่นั้นๆได้ และโปรตีนส่วนใหญ่จะอยู่ในน้ำ จึงพบว่าโปรตีนจะพยายามรักษาโครงสร้าง 3 มิติ โดยนำส่วนที่เป็น hydrophobic ไว้ด้านใน และนำส่วนที่เป็น hydrophilic ไว้ด้านนอก ดังรูป 3.4



รูปที่ 3.4 **รูปแสดงโครงสร้างของโปรตีนที่พยายามนำส่วนที่เป็น** hydrophobic**ไว้ด้านใน และนำส่วนที่เป็น** hydrophilic**ออกมาด้านนอก** (Bissery V., Gadomiaud C., Henrissat B., Lemesie L., Morgat V. and Momon. JP. 1990.)

หมายเหตุ ส่วนที่ไม่ชอบน้ำคือส่วน ที่เป็นจุดวงกลมสีดำ ส่วนที่ชอบน้ำคือส่วนที่เป็นรูปสามเหลี่ยม

และจะพบว่าในโปรตีนที่ฮอมอโลกัสกันจะมีลักษณะของกลุ่มของกรดอะมิโนที่เป็นพวก hydrophobic (ต่อไปจะเรียกว่า cluster)คล้ายๆกันดังนั้นในการพิจารณาความคล้ายกันของแผนภูมิ 2 มิติของกรดอะมิโน จะพิจารณาจากกลุ่มของ cluster ก่อนเป็นอันดับแรก แล้วจากนั้นค่อยพิจารณากรดอะมิโนรอบข้าง cluster

จากการศึกษาพบว่ามีการให้มาตรค่าของ hydrophobicity ไว้ด้วยกันทั้งหมด 40 มาตรด้วยกัน (Cornett J.L., 1987) ในที่นี่เลือกมาตราของ G Von Heijne และ C Blomberg (1971) สาเหตุเลือกมาตราของ G Von Heijne และ C Blomberg เพราะว่าคำนวณค่า hydrophobicity โดยพิจารณาจากการกระจายตัวของกรดอะมิโนในตัวทำละลายที่ต่างกันแล้วคำนวณค่า hydrophobicity สัมพัทธ์เทียบกับ glycine จาก surface area ของอะมิโนแต่ละตัว และพบว่าลำดับ hydrophobicity ของกรดอะมิโนที่มาตรานี้ ก็ตรงกับมาตราใหม่ๆที่คำนวณออกมาภายหลังซึ่งมีค่าดังตาราง 3.2

ชนิดของอะมิโน	hydrophobicity	ชนิดของอะมิโน	hidorphobicity
Arginine	-11.2	Threonine	0.4
Glutamine	-9.9	Histidine	0.5
Aspartic Acid	-7.4	Alanine	0.5
Lysine	-4.2	Methionine	1.3
Proline	-3.3	Valine	1.5
Cysteine	-2.8	Leucine	1.8
Serine	-0.3	Tyrosine	2.3
Glutamic acid	-0.3	Phenylalanine	2.5
Asparagine	-0.2	Isoleucine	2.5
Glycine	0	Tryptophan	3.4

ตารางที่ 3.2 ตารางแสดงค่า hydrophobicity สัมพัทธ์เทียบกับ glycine (Creighton T.E., 1984.)

และได้พิจารณาสมบัติทางเคมีอื่นๆ ไม่ว่าจะเป็น ความเป็นกรด เบส ค่าความหนาแน่นประจุ ลักษณะที่พบในโครงสร้างสามมิติ โครงการนี้จึงกรดอะมิโนออกเป็น 7 กลุ่ม และมีการกำหนดสีประจำกลุ่มเพื่อช่วยในการพิจารณาความคล้ายของแผนภูมิ 2 มิติ ดังนี้

กลุ่มที่	ลักษณะ	กรดอะมิโน	สี
1	Hydrophobic	W , I , F ,Y , L, V, M	เขียว
2	Acidic / Neutral	D , E / N,Q	แดง / ม่วง
3	Basic	R,K,H	น้ำเงิน
4	Disulfide bonds	C	Cyan
5	Always Loop	S,T	เหลือง
6	Breaker	G,P	เทา
7	Small Hydrophobic	A	เขียวอ่อน

