

ทำการสร้างแผนภูมิ 2 มิติของกรดอะมิโน จากลำดับของกรดอะมิโนซึ่งมาจากโปรตีนที่ขอมอโลกัสกัน และมี sequence identity ต่างๆกัน จำนวน 30 คู่ดังนี้

ID of PDB	Sequence identity(%)	Protein
1poh	35.29	Phosphotransferase (E. coli)
1ptf		Phosphotransferase (S. faecalis)
1nhk_R	43.7	Nucleoside Diphosphate Kinase (M. xanthus)
1ndc		Nucleoside Diphosphate Kinase (D. discoideum)
1bpt	44.64	Pancreatic Trypsin Inhibitor (BPTI) (B. taurus)
1aap_A		PIIn. Domain Of Alzheimer's Protein (H. sapiens)
5pal	44.86	Parvalbumin (T. semifasciata)
1omd		Oncomodulin (R. norvegicus)
1pza	45.00	Pseudoazurin (A. faecalis)
1pmy		Pseudoazurin (M. extorquens)
1thb_B	45.21	Hemoglobin (H. sapiens)
1pbx_B		Hemoglobin (P. bernacchii)
5hvp_B	48.48	HIV-1 Protease (HIV Type 1)
1ivp_A		HIV-2 Protease (HIV Type 2)
2sam	51.52	SIV-1 Protease (SIV Type 1)
4phv_B		HIV-1 Protease (HIV Type 1)
2cro	52.38	Cro Protein (Phage 434)
2or1_L		Repressor (Phage 434)
1crb	56.39	Cellular Retinol Binding Protein (R. rattus)
1opb_C		Cellular Retinol Binding Protein II (R. rattus)
1fkf	57.01	Fk-506 Binding Protein (H. sapiens)
1yat		Fk-506 Binding Protein (S. cerevisiae)

1pva_A 1cdp	62.04	Parvalbumin (E. lucius) Parvalbumin (C. carpio)
2ycc 5cyt_R	62.14	Cytochrome C (S. cerevisiae) Cytochrome C (T. alalunga)
1arz_A 1aiz_A	63.28	Azurin (Pseudomonas aeruginosa) Azurin (Alcaligenes denitrificans)
1mrj 1mom	65.04	Alpha-trichosanthin (T. kirilowii maxim) Momordin (M. charantia)
1cad 8rxn_A	66.67	Rubredoxin (P. furiosus) Rubredoxin (D. vulgaris)
1tad_B 1gia	69.35	Transducin-alpha (B. taurus) Gi Alpha 1 (R. rattus)
1has_A 1vaa_A	72.63	Human Class I HSA (H. sapiens) MHC Class I (M. musculus)
1gbt 1brc_E	73.09	Beta-trypsin (B. taurus) Trypsin Variant (R. rattus)
1bab_B 1fch_G11	73.29	Hemoglobin Thionville (H. sapiens) Hemoglobin (H. sapiens)
1dhf_A 1dr7	75.27	Dihydrofolate Reductase (H. sapiens) Dihydrofolate Reductase (G. gallus)
1hna 3gst_B	75.58	Glutathione S-transferase (H. sapiens) Glutathione S-transferase (R. rattus)

1ala	77.85	Annexin V (G. gallus)
1avr		Annexin V (H. sapiens)
1bra	79.82	Trypsin (R. rattus)
1mct_A		Trypsin (S. scrofa)
4p2p	84.17	Phospholipase A2 (S. scrofa)
2bpp		Phospholipase A2 (B. taurus)
135l	86.82	Lysozyme (M. gallopavo)
1hhl		Lysozyme (N. meleagris)
1ovb	90.57	Ovotransferrin (Duck)
1nnt		Ovotransferrin (G. gallus)
2lal_A	92.27	Lentil Lectin (L. culinaris)
2ltn_A		Pea Lectin (P. sativum)
2chf	97.62	Chey (S. typhimurium)
1chn		Chey (E. coli)
1etb_1	98.31	Transthyretin (H. sapiens)
1ttc_A		Transthyretin mutant (H. sapiens)

ตารางที่ 4.1 ตารางแสดงลำดับของอะมิโนที่นำมาทดสอบแยกตาม sequence identity (Hooft R.W.W., Rodriguez R., Sander C. and Vriend G. 1998.)

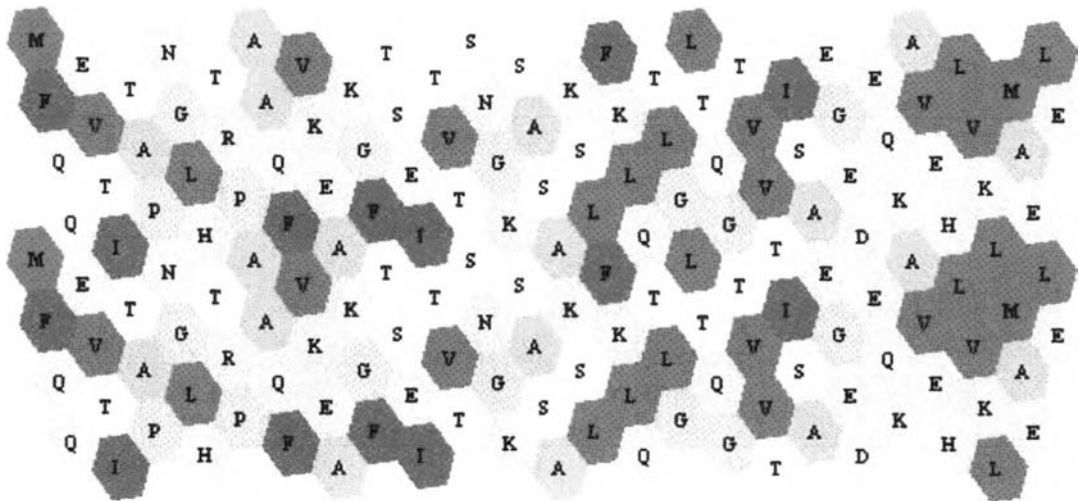
นำมาสร้างแผนภูมิ 2 มิติเพื่อทำการเปรียบเทียบ ได้ดังนี้

