CHAPTER II
LITERATURE REVIEW

Proteins

Proteins perform diverse biological functions. They store and transport a variety of particles ranging from macromolecules to electrons. They guide the flow of electrons in the vital process of photosynthesis. As hormones, they transmit information between specific cells and organs in complex organisms. Some proteins control the passage of molecules across the membranes that compartmentalize cells and organelles. Other proteins function in the immune systems of complex organisms to defend against intruders. Moreover, proteins also control gene expression by binding to specific sequences of nucleic acid, thereby turning genes on and off.

In spite of these diverse biological functions, proteins are a relatively homogeneous class of molecules. All are the same type of linear polymer, built of various combinations of the same 20 amino acids that are linked by peptide bonds. These polypeptide are different in amino acid composition. It has been shown that the functions of proteins are related to the three dimensional (3D) structure of such proteins, which in turn prescribed by the folding of the amino acid residues in the sequence (Voet, 1990; Creighton, 1993).

Proteins Structure

Primary structure

Primary structure of a protein includes all the covalent bonds between amino acids and is normally defined by the sequence of peptide-bonded amino acids and locations of disulfide bonds. The relative spatial arrangement of the linked amino acids is unspecified (Leghninger *et al.*, 1993).

Secondary structure

Secondary structure refers to regular, recurring arrangements in space of adjacent amino acid residues in a polypeptide chain. There are a few common types of secondary structure, the most prominent being the α helix and the β conformation. Different amino acids have different tendencies to form different types of secondary structure (Table 2.1) (Leghninger *et al.*, 1993; Creighton, 1993).

Tertiary structure

Tertiary structure refers to the spatial relationship among all amino acids in a polypeptide. It is the complete three-dimensional structure of the polypeptide. The boundary between secondary and tertiary structure is not always clear. Several different types of secondary structure are often found within the three-dimensional structure of a large protein (Leghninger *et al.*, 1993).

Quaternary structure

Quaternary structure refers to the spatial relationship of the polypeptides, or subunits, within the protein (Lehninger *et al.*, 1993).

Table 2.1 Relative frequency of occurrence of amino acid residues in the secondary structures of proteins

Amino acid	α helix	β sheet	β turn
Ala	1.29	0.90	0.78
Cys	1.11	0.74	0.80
Leu	1.30	1.02	0.59
Met	1.47	0.97	0.39
Glu	1.44	0.75	1.00
Gln	1.27	0.80	0.97
His	1.22	1.08	0.69
Lys	1.23	0.77	0.96
Val	0.91	1.49	0.47
Ile	0.97	1.45	0.51
Phe	1.07	1.32	0.58
Tyr	0.72	1.25	1.05
Trp	0.99	1.14	0.75
Thr	0.82	1.21	1.03
Gly	0.56	0.92	1.64
Ser	0.82	0.95	1.33
Asp	1.04	0.72	1.41
Asn	0.90	0.76	1.28
Pro	0.52	0.64	1.91
Arg	0.96	0.99	0.88

Source : Creighton, T.H. *Protein: Structures and Molecular Properties* (W.H. Freeman, 1983), p. 235.

Three - Dimensional structure of proteins

Each amino acid residue in a polypeptide chain contains three bonds of the polypeptide backbone, plus those of side chain. The peptide bond of the backbone has partial double-bond and is limited to planar *cis* or *trans* rotations. The other two bonds of the backbone have single-bond character which almost occur in the bond of various side chain. Rotation about any of these bonds produces different conformations which are stabilized by weak interactions. These different structures give rise to different functions of proteins. In the same way, the properties of a protein are largely determined by its three-dimensional (3D) structure.

Secondary structure

As previously described the secondary (2°) structure of a polymer is defined as the local conformation of its backbone. For proteins, this has come to mean the specification of regular polypeptide backbone folding patterns: helices, pleated sheets, and turns (Voet, 1990; Lehninger *et al.*, 1993).

A. Helical Structure

The helical structure is the result of the polypeptide chain twisted by the same amount about each of its C_α atoms given a constraint that the R groups of the amino acid residues protrude outward from the helical backbone. The α **helix**, a particular rigid arrangement of the polypeptide chain, was discovered by Pauling in 1951. The conformation of an α helix is characterized by having ψ = -45° to -50° and ϕ = -60°, and each helical turn includes 3.6 amino acids. The twisting of the helix has a right handed sense in the most common form of the helix, whereas a left handed sense has rarely found (Lehninger *et al.*, 1993). The hydrogen bond of the α helix are arranged such that the peptide C == O bond of the *n*th residues points along the helix towards the peptide N — H group of the $(n + 4)^{th}$ residue, which results in strong hydrogen bond. The

hydrogen bond linked the adjacent coil of α-helix together and give rise stable structure. Various amino acids are thought to have different tendencies to form α helices. Table 2.2, shows such a relative tendencies of amino acids measured using short peptides of defined sequences (Voet, 1990).