ตารางที่ 3.3 ตารางแสดงการแบ่งกลุ่มของกรดอะมิโนในการรวม cluster

โดยในกลุ่มของพวกมีสมบัติเป็น Hydrophobic ก็มีการกำหนดความเข้มสี ตามค่า hydrophobicityแต่ละตัวด้วยเพื่อแสดงลักษณะของ cluster

จากสมมติฐานที่ตั้งไว้ข้างต้นว่า แผนภูมิ 2 มิติของกรดอะมิโนที่สร้างขึ้นเป็นเครื่องมือที่ช่วยในการทดสอบฮอมอโลยีของโปรตีนโดนหลักการที่ว่าโปรตีนที่ฮอมอโลกัสกันจะให้แผนภูมิ 2 มิติของกรดอะมิโนที่คล้ายกันหรือเหมือนกัน และสามารถแก้ปัญหาในเรื่องของช่องว่างและsequence identity ของลำดับที่นำมาเปรียบเทียบมีค่าต่ำได้ จะทำการทดสอบโดยนำลำดับของกรดอะมิโนที่มาจากโปตีนที่ฮอมอโลยีกันจำนวน 30 คู่ ดังตารางที่ 4.1 ที่มี sequence identity ต่างๆกันตั้งแต่ 35%-98% มาสร้างแผนภูมิ 2 มิติของกรดอะมิโน แล้วพิจารณาแผนภูมิ 2 มิติของกรดอะมิโนที่สร้างขึ้นว่ามีลักษณะอย่างไร

การสร้างโครงข่ายประสาทเทียม

จากแผนภูมิสองมิติที่สร้างขึ้น พบว่ามีความยากลำบากในการพิจารณาความเหมือนหรือคล้ายกันของแผนภูมิที่สร้างขึ้นเนื่องจากผู้ใช้จะต้องมีประสบการณ์และความชำนาญสูง ดังนั้นในโครงงานนี้จึงได้ทำการทดลองนำโครงข่ายประสาทเทียมมาช่วยในการพิจารณาความเหมือนหรือคล้ายกันของแผนภูมิ 2 มิติ

การพิจารณาความเหมือนหรือคลายกันของแผนภูมิ 2 มิติของกรดอะมิโนจะพิจารณา แผนภูมิ 2 มิติที่เกิดจากลำดับของกรดอะมิโนจำนวน 18 ตัวสาเหตุที่เลือกรูปแบบที่เกิดจากลำดับของกรดอะมิโนจำนวน 18 ตัวเพราะการสร้าง แผนภูมิ 2 มิติเกิดจากการเอาลำดับของกรดอะมิโนมาพันเป็นเกลียวที่มีขนาด 18 อะมิโนต่อคาบ และสะดวกในการพิจารณาเลือกตัดมาจากโปรตีนต้นแบบ

การพิจารณาความคล้ายกันหรือเหมือนกันของกรดอะมิโนจะทำโดย จำแนกแผนภูมิ 2 มิติที่เกิดจากลำดับของกรดอะมิโนจำนวน 18 ตัว ออกเป็น pattern โดยแผนภูมิ 2 มิติที่เกิดจากลำดับของกรดอะมิโนจำนวน 18 ตัวที่เหมือนกันหรือคล้ายกันก็จะอยู่ใน pattern เดียวกัน จากนั้นเราจะทำการ train neural network ให้สามารถจำแนกแผนภูมิ 2 มิติที่เกิดจากลำดับของกรดอะมิโนจำนวน 18 ตัวที่ให้ว่าควรอยู่ว่าอยู่ใน pattern ไหน ถ้าแผนภูมิ 2 มิติที่เกิดจากลำดับของกรดอะมิโนจำนวน 18 ตัวที่ผ่านการจำแนกโดย neural network แล้วปรากฏว่าอยู่ใน pattern เดียวกันก็แสดงว่าแผนภูมิ 2 มิติที่เกิดจากลำดับของกรดอะมิโนจำนวน 18 ตัวนั้นเหมือนหรือคล้ายกัน

input

คือแผนภูมิ 2 มิติที่เกิดจากลำดับของกรดอะมิโนจำนวน 18 ตัว

เนื่องจากแผนภูมิ 2 มิติของกรดอะมิโนและลำดับของกรดอะมิโนมีความสัมพันธ์กัน จึงเลือกใช้ลำดับของกรดอะโนเพราะว่าสะดวกในการแปลงเป็นข้อมูลแบบทวิภาค (binary data)

โดยกำหนดให้กรดอะมิโนแต่ละชนิดแปลงเป็นข้อมูลแบบทวิภาคได้ดังนี้

ชนิดของอะมิโน	ข้อมูลแบบทวิภาค	ชนิดของอะมิโน	ข้อมูลแบบทวิภาค
Arginine	00000	Threonine	01010
Glutamine	00001	Histidine	01011
Aspartic Acid	00010	Alanine	01100
Lysine	00011	Methionine	01101
Proline	00100	Valine	01110
Cysteine	00101	Leucine	01111
Serine	00110	Tyrosine	10000
Glutamic acid	00111	Phenylalanine	10001
Asparagine	01000	Isoleucine	10010
Glycine	01001	Tryptophan	10011

ตารางที่ 3.4 ตารางแสดงการกำหนดค่าในขอมูลแบบทวิภาคให้กับกรดอะมิโนแต่ละตัว

output

คือ ชนิดของรูปแบบ(ในที่นี่กำหนดชนิดเป็นลำดับคือ ชนิดที่ 0, ชนิดที่ 1, ชนิดที่ 2,...) โดยจะเก็บอยู่ในรูปของข้อมูลทวิภาคที่มีขนาด 10 บิท

เช่น รูปแบบชนิดที่ 1 ก็จะเก็บในรูป 1000000000

รูปแบบชนิดที่ 5 ก็จะเก็บในรูป 1010000000

ตัวอย่าง input file

SNNS pattern definition file V3.2

generated at Wed Sep 1 15:58:23 1999

No. of patterns : 135

No. of input units : 90

No. of output units : 10

0 0 1 1 0 0 1 0 1 0 0 1 1 0 0 0 1 0 1 0 1 0 0 0 1

0 1 0 0 0 0 1 0 0 0 1 0 0 0 1 0 0 0 1 1 0 0 1 0 0

0 1 0 0 1 0 1 1 0 0 0 0 0 1 0 1 0 0 1 1 0 1 0 1 0

1 0 0 0 1 0 1 0 1 0 0 1 1 1 1

0 0 0 0 0 0 0 0 0 0

0 0 1 0 0 0 1 0 0 1 0 1 1 0 0 0 0 0 1 0 1 0 0 1 1

0 1 0 1 0 1 0 0 0 1 0 1 0 1 0 0 1 1 1 1 1 0 0 1 1

0 1 1 0 0 0 1 1 1 0 0 1 0 1 0 0 1 0 0 1 0 0 0 0 0

0 1 0 0 1 0 0 0 1 0 0 0 1 1 0

1 0 0 0 0 0 0 0 0 0

0 0 1 1 0 0 0 1 1 0 0 0 0 1 1 0 0 1 0 0 0 1 1 1 0

0 0 1 1 0 1 0 0 0 1 0 1 0 0 0 1 0 0 1 1 0 0 1 1 1

0 1 0 1 0 0 0 0 0 1 1 0 0 0 1 0 0 0 1 0 0 0 0 1 1

0 0 1 0 0 0 0 1 1 0 0 0 1 1 1

0 1 0 0 0 0 0 0 0 0

⋮ ⋮ ⋮ ⋮ ⋮

0 0 1 1 1 0 0 0 1 0 0 1 0 0 0 0 0 1 1 0 1 0 0 0 1

0 0 1 1 0 0 1 1 1 0 0 0 0 0 0 1 0 0 0 0 0 1 1 1 1

0 0 1 0 0 0 0 ์1 0 0 1 0 1 0 0 0 1 1 0 0 0 1 0 0

0 1 1 1 0 0 1 0 1 0 0 1 0 0 1

0 1 1 0 0 0 0 1 0 0

การสร้าง โครงข่ายประสาทเทียม

โดยปรกติแล้วการสร้างโครงข่ายประสาทเทียมขึ้นมามีความจำเป็นที่จะต้องสร้างหลายๆโครงข่ายโดยแต่ละโครงข่ายจะมีจำนวน hidden layer ต่างๆกัน เนื่องจากไม่สามารถบอกได้ว่าจำนวน hidden layer ที่เหมาะสมกับโครงข่ายประสาทเทียม นั้นๆ เป็นเท่าไร จึงมีความจำเป็นต้องสร้างโครงข่ายประสาทเทียม หลายๆตัวซึ่งมี hidden layer ที่ต่างๆกัน แล้วดูว่าตัวไหนมีเปอร์เซ็นต์ความถูกต้องสูงสุด ตัวนั้นก็จะเป็น โครงข่ายประสาทเทียม ที่มี hidden layer เหมาะสมกับงานนั้นที่สุด

ในงานวิจัยนี้ได้ทำการสร้าง โครงข่ายประสาทเทียมจำนวน 5 โครงข่ายบซึ่งมีจำนวน hidden layer 30, 40, 50, 60 และ 70

จากนั้นทำการทดสอบกับ test set 2 ชุด

ชุดที่ 1 มาจาก Training set 100 %

ชุดที่ 2 ไม่เคยผ่านการ Train มาก่อน 100 %

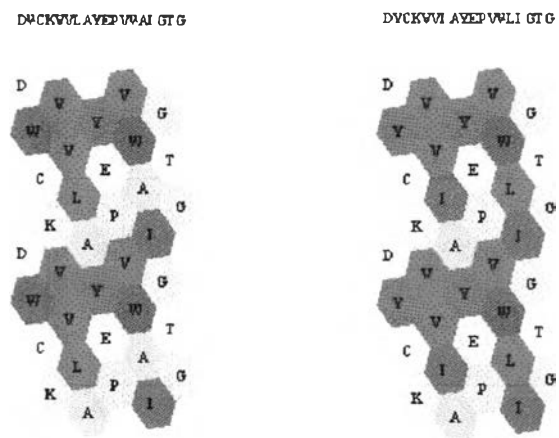
Training set and Test set

Training set and Test set ที่นำมาสร้าง neural network ทั้ง 5 ตัวที่กล้าวไว้ข้างต้นมาจากการตัดโปรตีนต้นแบบ 5 กลุ่ม (ภาคผนวกที่ 1) ออกเป็นรูปแบบชนิดต่างๆ โปรตีนทั้ง 5 กลุ่มนั้นได้แก่

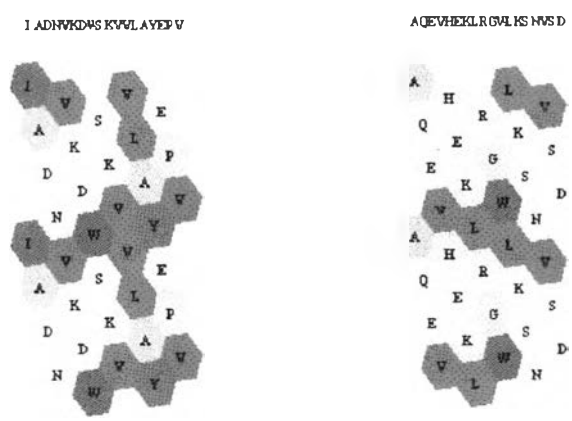
1. Myoglobin	จำนวน 17 ตัว
2. Fibronectin	จำนวน 14 ตัว
3. TBP	จำนวน 13 ตัว
4. Transmembran	จำนวน 13 ตัว
5. Triose phosphate Isomerase	จำนวน 14 ตัว

ในการตัดโปรตีนออกเป็นรูปแบบจะตัดเฉพาะบริเวณที่มี cluster เท่านั้น เพราะในการเปรียบเทียบความเหมือนหรือคล้ายกันของ แผนภูมิ 2 มิติของกรดอะมิโนเราจะพิจารณาจาก cluster เป็นอันดับแรก จากนั้นจึงพิจารณาสมบัติของกรดอะมิโนที่อยู่รอบข้าง cluster

จากโปรตีนทั้ง 5 กลุ่ม เราสามารถตัดออกมาเป็นรูปแบบได้ 262 รูปแบบ ซึ่งพิจารณาแล้วพบว่าเป็นรูปแบบที่แตกต่างกันเพียง 135 ชนิด เพราะบางรูปแบบที่ตัดออกมานั้นมีลักษณะคล้ายกันหรือเหมือนกัน จึงจัดให้เป็นชนิดเดียวกัน



รูปที่ 3.5 รูปแสดงรูปแบบของแผนภูมิ 2 มิติของกรดอะมิโนที่เกิดจากลำดับของกรดอะมิโนขนาด 18 ที่เป็นชนิดเดียวกัน



รูปที่ 3.6 รูปแสดงรูปแบบของแผนภูมิ 2 มิติของกรดอะมิโนที่เกิดจากลำดับของกรดอะมิโนขนาด 18 ต่างชนิดกัน

จึงได้ทำการแบ่ง รูปแบบที่ได้เป็น

1. Training set 212 รูปแบบ(135 ชนิด)

2 Test set ในชุดที่ 2 จำนวน 50 รูปแบบ

เปอร์เซ็นต์ความถูกต้อง

สามารถคำนวณได้จะสูตร

$$\text{เปอร์เซ็นต์ความถูกต้อง} = \frac{\text{จำนวนโปรตีนที่ทำนายชนิดได้ถูกต้อง}}{\text{จำนวนโปรตีนที่ใช้ทดสอบทั้งหมด}} \times 100\ \%$$

Network

ตัวอย่าง network file ที่ใช้ hidden 30 neuron(s)

SNNS network definition file V1.4-3D

generated at Wed Sep 22 01:02:27 1999

network name : train30

source files :

no. of units : 130

no. of connections : 3000

no. of unit types : 0

no. of site types : 0

learning function : Std_Backpropagation

update function : Topological_Order

unit default section :

act	bias	st	subnet	layer	act func	out func
0.00000	0.00000	h	0	1	Act_Logistic	Out_Identity

unit definition section :

no.	type Name	unit Name	act	bias	st	position	act func	out func	sites
1			0.00000	0.10311	i	0, 1,17136			
2			0.00000	-0.57433	i	0, 2,17136			
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	
130			0.00000	0.77494	o	20,10,17136			

connection definition section :

target	site	source:weight
91		90:-0.28176, 89:-0.00377, 88:-0.72144, 87:-0.33136, 86: 0.79794, 85:-0.68870, 84:-0.14213, 83:-0.23370,82:-0.11929, 81:-0.68651, 80:-0.83752, 79:-0.79634, 78: 0.97375, 77: 0.72623, 76:- 0.10891, 75:-0.86160,
⋮	⋮	⋮
130		120:-0.06222, 119: 0.69760, 118:-0.11232, 117: 0.79670, 116: 0.08628, 115:-0.02820, 114: 0.30983, 113: 0.11079, 112: 0.20106, 111:-0.66918, 110:-0.48789, 109: 0.46247, 108: 0.75637, 107:-0.00536, 106:-0.57289, 105: 0.85544, 104:-0.17328, 103: 0.57342, 102:-0.78603, 101: 0.01221, 100:-0.86883, 99: 0.47869, 98:-0.91968, 97: 0.79176,96:-0.93682, 95: 0.28500, 94:-0.04800, 93:-0.40099, 92: 0.56202, 91: 0.95544