หมายเหตุ ขอแสดงผลบางส่วนเนื่องจากลำดับของกรดอะมิโนในบางคู่มีขนาดยาวมากทำให้ไม่สะดวกในการแสดงผล

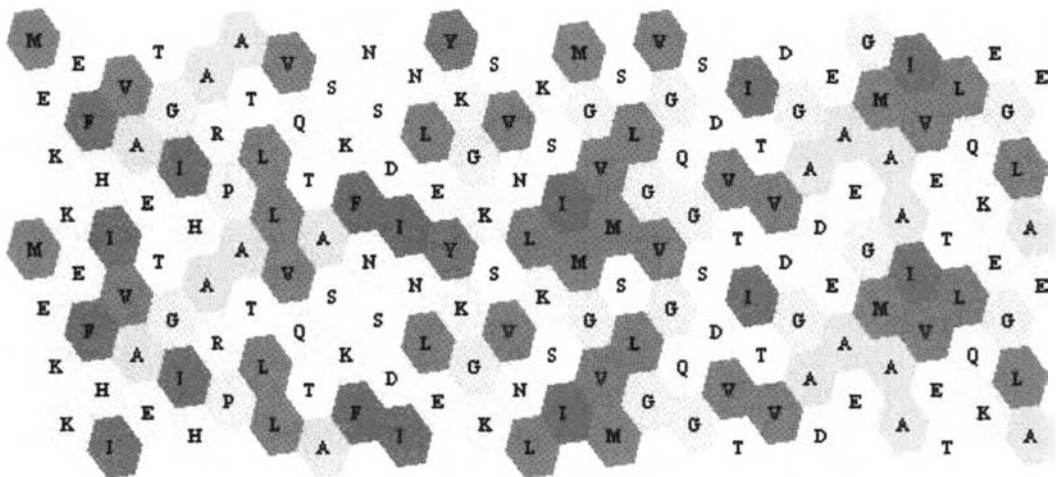
sequence identity 35.29 %

1.1 Phosphotransferase (E. coli)

MFQQEVITAPNGLHTRPAAQFVKEAKGFTS EITVTSNGKASAKSLFKLQTLGLTQGTWVTS AEGEDEQKAVEHLVKLMAELE



MEKKEFHI VAETGIHARPATLLVQTASKFNSDI NLEYKGRS VNLKSI NQVNSLGVGQGS DVTITVDGADEAEGRAAI VETLQKEGLAE



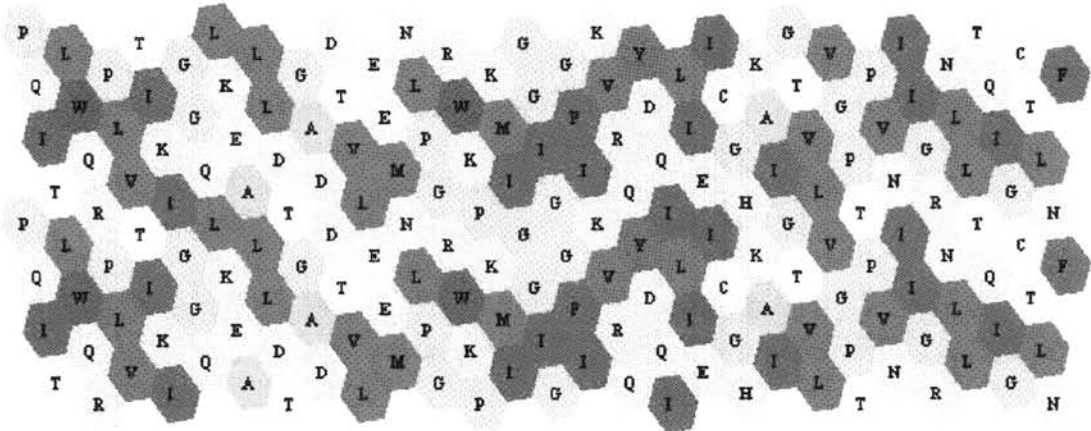
1.2 Phosphotransferase (S. faecalis)

รูปที่ 4.2 แผนภูมิ 2 มิติของกรดอะมิโน ของ Phosphotransferase ของ E. coli กับ S. faecalis

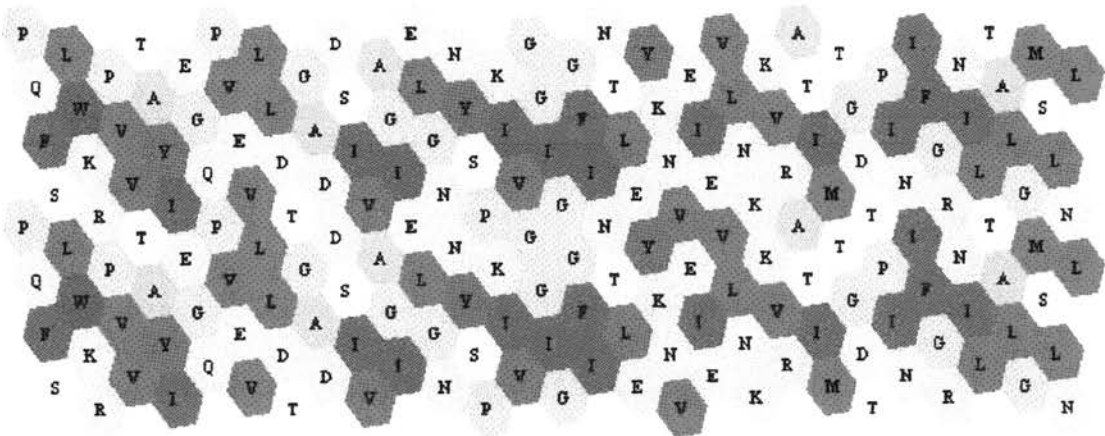
sequence identity 48.48 %

2.1 HIV-1 Protease (HIV Type 1)

PQI TLVQRPLVTI KI GGQLEKALLDT GADDTVLEEMNLPGRVQPKM GGI GGFI KVRQYDQI LI EI CGHKAI GTVLVGTTPVNI I GRNLTQI GCTLNF



PQFSLVWRPVTAVI EGQVVEULLDT GADDSI VAGI ELGRNVS PKI VGGI GGFI NTLEYKNVEI EWLNKKVRATI MT GDTPI NI FGRNI LTALGNSLNL



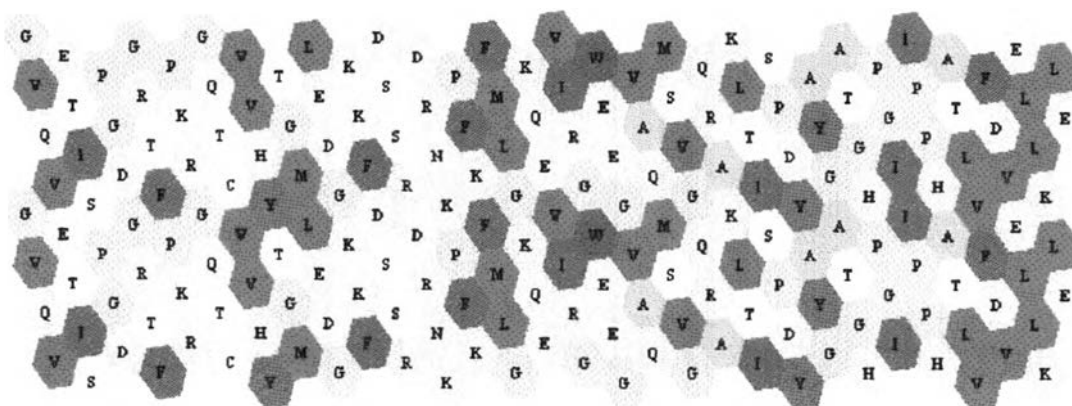
2.2 HIV-2 Protease (HIV Type 2)

รูปที่ 4.3 แผนภูมิ 2 มิติของกรดอะมิโนของ HIV-1Protease กับ HIV-2 Protease

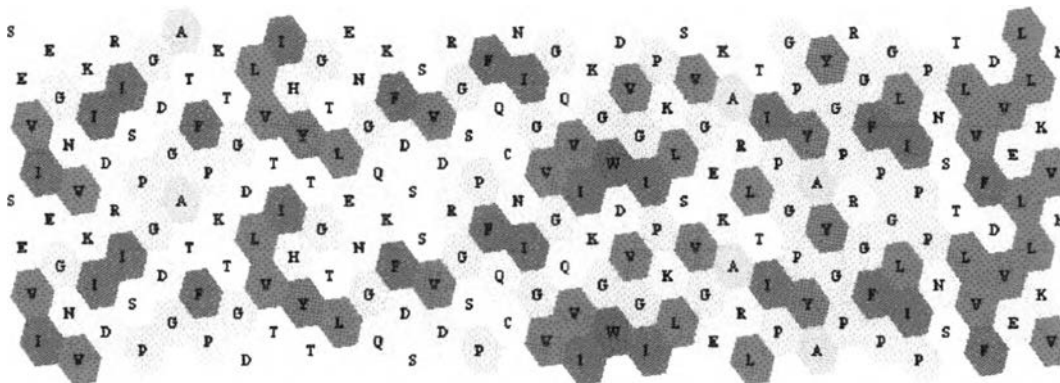
sequence identity 57.10 %

3.1 FK-506 Binding Protein (H. sapiens)

GVQVETI SP GDGRI FPKRGQI CVVHYT GML EDGKGFDS S RDRNKPFKFM LGKQVI RGVKKGVAQNS VGRQLKTI SPDYAYGAT GHPGI I PDHATLVFDVLLKLE



S EVI EGNVNI DRI SP GDNATFPRT GDLVVI HYT DTLENDQKFD S VDRDSPPQCN GVGVNI KQDVQI PKLS VQEKARTI I DGPYAVQRPFP GLI PDHSTLVFDVLLKVN

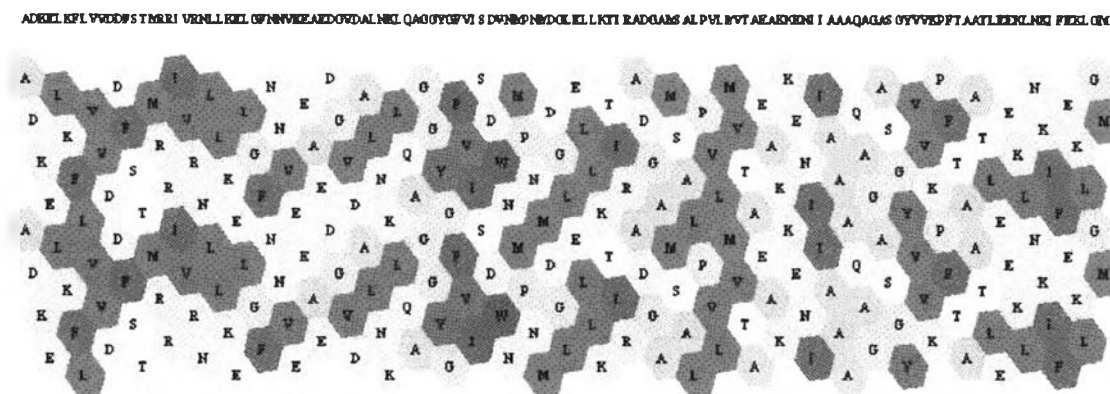
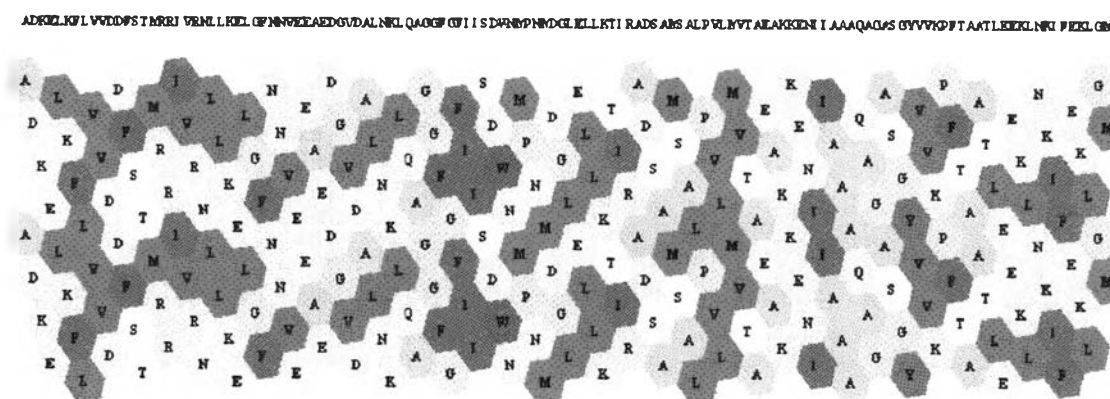


3.2 Fk-506 Binding Protein (S. cerevisiae)

รูปที่ 4.4 แผนภูมิ 2 มิติของกรดอะมิโน ของ FK-506 Binding Protein ของ *H. sapiens* กับ *S. cerevisiae*

sequence identity 97.62 %

4.1 Chey (*S. typhimurium*)



4.2 Chey (*E. coli*)

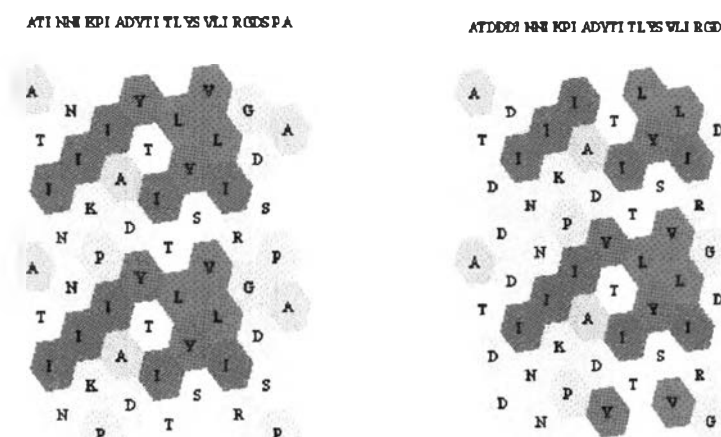
รูปที่ 4.5 แผนภูมิ 2 มิติของกรดอะมิโน ของ Chey ของ *S. typhimurium* กับ *E. coli*

ผลการทดลอง

1. จากแผนภูมิ 2 มิติของกรดอะมิโน ที่สร้างลำดับของกรดอะมิโนซึ่งมาจากโปรตีนที่ฮอมอโลจัสกันจากโปรตีนจำนวน 30 คู่พบว่า แต่ละคู่ให้แผนภูมิ 2 มิติของกรดอะมิโนที่คล้ายกันหรือเหมือนกัน
2. จากที่แผนภูมิ 2 มิติของกรดอะมิโน จากลำดับของกรดอะมิโนซึ่งมาจากโปรตีนที่มี sequence identity ต่างๆกันตั้งแต่ 35เปอร์เซ็นต์ - 98เปอร์เซ็นต์ พบว่าแผนภูมิ 2 มิติของกรดอะมิโนสามารถใช้ทดสอบฮอมอโลยีของโปรตีนได้ นั่นคือ ในโปรตีนที่ฮอมอโลจัสกัน ยังคงให้แผนภาพ 2 มิติของกรดอะมิโนที่คล้ายกัน

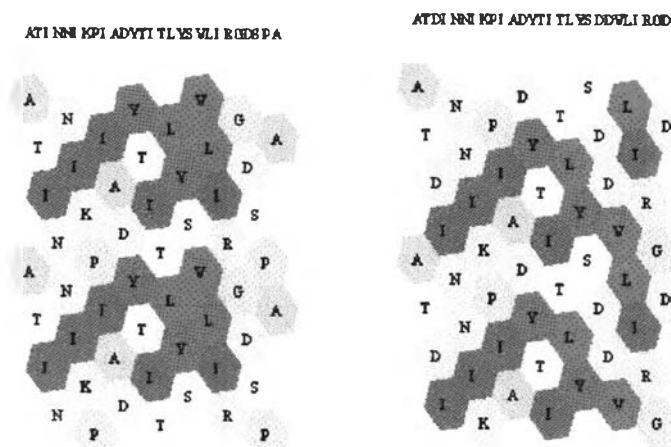
ทำการทดลองสร้างแผนภูมิ 2 มิติของกรดอะมิโนซึ่งมีช่องว่างแบ่งเป็น 2 แบบคือ

1 ช่องว่างอยู่ในบริเวณที่ไม่ใช่ cluster



รูปที่ 4.6 แผนภูมิ 2 มิติของกรดอะมิโน ที่มีช่องว่างในบริเวณที่ไม่ใช่ cluster

2 ช่องว่างอยู่ในบริเวณที่ที่เป็น cluster



รูปที่ 4.7 แผนภูมิ 2 มิติของกรดอะมิโน ที่มีช่องว่างในบริเวณที่เป็น cluster

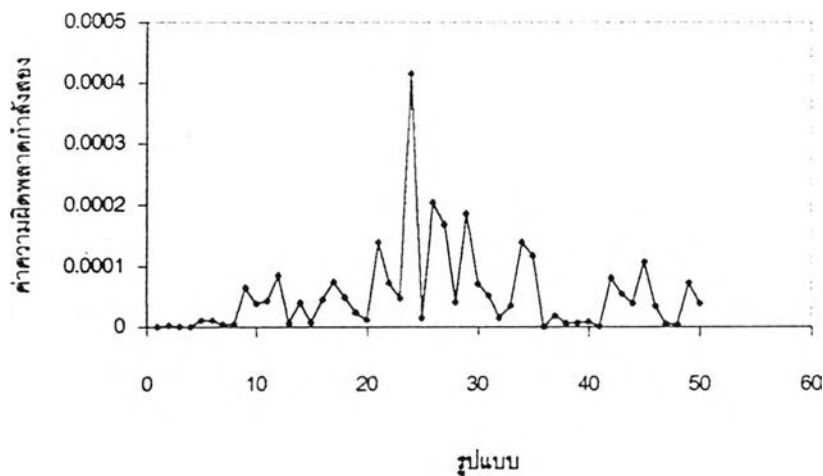
พบว่าในกรณีที่ 1 แผนภูมิ 2 มิติของกรดอะมิโน ที่มีช่องว่างในบริเวณที่ไม่ใช่ cluster ยังคงสามารถทดสอบฮอโมโลยีของลำดับของกรดอะมิโนได้อยู่เนื่องจากแผนภูมิ 2 มิติที่ได้ยังมีลักษณะคล้ายกันเพียงแต่มีการขยับหรือเลื่อนตำแหน่ง

ในกรณีที่ 2 แผนภูมิ 2 มิติของกรดอะมิโน ที่มีช่องว่างในบริเวณที่เป็น cluster ยังคงสามารถทดสอบฮอโมโลยีของลำดับของกรดอะมิโนได้อยู่เนื่องจากแผนภูมิ 2 มิติที่ได้ แต่จะมีความยุ่งยากในการพิจารณาเนื่องจาก cluster ได้มีการเปลี่ยนรูปร่างไป

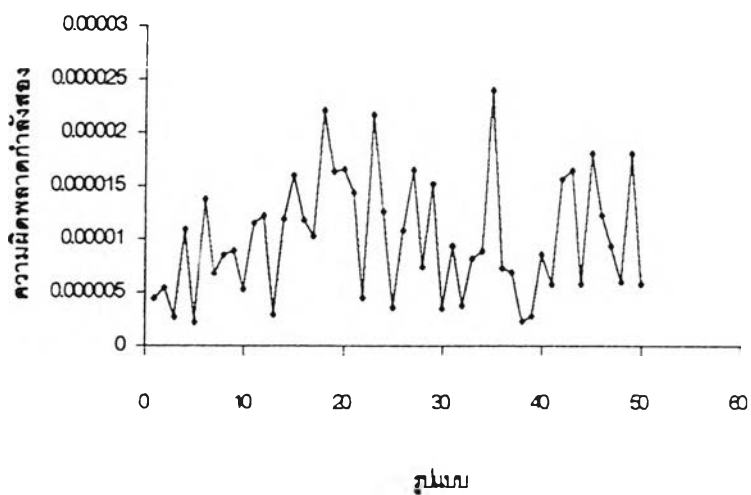
โครงข่ายประสาทเทียม

ในโครงการนี้ได้ทำการสร้างโครงข่ายประสาทเทียม สำหรับจำแนกชนิดของรูปแบบที่เกิดจากลำดับของกรดอะมิโนจำนวน 18 ตัว จำนวน 5 โครงข่าย จำแนกตามจำนวน hidden layer

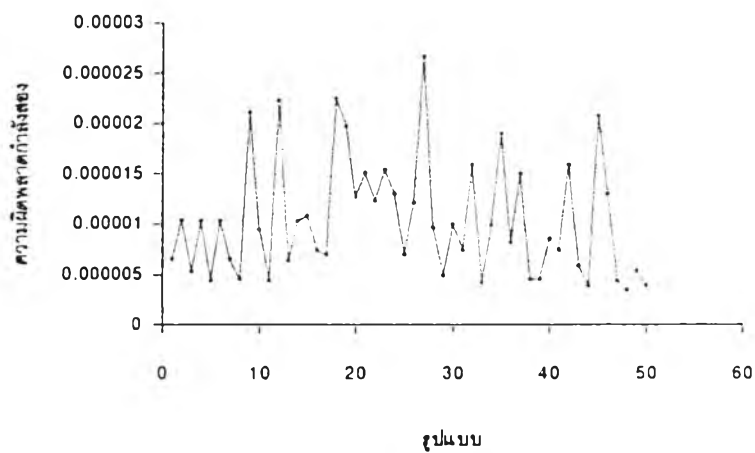
ผลที่ได้จากโครงข่ายประสาทเทียมแต่ละตัวโดยใช้ test set จำนวน 2 ชุด ชุดละ 50 ตัว นำมาเขียนกราฟผลรวมของความผิดพลาดกำลังสองผลเป็นดังนี้ test set ที่ 1 สมาชิกใน test set มาจาก training set 100 เปอร์เซ็นต์



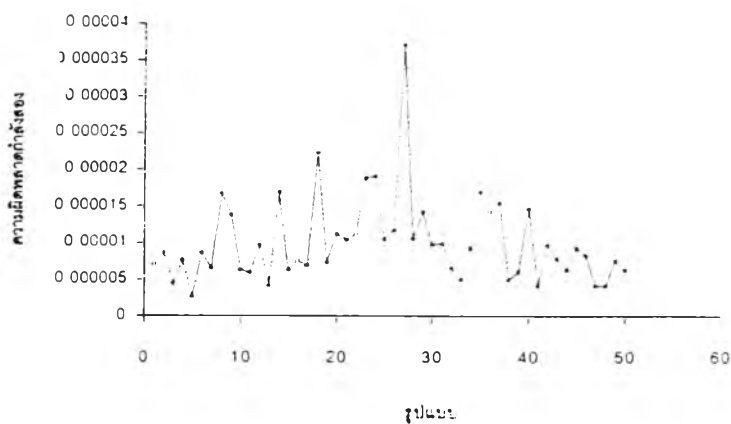
รูปที่ 4.8 กราฟระหว่างผลต่างกำลังสองกับรูปแบบที่ใช้ทดสอบเมื่อจำนวน hidden layer มีขนาด 30 สำหรับ test set ที่ 1



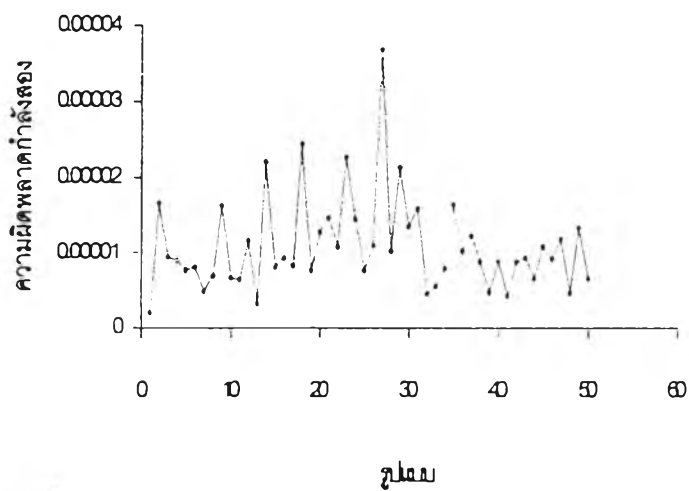
รูปที่ 4.9 กราฟระหว่างผลต่างกำลังสองกับรูปแบบที่ใช้ทดสอบเมื่อจำนวน hidden layer มีขนาด 40 สำหรับ test set ที่ 1



รูปที่ 4.10 กราฟระหว่างผลต่างกำลังสองกับรูปแบบที่ใช้ทดสอบเมื่อจำนวน hidden layer มีขนาด 50 สำหรับ test set ที่ 1

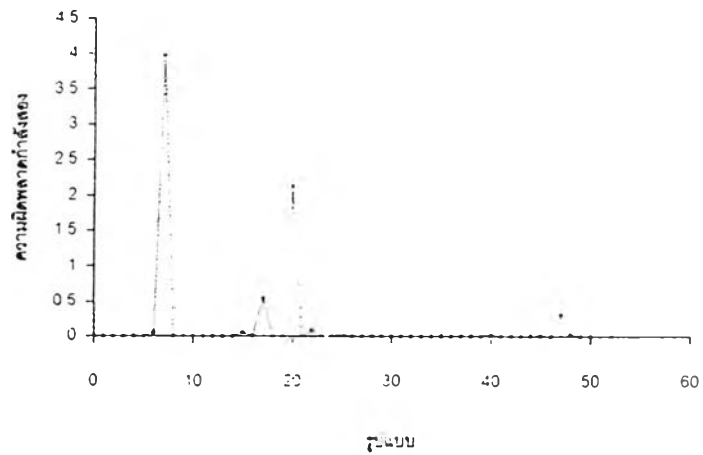


รูปที่ 4.11 กราฟระหว่างผลต่างกำลังสองกับรูปแบบที่ใช้ทดสอบเมื่อจำนวน hidden layer มีขนาด 60 สำหรับ test set ที่ 1

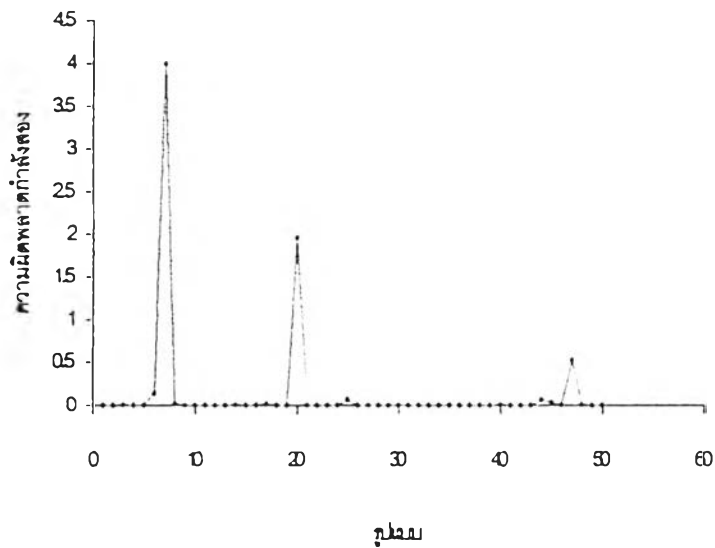


รูปที่ 4.12 กราฟระหว่างผลต่างกำลังสองกับรูปแบบที่ใช้ทดสอบเมื่อจำนวน hidden layer มีขนาด 70 สำหรับ test set ที่ 1

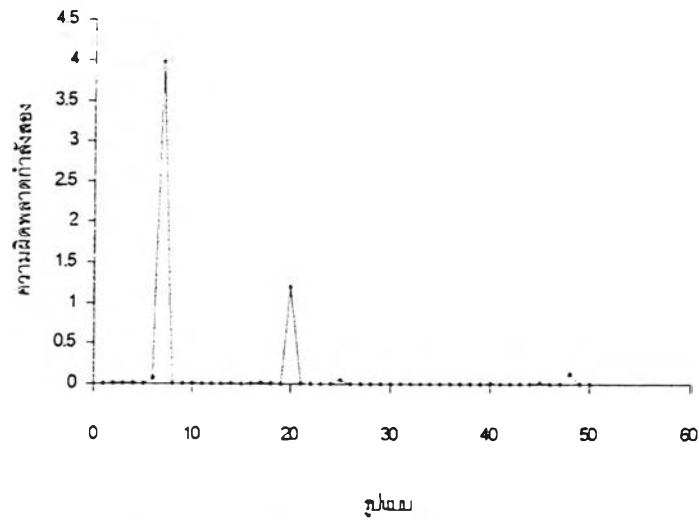
test set ที่ 2 สมาชิกใน test set เป็นรูปแบบที่ไม่ได้สอน 50 ตัว



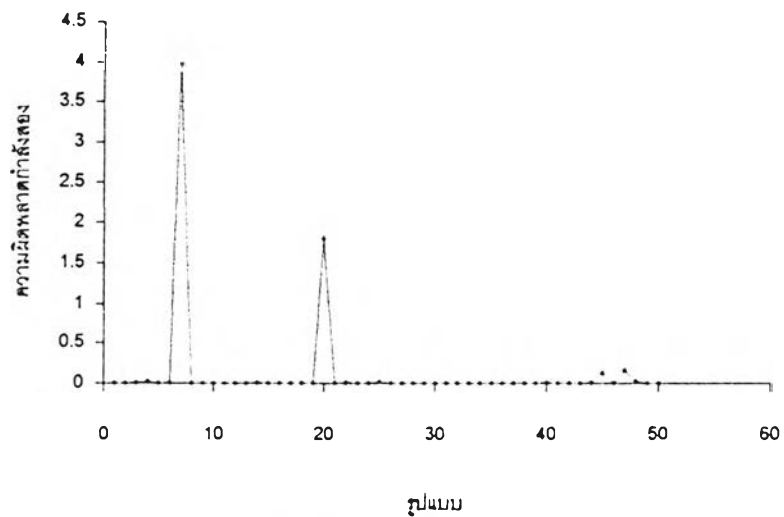
รูปที่ 4.13 กราฟระหว่างผลต่างกำลังสองกับรูปแบบที่ใช้ทดสอบเมื่อจำนวน hidden layer มีขนาด 30 สำหรับ test set ที่ 2



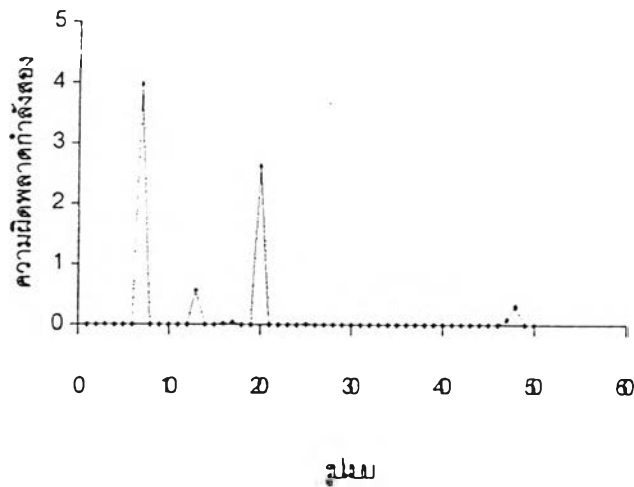
รูปที่ 4.14 กราฟระหว่างผลต่างกำลังสองกับรูปแบบที่ใช้ทดสอบเมื่อจำนวน hidden layer มีขนาด 40 สำหรับ test set ที่ 2



รูปที่ 4.15 กราฟระหว่างผลต่างกำลังสองกับรูปแบบที่ใช้ทดสอบเมื่อจำนวน hidden layer มีขนาด 50 สำหรับ test set ที่ 2



รูปที่ 4.16 กราฟระหว่างผลต่างกำลังสองกับรูปแบบที่ใช้ทดสอบเมื่อจำนวน hidden layer มีขนาด 60 สำหรับ test set ที่ 2



รูปที่ 4.17 กราฟระหว่างผลต่างกำลังสองกับรูปแบบที่ใช้ทดสอบเมื่อจำนวน hidden layer มีขนาด 70 สำหรับ test set ที่ 2

ผลจากการนำค่าความผิดพลาดมาเขียนกราฟพบว่าในกรณีชุดทดสอบที่ 1

จำนวน hidden	ผลรวมค่าความผิดพลาดกำลัง 2 ใช้ test set ชุดที่ 1	ผลรวมค่าความผิดพลาดกำลัง 2 ใช้ test set ชุดที่ 2
30	0-0.0005	0-4
40	0-0.00003	0-4
50	0-0.00003	0-4
60	0-0.00004	0-4
70	0-0.00004	0-4

ตารางที่ 4.2 ตารางแสดงผลรวมของความผิดพลาดจำแนกตามใน hidden layer ใน test set ที่ 1 และ test set ที่ 2

ใน test set ที่ 1 พบว่าโครงข่ายประสาทเทียม มีค่าความผิดพลาดในการจำแนกน้อยมาก

ใน test set ที่ 2 พบว่าโครงข่ายประสาทเทียมมีความผิดพลาดในการจำแนกบางรูปแบบมาก

ต่อไปจะพิจารณาความถูกต้องแม่นยำของการจำแนกรูปแบบของแต่ละ hidden layer

การพิจารณาความแม่นยำของโครงข่ายประสาทเทียมจะพิจารณาจาก

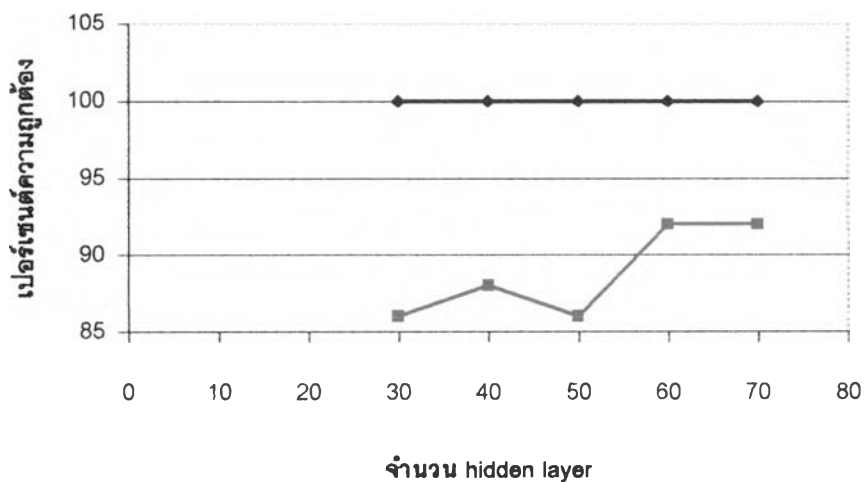
$$\text{เปอร์เซ็นต์ความถูกต้อง} = \frac{\text{จำนวนโปรตีนที่ทำนายชนิดได้ถูกต้อง}}{\text{จำนวนโปรตีนที่ใช้ทดสอบทั้งหมด}} \times 100\%$$

แสดงดังตาราง

จำนวน hidden layer	เปอร์เซ็นต์ความถูกต้องในชุดทดสอบที่ 1	เปอร์เซ็นต์ความถูกต้องในชุดทดสอบที่ 2
30	100	86
40	100	88
50	100	86
60	100	92
70	100	92

ตารางที่ 4.3 ตารางแสดงเปอร์เซ็นต์ความถูกต้องเทียบกับจำนวน hidden layer ในแต่ละ test set ที่ 1 และ test set ที่ 2

ซึ่งนำมาเขียนกราฟได้ดังนี้



รูปที่ 4.18 กราฟระหว่างจำนวน hidden layer กับเปอร์เซ็นต์ความถูกต้องใน test set ชุดที่ 1 และ test set ชุดที่ 2

จะได้ว่า โครงข่ายประสาทเทียมที่มี hidden layer จำนวน 60 และ 70 ให้ค่าเปอร์เซ็นต์ความถูกต้องสูงสุด