B. Beta Structure

The β **pleated sheet** was also discovered by Pauling and Coring in 1951. It has repeating ϕ and ψ angles and utilizes the full hydrogen bonding capacity of the polypeptide backbone. There are two types of β pleated sheet (Voet, 1990):

1. **The antiparallel β pleated sheet**, in which the hydrogen bonded polypeptide chains run in opposite directions.
2. The **parallel β pleated sheet**, in which the hydrogen bonded chains extend in the same direction .

C. Nonrepetitive Structure

The α helix and the β pleated sheet are the major repetitive secondary structure mostly found in the wide variety of proteins. Other repetitive structures exist, often in one or a few specialized proteins. β **bend** or β **turn** structure is the result of abruptly reversed direction of polypeptide chain which always connected with antiparallel β sheet. When the CO group of the first residue hydrogen bonded with the forth residue of NH group that resulting in a hairpin turn structure. Gly and Pro residues often occur in β turns because they are small and flexible. About 6% of the peptide bonds involving the imino nitrogen of the Pro are in cis configuratins which many of these occur in a tight turn. β Turns are often found near the surface of the protein (Lehninger *et al.*, 1993)

Tertiary Structure

The tertiary structure of a protein is the folding of its secondary structure elements, together with the spatial dispositions of its side chains (Voet, 1995). Amino acids that are far apart in the polypeptide sequence and residue in different types of secondary structure may interact when the protein is folded. The formation of bends in the polypeptide chain during folding and the direction and angle of these bends are determined by the number and location of specific bend-producing amino acids such as Pro, Thr, Ser, and Gly residues. Moreover, loop of highly folded polypeptide chain are held in their characteristic tertiary position by different kinds of weak-bonding interactions between R groups of adjacent loops (Lehninger *et al.*, 1993). The tertiary structure can be classified into four classes base on the folding class. There are four folding classes defined as follow (Levitt and Chothia, 1976):

1. *All α proteins* have mainly α helix secondary structure, for example, Myohaemerythrin, Myogen and Myoglobin.
2. *All β proteins* have mainly β sheet secondary structure, for example, Rubredoxin, Prealbumin, concanavalin and Chymotrypsin.
3. *α+β proteins* have α-helix and *β-strand* secondary structure segments that do not mix but tend to segregate along the polypeptide chain, for example, T4 Lysozyme, Papain and thermolysin.
4. *α/β proteins* have mixed or approximately alternating segments of α-helical and β-strand secondary structure, for example, Thioredoxin, Flavodoxin and Triose phosphate isomerase.

Table 2.3 shows the quantitative criteria for folding types classification in terms of secondary structure content from various references. The criteria of Nishikawa *et al.* (1983) are the most often used (Eisenhaber *et al.*, 1995).

Table 2.2 Relative helical tendencies of the amino acids measured in one peptide.

Amino acid residue	**Relative stabilization of α-helical conformation (kcal/mol)**
Ala	-0.77
Arg	-0.68
Lys	-0.65
Leu	-0.62
Met	-0.50
Trp	-0.45
Phe	-0.41
Ser	-0.35
Gln	-0.33
Glu	-0.27
Cys	-0.23
Ile	-0.23
Tyr	-0.17
Asp	-0.15
Val	-0.14
Thr	-0.11
Asn	-0.07
His	-0.06
Gly	0
Pro	~3

Source : . O'Neil, K.T and DeGrado, W.F. 1990. *Science.* 250 : 646-651.

Creighton, T.H. 1993. *PROTEINS : Structures and Molecular Properties.* 2 ed. (W.H.Freeman and Company) p. 186.

Table2.3 Definition of secondary structural classes of proteins.

Class		Definition	
	α-content	β-content	References
All-α	> 15%	< 10%	Nishikawa and Ooi, 1982
All-β	< 10%	> 15%	
Mixed	> 15%	> 10%	
Irregular	< 15%	< 15%	
All-α	$\alpha > \beta$	$\alpha < \beta$	Sheridan *et al.*, 1985
All-β	$\alpha < \beta$	$\alpha > \beta$	
Parallel	With parallel β-sheet	With parallel β-sheet	
Irrecular	With 4.5% Cys	With 4.5% Cys	
All-α	>40%	<5%	Klein and DeLisi 1986
all-β	<10%	>30%	
Mixed	≥15%	≥15%	
Irrecular		$\alpha + \beta$ <20%	
All-α	≥30%	$\geq 0.15 \cdot (\alpha + \beta)$	Kneller *et al.*, 1990
All-β	≤10%		
Mixed	>15%	>5%	
Irregular	All remaining proteins	All remaining proteins	

Source : Eisenhaber, F. *et al.* 1995. *Critical Review in Biochemistry and Molecular Biology.* 30(1), p. 25

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

The amino acid of proteins

Most protein are composed of 20 standard amino acids (Table 2.4). The 20 standard amino acids are α-amino acids except proline because a primary amino group and a carboxylic acid group substitute on the same carbon atom. The general structure of α-amino acid is shown in Figure 2.1. These amino acid differ in side chains (R groups) which can be divided into 8 major types (Table 2.4): **Aliphatics, Aromatics, Imino, Sulfur, Hydroxy, Basics, Acidics, Amides** (Stryer, 1988). Generally, 50 up to more than 2000 amino acids are linked by **peptide bonds** (Figure 2.2) to form a polypeptide chain, each of which is a polymer of amino acid residues (Creighton, 1993). The properties of amino acid side chains depend inpart on their physical and chemical properties such as size, shape, chemical reactivity, H-bonding capability and the presence of any ionizable groups.

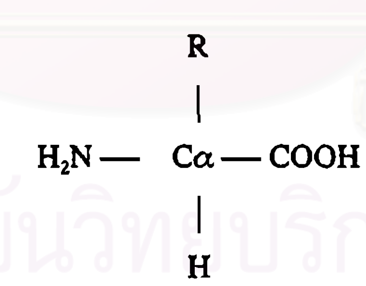

Figure 2.1 The general structural formula of α-amino acid.

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

Table 2.4 Covalent structure and abbreviations of 20 standard amino acids of proteins.

Group	Name Three letter symbol One letter symbol	Structure formula
Aliphatics	Glycine, Gly, G	COO^- H—C—H NH_3^+
	Alanine, Ala, A	COO^- H—C—CH_3 NH_3^+
	Valine, Val, V	COO^- CH_3 H—C—CH NH_3^+ CH_3
	Leucine, Leu, L	COO^- CH_3 H—C—CH_2—CH NH_3^+ CH_3
	Isoleucine, Ile, I	COO^- CH_3 H—C——C—CH_2—CH NH_3^+ H

สถาบันวิทยบริการ จุฬาลงกรณ์มหาวิทยาลัย

Table 2.4 (continued)

Group	Name Three letter symbol One letter symbol	Structure formula
Aromatics	Phenylalanine, Phe, F	COO^- H—C—CH_2—(ring)—OH NH_3^+
	Tryptophan, Trp, W	COO^- H—C—CH_2—(indole ring, N–H) NH_3^+
	Tyrosine, Tyr, Y	COO^- H—C—CH_2—(ring) NH_3^+
Imino	Proline, Pro, P	COO^-, C, H, H_2C, CH_2, CH_2, N–H (ring)
Sulfur	Cysteine, Cys, C	COO^- H—C—CH_2—S NH_3^+
	Methionine, Met, M	COO^- H—C—CH_2—CH_2—S—CH_3 NH_3^+

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

Table 2.4 (continued)

Group	Name Three letter symbol One letter symbol	Structure formula
Hydroxy	Serine, Ser, S	$H-C(COO^-)(NH_3^+)-CH_2-OH$
	Threonine, Thr, T	$H-C(COO^-)(NH_3^+)-C(H)(OH)-CH_3$
Basics	Lysine, Lys, K	$H-C(COO^-)(NH_3^+)-CH_2-CH_2-CH_2-CH_2-NH_3^+$
	Arginine, Arg, R	$H-C(COO^-)(NH_3^+)-CH_2-CH_2-CH_2-NH-C(NH_2)=NH_2^+$
	Histidine, His, H	$H-C(COO^-)(NH_3^+)-CH_2-$ (imidazole ring: N, N–H)

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

Table 2.4 (continued)

Group	Name Three letter symbol One letter symbol	Structure formula
Acidics	Aspartic acid	$H-C(COO^-)(NH_3^+)-CH_2-C(=O)O^-$
	Glutamic acid, Glu, E	$H-C(COO^-)(NH_3^+)-CH_2-CH_2-C(=O)O^-$
Amides	Asparagine, Asn, N	$H-C(COO^-)(NH_3^+)-CH_2-C(=O)NH_2$
	Glutamine, Gln, Q	$H-C(COO^-)(NH_3^+)-CH_2-CH_2-C(=O)NH_2$

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

Peptide

A peptide is composed of amino acid residues joined covalently through peptide bonds (Creighton, 1993).

Peptide Bonds

A Peptide bond is the CO-NH linkage resulting from α-amino acids polymerized through the elimination of a water molecule. Each amino acid residue is linked to its neighbors in a head-to-tail direction rather than forming branched chains. Thus most peptides are linear polymers (Lehninger *et al.*, 1993)

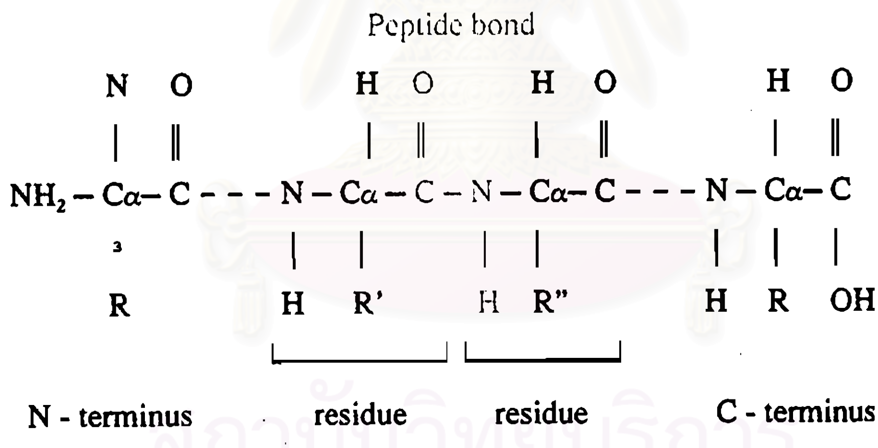

Figure 2.2 The peptide bond

Acid -Base Properties

Amino acids and peptides have acid-base properties (Voet, 1990). The α- amino acids have one or two acid - base groups. The relationship between pH and base to acid ratio of an acidic is known as Henderson - Hasselbalch equation:

$$pH = pK + \log \frac{[A^-]}{[HA]}$$

where pK = -log dissociation constant, [HA] is the concentration of the acidic specie and $[A^-]$ is the concentration of its conjugate base.

The noncovalent interactions that determine the properties of proteins

The three-dimensional structures of protein result from the folding of the polypeptide chains. Many simultaneous interactions take place among different parts of the molecule. The biological activities of protein are also mediated by their interactions with their environment. All of these interactions arise from a limited set of fundamental noncovalent forces, which are:

Short-Range Repulsion

The repulsion always takes place between atoms and between molecules where closing to each other. When atom or molecules come near enough for their electron orbitals to begin to overlap, the repulsion increases enormously because the electrons on the different molecules cannot be in the same part of space at the same time. The repulsive energy increases with the inverse of the 12th power of the distance between the centers of the two atoms (Creighton, 1993).

Electrostatic Forces

Molecules are collections of electrically charged particles. Thus, their interactions are determined by the laws of classical electrostatics. The energy of association, U, of two electric charges, q_1 and q_2, that are separated by the distance r, is found by integrating the

หอสมุดกลาง สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

expression for Coulomb's law, $F = kq_1q_2/Dr^2$, to determine the work necessary to separate these charges by an infinite distance:

$$U = \frac{kq_1q_2}{Dr}$$

which $k = 9.0 \times 10^9$ j.m.c^{-2} and D is the dielectric constant of the medium in which the charges are immersed.

Van Der Waals Interactions

Van Der Waals interactions are a weak force where close range of atoms and molecules attract each other, even in the absence of charged groups. These interaction result from electrostatic interactions among permanent and/or induced dipoles. These forces are responsible for numerous interaction of varying strengths between nonbonded neighboring atoms.

Hydrogen Bonding

Hydrogen bonds are electrostatic interaction between a dipole of covalent bond to the hydrogen atom, in which the hydrogen atom has a partial positive charge, and a partial negative charge on the other electronegative atom.

D — H···· A

It is the interaction of weakly acidic donor group (D — H) and a lone pair electrons acceptor (A). (Creighton, 1993).

Hydrophobic Forces

The hydrophobic effect arise from interaction of nonpolar substances which trying to minimize their contacts with water. Hydrophobic interactions are a major influence that causes proteins to fold into their native conformations (Kauzmann, 1950). Table 2.5 shows an index of combined hydrophobic and hydrophilic tendencies. The hydropathies are a good index for the prediction of which portions of a polypeptide chain are inside a

protein, out of contact with the aqueous solvent, and which portions are outside, in contact with the aqueous solvent. In proteins, the effects of hydrophobic forces are often termed *hydrophobic bonding* because that bonding does not generated the directionally specific interaction (Voet, 1990).

Disulfide Bond

Disulfide bonds are covalent and can be kept intact under appropriate conditions. Proteins that contain disulfide bonds can be cleaved into peptides under appropriate conditions. From this reaction, the peptide that contain disulfide bond or are linked by them can be identified (Creighton, 1993)

The Protein Data Bank

The Protein Data Bank (PDB) is the collection of experimentally-determined three-dimensional structures of biological macromolecules. The archive contains atomic coordinates, bibliographic citations, primary and secondary structure information, as well as crytallographic structure factors and NMR experimental data. The PDB are under the direction of Dr. Joel L. Sussman and the staffs at Brookhaven National Laboratory New York State, USA. The data are available over the Internet. The protein database always include sequence, references (sequencing papers, X-ray, NMR papers, other papers), taxonomic data, annotations (function(s), Post-translation modification, domain, Quaternary structure, similarities to other proteins, disease(s) associated with deficiency(ies), variances, etc.), keywords, cross reference to other data banks (PDB; http:\\www.pdb.bnl.gov)

Table 2.5 Hydropathy associated with standard amino acids.

Amino acids	Hydropathy
Ile	4.5
Val	4.2
Leu	3.8
Phe	2.8
Cys	2.5
Met	1.9
Ala	1.8
Gly	-0.4
Thr	-0.7
Ser	-0.8
Trp	-0.9
Tyr	-1.3
Pro	-1.6
His	-3.2
Glu	-3.5
Gln	-3.5
Asp	-3.5
Asn	-3.5
Lys	-3.9
Arg	-4.5

Source: Kyte, J. and Doolitle, R. (1982) , *J.Mol.Biol.*, 157. p. 110.
Voet. (1990). Biochemistry. (John Wiley& Sons). p. 179.

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

Artificial neural network

Biological neural system

The brain is composed of about 10^{11} neuron (nerve cells) of many different types. *Dendrites* are single tubular fibers forming tree-like networks of nerve fiber. These dendrites are connected to the *cell body* or center part, where the cell nucleus is located. A single fiber which is extending from the cell body is called on *axon*. This fiber serves to transmit the generated neural activity to other nerve cell. At the ends of these are the transmitting ends of the synaptic junctions or *synapses*, to other neuron. The receiving ends of these junctions on other cells can be found both on the dendrites and on the cell bodies themselves. The axon of a typical neuron makes a few thousand synapses with other neurons. The transmission of a signal from one cell to another at a synapse is either electrical or chemical. In chemical process, specific transmitter substances are released from the sending site of the junction. The effect is to raise or lower the electrical potential inside the body of the receiving cell. If this potential reaches a threshold, a pulse or action potential of fixed strength and duration is sent to down the axon (Muller & Reinhardt, 1990; Hertz *et al.*, 1991)

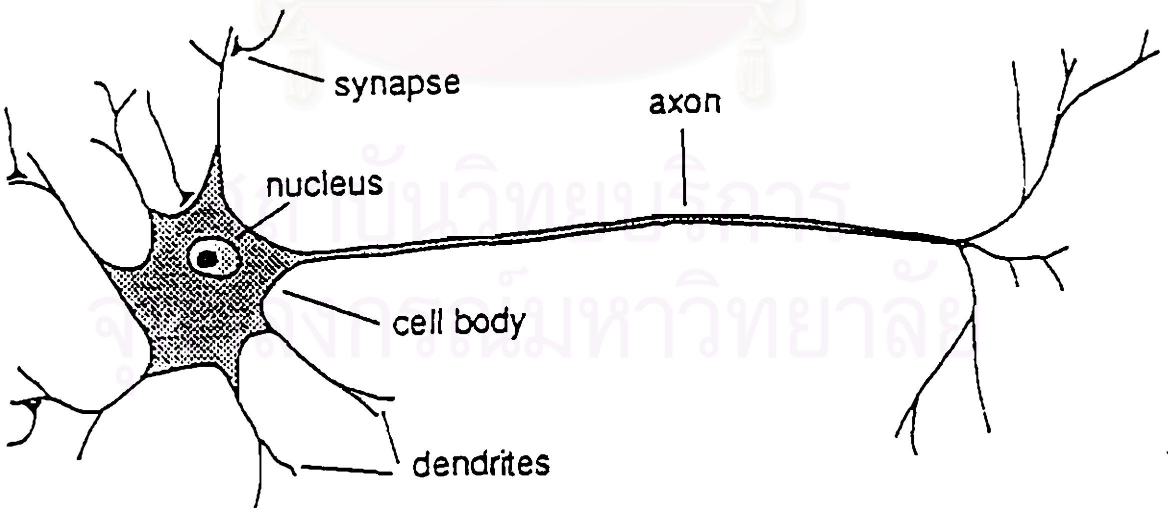

Figure 2.3 Schematic drawing of a typical neuron.

Source: Hertz, *et al.*, 1991. *Introduction to the theory of neural computation*. (Addison-Wesley Publishing) p. 2.

The basic structure of artificial neural network

A neural network is a computational model composed of nodes (units or neuron) and connection between nodes. The numbers associated with the inputs to each node are its weights. These are very roughly based on the firing rate of a biological neuron and the strength of a synapse (connection) between two neurons in the human brain (Alexander & Morton, 1990; Zeidenberg, 1990).

The topology of a neural network refers to its interconnection scheme which is specified by the number of *layers* and the number of nodes per layer (Fig. 1.4). The type of layers include (Fu. 1994):

The input layer: This layer consists of *input units*, which encode the information presented to the network for processing. Theses units perform no processing but only distribute information to other units.

The hidden layer: This layer consists of hidden units, which can not be seen from outside the network.

The output layer: This layer consists of output units, which encode possible values with consideration of input information.

According to the connection scheme, a connection between nodes in different layers is called an *interlayer connection*. A *intralayer connection*, on the other hand, is a connection between nodes within the same layer.

Learning in neural network is the adjustment of weights via a learning rule. The adjustment is made when the network is trained until the desired output is obtained. When leaning is complete, the weights are fixed, unless there is further learning.

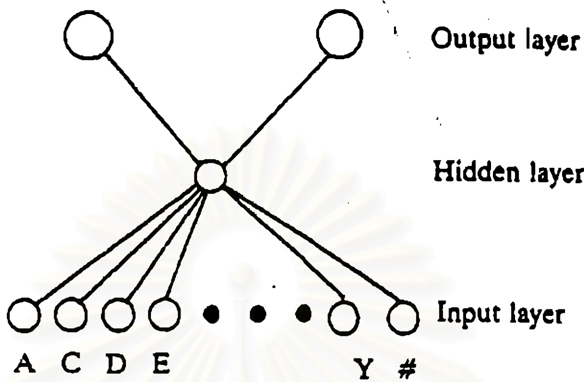

Figure 2.4 The basic structure of neural networks.
Source: Dubchak, *et al.*, 1995. *Proc. Nalt. Acad. Sci.* Vol 92 : p. 8701.

McCulloch and Pitts Model

This McCulloch and Pitts neuron (MCP) is a simple model of a neuron as a binary threshold unit. Specifically, the model neuron computes a weighted sum of its inputs from other units, and outputs a one or a zero according to whether this sum is above or below a certain threshold:

$$n_i\,(t+1) = \Theta\,(\textstyle\sum_j w_{ij} n_j\,(t) - \mu_i) \qquad (1)$$

and $\Theta\,(x)$ is a step function

$$\Theta\,(x) = \begin{cases} 1 & \text{if } X \geq 0\,; \\ 0 & \text{otherwise} \end{cases} \qquad (2)$$

The weight, w_{ij}, is the strength of the synapse connecting neuron j to neuron i. It can be positive or negative corresponding to an excitatory or inhibitory synapse respectively. If there is no synapse between i and j, $w_{ij} = 0$

μ_i is the threshold value for unit i; the weighted sum of inputs must reach or exceed the threshold for the neuron to fire.

n_j is the state of neuron i either *firing* or *not firing* as 1 or 0 respectively.

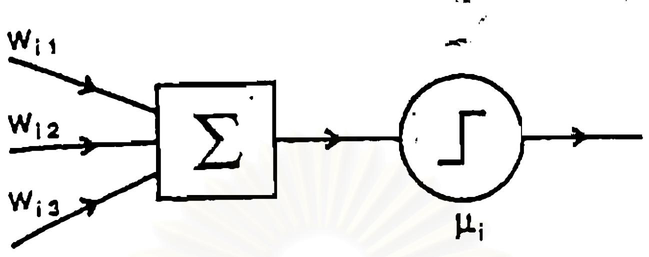

Figure 2.5 Schematic diagram of McCulloch and Pitts.

Source: Hertz, *et al.*, 1991. *Introduction to the theory of neural computation*. (Addison-Wesley Publishing) p. 3.

A simple generalized of MCP equation is

$$n_i = g(\Sigma_j w_{ij}n_{ij} - \mu_{ij}) \qquad (3)$$

n_i is a continuos value and is called the state or activation of unit i

$g(x)$ is a nonlinear function called the activation function, gain function, transfer function, or squashing function. This replaces the threshold function $\Theta(x)$ of (1). So that can simply give a rule for updating whenever that occurs rather than writing the time t or $t+1$ explicitly as in equation (1). Units are often updated asynchronously: in random order at random times.

Feed-Forward Networks

Layered feed-forward networks or perceptrons (Rosenblatt, 1962) are composes of a set of input terminals whose only role is to feed input patterns into the rest of the networks. After this can come one or more intermediate layers of units, followed by a final output layer where the result of the computation is read off. The units in intermediate layers are often called *hidden units* because they have no direct connection to the outside word, neither input nor output.

One layer feed forward network or simple perceptrons. There is a set of *n,* inputs and output layers, but no hidden layers. Its computation is describe by

$$O_i = g(h_i) = g(\Sigma_j W_{ij}X_j) \qquad (4)$$

O_i = output of unit *i*

X_j = input of unit *i*

g(h) is activation function computed by the units and usually nonlinear. It can be used as a threshold function or sigmoid function.

If θ_i is a threshold of output unit *i* which can be treated as connection to an input terminal that is permanently clamped at -1. If $X_0 = -1$ and choose connection strengths $W_{i0} = \theta_i$ to obtain

$$O_i = g(\Sigma_{j=0} W_{ik}X_k) = g(\Sigma_{j=1} W_{ij}X_j - \theta_i) \qquad (5)$$

The output of each input pattern μ is

$$O_i^{\mu} = g(\Sigma_{j=0} W_{ij}X_j^{\mu}) \qquad (6)$$

by fixed *Ti* as a target output. Thus, the desired output from the neural net is

$$O_i^{\mu} = T_i^{\mu} \qquad (7)$$

for each *i* and μ

The learning algorithm for feed -forward networks are as follow :

Weight Initialization. Weights and node thresholds are to first act small random numbers

Calculation of Activation.

1. The activation level of an input unit is determined by the instance presented to the network.
2. The activation level O_i of an output unit is determined by

$$O_i = g(\Sigma W_{ij}X_j - \theta_i) \qquad (8)$$

Weight training

1. The weight is adjusted by

$$W_{ij}^{new} = W_{ij}^{old} + \Delta W_{ij} \quad (9)$$

where W_{ij}^{old} is the weight from unit i to unit j and ΔW_{ij} is the weight adjustment.

2. The weight change may be computed by the delta rule:

$$\Delta W_{ij} = \eta \delta_j X_i \quad (10)$$

where η is a trial-independent learning rate $(0 < \eta < 1)$ and δ_j is the error at unit j:

$$\delta_j = T_j - O_j \quad (11)$$

T_j is the desired (target) output activation and Oj is the actual output activation at output unit j.

3. The iterations are repeated until covergence is obtained.

The perceptron may use the delta rule to perform weight training. As seen in equation (10), the weight adjustment is proportional to the error. The adjusted weight will make the network output closer to the desired output. Furthermore, the weight adjustment also depends on the input (Xi). If the input is zero, there will be no adjustment. In this case, the weight does not contribute to the network output, so it should not be blamed. The learning rate η sets the step size. If η is too small, the convergence is unnecessarily slow, whereas if η is too large learning process may diverge (Fu, 1994).

Multi Layers feed forward neural network

A multilayer perceptrons is a feedforward neural network with at least one hidden or intermediate layer between input and output (Rumelhard *et al*, 1986). It can deal with nonlinear classification problems because it can form more complex decision regions. Each node in the first layer can create a hyperplane. Each node in the second layer can combine hyperplanes to create convex decision regions. Each node in the third layer combine convex regions to form concave regions (Fig. 2.6). It is thus possible to form any arbitrary regions with sufficient layers and sufficient hidden units (Fu, 1994).

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

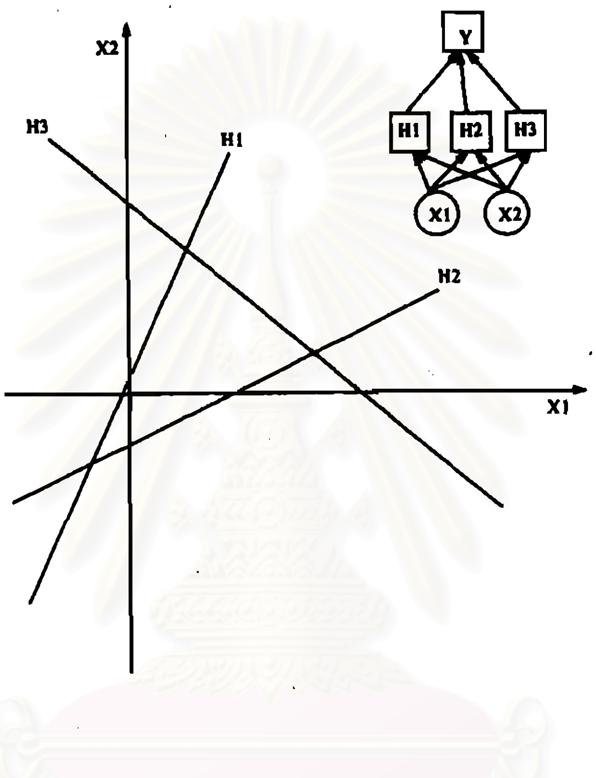

Figure. 2.6 Decision regions created by a multilayer perceptrons

Source : Fu, 1994. *Neural Networks in Computer Intelligence.* (McGraw-Hill).

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

Back-propagation

A back-propagation algorithm was invented independently several times, by Bryson and Ho (1969), Werbos (1974), Parker (1985) and Rumelhart *et al.* (1986). A closely related approach was proposed by Le Cun (1985). The algorithm gives a prescription for changing the weight in any feedforward network to learn a training set of input-output pairs. Typically, back-propagation employs three or more layers of processing units.

Figure 2.7 shows the topology for a three-layer propagation network. The bottom layer of units is the input layer, the only units in the network that receive external input. The layer above is the hidden layer, in which the processing units are interconnected to layers above and below. The top layer is the output layer. All layers are fully interconnect. Units are not connected to other units in the same layer (Dayhoff, 1990).

The backpropagation network in essence learns a mapping from a set of input patterns (e.g., extracted features) to a set of output patterns (e.g., class information). This network can be designed and trained to accomplish a wide variety of mappings. This ability comes from the nodes in hidden layers or layers of the network which learn to respond to features found in the input patterns. The features recognized or extracted by the hidden units (nodes) correspond to the correlation of activity among different input units. As the network is trained with different examples, the network has the ability to generalize over similar features found in different patterns. Thus the hidden units must be trained to extract a sufficient set of general features applicable to both seen and unseen instances.

The back-propagation network is capable of approximating arbitrary mapping giving a set of examples. The sigmoid function guarantees that the outputs are bounded between 0 and 1 (Rumelhart *et al.*, 1986; Fu, 1994).

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

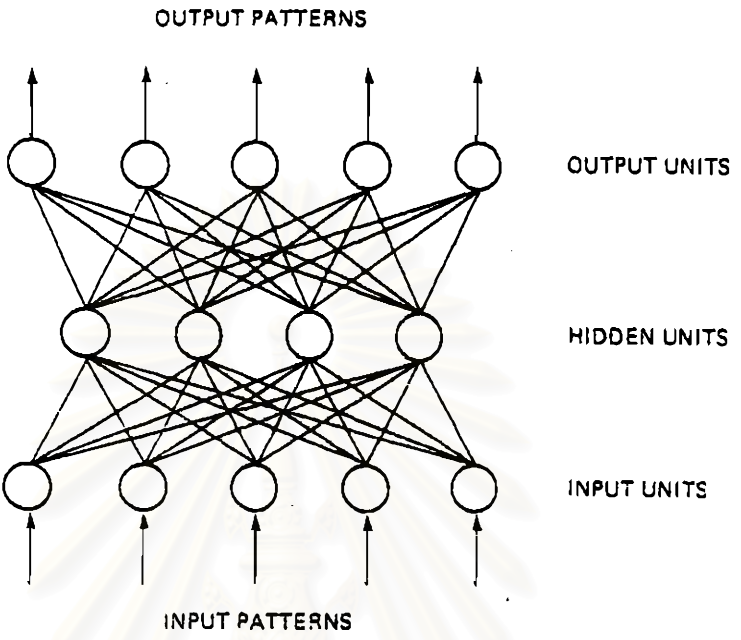

Figure 2.7 A three layer back propagation network fully interconnection.
Source: Dayhoff, 1990. *Neural Network Architectures*. (Van nostrand reihold). p.60

The learning algorithm for back-propagation network are as follow :

Weight Initialization

The weights and node thresholds are first set to small random numbers. Note that the note threshold is the negative of the weight from the bias unit (whose activation level is fixed at 1).

Calculation of Activation

1. The activation level of an input unit is determined by the instance present of network.
2. The activation level O_i of a hidden unit and output unit is determined by node

$$O_i = g(\Sigma W_{ij} O_j - \theta_i) \qquad (12)$$

where W_i is the weight from an input , θ_J is the node thershold and g(h) is the sigmoid function:

$$g(h) = \frac{1}{1+e^{-h}} \qquad (13)$$

Weight Training.

1. Starting at the output units and work backward to the hidden layer recursively, the weights are adjusted by

   $$W_i^{(t+1)} = W_{ij}^{t} + \Delta W_{ij} \qquad (14)$$

   Where W_{ij}^{t} is the weight from unit *i* to unit *j* at time *t* (or t^{th} iteration) and ΔW_{ij} is the weight adjustment.

2. The weight change is computed by

   $$\Delta W_{ij} = \eta \delta_i O_j \qquad (15)$$

   Where η is a trial-independent learning rate ($0 < \eta < 1$) and δ_i is the error gradient at unit *j*. Convergence is sometimes faster by adding a momentum termm :

   $$W_i^{(t+1)} = W_{ij}^{t} + \eta \delta_i O_j + \alpha [W_{ij}^{t} - W_i^{(t-1)}] \qquad (16)$$

   where $0 < \alpha < 1$.

3. The error gradient is given by :

   - For the output units :

   $$\delta_i = O_i(1-O_i)(T_i - O_i) \qquad (17)$$

   Where T_i is the desired (target) output activation and O_i is the actual output activation at output unit *j*.

   - For hidden units:

   $$\delta_i = O_i(1-O_i)\Sigma_k \delta_k W_{ki} \qquad (18)$$

   where δ_k is the error gradient at unit *k* to which a connection points from hidden unit *j*.

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

4. The iterations are repeated until convergence in terms of the selected error criterion is obtained. An iteration includes presenting an instance, calculating activation, and modifying weights.

The name "back-propagation" comes from the fact that the error (gradient) of hidden units are derived from propagating backward the errors associated with output units (as Eq.18) since the target values for the hidden unit are not given. In the back-propagation network, the activation function chosen is the sigmoid function, which compresses the output value into the range between 0 and 1. The sigmoid function is advantageous in that it can accommodate large signals without saturation while allowing the passing of small signals without attenuation. Also, it is a smooth function so that gradients can be calculated, which are required for a gradient descent search.

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย