การทำนายสัดส่วนของโครงสร้างทุติยภูมิในโปรตีนโดยใช้ข่ายงานประสาท

นางสาวจุไรรัตน์ พรหมใจ

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาชีวเคมี ภาควิชาชีวเคมี
บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2541

974-332-423-2

ลิขสิทธิ์ของบัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย

I18903661

PREDICTION OF THE EXTENT OF PROTEIN SECONDARY STRUCTURES USING NEURAL NETWORKS

Miss Jurairat Phromjai

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

A Thesis Submitted in Partial Fulfillment of the Requirements

for the Degree of Master of Science in Biochemistry

Department of Biochemistry

Graduate School

Chulalongkorn University

Academic year 1998

974-332-423-2

Thesis Title	Prediction of the Extent of Protein Secondary Structures Using Neural Networks
By	Miss Jurairat Phromjai
Department	Biochemistry
Thesis Advisor	Lerson Tanasugarn, Ph.D.

Accepted by the Graduate School, Chulalongkorn University in Partial Fulfillment of the Requirements for the Master's Degree.

.....Dean of Graduate School

(Professor Supavat Chutivongse, M.D.)

Thesis Committee

.....Chairman

(Assistant Professor Suganya Soontaros, Ph.D.)

.....Thesis Advisor

(Lerson Tanasugarn, Ph.D.)

.....Thesis Co-Advisor

(Associate Professor Chidchanok Lursinsap, Ph.D.)

.....Member

(Rath Pichyangkura, Ph.D.)

พิมพ์ต้นฉบับบทคัดย่อวิทยานิพนธ์ภายในกรอบสีเขียวนี้เพียงแผ่นเดียว

จุไรรัตน์ พรหมใจ : การทำนายสัดส่วนของโครงสร้างทุติยภูมิโปรตีนโดยใช้ข่ายประสาท (PREDICTION OF THE EXTENT OF PROTEIN SECONDARY STRUCTURE USING NEURAL NETWORKS) อ. ที่ปรึกษา : อ. ดร. เลอสรร ธนสุกาญจน์, อ. ที่ปรึกษาร่วม : รศ. ดร. ชิดชนก เหลือสินทรัพย์, 241 หน้า. ISBN 974-332-423-2

งานวิจัยนี้เป็นการศึกษาการทำนายโครงสร้างสามมิติของโปรตีนโดยใช้ระบบประมวลผลข้อมูลแบบข่ายประสาท (Neural Networks) และใช้สมบัติของลำดับกรดอะมิโนที่มีผลต่อโครงสร้างสามมิติของโปรตีนเป็นข้อมูลโดยคาดว่าสมบัติเหล่านี้จะทำให้การทำนายโครงสร้างสามมิติของโปรตีนมีความถูกต้องแม่นยำสูงขึ้น สมบัติที่ใช้ในการสอน Neural Networks ในการศึกษานี้ ได้แก่ hydropathy, hydrophobicity, helical tendencies และ amino acid side chain properties โดยสมบัติเหล่านี้ถูกใช้แทนที่ลำดับของกรดอะมิโนแล้วถูกป้อนเข้าไปใน Neural Networks โปรตีน 98 ตัว ได้ถูกจัดอยู่ในกลุ่มที่ใช้ในการสอน 70 ตัว และจัดอยู่ในกลุ่มที่ใช้ในการทดสอบ 28 ตัว

จากผลการทดลองการทำนายว่ามีโครงสร้าง helix, sheet และ turn ในโปรตีนหรือไม่ โดยใช้ network เดียวกันให้ผลของการทำนายถูกต้องน้อยกว่าการทำนายโครงสร้าง helix, sheet และ turn ที่มีการแยก network กัน และแต่ละสมบัติจะให้ผลการทำนายโครงสร้าง helix ถูกต้องสูงสุดเมื่อเทียบกับการทำนายโครงสร้างชนิดอื่น ความสามารถของสมบัติๆ ในการทำนายโครงสร้างทุติยภูมิของโปรตีนจัดลำดับได้ดังนี้ (1.1) สมบัติ amino acid side chain ให้ผลการทำนายโครงสร้างว่ามีโครงสร้าง helix, sheet และ turn หรือไม่จาก network เดียวกัน, การทำนายว่ามีโครงสร้าง sheet หรือไม่, การทำนายว่ามีโครงสร้าง turn หรือไม่, ร้อยละของ โครงสร้าง helix และ sheet ที่แบ่งเป็น 3 กลุ่ม และ การทำนายร้อยละของ helix ที่แบ่งเป็น 2 กลุ่ม ถูกต้องสูงสุด (1.2) สมบัติ hydropathy ที่แบ่งออกเป็น 2 กลุ่ม ให้ผลการทำนายว่ามีโครงสร้าง helix หรือไม่, การทำนายร้อยละของ helix และ sheet ที่แบ่งเป็น 6 กลุ่ม และ การทำนายร้อยละของ turn และ sheet ที่แบ่งออกเป็น3 กลุ่ม และ 2 กลุ่ม ตามลำดับ ถูกต้องสูงสุด (2) สมบัติ hydropathy ที่แบ่งออกเป็น 7 กลุ่มให้ผลการทำนายร้อยละของ helix และ turn ที่แบ่งออกเป็น 6 กลุ่ม, ร้อยละของ sheet ที่แบ่งออกเป็น 3 กลุ่ม ถูกต้องสูงสุด (3) สมบัติ hydrophobicity ให้ผลการทำนายร้อยละของ helix ที่แบ่งเป็น 2 กลุ่มถูกต้องสูงสุด

การทำนายว่ามีโครงสร้างของ helix, sheet และ turn หรือไม่ ให้ผลการทำนายถูกต้องในช่วง 85-100%, 70-85% และ 45-70% ตามลำดับ ส่วนการทำนายจำนวนร้อยละของ โครงสร้าง helix, sheet และ turn ในกรณีที่แบ่งร้อยละออกเป็น 6 กลุ่ม (0%, 1-20%, 21-40%, 41-60%, 61-80%, 81-100%) ให้ผลของการทำนายร้อยละของ helix, sheet และ turn มีความถูกต้องในช่วง 35-65%, 30-50% และ 25-50% ตามลำดับ ในการทำนายร้อยละโดยแบ่งร้อยละออกเป็น 3 กลุ่ม (0%, 1-50%, 51-100%) ให้ผลการทำนายร้อยละของโครงสร้าง helix, sheet และ turn มีความถูกต้องในช่วง 65-85%, 60-80% และ 50-65% ตามลำดับ ส่วนการทำนายว่ามีร้อยละของ helix, sheet และ turn ว่าอยู่ในกลุ่มน้อยกว่าหรือเท่ากับ 15% หรือ อยู่ในกลุ่มมากกว่า 15% ให้ผลการทำนายถูกต้องในช่วง 60-80% สำหรับโครงสร้าง helix และให้ผลการทำนายถูกต้องในช่วง 60-75% สำหรับโครงสร้าง sheet

ภาควิชา ชีวเคมี ลายมือชื่อนิสิต .....

สาขาวิชา ชีวเคมี ลายมือชื่ออาจารย์ที่ปรึกษา .....

ปีการศึกษา 2541 ลายมือชื่ออาจารย์ที่ปรึกษาร่วม .....

พิมพ์ต้นฉบับบทคัดย่อวิทยานิพนธ์ภายในกรอบสีเขียวนี้เพียงแผ่นเดียว

C826167 : MAJOR BIOCHEMISTRY
KEY WORD: PROTEIN / STRUCTURE / AMINO ACID PROPERTIES / NEURAL NETWORKS

JURAIRAT PHROMJAI : PREDICTION OF THE EXTENT OF PROTEIN SECONDARY STRUCTURE USING NEURAL NETWORK. THESIS ADVISOR : DR. LERSON TANASUKARN, Ph.D. THESIS COADVISOR : ASSO. PROF. CHIDCHANOK LURSINSAP, Ph.D. 241 pp. ISBN 974-332-423-2

We present a method for predicting protein structures based on a digital computer of neural networks. The neural networks learned from existing protein how to predict the secondary structure of amino acid sequences. The amino acid properties of amino acids such as hydropathy, hydrophobicity, helical tendencies and amino acid side chain properties were used as input vector. These properties were coded into the amino acid sequences and used as input patterns for both training and testing. Seventy amino acid sequences and twenty-eight amino acid sequences from different proteins were used for training and testing respectively.

The percent predictions accuracies of the existence of helix, sheet and turn structures using in the same network were lower than the prediction from separate networks. Each property gave the highest prediction accuracies for helix structure prediction. Properties can be ranked by their abilities to predict protein secondary structures as follow: (1.1) Amino acid side chain properties gave the highest accuracy for the prediction of the existence of helix, sheet turn in the same network, the existence of sheet and turn structure, percent helix, sheet (3 groups) and percent helix (2 groups). (1.2) Hydropathy (2 groups) gave the highest accuracy for the prediction of the existence of helix structure, percent helix and percent sheet (6 groups), percent turn (3 groups) and percent sheet (2 groups). (2) Hydropathy (7 groups) gave the highest accuracy for the prediction of percent sheet and turn (6 groups) and percent sheet (3 groups). (3) Hydrophobicity gave the highest accuracy for the prediction of percent helix (2 groups).

The range of percent accuracy prediction from all properties for helix, sheet and turn were between 85-100%, 70-85% and 45-70% respectively. The range of percent accuracy from all properties for predictions of percent helix, sheet and turn (6 groups of 0%, 1-20%, 21-40%, 41-60%, 61-80% and 81-100%) were 35-65%, 30-50%% and 25-50% respectively. The percent accuracies for percent helix, sheet and turn (3 groups of 0%, 1-50% and 51-100%) were 65-85%, 60-80% and 50-65% respectively. The percent accuracies for percent helix and sheet (2 groups of $\leq$ 15% and > 15%) were 60-80% and 60-75% respectively. The percent of secondary structure prediction is useful for the folding classes prediction.

ภาควิชา ชีวเคมี ลายมือชื่อนิสิต .....

สาขาวิชา ชีวเคมี ลายมือชื่ออาจารย์ที่ปรึกษา .....

ปีการศึกษา 2541 ลายมือชื่ออาจารย์ที่ปรึกษาร่วม .....

ACKNOWLEDGMENT

I would like to express my deepest gratitude and appreciation to Dr. Lerson Tanasugarn, my advisor for his excellent guidance, encouragement, suggestions and discussion throughout this study.

My appreciation is also expressed to my Co-advisor, Assoc. Prof. Dr. Chidchanok Lursinsap, for his guidance, advise and encouragement throughout my study. All the kindness and helps I have received will be long remembered with gratefulness and respectfulness.

I am most grateful to all members of the thesis examination committee : Asst. Prof. Dr. Sukanya Soontaros and Dr. Rath Pichyangkura for their kindness and valuable suggestions.

Special thanks is expressed to Miss Sompong Sapsutthipas for her encouragement and help in the preparation of the thesis.

Special thanks must be extended to Miss Pilailuk Akkapaiboon and .Miss Kingkarn Sookhanaphibarn for their gracious assistance.

Finally, the greatest indebtedness is expressed to my parents and my family for their tender love, understanding and encouragement.

Jurairat Phromjai

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

CONTENTS

	Page
THAI ABSTRACT	iv
ENGLISH ABSTRACT	v
ACKNOWLEDGEMENT	vi
CONTENTS	vii
LIST OF TABLES	xvi
LIST OF FIGURES	xxii
CHAPTER I INTRODUCTION	1
CHAPTER II LITERATURE REVIEWS	6
Proteins	6
Proteins Structure	7
Three-Dimensional structuce of proteins	9
Secondary structure	9
Tertiary Structure	11
The amino acid of proteins	14
Peptide	19
Peptide Bonds	19
Acid-Base Properties	19
The noncovalent interactions that determine the properties of proteins	20
Short-Range Repulsion	20
Electrostatic Forces	20
Van Der Waals Interactions	21
Hydrogen Bonding	21
Hydrophobic Forces	21
Disulfide Bond	22
The Protein Data Bank	22

Page

Artificial neural network 24

Biological neural system 24

The basic structure of artificial neural network 25

McCulloch and Pitts Model 26

Feed-Forward Networks 26

Multi Layer feed forward neural network 29

Back-propagation 30

CHAPTER III MATERIAL AND METHODS 35

3.1 Materials 35

3.1.1 Protein Database 35

3.1.2 Computer Hardware 35

3.1.3 Computer Software 35

Linux 35

SNNS (Stuttgart Neural Network Simulator) 36

Molecular visualization 40

3.2 Methods 41

3.2.1 Protein database 41

3.2.2 Properties of amino acid used 41

3.2.3 Input pattern and output pattern construction 50

3.2.4 Neural Networks 52

3.2.5 Training and Testing 53

3.2.6 Measurement of accuracy prediction 53

3.3 Structure prediction problem as a classification problem 54

3.3.1 Geometric meaning of neural function 54

3.3.2 Classification of protein structure 55

CHAPTER IV RESULTS 56

4.1 Secondary structure, helix, sheet, turn, prediction in amino acid sequence of proteins 56

Page

4.1.1 Prediction of helix, sheet, and turn in a same network using 20 numbers substituted 20 amino acids..... 56
4.1.2 Prediction of helix, sheet and turn in a same network using hydropathy (2 groups)..... 60
4.1.3 Prediction of helix, sheet and turn in the same network using hydropathy classified into 7 groups..... 64
4.1.4 Prediction of helix, sheet, and turn in the same network using 8 groups of amino acid side chain properties..... 68
4.1.5 Prediction of helix, sheet and turn in the same network using hydrophobicity (3 groups)..... 72
4.1.6 The comparison of the existence of helix, sheet, and turn in when using 2 groups of hydropathy, 7 groups of hydropathy, 8 groups of amino acid side chain properties and 3 groups of hydrophobicity..... 76
4.2 Helix structure prediction in amino acid sequence of proteins..... 78
4.2.1 Prediction of helix structure using hydropathy (2 groups)..... 79
4.2.2 Prediction of helical structure using hydropathy (7 groups).... 83
4.2.3 Prediction of helix structure using amino acid side chain properties (8 groups)..... 86
4.2.4 Prediction of helix structure using hydrophobicity (3 groups)..... 90
4.2.5 The comparison of the existence of helical structure using 2 and 7 groups of hydropathy, 8 groups of amino acid side chain properties and 3 groups hydrophobicity..... 94
4.3 Sheet structure prediction in amino acid sequence of proteins..... 95
4.3.1 Prediction of sheet structure using hydropathy (2 groups)..... 97
4.3.2 Prediction of sheet structure using hydropathy (7 groups)..... 98

Page

4.3.3 Prediction of sheet structure using amino acid side chain properties (8 groups)..... 100
4.3.4 Prediction of sheet structure using hydrophobicity (3groups)..... 101
4.3.5 A comparison of sheet structure prediction in separate networks using hydropathy (2 groups), hydropathy (7 groups), amino acid side chain properties (8 groups) and hydrophobicity (3 groups)..... 103
4.4 Turn structure prediction in amino acid sequence of proteins..... 104
4.4.1 Prediction of turn structure using hydropathy (2 groups)..... 106
4.4.2 Prediction of turn structure using hydropathy (7 groups)..... 107
4.4.3 Prediction of turn structure using amino acid side chain properties (8 groups)..... 109
4.4.4 Prediction of turn structure using hydrophobicity (3 groups). 110
4.4.5 Comparison of turn structure prediction in separate networks using hydropathy (2 groups), hydropathy (7 groups), amino acid side chain properties and hydrophobicity..... 112
4.5 The existence of helix, sheet and turn structures prediction..... 113
4.5.1 The existence of helix, sheet, and turn structures prediction using hydropathy (2 groups)..... 114
4.5.2 The existence of helix, sheet, and turn structures prediction using hydropathy (7 groups)..... 114
4.5.3 The existence of helix, sheet and turn structures prediction using amino acid side chain properties (8 groups)..... 115
4.5.4 The existence of helix, sheet, and turn structures prediction using hydrophobicity (3 groups)..... 116

Page

4.6 Percent helix structure (6 groups) prediction in amino acid sequence of proteins 117

4.6.1 Prediction of percent helix structure (6 groups) using hydropathy (2 groups) 119

4.6.2 Prediction of percent helix structure (6 groups) using hydropathy (7 groups) 122

4.6.3 Prediction of percent helix structure (6 groups) using amino acid side chain properties (8 groups) 126

4.6.4 Prediction of percent helix structure (6 groups) using hydrophobicity (3 groups) 129

4.6.5 Comparison of the percent helical structure (6 groups) in amino acid sequence prediction using 2 and 7 groups of hydropathy, 8 groups of amino acid side chain properties and 3 groups of hydrophobicity 133

4.7 Percent sheet structure (6 groups) prediction in amino acid sequence of proteins 134

4.7.1 Prediction of percent sheet structure (6 groups) using hydropathy (2 groups) 136

4.7.2 Prediction of percent sheet structure (6 groups) using hydropathy (7 groups) 137

4.7.3 Prediction of percent sheet structure (6 groups) using amino acid side chain properties (8 groups) 139

4.7.4 Prediction of percent sheet structure (6 groups) using hydrophobicity (3 groups) 140

4.7.5 The comparison of percent of sheet structure (6 groups) predictions using 2 and 7 groups of hydropathy, 8 groups of amino acid side chain properties and 3 groups of hydrophobicity 142

Page

4.8 Percent turn structure (6 groups) prediction in amino acid sequence of proteins 143

4.8.1 Prediction of percent turn structure (6 groups) using hydropathy (2 groups) 145

4.8.2 Prediction of percent turn structure (6 groups) using hydropathy (7 groups) 146

4.8.3 Prediction of percent turn structure (6 groups) using amino acid side chain properties (8 groups) 148

4.8.4 Prediction of percent turn structure (6 groups) using hydrophobicity (3 groups) 149

4.8.5 The comparison of percent of turn structure (6 groups) predictions using 2 and 7 groups of hydropathy, 8 groups of amino acid side chain properties and 3 groups of hydrophobicity 151

4.9 Percent helix, sheet and turn structures (6 groups) predictions 152

4.9.1 Percent helix, sheet and turn structures (6 groups) predictions using hydropathy (2 groups) 152

4.9.2 Percent helix, sheet and turn structures (6 groups) predictions using hydropathy (7 groups) 153

4.9.3 Percent helix, sheet and turn structures (6 groups) predictions using amino acid side chain properties 154

4.9.4 Percent helix, sheet and turn structure prediction (6 groups) using hydrophobicity 155

4.10 Percent helix structure (3 groups) prediction in amino acid sequence of proteins 156

4.10.1 Prediction of percent helix structure (3 groups) using hydropathy (2 groups) 160

4.10.2 Prediction of percent helix structure (3 groups) using hydropathy (7 groups) 161

Page

4.10.3 Prediction of percent helix structure (3 groups) using amino acid side chain properties (8 groups)..... 165

4.10.4 Prediction of percent helix structure (3 groups) using hydrophobicity (3 groups)..... 168

4.10.5 Comparison of the percent helix (3 groups) prediction using 2 and 7 groups of hydropathy, of amino acid side chain and hydrophobicity..... 172

4.11 Percent sheet structure (3 groups) prediction in amino acid sequence of proteins..... 173

4.11.1 Prediction of percent sheet structure (3 groups) using hydropathy (2 groups)..... 175

4.11.2 Prediction of percent sheet structure (3 groups) using hydropathy (7 groups)..... 176

4.11.3 Prediction of percent sheet structure (3 groups) using amino acid side chain properties (8 groups)..... 178

4.11.4 Prediction of percent sheet structure (3 groups) using hydrophobicity (3 groups)..... 179

4.11.5 Comparison of the percent sheet structure (3 groups) in amino acid sequence prediction using 2 and 7 groups of hydropathy, 8 groups of amino acid side chain properties and 3 groups of hydrophobicity..... 181

4.12 Percent turn structure (3 groups) prediction in amino acid sequence of proteins..... 182

4.12.1 Prediction of percent turn structure (3 groups) using hydropathy (2 groups)..... 184

4.12.2 Prediction of percent turn structure (3 groups) using hydropathy (7 groups)..... 185

4.12.3 Prediction of percent turn structure (3 groups) using amino acid side chain properties (8 groups)..... 187

Page

4.12.4 Prediction of percent sheet structure (3 groups) using hydrophobicity (3 groups)..... 188

4.12.5 Comparison of the percent helical structure (3 groups) in amino acid sequence prediction using 2 and 7 groups of hydropathy, 8 groups of amino acid side chain properties and 3 groups of hydrophobicity..... 190

4.13 Percent helix, sheet and turn structures (3 groups) predictions.. 191

4.13.1 Percent helix, sheet and turn structures (3 groups) predictions using hydropathy (2 groups)..... 191

4.13.2 Percent helix, sheet and turn (3 groups) predictions using hydropathy (7 groups)..... 192

4.13.3 Percent of helix, sheet and turn (3 groups) predictions using amino acid side chain properties..... 193

4.13.4 Percent of helix, sheet and turn (3 groups) predictions using hydrophobicity..... 194

4.14 Percent helix structure (2 groups) prediction in amino acid sequence of proteins..... 195

4.14.1 Prediction of percent helix structure (2 groups) using hydropathy (2 groups)..... 197

4.14.2 Prediction of percent helix structure (2 groups) using helical tendencies (5 groups)..... 198

4.14.3 Prediction of percent helix structure (2 groups) using amino acid side chain properties (8 groups)..... 200

4.14.4 Prediction of percent helix structure (2 groups) using hydrophobicity (3 groups)..... 201

4.14.5 Comparison of the percent helical structure (2 groups) in amino acid sequence prediction using 2 of hydropathy, 5 groups of helical tendencies, 8 groups of amino acid side chain properties and 3 groups of hydrophobicity..... 203

Page

4.15 Percent sheet structure (2 groups) prediction in amino acid sequence of proteins 204

4.15.1 Prediction of percent sheet structure (2 groups) using hydropathy (2 groups) 206

4.15.2 Prediction of percent sheet structure (2 groups) using amino acid side chain properties (8 groups) 207

4.15.3 Prediction of percent sheet structure (2 groups) using hydrophobicity (3 groups) 209

4.15.4 Comparison of the percent sheet structure (2 groups) in amino acid sequence prediction using 2 hydropathy, 8 groups of amino acid side chain properties and 3 groups of hydrophobicity 211

CHAPTER V DISCUSSION 212

5.1 Motivation and Benefits of this study 212

5.2 Evaluation of secondary structure prediction accuracy 215

5.3 Training and Testing on Neural Networks 216

5.4 Prediction accuracy 218

5.5 Further study 223

CHAPTER VI CONCLUSIONS 224

REFERENCES 226

APPENDIX A 234

APPENDIX B 239

BIOGRAPHY 241

LIST OF TABLES

Table		Page
2.1	Relative frequency of occurrence of amino acid residues in the secondary structures of proteins	8
2.2	Relative helical tendencies of the amino acids measured in one peptide	12
2.3	Definition of secondary structural classes of proteins	13
2.4	Covalent structure and abbreviations of 20 standard amino acids of proteins	15
2.5	Hydropathy associated with standard amino acids	23
3.1	Database of protein(s) used in protein structures prediction by NNs	43
3.2	Two groups of amino acids which were divided by hydropathy property	47
3.3	Seven group of amino acids which were divided by the hydropathy property	48
3.4	Five groups of amino acids which were divided by relative helical tendencies	49
4.1	The training set, testing set and possible outputs of secondary structures prediction	57
4.2	The trained network file, testing result file and percent accuracy of helix sheet and turn structure prediction in the same network using hydropathy (2 groups)	63
4.3	The trained network file, testing result file and percent accuracy of helix sheet and turn structure prediction in the same network using hydropathy (7 groups)	67

Table		Page
4.4	The trained network file, testing result file and percent accuracy of helix sheet and turn structure prediction in the same network using amino acid side chain properties.....	71
4.5	The trained network file, testing result file and percent accuracy of helix sheet and turn structure prediction in the same network using hydrophobicity.....	75
4.6	The training set, testing set and output of helix structure prediction in separate networks.....	79
4.7	The trained network, testing result files and percent accuracy prediction of helix structure using hydropathy (2 groups).....	82
4.8	The trained network, testing result files and percent accuracy prediction of helix structure using hydropathy (7groups).....	83
4.9	The trained network, testing result files and percent accuracy prediction of helix structure using 8 groups of amino acid side chain properties.....	89
4.10	The trained network, testing result files and percent accuracy prediction of helix structure using hydrophobicity	90
4.11	The training set, testing set and outputs of sheet structure prediction in separate networks.....	96
4.12	The trained network file, testing result file and percent accuracy of sheet structure prediction using hydropathy (2 groups).....	98
4.13	The trained network file, testing result file and percent accuracy of sheet structure prediction using hydropathy (7 groups).....	99
4.14	The trained network file, testing result file and percent accuracy of sheet structure prediction using amino acid side chain properties (8 groups).....	100
4.15	The trained network file, testing result file and percent accuracy of sheet structure prediction using hydrophobicity (3 groups).....	102

Table		Page
4.16	The training set, testing set and outputs of turn structure prediction in separate networks.....	105
4.17	The trained network, testing result files and percent accuracy of turn structure prediction using hydropathy (2 groups).....	106
4.18	The trained network, testing result files and percent accuracy of turn structure prediction using hydropathy (7 groups).....	108
4.19	The trained network, testing result files and percent accuracy of turn structure prediction using amino acid side chain properties (8 groups).....	109
4.20	The trained network file, testing result file and percent accuracy of sheet structure prediction using hydrophobicity (3 groups).....	111
4.21	The training set, testing set and possible outputs of percent helix Structure (6 groups) prediction.....	118
4.22	The trained network file, testing result file and percent accuracy of the of percent helix structure (6 groups) prediction using hydropathy (2 groups).....	119
4.23	The trained network file, testing result file and percent accuracy of percent helix structure prediction (6 groups) using hydropathy (7 groups).....	125
4.24	The trained network file, testing result file and percent accuracy of percent helix structure prediction (6 groups) using 8 groups of amino acid side chain properties.....	126
4.25	The trained network file, testing result file and percent accuracy of percent helix structure prediction (6 groups) using hydrophobicity.....	132
4.26	The training set, testing set and possible outputs of percent sheet structure (6 groups) prediction.....	135
4.27	The trained network file, testing result files and percent accuracy of percent sheet structure (6 groups) prediction using hydropathy (2 groups).....	136

Table **Page**

4.28 The trained network file, testing result file and percent accuracy of percent sheet structure (6 groups) prediction using hydropathy (7 groups)..... 138

4.29 The trained network file, testing result file and percent accuracy of percent sheet structure (6 groups) prediction using amino acid side chains (8 groups)..... 139

4.30 The trained network file, testing result file and percent accuracy of percent sheet structure (6 groups) prediction using hydrophobicity..... 141

4.31 The training set, testing set and possible outputs of percent turn structure (6 groups) prediction..... 144

4.32 The trained network file, testing result file and percent accuracy of percent turn structure (6 groups) prediction using hydropathy (2 groups)..... 146

4.33 The trained network file, testing result file and percent accuracy of percent turn structure (6 groups) prediction using hydropathy (7 groups)..... 147

4.34 The trained network file, testing result file and percent accuracy prediction of percent turn structure (6 groups) using amino acid side chains properties (8 groups)..... 149

4.35 The trained network file, testing result file and percent accuracy of percent turn structure (6 groups) prediction using hydrophobicity..... 150

4.36 The training set, testing set and possible outputs of percent helix structure (3 groups) prediction..... 157

4.37 The trained network file, testing result file and percent accuracy of the percent helix structure (3 groups) prediction using hydropathy (2 groups)..... 161

4.38 The trained network file, testing result file and percent accuracy of percent helix structure prediction (3 groups) using hydropathy (7 groups)..... 164

Table		Page
4.39	The trained network file, testing result file and percent accuracy of percent helix structure prediction (3 groups) using 8 groups of amino acid side chain properties.....	165
4.40	The trained network file, testing result file and percent accuracy of percent helix structure prediction (3 groups) using hydrophobicity.....	171
4.41	The training set, testing set and possible outputs of percent sheet structure (3 groups) prediction.....	174
4.42	The trained network file, testing result file and percent accuracy of the prediction of percent sheet structure (2 groups) using hydropathy (2 groups).....	175
4.43	The trained network file, testing result file and percent accuracy of percent sheet structure prediction (3 groups) using hydropathy (7 groups).....	177
4.44	The trained network file, testing result file and percent accuracy of percent sheet structure prediction (3 groups) using 8 groups of amino acid side chain properties.....	178
4.45	The trained network file, testing result file and percent accuracy of percent sheet structure prediction (3 groups) using hydrophobicity.....	180
4.46	The training set, testing set and possible outputs of percent turn structure (3 groups) prediction.....	183
4.47	The trained network file, testing result file and percent accuracy of the prediction of percent turn structure (3 groups) using hydropathy (2 groups).....	184
4.48	The trained network file, testing result file and percent accuracy of percent turn structure prediction (3 groups) using hydropathy (7 groups).....	186
4.49	The trained network file, testing result file and percent accuracy of percent turn structure prediction (3 groups) using 8 groups of amino acid side chain properties.....	188

Table		Page
4.50	The trained network file, testing result file and percent accuracy of percent turn structure prediction (3 groups) using hydrophobicity.....	189
4.51	The training set, testing set and possible outputs of percent helix structure (2 groups) prediction.....	196
4.52	The trained network file, testing result file and percent accuracy of the prediction of percent helix structure (2 groups) using hydropathy (2 groups).....	198
4.53	The trained network file, testing result file and percent accuracy of percent helix structure prediction (2 groups) using helical tendencies (5 groups).....	199
4.54	The trained network file, testing result file and percent accuracy of percent helix structure prediction (2 groups) using 8 groups of amino acid side chain properties.....	201
4.55	The trained network file, testing result file and percent accuracy of percent helix structure prediction (2 groups) using hydrophobicity.....	202
4.56	The training set, testing set and possible outputs of percent sheet structure (2 groups) prediction.....	205
4.57	The trained network file, testing result file and percent accuracy of the prediction of percent sheet structure (2 groups) using hydropathy (2 groups).....	207
4.58	The trained network file, testing result file and percent accuracy of percent sheet structure prediction (2 groups) using 8 groups of amino acid side chain properties.....	208
4.59	The trained network file, testing result file and percent accuracy of percent sheet structure prediction (2 groups) using hydrophobicity.....	210

LIST OF FIGURES

Figure		Page
1.1	The gab between the number of protein sequences and three dimensional structures	5
2.1	The general structural formula of α-amino acid	14
2.2	The peptide bond	19
2.3	Schematic drawing of a typical neuron	24
2.4	The basic structure of neural networks	26
2.5	Schematic diagram of McCulloch and Pitts	27
2.6	Decision regions created by a multilayer perceptrons	30
2.7	A three layers back propagation network fully interconnection	32
3.1	SNNS components	37
3.2	A simple network, a feed-forward net with three layer of units	38
3.3	Protein structure prediction networks	51
4.1	The example of input and output patterns for training and testing. The input patterns consisted of 481 units representing the numbers coded amino acid residues	58
4.2	The example of incorrect result pattern. The input patterns were the number coded amino acid residues	59
4.3	The example of input and output patterns for training and testing. The input patterns consisted of 481 units representing amino acid sequences which were coded by 2 groups of hydropathy	61
4.4	The example of correct result pattern after testing with a trained network The input patterns were hydropathy (2 groups) coded amino acid residues. The observed outputs were 1 1 0 representing helix, sheet, turn structures respectively	62

Figure		Page
4.5	Percent accuracy prediction of helix, sheet, and turn in the same network using 2 groups of hydropathy as input vector with various numbers of hidden units.....	63
4.6	The example of input and output patterns for training and testing. The input pattern consisted of 481 units representing amino acid sequences which were coded by 7 groups of hydropathy. The outputs consisted of 3 units representing the helix, sheet, turn structures respectively.....	65
4.7	The example of correct result pattern after testing with a trained network. The input patterns were hydropathy (7 groups) coded amino acid residues. The observed outputs were 1 1 1 representing helix, sheet, turn structures respectively.....	66
4.8	Percent accuracy prediction of helix, sheet, and turn in the same network using 7 groups of hydropathy as input vector with various numbers of hidden units.....	67
4.9	The example of input and output patterns for training and testing. The input patterns consisted of 481 units representing amino acid sequences which were coded by 8 groups of amino acid side chain properties. The outputs consisted of 3 units representing the helix, sheet, turn structures respectively.....	69
4.10	The example of result pattern after testing with a trained network. The input patterns were amino acid side chain properties coded amino acid residues. The observed outputs were 1 1 1 representing helix, sheet, turn structures respectively.....	70
4.11	Percent accuracy prediction of helix, sheet, and turn in the same network using 8 groups of amino acid side chain properties as input vector with various numbers of hidden.....	71

Figure		Page
4.12	The example of input and output patterns for training and testing. The input patterns were amino acid sequences which coded by hydrophobicity. The outputs representing the helix, sheet, turn.....	73
4.13	The example of result pattern after testing with a trained network. The input patterns were hydrophobicity coded amino acid residues. The observed outputs were 1 1 1 representing helix, sheet, turn structures respectively.....	74
4.14	Percent accuracy prediction of helix, sheet, and turn in the same network using 3 groups of hydrophobicity as input vector with various numbers of hidden units.....	75
4.15	The comparison of percent accuracy prediction of helix, sheet, and turn in the same network using different input properties.....	77
4.16	The example of input and output patterns for training and testing. The input patterns consisted of 481 units representing amino acid sequences which were coded by 2 groups of hydropathy. The output consisted of 1 unit representing the helix or sheet or turn structure.....	80
4.17	The example of result pattern after testing with a trained network. The input patterns were the hydropathy (2 groups) coded amino acid residues. The observed output was 1 representing helix or sheet or turn structure.....	81
4.18	Percent accuracy prediction of helix structure using 2 groups of hydropathy as input vector with various numbers of hidden units....	82
4.19	The example of input and output patterns for training and testing. The input patterns consisted of 481 units representing amino acid sequences which were coded by 7 groups of hydropathy. The output consisted of 1 unit representing the helix or sheet or turn structure.....	84

Figure **Page**

4.20 The example of correct result pattern after testing with a trained network. The input patterns were hydropathy (7 groups) coded amino acid residues. The observed output was 1 representing helix or sheet or turn structure..... 85

4.21 Percent accuracy prediction of helix structure using 7 groups of hydropathy as input vector with various numbers of hidden units.... 86

4.22 The example of input and output patterns for training and testing. The input patterns consisted of 481 units representing amino acid sequences which were coded by 8 groups of amino acid side chain properties. The output consisted of 1 unit representing the helix or sheet or turn structure..... 87

4.23 The example of correct result pattern after testing with a trained network. The input patterns were amino acid side chain properties coded amino acid residues. The observed output was 1 unit representing helix or sheet or turn..... 88

4.24 Percent accuracy prediction of helix structure using 8 groups of amino acid side chain properties as input vector with various numbers of hidden units..... 89

4.25 The example of input and output patterns for training and testing. The input patterns consisted of 481 units representing amino acid sequences which were coded by hydrophobicity. The output consisted of 1 unit representing helix or sheet or turn structure..... 91

4.26 The example of result pattern after testing with a trained network. The input patterns were hydrophobicity coded amino acid residues. The observed output was 1 representing helix or sheet or turn structure..... 92

4.27 Percent accuracy prediction of helix structure using 3 groups of hydrophobicity as input vector with various numbers of hidden units..... 93

Figure		Page
4.28	The comparison of percent accuracy prediction of helix structure using various types of input properties.....	94 97
4.29	Percent accuracy prediction of sheet structure using 2 groups of hydropathy as input vector with various numbers of hidden units....	
4.30	Percent accuracy prediction of sheet structure using 7 groups of hydropathy as input vector with various numbers of hidden units....	99
4.31	Percent accuracy prediction of sheet structure using 8 groups of amino acid side chain properties as input vector with various numbers of hidden units.....	101
4.32	Percent accuracy prediction of sheet structure using 3 groups of hydrophobicity as input vector with various numbers of hidden units.....	102
4.33	The comparison of percent accuracy prediction of sheet structure using various types of input properties.....	103
4.34	Percent accuracy prediction of turn structure using 2 groups of hydropathy as input vector with various numbers of hidden units....	107
4.35	Percent accuracy prediction of turn structure using 7 groups of hydropathy as input vector with various numbers of hidden units....	108
4.36	Percent accuracy prediction of turn structure using 8 groups of amino acid side chain properties as input vector with various numbers of hidden units.....	110
4.37	Percent accuracy prediction of turn structure using 3 groups of hydrophobicity as input vector with various numbers of hidden units.....	111
4.38	The comparison of percent accuracy prediction of turn structure using various types of input properties.....	112
4.39	The percent accuracy predictions of the existence of helix, sheet, turn in separate networks and in the same network using hydropathy (2 groups).....	113

Figure		Page
4.40	The percent accuracy predictions of the existence of helix, sheet, turn in separate networks and in the same network using hydropathy (7 groups).....	114
4.41	The percent accuracy predictions of the existence of helix, sheet, turn in separate networks and in the same network using amino acid side chain properties (8 groups).....	115
4.42	The percent accuracy predictions of the existence of helix, sheet, turn in separate networks and in the same network using hydrophobicity (3 groups).....	116
4.43	The example of input and output patterns for training and testing. The input patterns consisted of 481 units representing amino acid sequences which were coded by 2 groups of hydropathy. The outputs consisted of 3 units representing percent helix or sheet or turn structures (6 groups).....	120
4.44	The example of result pattern after testing with a trained network. The input patterns were hydropathy (2 groups) coded amino acid residues. The observed outputs were 0 0 1 representing percent helix or sheet or turn (6 groups).....	121
4.45	Percent accuracy prediction of percent helix structure (6 groups) using 2 groups of hydropathy as input vector with various numbers of hidden units.....	122
4.46	The example of input and output patterns for training and testing. The input patterns consisted of 481 units representing amino acid sequences which were coded by 7 groups of hydropathy. The outputs consisted of 3 units representing the percent helix or sheet or turn structures (6 groups).....	123
4.47	The example of result pattern after testing with a trained network. The input patterns were hydropathy (7 groups) coded amino acid residues. The output was percent helix or sheet or turn (6 groups)..	124

Figure		Page
4.48	Percent accuracy prediction of percent helix structure (6 groups) using 7 groups of hydropathy as input vector with various numbers of hidden units.....	125
4.49	The example of input and output patterns for training and testing. The input patterns consisted of 481 units representing amino acid sequences which were coded by 8 groups of amino acid side chain properties. The outputs consisted of 3 units representing percent of helix or sheet or turn structures (6 groups).....	127
4.50	The example of result pattern after testing with a trained network. The input patterns were amino acid side chain properties coded amino acid residues. The observed outputs were 0 0 1 representing percent helix or sheet or turn structures (6 groups).....	128
4.51	Percent accuracy prediction of percent helix structure (6 groups) using 8 groups of amino acid side chain properties as input vector with various numbers of hidden units.....	129
4.52	The example of input and output patterns for training and testing. The input patterns consisted of 481 units representing amino acid sequences which were coded by hydrophobicity. The outputs consisted of 3 units representing percent helix or sheet or turn structures (6 groups) respectively.....	130
4.53	The example of correct result pattern after testing with a trained network. The input patterns were hydrophobicity coded amino acid residues. The observed outputs were 0 0 1 representing percent helix or sheet or turn structures (6 groups).....	131
4.54	Percent accuracy prediction of percent helix structure (6 groups) using hydrophobicity (3 groups) as input vector with various numbers of hidden units.....	132
4.55	The comparison of percent accuracy prediction of percent helix structure (6 groups) using various types of input properties.....	133

Figure		Page
4.56	Percent accuracy prediction of percent sheet structure (6 groups) using 2 groups of hydropathy as input vector with various numbers of hidden units.....	137
4.57	Percent accuracy prediction of percent sheet structure (6 groups) using 7 groups of hydropathy as input vector with various numbers of hidden unit.....	138
4.58	Percent accuracy prediction of percent sheet structure (6 groups) using 8 groups of amino acid side chain properties as input vector with various numbers of hidden units.....	140
4.59	Percent accuracy prediction of percent sheet structure (6 groups) using hydrophobicity (3 groups) as input vector with various numbers of hidden units.....	141
4.60	The comparison of percent accuracy prediction of percent sheet structure (6 groups) using various types of input properties.....	142
4.61	Percent accuracy prediction of percent turn structure (6 groups) using 2 groups of hydropathy as input vector with various numbers of hidden units.....	145
4.62	Percent accuracy prediction of percent turn structure (6 groups) using 7 groups of hydropathy as input vector with various numbers of hidden.....	147
4.63	Percent accuracy prediction of percent turn structure (6 groups) using 8 groups of amino acid side chain properties as input vector with various numbers of hidden units.....	148
4.64	Percent accuracy prediction of percent turn structure (6 groups) using 3 groups of hydrophobicity as input vector with various numbers of hidden units.....	150
4.65	The comparison of percent accuracy prediction of percent turn structure (6 groups) using various types of input properties.....	151

Figure		Page
4.66	The percent accuracy predictions of the percent of helix, sheet, turn (6 groups) in separate networks using hydropathy (2 groups)...	152
4.67	The percent accuracy predictions of the percent of helix, sheet, turn (6 groups) in separate networks using hydropathy (7 groups)...	153
4.68	The percent accuracy predictions of the percent of helix, sheet, turn (6 groups) in separate networks using amino acid side chain properties (8 groups).....	154
4.69	The percent accuracy predictions of the percent of helix, sheet, turn (6 groups) in separate networks using hydrophobicity.....	155
4.70	The example of input and output patterns for training and testing. The input patterns were 2 groups of hydropathy coded amino acid sequences. The outputs consisted of 2 units representing the percent helix or sheet or turn (3 groups).....	158
4.71	The example of correct result after testing with a trained network. The output patterns were the hydropathy (2 groups) coded amino acid residues. The observed outputs were 0 1 representing percent helix or sheet or turn structure (3 groups).....	159
4.72	Percent accuracy prediction of percent helix structure (3 groups) using 2 groups of hydropathy as input vector with various numbers of hidden units.....	160
4.73	The example of input and output patterns for training and testing. The inputs were 7 groups of hydropathy coded amino acid sequences. The outputs consisted of 2 units representing percent helix or sheet or turn structure (3 groups).....	162
4.74	The example of correct result after testing with a trained network. The output patterns were the hydropathy (7 groups) coded amino acid residues. The observed outputs were 0 1 representing helix or sheet or turn structure respectively.....	163

Figure		Page
4.75	Percent accuracy prediction of percent helix structure (3 groups) using 7 groups of hydropathy as input vector with various numbers of hidden units.....	164
4.76	The example of input and output patterns. The inputs were amino acid side chain properties coded amino acid sequences. The outputs consisted of 2 units representing the helix or sheet or turn..	166
4.77	The example of correct result after testing with a trained network. The output patterns were amino acid side chain properties coded amino acid residues. The observed outputs were 0 1 representing helix or sheet or turn structures.....	167
4.78	Percent accuracy prediction of percent helix structure (3 groups) using 8 groups of amino acid side chain properties as input vector with various numbers of hidden units.....	168
4.79	The example of input and output patterns for training and testing . The inputs were hydrophobicity coded amino acid sequences. The outputs consisted of 2 units representing percent helix or sheet or turn structure (3 groups).....	169
4.80	The example of correct result pattern after testing with a trained network. The input patterns were hydrophobicity coded amino acid residues. The observed outputs were 0 1 representing percent helix or sheet or turn (3 groups).....	170
4.81	The comparison of percent accuracy prediction of percent turn structure (3 groups) using various types of input properties.....	171
4.82	The comparison of percent accuracy prediction of percent turn structure (3 groups) using various types of input properties.....	172
4.83	Percent accuracy prediction of percent sheet structure (3 groups) using 2 groups of hydropathy as input vector with various numbers of hidden units.....	176

Figure		Page
4.84	Percent accuracy prediction of percent sheet structure (3 groups) using 7 groups of hydropathy as input vector with various numbers of hidden units	177
4.85	Percent accuracy prediction of percent sheet structure (3 groups) using 8 groups of amino acid side chain properties as input vector with various numbers of hidden units	179
4.86	Percent accuracy prediction of percent sheet structure (3 groups) using hydrophobicity as input vector with various numbers of hidden units	180
4.87	The comparison of percent accuracy prediction of percent sheet structure (3 groups) using various types of input properties	181
4.88	Percent accuracy prediction of percent turn structure (3 groups) using 2 groups of hydropathy as input vector with various numbers of hidden units	185
4.89	Percent accuracy prediction of percent turn structure (3 groups) using 7 groups of hydropathy as input vector with various numbers of hidden units	186
4.90	Percent accuracy prediction of percent turn structure (3 groups) using 8 groups of amino acid side chain properties as input vector with various numbers of hidden units	187
4.91	Percent accuracy prediction of percent turn structure (3 groups) using hydrophobicity as input vector with various numbers of hidden units	189
4.92	The comparison of percent accuracy prediction of percent turn structure (3 groups) using various types of input properties	190
4.93	The percent accuracy predictions of the percent of helix, sheet, turn (3 groups) in separate networks using hydropathy (2 groups)...	191

Figure		Page
4.94	The percent accuracy predictions of the percent of helix, sheet, turn (3 groups) in separate networks using hydropathy (7 groups)...	192
4.95	The percent accuracy predictions of the percent of helix, sheet, turn (3 groups) in separate networks using amino acid side chain properties.....	193
4.96	The percent accuracy predictions of the percent of helix, sheet, turn (3 groups) in separate networks using hydrophobicity.....	194
4.97	Percent accuracy prediction of percent helix structure (2 groups) using 2 groups of hydropathy as input vector with various numbers of hidden units.....	197
4.98	Percent accuracy prediction of percent helix structure (2 groups) using helical tendencies as input vector with various numbers of hidden.....	199
4.99	Percent accuracy prediction of percent helix structure (2 groups) using 8 groups of amino acid side chain properties as input vector with various numbers of hidden units.....	200
4.100	Percent accuracy prediction of percent helix structure (2 groups) using hydrophobicity as input vector with various numbers of	
4.101	hidden units.....	202
	The comparison of percent accuracy prediction of percent helix structure (2 groups) using various types of input properties.....	203
4.102	Percent accuracy prediction of percent sheet structure (2 groups) using 2 groups of hydropathy as input vector with various numbers of hidden units.....	206
4.103	Percent accuracy prediction of percent sheet structure (2 groups) using 8 groups of amino acid side chain properties as input vector with various numbers of hidden units.....	208

Figure		Page
4.104	Percent accuracy prediction of percent sheet structure (2 groups) using hydrophobicity as input vector with various numbers of hidden units.....	209
4.105	The comparison of percent accuracy prediction of percent sheet structure (2 groups) using various types of input properties.....	211

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย