CHAPTER IV

RESULTS

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

Training set.

Seventy patterns of amino acid sequences were saved as a learn-pattern file for training (Table 4.1). Using these learn-pattern file, the network was trained for the prediction of helix, sheet and turn in the same network by amino acid properties in section 4.1.1, 4.1.2, 4.1.3, 4.1.4 and 4.1.5.

Testing set.

Twenty eight patterns of amino acid sequences were saved as a test-pattern file for testing (Table 4.1). Using this test pattern file, the network was tested for the prediction of helix, sheet and turn in the same network by amino acid properties in section 4.1.1, 4.1.2, 4.1.3, 4.1.4 and 4.1.5.

4.1.1 Prediction of helix, sheet and turn in the same network using numbers substituted amino acids

The input pattern was composed of 481 units of amino acids which were coded with the following symbols:

GLY = 0.01 ALA = 0.06 VAL = 0.11 LEU = 0.16
ILE = 0.21 PRO = 0.26 PHE = 0.31 TYR = 0.36
TRP = 0.41 SER = 0.46 THR = 0.51 CYS = 0.56
MET= 0.61 ASN= 0.66 GLN = 0.71 ASP = 0.76
GLU= 0.81 LYS = 0.86 ARG = 0.91 HIS = 0.96

These numbers did not stand for any significant property of amino acids. The output pattern was composed of 3 units of helix, sheet and turn. Symbol “1” and “0” represented an existence and non-existence of these three structures, respectively (Figure 4.1). The networks were trained with 7 and 70 hidden units. For testing, 28 input patterns were tested with the trained network. The accuracy of prediction obtained from the networks with 7 and 70 hidden units were both 28 percent. The example of result patterns id shown in Figure 4.2.

Table 4.1 The training set, testing set and possible outputs of secondary structures prediction.

Training set (Input No.)*	**Testing set (Input No.)***	**Output patterns**	**Structure type existing in amino acid sequence**
7, 13, 18, 25, 48, 59, 73, 78	77, 83, 87	1 0 0	only helix
4, 43	–	0 1 0	only sheet
–	–	0 0 1	only turn
1, 5, 8, 9, 15, 24, 26, 33, 34, 37, 38, 39, 41, 45 46, 50, 52, 55, 75	57, 67, 69, 76, 89, 91, 93	1 1 0	helix and sheet
23, 28, 40, 53, 61, 70, 84	86, 88, 94	1 0 1	helix and turn
6, 10, 11, 64	68	0 1 1	sheet and turn
2, 3, 12, 14, 16, 17, 19, 20, 21, 22, 27, 29, 30, 31, 32, 35, 36, 42, 44, 47, 49, 51, 54, 56, 58, 60, 62, 63, 65, 66	71, 72, 74, 79, 80, 81, 82, 85, 90, 92, 95, 96, 97, 98	1 1 1	helix, sheet and turn

```
Names of input no. were listed in Table 3.1
SNNS pattern definition file V1.4

No. of patterns     : 70
No. of input units  : 481
No. of output units : 3

#input1 lalc
 0.86 0.76 0.31 0.51 0.86 0.56 0.81 0.16 0.46 0.71 0.66 0.16 0.36
 0.76 0.21 0.76 0.01 0.36 0.01 0.91 0.21 0.06 0.16 0.26 0.81 0.16
 0.21 0.56 0.51 0.61 0.31 0.96 0.51 0.46 0.01 0.36 0.76 0.51 0.71
 0.06 0.21 0.11 0.81 0.66 0.76 0.81 0.46 0.51 0.81 0.36 0.01 0.16
 0.31 0.71 0.21 0.46 0.66 0.06 0.16 0.41 0.56 0.86 0.46 0.46 0.71
 0.46 0.26 0.71 0.46 0.91 0.66 0.21 0.56 0.76 0.21 0.51 0.56 0.76
 0.86 0.31 0.16 0.76 0.76 0.76 0.21 0.51 0.76 0.76 0.21 0.61 0.56
 0.06 0.86 0.86 0.21 0.16 0.76 0.21 0.86 0.01 0.21 0.76 0.36 0.41
 0.21 0.06 0.96 0.86 0.06 0.16 0.56 0.51 0.81 0.86 0.16 0.81 0.71
 0.41 0.16 0.56 0.81 0.86 0.81 0.00 0.00 0.00 0.00 0.00 0.00 0.00
 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
#output1 lalc
  1    1    0

#input2
 0.06 0.16 0.41 0.71 0.31 0.66 0.01 0.61 0.21 0.86 0.56 0.86 0.21
 0.26 0.46 0.46 0.81 0.26 0.16 0.16 0.76 0.31 0.66 0.66 0.36 0.01
 0.56 0.36 0.56 0.01 0.16 0.01 0.01 0.46 0.01 0.51 0.26 0.11 0.76
 .... .... .... .... .... .... .... .... .... .... .... .... ....
 .... .... .... .... .... .... .... .... .... .... .... .... ....

#output2 lbp2
 ....  ....  ....

#input70 8adh
 .... .... .... .... .... .... .... .... .... .... .... .... ....
 .... .... .... .... .... .... .... .... .... .... .... .... ....
```

Figure 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. The outputs consisted of 3 units representing the helix, sheet and turn structures respectively.

```
SNNS result  file V1.4-3D

No. of patterns       : 70
No. of input units    : 481
No. of output units   : 3
start pattern         : 1
end pattern           : 28
input patterns included
teaching outputs included

#input1
 0.46 0.96 0.96 0.41 0.01 0.36 0.01 0.86 0.96 0.66 0.01 0.26 0.81
 0.96 0.41 0.96 0.86 0.76 0.31 0.26 0.21 0.06 0.86 0.01 0.81 0.91
 0.71 0.46 0.26 0.11 0.76 0.21 0.76 0.51 0.96 0.51 0.06 0.86 0.36
 0.76 0.26 0.46 0.16 0.86 0.26 0.16 0.46 0.11 0.46 0.36 0.76 0.71
 0.06 0.51 0.46 0.16 0.91 0.21 0.16 0.66 0.66 0.01 0.96 0.06 0.31
 0.66 0.11 0.81 0.31 0.76 0.76 0.46 0.71 0.76 0.86 0.06 0.11 0.16
 0.86 0.01 0.01 0.26 0.16 0.76 0.01 0.51 0.36 0.91 0.16 0.21 0.71
 0.31 0.96 0.31 0.96 0.41 0.01 0.46 0.16 0.76 0.01 0.71 0.01 0.46
 0.81 0.96 0.51 0.11 0.76 0.86 0.86 0.86 0.36 0.06 0.06 0.81 0.16
 0.96 0.16 0.11 0.96 0.41 0.66 0.51 0.86 0.36 0.01 0.76 0.31 0.01
 0.86 0.06 0.11 0.71 0.71 0.26 0.76 0.01 0.16 0.06 0.11 0.16 0.01
 0.21 0.31 0.16 0.86 0.11 0.01 0.46 0.06 0.86 0.26 0.01 0.16 0.71
 0.86 0.11 0.11 0.76 0.11 0.16 0.76 0.46 0.21 0.86 0.51 0.86 0.01
 0.86 0.46 0.06 0.76 0.31 0.51 0.66 0.31 0.76 0.26 0.91 0.01 0.16
 0.16 0.26 0.81 0.46 0.16 0.76 0.36 0.41 0.51 0.36 0.26 0.01 0.46
 0.16 0.51 0.51 0.26 0.26 0.16 0.16 0.81 0.56 0.11 0.51 0.41 0.21
 0.11 0.16 0.86 0.81 0.26 0.21 0.46 0.11 0.46 0.46 0.81 0.71 0.11
 0.16 0.86 0.31 0.91 0.86 0.16 0.66 0.31 0.66 0.01 0.81 0.01 0.81
 0.26 0.81 0.81 0.16 0.61 0.11 0.76 0.66 0.41 0.91 0.26 0.06 0.71
 0.26 0.16 0.86 0.66 0.91 0.71 0.21 0.86 0.06 0.46 0.31 0.86 0.00
 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
#output3
   1       1       1
0.9890  0.9800  0.0123

#input2
 0.86 0.86 0.11 0.11 0.16 0.01 0.86 0.86 0.01 0.76 0.51 0.11 0.81
 0.16 0.51 0.56 0.51 0.06 0.46 0.71 0.86 0.86 0.46 0.21 0.71 0.31
 .... .... .... .... .... .... .... .... .... .... .... .... ....
 .... .... .... .... .... .... .... .... .... .... .... .... ....

#output2
 ....  ....  ....
 ....  ....  ....
#input28
 .... .... .... .... .... .... .... .... .... .... .... .... ....
 .... .... .... .... .... .... .... .... .... .... .... .... ....
```

Figure 4.2 The example of incorrect result after testing with a trained network. The input patterns were the number coded amino acid residues. The observed outputs were 1 1 1 representing helix, sheet, turn structures respectively. The predicted outputs were the three floating numbers, 0.9890, 0.9800 and 0.0123.

4.1.2 Prediction of helix, sheet and turn in the same network using hydropathy (2 groups)

The hydropathy tendency of each amino acid was classified into 2 groups (Table 3.1). All amino acid sequences were coded by the symbols representing these two groups.

Input and output patterns for training and testing

Each input pattern was composed of 481 input units of amino acid which coded by 2 groups of hydropathy. The output pattern was composed of 3 units of helix sheet and turn. Symbol "1" and "0" represented an existence and non-existence of these 3 structures, respectively (Figure 4.3). There were 7 possible output results from the training set (Table 4.1).

Training and testing.

The learn-pattern file, *thy2gr3op.pat)*, with 70 patterns was used in training the networks with various hidden units. The network file of each trial (one for each number of hidden units) was shown in Table 4.2. The maximum value for cycle was set to 100000 and the maximum error to stop was set to 0.01.

The test-pattern file, *tehy2gr3op.pat*, with 28 patterns was used in testing the trained network as described above. The trained network files and the testing result files are also shown in Table 4.2 and the example of result patterns is shown in Figure 4.4. Figure 4.5 and Table 4.2 show that both *hy2gr3op70.net* and *hy2gr3op100.net* which had 70 and 100 hidden units respectively, gave the highest percent accuracy prediction, 50%. The percent accuracy prediction decreased to 42.86% where the network had 120 units which was the highest unit in this trial. Although both networks gave the same highest percent accuracy prediction, *hy2gr3op70.net* was suitable for using as the network prediction because it was more compact and trained faster than *hy2gr3op100.net*. The network with140 hidden units was also trained and gave the same result as 120 hidden network (data not show). However, this network with 140 hidden units spent too much time for training. Thus, the network with 140 hidden units did not performed in others study.

```
SNNS pattern definition file V1.4
No. of patterns        : 70
No. of input units     : 481
No. of output units    : 3

#input1 1alc
       0.2   0.2   0.1   0.2   0.2   0.1   0.2   0.1   0.2   0.2   0.2   0.1   0.2
       0.2   0.1   0.2   0.2   0.2   0.2   0.2   0.1   0.1   0.1   0.2   0.2   0.1
       0.1   0.1   0.2   0.1   0.1   0.2   0.2   0.2   0.2   0.2   0.2   0.2   0.2
       0.1   0.1   0.1   0.2   0.2   0.2   0.2   0.2   0.2   0.2   0.2   0.2   0.1
       0.1   0.2   0.1   0.2   0.2   0.1   0.1   0.2   0.1   0.2   0.2   0.2   0.2
       0.2   0.2   0.2   0.2   0.2   0.2   0.1   0.1   0.2   0.1   0.2   0.1   0.2
       0.2   0.1   0.1   0.2   0.2   0.2   0.1   0.2   0.2   0.2   0.1   0.1   0.1
       0.1   0.2   0.2   0.1   0.1   0.2   0.1   0.2   0.2   0.1   0.2   0.2   0.2
       0.1   0.1   0.2   0.2   0.1   0.1   0.1   0.2   0.2   0.2   0.1   0.2   0.2
       0.2   0.1   0.1   0.2   0.2   0.2   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
#output1 1alc
      1     1  0


#input2 1bp2
       0.1   0.1   0.2   0.2   0.1   0.2   0.2   0.1   0.1   0.2   0.1   0.2   0.1
       0.2   0.2   0.2   0.2   0.2   0.1   0.1   0.2   0.1   0.2   0.2   0.2   0.2
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
#outut2 1bp2
      ...    ....  ....

#input70 8adh

       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
```

Figure 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. The outputs consisted of 3 units representing the helix, sheet, turn structures respectively.

```
SNNS result file V1.4-3D

No. of hidden units  : 70
No. of patterns      : 28
No. of input units   : 481
No. of output units  : 3
start pattern        : 1
end pattern          : 28
input patterns included
teaching outputs included

#input1
  0.2   0.2   0.1   0.1   0.1   0.2   0.2   0.2   0.2   0.2   0.2   0.1   0.2
  0.1   0.2   0.1   0.2   0.1   0.2   0.2   0.2   0.2   0.2   0.1   0.2   0.1
  0.2   0.2   0.2   0.2   0.2   0.2   0.2   0.1   0.2   0.1   0.1   0.2   0.2
  0.2   0.2   0.2   0.1   0.1   0.2   0.2   0.2   0.2   0.2   0.2   0.1   0.2
  0.2   0.2   0.1   0.2   0.2   0.2   0.2   0.2   0.1   0.2   0.2   0.2   0.2
  0.2   0.1   0.2   0.1   0.1   0.1   0.2   0.2   0.1   0.2   0.1   0.2   0.2
  0.2   0.2   0.2   0.2   0.1   0.1   0.2   0.1   0.2   0.2   0.2   0.2   0.2
  0.2   0.1   0.2   0.1   0.1   0.1   0.1   0.2   0.1   0.2   0.1   0.2   0.2
  0.2   0.2   0.2   0.1   0.1   0.2   0.2   0.2   0.2   0.1   0.2   0.1   0.2
  0.1   0.2   0.2   0.2   0.2   0.2   0.2   0.2   0.2   0.2   0.1   0.2   0.1
  0.2   0.2   0.2   0.2   0.2   0.2   0.2   0.1   0.2   0.2   0.2   0.2   0.2
  0.1   0.2   0.1   0.2   0.2   0.1   0.2   0.1   0.2   0.2   0.2   0.2   0.2
  0.2   0.2   0.1   0.2   0.1   0.1   0.2   0.2   0.2   0.2   0.2   0.1   0.2
  0.1   0.2   0.1   0.2   0.1   0.1   0.1   0.1   0.1   0.0   0.0   0.0   0.0
  0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
  0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
  0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
  0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
  0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
  0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
  0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
  0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
  0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
  0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
  0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
  0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
  0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
  0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
  0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
  0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
  0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
  0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
  0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
  0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
  0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
  0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
  0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
#output1
    1        1        0
 0.9941   0.9873   0.0011

#input2
  0.2   0.2   0.1   0.1   0.2   0.1   0.2   0.1   0.2   0.2   0.2   0.1   0.2
  0.2   0.2   0.2   0.1   0.2   0.2   0.1   0.2   0.1   0.2   0.2   0.2   0.2
  0.2   0.1   0.2   0.1   0.1   0.1   0.2   0.2   0.2   0.2   0.2   0.2   0.1
  ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
  ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
#output2
   ....    ....    ....
   ....    ....    ....

#input28
        ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
        ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
```

Figure 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. The predicted outputs were the three floating numbers, 0.9941, 0.9873 and 0.0011.

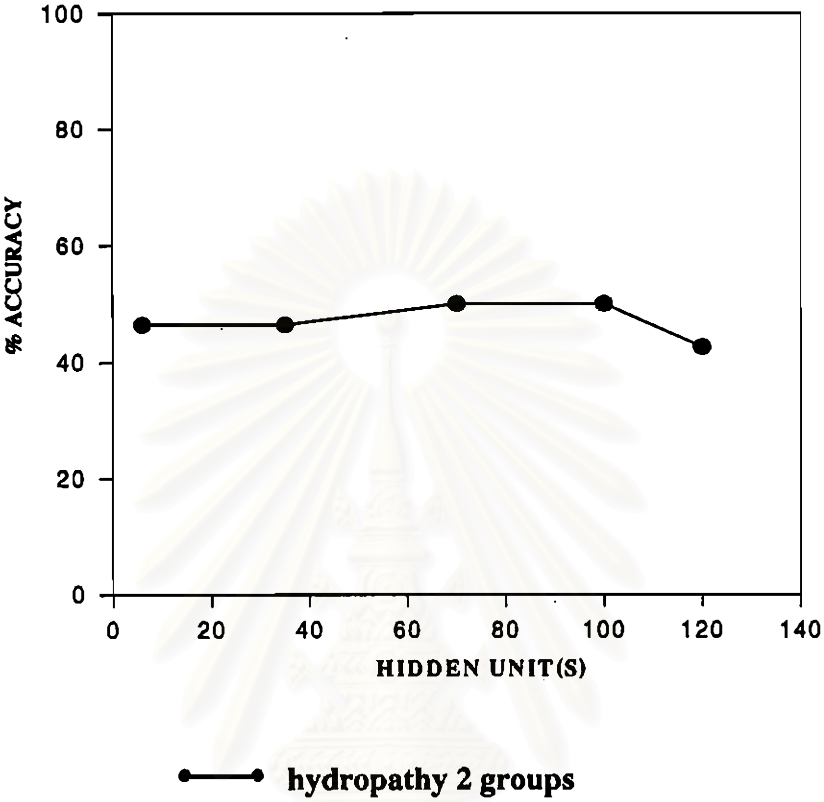

Figure 4.5 Percent accuracy prediction of helix, sheet, and turn in the same network using 2 groups of hydropathy as input vector with various numbers of hidden units (7, 35, 70, 100, 120).

Table 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).

Hidden units	No. of units	No. of links	Training network file	Testing result file	% accuracy of prediction
7	491	3388	hy2gr3op7.net	tehy2gr3op7.res	46.43
35	519	16940	hy2gr3op35.net	tehy2gr3op35.res	46.43
70	554	33880	hy2gr3op70.net	tehy2gr3op70.res	50.00
100	584	48400	hy2gr3op100.net	tehy2gr3op100.res	50.00
120	604	58080	hy2gr3op120.net	tehy2gr3op120.res	42.86

4.1.3 Prediction of helix, sheet and turn in the same network using hydropathy classified into 7 groups.

The hydropathy tendency of each amino acid was classified into 7 groups (Table 3.2). All amino acid sequence were coded by the symbol which representing for these seven groups of hydropathy.

Input and output patterns for training and testing.

Each input pattern was composed of 481 input units of amino acid residue which were coded by 7 groups of hydropathy. The output pattern consisted of 3 units of helix, sheet and turn. Symbol "1" and "0" represented an existence and non-existence of these structure respectively (Figure 4.6). The training and testing sets are shown in Table 4.1.

Training and testing.

The learn-pattern file, *thy7gr3op.pat,* with 70 patterns was used in training the network with various hidden units (Figure 4.6) The trained networks and testing result files are shown in Table 4.3. The maximum learn cycles was set to 100000 and the maximum error to stop was set to 0.01.

The test-pattern file, *tehy7gr3op.pat,* with 28 patterns was used in the testing. The trained network, *hy7gr3op7.net*, gave the highest accuracy prediction, 46.43%. Whereas, the network with 35 and 70 hidden units gave the lowest accuracy prediction, 39.29%. Thus, the network with 7 hidden unit was the best network for the prediction the prediction of the existence of these 3 structures using hydropathy classified into 7 groups (Table 4.3 and Figure 4.8). The example of testing result patterns is shown in (Figure 4.7)

```
SNNS pattern definition file V1.4

No. of patterns        : 70
No. of input units     : 481
No. of output units    : 3

#input1 1alc
       0.7     0.6     0.2     0.4     0.7     0.2     0.6     0.1     0.4     0.6     0.6     0.1     0.5
       0.6     0.1     0.6     0.4     0.5     0.4     0.7     0.1     0.3     0.1     0.5     0.6     0.1
       0.1     0.2     0.4     0.3     0.2     0.6     0.4     0.4     0.4     0.5     0.6     0.4     0.6
       0.3     0.1     0.1     0.6     0.6     0.6     0.6     0.4     0.4     0.6     0.5     0.4     0.1
       0.2     0.6     0.1     0.4     0.6     0.3     0.1     0.4     0.2     0.7     0.4     0.4     0.6
       0.4     0.5     0.6     0.4     0.7     0.6     0.1     0.2     0.6     0.1     0.4     0.2     0.6
       0.7     0.2     0.1     0.6     0.6     0.6     0.1     0.4     0.6     0.6     0.1     0.3     0.2
       0.3     0.7     0.7     0.1     0.1     0.6     0.1     0.7     0.4     0.1     0.6     0.5     0.4
       0.1     0.3     0.6     0.7     0.3     0.1     0.2     0.4     0.6     0.7     0.1     0.6     0.6
       0.4     0.1     0.2     0.6     0.7     0.6     0.0     0.0     0.0     0.0     0.0     0.0     0.0
       0.0     0.0     0.U     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0
       0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0
       0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0
       0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0
       0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0
       0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0
       0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0
       0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0
       0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0
       0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0
       0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0
       0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0
       0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0
       0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0
       0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0
       0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0
       0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0
       0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0
       0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0
       0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0
       0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0
       0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0
       0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0
       0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0
       0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0
       0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0
       0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0
#output1 1alc
     1  1  0

#input2
       0.3     0.1     0.4     0.6     0.2     0.6     0.4     0.3     0.1     0.7     0.2     0.7     0.1
       0.5     0.4     0.4     0.6     0.5     0.1     0.1     0.6     0.2     0.6     0.6     0.5     0.4
       0.2     0.5     0.2     0.4     0.1     0.4     0.4     0.4     0.4     0.4     0.5     0.1     0.6
       ...     ...     ...     ...     ...     ...     ...     ...     ...     ...     ...     ...     ...
       ...     ...     ...     ...     ...     ...     ...     ...     ...     ...     ...     ...     ...
#outut2 1bp2
       ...   ....   ....

#input70 8adh

       ...     ...     ...     ...     ...     ...     ...     ...     ...     ...     ...     ...     ...
       ...     ...     ...     ...     ...     ...     ...     ...     ...     ...     ...     ...     ...
       ...     ...     ...     ...     ...     ...     ...     ...     ...     ...     ...     ...     ...
```

Figure 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.

```
SNNS result file V1.4-3D

No. of hidden units : 70
No. of patterns     : 28
No. of input units  : 481
No. of output units : 3
startpattern        : 1
endpattern          : 28
input patterns included
teaching outputs included

#input1
       0.4   0.6   0.6   0.4   0.4   0.5   0.4   0.7   0.6   0.6   0.4   0.5   0.6
       0.6   0.4   0.6   0.7   0.6   0.2   0.5   0.1   0.3   0.7   0.4   0.6   0.7
       0.6   0.4   0.5   0.1   0.6   0.1   0.6   0.4   0.6   0.4   0.3   0.7   0.5
       0.6   0.5   0.4   0.1   0.7   0.5   0.1   0.4   0.1   0.4   0.5   0.6   0.6
       0.3   0.4   0.4   0.1   0.7   0.1   0.1   0.6   0.6   0.4   0.6   0.3   0.2
       0.6   0.1   0.6   0.2   0.6   0.6   0.4   0.6   0.6   0.7   0.3   0.1   0.1
       0.7   0.4   0.4   0.5   0.1   0.6   0.4   0.4   0.5   0.7   0.1   0.1   0.6
       0.2   0.6   0.2   0.6   0.4   0.4   0.4   0.1   0.6   0.4   0.6   0.4   0.4
       0.6   0.6   0.4   0.1   0.6   0.7   0.7   0.7   0.5   0.3   0.3   0.6   0.1
       0.6   0.1   0.1   0.6   0.4   0.6   0.4   0.7   0.5   0.4   0.6   0.2   0.4
       0.7   0.3   0.1   0.6   0.6   0.5   0.6   0.4   0.1   0.3   0.1   0.1   0.4
       0.1   0.2   0.1   0.7   0.1   0.4   0.4   0.3   0.7   0.5   0.4   0.1   0.6
       0.7   0.1   0.1   0.6   0.1   0.1   0.6   0.4   0.1   0.7   0.4   0.7   0.4
       0.7   0.4   0.3   0.6   0.2   0.4   0.6   0.2   0.6   0.5   0.7   0.4   0.1
       0.1   0.5   0.6   0.4   0.1   0.6   0.5   0.4   0.4   0.5   0.5   0.4   0.4
       0.1   0.4   0.4   0.5   0.5   0.1   0.1   0.6   0.2   0.1   0.4   0.4   0.1
       0.1   0.1   0.7   0.6   0.5   0.1   0.4   0.1   0.4   0.4   0.6   0.6   0.1
       0.1   0.7   0.2   0.7   0.7   0.1   0.6   0.2   0.6   0.4   0.6   0.4   0.6
       0.5   0.6   0.6   0.1   0.3   0.1   0.6   0.6   0.4   0.7   0.5   0.3   0.6
       0.5   0.1   0.7   0.6   0.7   0.6   0.1   0.7   0.3   0.4   0.2   0.7   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
#output1
     1       1        1
  0.9900  0.8750   0.9540

#input2
       0.7   0.7   0.1   0.1   0.1   0.4   0.7   0.7   0.4   0.6   0.4   0.1   0.6
       0.1   0.4   0.2   0.4   0.3   0.4   0.6   0.7   0.7   0.4   0.1   0.6   0.2
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
#output2
   ....   ....   ....
   ....   ....   ....

#input28
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
```

Figure 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. The predicted outputs were the three floating numbers, 0.99, 0.875 and 0.954

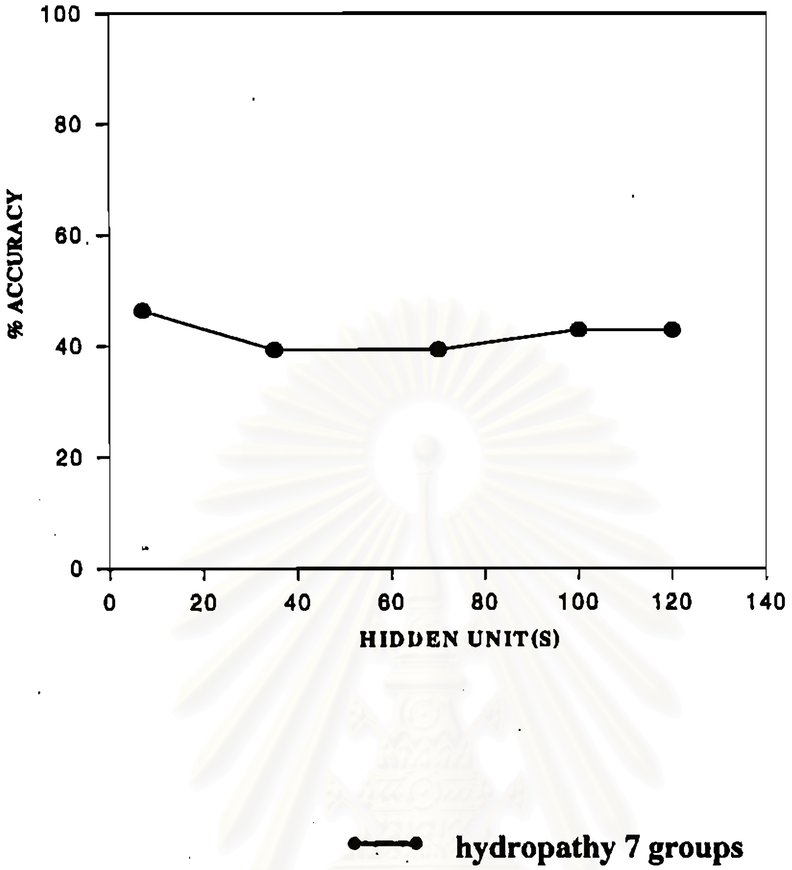

Figure 4.8 Percent accuracy prediction of helix, sheet, and turn in the same network using 7 groups of hydropathy as input vector with various numbers of hidden units (7, 35, 70, 100, 120).

Table 4.3 The trained network file, testing result file and percent accuracy of helix, sheet and turn prediction in the same network using hydropathy (7 groups).

Hidden units	No. of units	No. of links	Training network file	Testing result file	% accuracy of prediction
7	491	3388	hy7gr3op7.net	tehy7gr3op7.res	46.43
35	519	16940	hy7gr3op35.net	tehy7gr3op35.res	39.29
70	554	33880	hy7gr3op70.net	tehy7gr3op70.res	39.29
100	584	48400	hy7gr3op100.net	tehy7gr3op100.res	42.86
120	604	58080	hy7gr3op120.net	tehy7gr3op120.res	42.86

4.1.4 Prediction of helix, sheet, and turn in the same network using 8 groups of amino acid side chain properties

The amino acid sequences of proteins were substituted by the 8 groups of amino acid side chain as input patterns for training and testing.

Input and output pattern for training and testing.

Each input pattern was composed of 481 input units of amino acid which coded by 8 groups of the side chain properties. The output pattern consisted of 3 units of helix, sheet and turn. A symbol "1" and a symbol "0" were represented for an existence and non-existence of these structure respectively. The training set and testing set were shown in Table 4.1.

Training and testing.

The learn-pattern file, *aa8gr3op.pat,* with 70 patterns was used in training the network with various hidden units (Figure 4.9) The trained networks and testing result files was shown in Table 4.4. The Max learn cycles was set to 100000 and the Max error to stop was set to 0.01.

The test-pattern file, *teaa8gr3op.pat,* with 28 patterns (Figure 4.10) was used in the testing. Both *aa8gr3op70.net* and *aa8gr3op100.net*, gave the highest accuracy of 52.63%. Whereas, the networks with 7 and 35 hidden units gave 50 % accuracy prediction. The lowest accuracy prediction resulted from *aa8gr3op120.net* (44.74%). Thus, the network with 70 hidden units was the best network for the prediction of the existence of these 3 structure using the amino acid side chain properties which classified into 8 groups (Table 4.4 and Figure 4.11).

```
SNNS pattern definition file V1.4

No. of patterns       : 70
No. of input units    : 481
No. of output units   : 3

#input1 1alc
       0.1   0.5   0.3   0.7   0.2   0.1   0.2   0.7   0.7   0.6   0.8   0.1   0.0
       0.3   0.7   0.8   0.2   0.5   0.6   0.1   0.2   0.3   0.1   0.1   0.8   0.1
       0.8   0.8   0.8   0.5   0.3   0.1   0.7   0.1   0.6   0.5   0.5   0.7   0.5
       0.6   0.6   0.7   0.5   0.5   0.1   0.6   0.3   0.1   0.5   0.1   0.5   0.2
       0.8   0.3   0.1   0.5   0.1   0.6   0.7   0.1   0.1   0.8   0.1   0.1   0.6
       0.5   0.2   0.6   0.1   0.8   0.2   0.7   0.7   0.8   0.8   0.7   0.6   0.5
       0.1   0.1   0.6   0.1   0.1   0.3   0.2   0.5   0.7   0.5   0.2   0.6   0.1
       0.2   0.8   0.2   0.6   0.2   0.6   0.2   0.1   0.5   0.5   0.8   0.7   0.6
       0.1   0.5   0.7   0.6   0.5   0.1   0.7   0.1   0.1   0.6   0.2   0.5   0.1
       0.7   0.1   0.6   0.1   0.1   0.6   0.2   0.8   0.5   0.1   0.6   0.2   0.5
       0.5   0.1   0.1   0.7   0.1   0.1   0.5   0.6   0.1   0.7   0.1   0.1   0.1
       0.1   0.1   0.1   0.1   0.1   0.4   0.6   0.1   0.1   0.7   0.1   0.8   0.3
       0.6   0.1   0.8   0.6   0.1   0.1   0.7   0.1   0.1   0.8   0.1   0.1   0.6
       0.5   0.6   0.1   0.6   0.6   0.1   0.3   0.2   0.5   0.8   0.2   0.7   0.3
       0.5   0.5   0.1   0.1   0.3   0.5   0.5   0.1   0.7   0.2   0.2   0.5   0.2
       0.3   0.1   0.5   0.1   0.5   0.6   0.3   0.3   0.1   0.2   0.7   0.5   0.1
       0.5   0.2   0.1   0.1   0.4   0.6   0.7   0.5   0.1   0.5   0.1   0.5   0.5
       0.7   0.8   0.1   0.1   0.8   0.2   0.6   0.5   0.1   0.1   0.5   0.8   0.1
       0.7   0.1   0.7   0.8   0.1   0.1   0.3   0.4   0.8   0.6   0.8   0.8   0.6
       0.3   0.5   0.8   0.3   0.1   0.6   0.1   0.6   0.5   0.1   0.6   0.1   0.5
       0.2   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
#output1 1alc
    1   1   1


#input2 1bp2
       0.1   0.3   0.6   0.1   0.3   0.1   0.7   0.1   0.1   0.6   0.4   0.7   0.6
       0.5   0.6   0.8   0.3   0.1   0.1   0.2   0.8   0.6   0.5   0.5   0.6   0.6
       0.1   0.1   0.6   0.4   0.1   0.7   0.4   0.6   0.6   0.3   0.1   0.8   0.1
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
#outut2 1bp2
      ...  ....  ....


#input70 8adh
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
```

Figure 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.

```
SNNS result file V1.4-3D

No. of hidden units : 70
No. of patterns     : 28
No. of input units  : 481
No. of output units : 3
startpattern        : 1
endpattern          : 28
input patterns included
teaching outputs included

#input1
       0.4   0.8   0.5   0.1   0.5   0.7   0.6   0.1   0.6   0.6   0.6   0.6   0.1
       0.1   0.1   0.6   0.4   0.5   0.8   0.5   0.7   0.1   0.1   0.5   0.6   0.1
       0.1   0.1   0.6   0.8   0.8   0.5   0.1   0.8   0.1   0.1   0.7   0.1   0.1
       0.1   0.5   0.6   0.6   0.3   0.6   0.2   0.1   0.2   0.7   0.1   0.1   0.4
       0.1   0.1   0.8   0.4   0.7   0.3   0.1   0.2   0.1   0.8   0.2   0.1   0.5
       0.6   0.6   0.1   0.6   0.1   0.1   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0

#output1
   1        1        1
0.9452   0.8761   0.9235

#input2
       0.3   0.6   0.2   0.5   0.1   0.1   0.7   0.6   0.7   0.5   0.4   0.1   0.1
       0.4   0.1   0.1   0.4   0.1   0.1   0.1   0.1   0.3   0.7   0.1   0.2   0.7
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
#output2
   ....     ....     ....
   ....     ....     ....

#input28
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
```

Figure 4.10 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 outputs were 1 1 1 representing helix, sheet, turn structures respectively. The predicted outputs were the three floating numbers, 0.9452, 0.8761 and 0.9235.

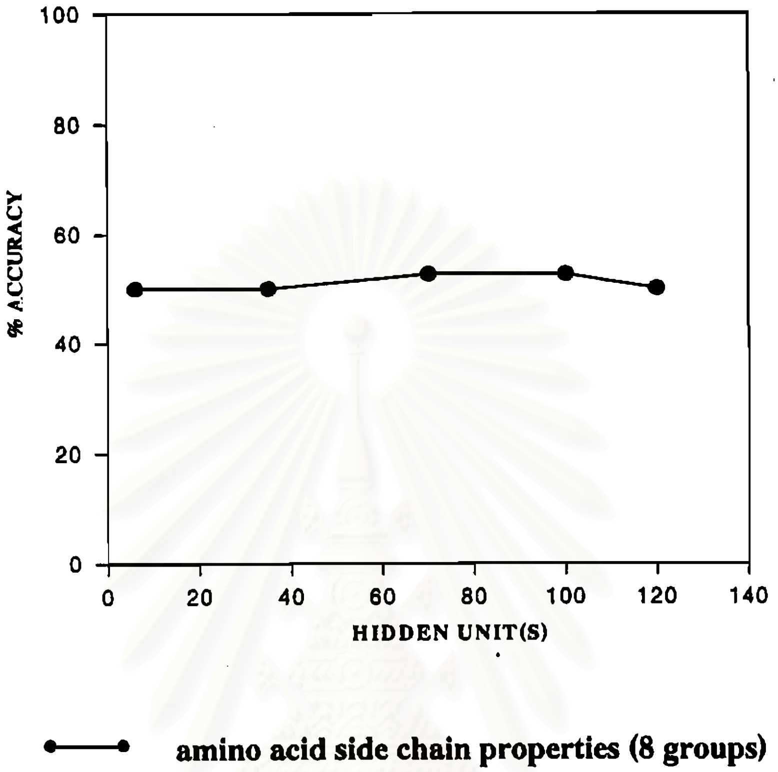

Figure 4.11 Percent accuracy prediction of helix, sheet, and turn in the same network using 8 groups of amino acid side chain properties as input vector with various numbers of hidden units (7, 35, 70, 100, 120).

Table 4.4 The trained network file, testing result file and percent accuracy of helix, sheet and turn prediction in the same network using amino acid side chain properties.

Hidden units	No. of units	No. of links	Training network file	Testing result file	% accuracy of prediction
7	491	3388	aa8gr3op7.net	teaa8gr3op7.res	50.00
35	519	16940	aa8gr3op35.net	teaa8gr3op35.res	50.00
70	554	33880	aa8gr3op70.net	teaa8gr3op70.res	52.63
100	584	48400	aa8gr3op100.net	teaa8gr3op100.res	52.63
120	604	58080	aa8gr3op120.net	teaa8gr3op120.res	44.74

4.1.5 Prediction of helix, sheet and turn in the same network using hydrophobiciy (3 groups)

The relative hydrophobicity of amino acids was classified into 3 groups (Section 3.2.2). All amino acid sequences were coded by the symbols representing these 3 groups.

Input and output pattern for training and testing.

Each input pattern was composed of 481 input units of amino acid which coded by 3 groups of the hydrophobicity. The output pattern consisted of 3 units of helix, sheet and turn. A symbol "1" and a symbol "0" were represented for an existence and non-existence of these structure respectively. The training set and testing set are shown in Table 4.1.

Training and testing.

The learn-pattern file, *thydro3op.pat,* with 70 patterns was used in training the network with various hidden units (Figure 4.12) The trained networks and testing result files are shown in Table 4.5. The Max learn cycles was set to 100000 and the Max error to stop was set to 0.01.

The test-pattern file, *tehydro3op.pat,* with 28 patterns was used in the testing. Both *hydro3op100.net* and *hydro3op120.net*, gave the highest accuracy of 35.71%. Whereas, the networks with 70 and 35 hidden units gave 32.14 % accuracy. The lowest accuracy prediction resulted from *hydro3op7.net*. Thus, The network with 70 hidden units was the best network for the prediction of the existence of these 3 structure using the amino acid side chain properties which classified into 8 group (Table 4.5 and Figure 4.14). The example of testing result patterns is shown in (Figure 4.15)

```
SNNS pattern definition file V1.4

No. of patterns       : 70
No. of input units    : 481
No. of output units   : 3


#input1 lalc
       0.1   0.1   0.3   0.2   0.1   0.3   0.1   0.3   0.2   0.1   0.1   0.3   0.2
       0.1   0.3   0.1   0.2   0.2   0.2   0.1   0.3   0.2   0.3   0.2   0.1   0.3
       0.3   0.3   0.2   0.3   0.3   0.2   0.2   0.2   0.2   0.2   0.1   0.2   0.1
       0.2   0.3   0.3   0.1   0.1   0.1   0.1   0.2   0.2   0.1   0.2   0.2   0.3
       0.3   0.1   0.3   0.2   0.1   0.2   0.3   0.3   0.3   0.1   0.2   0.2   0.1
       0.2   0.2   0.1   0.2   0.1   0.1   0.3   0.3   0.1   0.3   0.2   0.3   0.1
       0.1   0.3   0.3   0.1   0.1   0.1   0.3   0.2   0.1   0.1   0.3   0.3   0.3
       0.2   0.1   0.1   0.3   0.3   0.1   0.3   0.1   0.2   0.3   0.1   0.2   0.3
       0.3   0.2   0.2   0.1   0.2   0.3   0.3   0.2   0.1   0.1   0.3   0.1   0.1
       0.3   0.3   0.3   0.1   0.1   0.1   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
#output1 lalc
  1   1   1


#input2 lbp2
       0.2   0.3   0.3   0.1   0.3   0.1   0.2   0.3   0.3   0.1   0.3   0.1   0.3
       0.2   0.2   0.2   0.1   0.2   0.3   0.3   0.1   0.3   0.1   0.1   0.2   0.2
       0.3   0.2   0.3   0.2   0.3   0.2   0.2   0.2   0.2   0.2   0.2   0.3   0.1
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
#outut2 lbp2
      ...   ....  ....


#input70 8adh
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
```

Figure 4.12 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 the helix, sheet, turn structures respectively.

```
SNNS result file V1.4-3D

No. of hidden units : 70
No. of patterns     : 28
No. of input units  : 481
No. of output units : 3
startpattern        : 1
endpattern          : 28
input patterns included
teaching outputs included

#input1
       0.2   0.2   0.2   0.3   0.2   0.2   0.2   0.1   0.2   0.1   0.2   0.2   0.1
       0.2   0.3   0.2   0.1   0.1   0.3   0.2   0.3   0.2   0.1   0.2   0.1   0.1
       0.1   0.2   0.2   0.3   0.1   0.3   0.1   0.2   0.2   0.2   0.2   0.1   0.2
       0.1   0.2   0.2   0.3   0.1   0.2   0.3   0.2   0.3   0.2   0.2   0.1   0.1
       0.2   0.2   0.2   0.3   0.1   0.3   0.3   0.1   0.1   0.2   0.2   0.2   0.3
       0.1   0.3   0.1   0.3   0.1   0.1   0.2   0.1   0.1   0.1   0.2   0.3   0.3
       0.1   0.2   0.2   0.2   0.3   0.1   0.2   0.2   0.2   0.1   0.3   0.3   0.1
       0.3   0.2   0.3   0.2   0.3   0.2   0.2   0.3   0.1   0.2   0.1   0.2   0.2
       0.1   0.2   0.2   0.3   0.1   0.1   0.1   0.1   0.2   0.2   0.2   0.1   0.3
       0.2   0.3   0.3   0.2   0.3   0.1   0.2   0.1   0.2   0.2   0.1   0.3   0.2
       0.1   0.2   0.3   0.1   0.1   0.2   0.1   0.2   0.3   0.2   0.3   0.3   0.2
       0.3   0.3   0.3   0.1   0.3   0.2   0.2   0.2   0.1   0.2   0.2   0.3   0.1
       0.1   0.3   0.3   0.1   0.3   0.3   0.1   0.2   0.3   0.1   0.2   0.1   0.2
       0.1   0.2   0.2   0.1   0.3   0.2   0.1   0.3   0.1   0.2   0.1   0.2   0.3
       0.3   0.2   0.1   0.2   0.3   0.1   0.2   0.3   0.2   0.2   0.2   0.2   0.2
       0.3   0.2   0.2   0.2   0.2   0.3   0.3   0.1   0.3   0.3   0.2   0.3   0.3
       0.3   0.3   0.1   0.1   0.2   0.3   0.2   0.3   0.2   0.2   0.1   0.1   0.3
       0.3   0.1   0.3   0.1   0.1   0.3   0.1   0.3   0.1   0.2   0.1   0.2   0.1
       0.2   0.1   0.1   0.3   0.3   0.3   0.1   0.1   0.3   0.1   0.2   0.2   0.1
       0.2   0.3   0.1   0.1   0.1   0.1   0.3   0.1   0.2   0.2   0.3   0.1   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
#output1
      1       1        1
  0.8543  0.7500   0.9834

#input2
       0.1   0.1   0.3   0.3   0.3   0.2   0.1   0.1   0.2   0.1   0.2   0.3   0.1
       0.3   0.2   0.3   0.2   0.2   0.2   0.1   0.1   0.1   0.2   0.3   0.1   0.3
       0.2   0.3   0.1   0.1   0.2   0.1   0.1   0.3   0.1   0.3   0.3   0.2   0.1
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
#output2
   ....   ....   ....
   ....   ....   ....

#input28
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
```

Figure 4.13 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 1 1 1 representing helix, sheet, turn structures respectively. The predicted outputs were the three floating numbers, 0.8543, 0.75 and 0.9834.

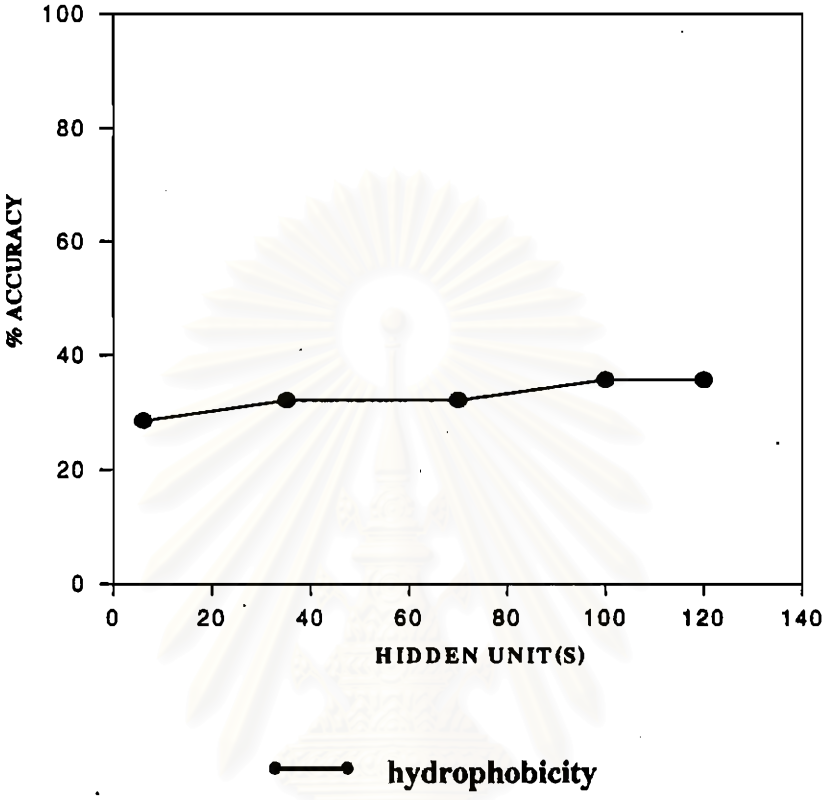

Figure 4.14 Percent accuracy prediction of helix, sheet, and turn in the same network using 3 groups of hydrophobicity as input vector with various numbers of hidden units (7, 35, 70, 100, 120).

Table 4.5 The trained net work file, and testing result file and percent accuracy of helix, sheet and turn prediction in the same network using hydrophobicity.

Hidden units	No. of units	No. of links	Training network file	Testing result file	% accuracy of prediction
7	491	3388	hydro3op7.net	tehydro3op7.res	28.57
35	519	16940	hydro3op35.net	tehydro3op35.res	32.14
70	554	33880	hydro3op70.net	tehydro3op70.res	32.14
100	584	48400	hydro3op100.net	tehydro3op100.res	35.71
120	604	58080	hydro3op120.net	tehydro3op120.res	35.71

4.1.6 Comparison of the existence of helix, sheet and turn in amino acid sequences prediction when using 2 groups of hydropathy, 7 groups of hydropathy, 8 groups of amino acid side chain properties and 3 groups of hydrophobicity

Figure 4.15 shows that each properties vector gave a different percent accuracy prediction. The amino acid side chain properties (8 groups) were the best properties which gave the highest prediction accuracy whereas hydrophobicity gave the lowest accuracy prediction. The overall results of the prediction were not significantly different in all properties except hydrophobicity, which gave the very low percent accuracy. However, the overall predictions accuracy were not quite good. Most of these networks still gave prediction accuracy lower than 50 %. This problem might be the network were not specific for all three patterns (helix, sheet and turn). Moreover, the number of patterns for each possible output may not be enough to give a good training network for a good prediction.

Thus, separate networks for prediction of helix, sheet and turn were constructed, trained and tested as reported in the next sections (section 4.2, 4.3 and 4.4).

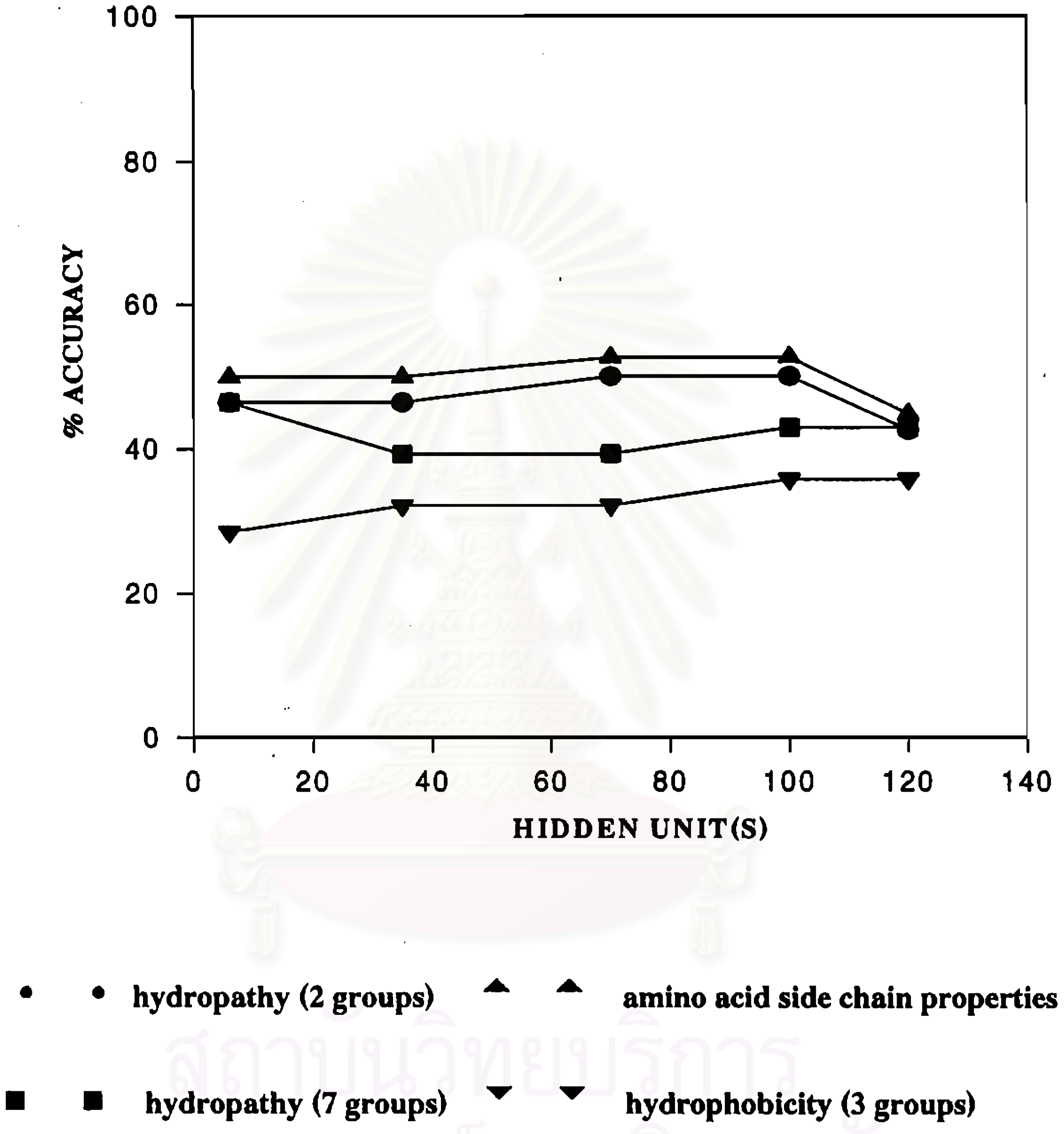

Figure 4.15 The comparison of percent accuracy prediction of helix, sheet, and turn in the same network using different input properties.

4.2 Helix structure prediction in amino acid sequence of proteins

Training set.

Seventy patterns of amino acid sequences (Table 4.6) were saved as a learn pattern file for use in training. Using this pattern file, the networks were trained for the prediction of only helix structure using amino acid properties in the section 4.2.1, 4.2.2, 4.2.3 and 4.2.4.

Testing set.

Twenty eight patterns of amino acid sequences (Table 4.6) were saved as a test pattern file. Using this test pattern file, the networks were tested for prediction of only helix structure using amino acid properties in the section 4.2.1, 4.2.2, 4.2.3 and 4.2.4.

Input and output patterns for training and testing.

An input pattern was composed of 481 input units of amino acid which were coded with properties of amino acid as previously described. An output pattern had one unit of helix structure. Symbol "1" and "0" represented an existence and non-existence of helix structure.

A learn-pattern file with 70 patterns was used in training the networks with various hidden units, 7, 35, 70, 100 and 120. The maximum value for learn cycles was set to 100000 and the maximum error to stop was set to 0.01. A test-pattern file with 28 patterns was used for testing the trained networks.

Table 4.6 The training set, testing set and outputs of helix structure prediction in separate networks.

Training set (Input No.)*	Testing set (Input No.)*	Output patterns	Structure type existing in amino acid sequence
1, 2, 3, 5, 8, 9 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 65, 66, 67, 69, 70, 71	72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98	1	helix
4, 6, 10, 11, 43, 64	68	0	no helix

* Names of input no. were listed in Table 3.1

4.2.1 Prediction of helix structure using hydropathy (2 groups)

The learn-pattern file was *thy2gr1h.pat* and the testing-pattern file was *tehy2gr1h.pat* (Figure 4.16). The trained network and testing result files were shown in Table 4.7 and the example of result patterns was shown in Figure 4.17. The prediction with highest accuracy (100%) was the result from *hy2gr1h100.net* which had 100 hidden units. Whereas the prediction accuracy from others were lower.. The networks with 7, 35, 70 and 120 hidden units gave 82.14 %, 78.57 %, 89.29 % and 92.90% accuracy respectively (Table 4.7 and Figure 4.18).

```
SNNS pattern definition file V1.4
No. of patterns       : 70
No. of input units    : 481
No. of output units   : 1

#input1 1alc
       0.2    0.2    0.1    0.2    0.2    0.1    0.2    0.1    0.2    0.2    0.2    0.1    0.2
       0.2    0.1    0.2    0.2    0.2    0.2    0.2    0.1    0.1    0.1    0.2    0.2    0.1
       0.1    0.1    0.2    0.1    0.1    0.2    0.2    0.2    0.2    0.2    0.2    0.2    0.2
       0.1    0.1    0.1    0.2    0.2    0.2    0.2    0.2    0.2    0.2    0.2    0.2    0.1
       0.1    0.2    0.1    0.2    0.2    0.1    0.1    0.2    0.1    0.2    0.2    0.2    0.2
       0.2    0.2    0.2    0.2    0.2    0.2    0.1    0.1    0.2    0.1    0.2    0.1    0.2
       0.2    0.1    0.1    0.2    0.2    0.2    0.1    0.2    0.2    0.2    0.1    0.1    0.1
       0.1    0.2    0.2    0.1    0.1    0.2    0.1    0.2    0.2    0.1    0.2    0.2    0.2
       0.1    0.1    0.2    0.2    0.1    0.1    0.1    0.2    0.2    0.2    0.1    0.2    0.2
       0.2    0.1    0.1    0.2    0.2    0.2    0.0    0.0    0.0    0.0    0.0    0.0    0.0
       0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
       0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
       0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
       0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
       0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
       0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
       0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
       0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
       0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
       0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
       0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
       0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
       0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
       0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
       0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
       0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
       0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
       0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
       0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
       0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
       0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
       0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
       0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
       0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
       0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
       0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
       0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
#output1 1alc
      1

#input2 1bp2
       0.1    0.1    0.2    0.2    0.1    0.2    0.2    0.1    0.1    0.2    0.1    0.2    0.1
       0.2    0.2    0.2    0.2    0.2    0.1    0.1    0.2    0.1    0.2    0.2    0.2    0.2
       ...    ...    ...    ...    ...    ...    ...    ...    ...    ...    ...    ...    ...
       ...    ...    ...    ...    ...    ...    ...    ...    ...    ...    ...    ...    ...
#outut2 1bp2
      ...

#input70 8adh

       ...    ...    ...    ...    ...    ...    ...    ...    ...    ...    ...    ...    ...
       ...    ...    ...    ...    ...    ...    ...    ...    ...    ...    ...    ...    ...
       ...    ...    ...    ...    ...    ...    ...    ...    ...    ...    ...    ...    ...
```

Figure 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.

```
SNNS result file V1.4-3D
No. of hidden units : 70
No. of patterns     : 28
No. of input units  : 481
No. of output units : 1
startpattern        : 1
endpattern          : 28
input patterns included
teaching outputs included

#input1
       0.2   0.2   0.1   0.2   0.2   0.1   0.2   0.1   0.2   0.2   0.2   0.1   0.2
       0.2   0.1   0.2   0.2   0.2   0.2   0.2   0.1   0.1   0.1   0.2   0.2   0.1
       0.1   0.1   0.2   0.1   0.1   0.2   0.2   0.2   0.2   0.2   0.2   0.2   0.2
       0.1   0.1   0.1   0.2   0.2   0.2   0.2   0.2   0.2   0.2   0.2   0.2   0.1
       0.1   0.2   0.1   0.2   0.2   0.1   0.1   0.2   0.1   0.2   0.2   0.2   0.2
       0.2   0.2   0.2   0.2   0.2   0.2   0.1   0.1   0.2   0.1   0.2   0.1   0.2
       0.2   0.1   0.1   0.2   0.2   0.2   0.1   0.2   0.2   0.2   0.1   0.1   0.1
       0.1   0.2   0.2   0.1   0.1   0.2   0.1   0.2   0.2   0.1   0.2   0.2   0.2
       0.1   0.1   0.2   0.2   0.1   0.1   0.1   0.2   0.2   0.2   0.1   0.2   0.2
       0.2   0.1   0.1   0.2   0.2   0.2   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
#output1
    1
  0.9564

#input2
       0.1   0.1   0.2   0.2   0.1   0.2   0.2   0.1   0.1   0.2   0.1   0.2   0.1
       0.2   0.2   0.2   0.2   0.2   0.1   0.1   0.2   0.1   0.2   0.2   0.2   0.2
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
#output
   ....
   ....

#input70
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
```

Figure 4.17 The example of correct 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. The predicted output was the floating number, 0.9564.

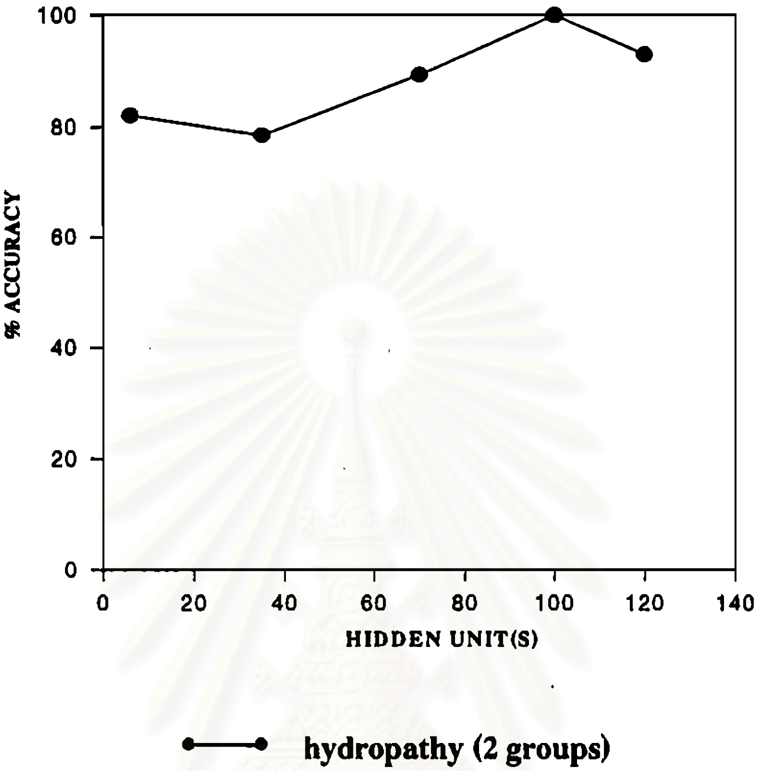

Figure 4.18 Percent accuracy prediction of helix structure using 2 groups of hydropathy as input vector with various numbers of hidden units (7, 35, 70, 100, 120).

Table 4.7 The trained network file, testing result file and percent accuracy of helix structure prediction using hydropathy (2 groups).

Hidden units	No. of units	No. of links	Training network file	Testing result file	% accuracy of prediction
7	489	3374	hy2gr1h7.net	tehy2gr1h7.res	82.14
35	517	16870	hy2gr1h35.net	tehy2gr1h35.res	78.57
70	552	33730	hy2gr1h70.net	tehy2gr1h70.res	89.29
100	582	48200	hy2gr1h100.net	tehy2gr1h100.res	100.00
120	602	57840	hy2gr1h120.net	tehy2gr1h120.res	92.90

4.2.2 Prediction of helical structure using hydropathy (7 groups)

The learn-pattern file, *thy7gr1h.pat*, was used in training the networks with various hidden units. Figure 4.19 shows the example of learn-patterns. Test-pattern file, *tehy7gr1h.pat*, was used in testing the trained network. The trained network and testing result file are shown in Table 4.8 and the example of result patterns is shown in Figure 4.20. Table 4.8 and Figure 4.21 show that *hy7gr1h35.net*, *hy7gr1h100.net* and *hy7gr1h120* which had 35, 100 and 120 respectively were the best networks for helix structure prediction. These network gave 92.86% accuracy. Whereas, the networks with 7 and 70 hidden units gave 89.29 % accuracy.

Table 4.8 The trained network file, testing result files and percent accuracy of helix structure prediction using hydropathy (7 groups).

Hidden units	No. of units	No. of links	Training network file	Testing result file	% accuracy of prediction
7	489	3374	hy7gr1h7.net	tehy7gr1h7.res	89.29
35	517	16870	hy7gr1h35.net	tehy7gr1h35.res	92.86
70	552	33730	hy7gr1h70.net	tehy7gr1h70.res	89.29
100	582	48200	hy7gr1h100.net	tehy7gr1h100.res	92.86
120	602	57840	hy7gr1h120.net	tehy7gr1h120.res	92.86

```
SNNS pattern definition file V1.4

No. of patterns       : 70
No. of input units    : 481
No. of output units   : 1

#input1 lalc
       0.7   0.6   0.2   0.4   0.7   0.2   0.6   0.1   0.4   0.6   0.6   0.1   0.5
       0.6   0.1   0.6   0.4   0.5   0.4   0.7   0.1   0.3   0.1   0.5   0.6   0.1
       0.1   0.2   0.4   0.3   0.2   0.6   0.4   0.4   0.4   0.5   0.6   0.4   0.6
       0.3   0.1   0.1   0.6   0.6   0.6   0.6   0.4   0.4   0.6   0.5   0.4   0.1
       0.2   0.6   0.1   0.4   0.6   0.3   0.1   0.4   0.2   0.7   0.4   0.4   0.6
       0.4   0.5   0.6   0.4   0.7   0.6   0.1   0.2   0.6   0.1   0.4   0.2   0.6
       0.7   0.2   0.1   0.6   0.6   0.6   0.1   0.4   0.6   0.6   0.1   0.3   0.2
       0.3   0.7   0.7   0.1   0.1   0.6   0.1   0.7   0.4   0.1   0.6   0.5   0.4
       0.1   0.3   0.6   0.7   0.3   0.1   0.2   0.4   0.6   0.7   0.1   0.6   0.6
       0.4   0.1   0.2   0.6   0.7   0.6   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
#output1 lalc
      1

#input2
       0.3   0.1   0.4   0.6   0.2   0.6   0.4   0.3   0.1   0.7   0.2   0.7   0.1
       0.5   0.4   0.4   0.6   0.5   0.1   0.1   0.6   0.2   0.6   0.6   0.5   0.4
       0.2   0.5   0.2   0.4   0.1   0.4   0.4   0.4   0.4   0.4   0.5   0.1   0.6
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
#outut2 lbp2
      ...

#input70 8adh

       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
```

Figure 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.

```
SNNS result file V1.4-3D

No. of hidden units : 70
No. of patterns     : 28
No. of input units  : 481
No. of output units : 1
startpattern        : 1
endpattern          : 28
input patterns included
teaching outputs included

#input1
       0.4   0.6   0.6   0.4   0.4   0.5   0.4   0.7   0.6   0.6   0.4   0.5   0.6
       0.6   0.4   0.6   0.7   0.6   0.2   0.5   0.1   0.3   0.7   0.4   0.6   0.7
       0.6   0.4   0.5   0.1   0.6   0.1   0.6   0.4   0.6   0.4   0.3   0.7   0.5
       0.6   0.5   0.4   0.1   0.7   0.5   0.1   0.4   0.1   0.4   0.5   0.6   0.6
       0.3   0.4   0.4   0.1   0.7   0.1   0.1   0.6   0.6   0.4   0.6   0.3   0.2
       0.6   0.1   0.6   0.2   0.6   0.6   0.4   0.6   0.6   0.7   0.3   0.1   0.1
       0.7   0.4   0.4   0.5   0.1   0.6   0.4   0.4   0.5   0.7   0.1   0.1   0.6
       0.2   0.6   0.2   0.6   0.4   0.4   0.4   0.1   0.6   0.4   0.6   0.4   0.4
       0.6   0.6   0.4   0.1   0.6   0.7   0.7   0.7   0.5   0.3   0.3   0.6   0.1
       0.6   0.1   0.1   0.6   0.4   0.6   0.4   0.7   0.5   0.4   0.6   0.2   0.4
       0.7   0.3   0.1   0.6   0.6   0.5   0.6   0.4   0.1   0.3   0.1   0.1   0.4
       0.1   0.2   0.1   0.7   0.1   0.4   0.4   0.3   0.7   0.5   0.4   0.1   0.6
       0.7   0.1   0.1   0.6   0.1   0.1   0.6   0.4   0.1   0.7   0.4   0.7   0.4
       0.7   0.4   0.3   0.6   0.2   0.4   0.6   0.2   0.6   0.5   0.7   0.4   0.1
       0.1   0.5   0.6   0.4   0.1   0.6   0.5   0.4   0.4   0.5   0.5   0.4   0.4
       0.1   0.4   0.4   0.5   0.5   0.1   0.1   0.6   0.2   0.1   0.4   0.4   0.1
       0.1   0.1   0.7   0.6   0.5   0.1   0.4   0.1   0.4   0.4   0.6   0.6   0.1
       0.1   0.7   0.2   0.7   0.7   0.1   0.6   0.2   0.6   0.4   0.6   0.4   0.6
       0.5   0.6   0.6   0.1   0.3   0.1   0.6   0.6   0.4   0.7   0.5   0.3   0.6
       0.5   0.1   0.7   0.6   0.7   0.6   0.1   0.7   0.3   0.4   0.2   0.7   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
#output1
     1
  0.9954

#input2
       0.7   0.7   0.1   0.1   0.1   0.4   0.7   0.7   0.4   0.6   0.4   0.1   0.6
       0.1   0.4   0.2   0.4   0.3   0.4   0.6   0.7   0.7   0.4   0.1   0.6   0.2
       0.6   0.4   0.7   0.6   0.4   0.6   0.6   0.1   0.7   0.1   0.1   0.4   0.6
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
#output2
   ....
   ....
#input28
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
```

Figure 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. The predicted output was the floating number, 0.9954.

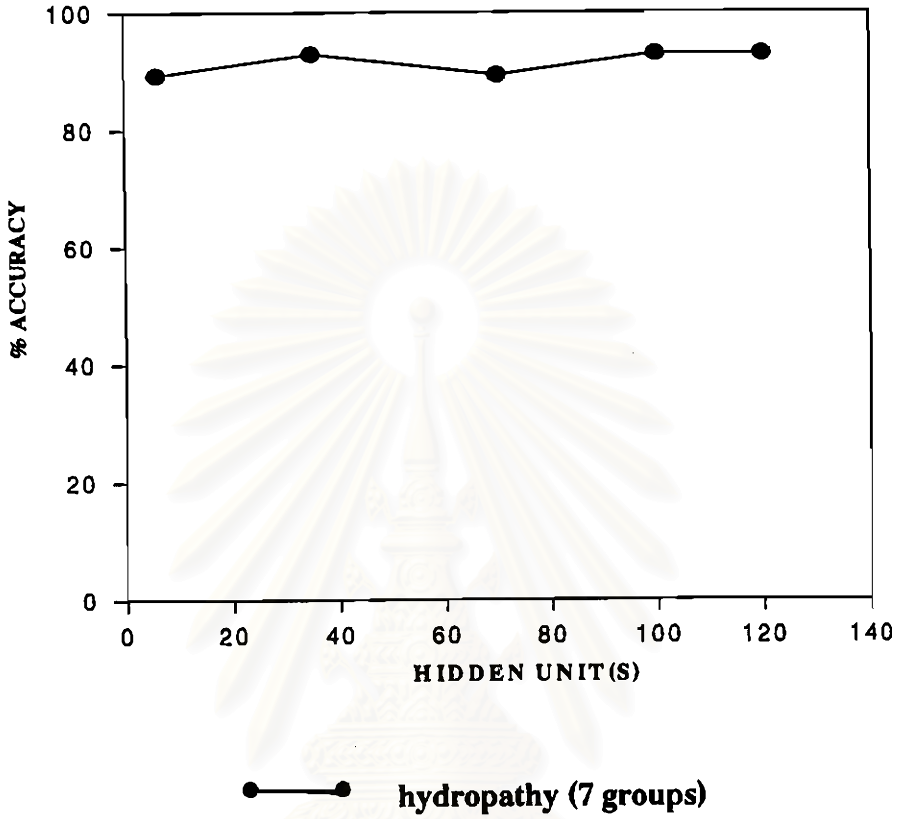

Figure 4.21 Percent accuracy prediction of helix structure using 7 groups of hydropathy as input vector with various numbers of hidden units(7, 35, 70,100, 120).

4.2.3 Prediction of helix structure using amino acid side chain properties (8 groups)

The learn-pattern file, *taa8gr1h.pat*, and the test-pattern file, *teaa8gr1h.pat*, were used in training network and the testing respectively. The example of training or testing patterns is shown in Figure 4.22. The trained networks and testing result files are shown in Table 4.9 and the example of the testing result pattern was shown in Figure 4.23. Figure 4.24 and Table 4.9 show that *aa8gr1h7.net* gave the highest percent accuracy prediction, 100%. The trained network, *aa8gr1h35.net*, *aa8gr1h70.net* and *aa8gr1h100.net* gave the same 96.42% accuracy prediction. Whereas *aa8gr1h120.net gave* the lowest accuracy prediction, 92.85%.

```
SNNS pattern definition file V1.4

No. of patterns        : 70
No. of input units     : 481
No. of output units    : 1

#input1 1alc
       0.1    0.5    0.3    0.7    0.2    0.1    0.2    0.7    0.7    0.6    0.8    0.1    0.0
       0.3    0.7    0.8    0.2    0.5    0.6    0.1    0.2    0.3    0.1    0.1    0.8    0.1
       0.8    0.8    0.8    0.5    0.3    0.1    0.7    0.1    0.6    0.5    0.5    0.7    0.5
       0.6    0.6    0.7    0.5    0.5    0.1    0.6    0.3    0.1    0.5    0.1    0.5    0.2
       0.8    0.3    0.1    0.5    0.1    0.6    0.7    0.1    0.1    0.8    0.1    0.1    0.6
       0.5    0.2    0.6    0.1    0.8    0.2    0.7    0.7    0.8    0.8    0.7    0.6    0.5
       0.1    0.1    0.6    0.1    0.1    0.3    0.2    0.5    0.7    0.5    0.2    0.6    0.1
       0.2    0.8    0.2    0.6    0.2    0.6    0.2    0.1    0.5    0.5    0.8    0.7    0.6
       0.1    0.5    0.7    0.6    0.5    0.1    0.7    0.1    0.1    0.6    0.2    0.5    0.1
       0.7    0.1    0.6    0.1    0.1    0.6    0.2    0.8    0.5    0.1    0.6    0.2    0.5
       0.5    0.1    0.1    0.7    0.1    0.1    0.5    0.6    0.1    0.7    0.1    0.1    0.1
       0.1    0.1    0.1    0.1    0.1    0.4    0.6    0.1    0.1    0.7    0.1    0.8    0.3
       0.6    0.1    0.8    0.6    0.1    0.1    0.7    0.1    0.1    0.8    0.1    0.1    0.6
       0.5    0.6    0.1    0.6    0.6    0.1    0.3    0.2    0.5    0.8    0.2    0.7    0.3
       0.5    0.5    0.1    0.1    0.3    0.5    0.5    0.1    0.7    0.2    0.2    0.5    0.2
       0.3    0.1    0.5    0.1    0.5    0.6    0.3    0.3    0.1    0.2    0.7    0.5    0.1
       0.5    0.2    0.1    0.1    0.4    0.6    0.7    0.5    0.1    0.5    0.1    0.5    0.5
       0.7    0.8    0.1    0.1    0.8    0.2    0.6    0.5    0.1    0.1    0.5    0.8    0.1
       0.7    0.1    0.7    0.8    0.1    0.1    0.3    0.4    0.8    0.6    0.8    0.8    0.6
       0.3    0.5    0.8    0.3    0.1    0.6    0.1    0.6    0.5    0.1    0.6    0.1    0.5
       0.2    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
       0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
       0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
       0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
       0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
       0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
       0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
       0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
       0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
       0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
       0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
       0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
       0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
       0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
       0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
       0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
       0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
#output1 1alc
    1


#input2 1bp2
       0.1    0.3    0.6    0.1    0.3    0.1    0.7    0.1    0.1    0.6    0.4    0.7    0.6
       0.5    0.6    0.8    0.3    0.1    0.1    0.2    0.8    0.6    0.5    0.5    0.6    0.6
       0.1    0.1    0.6    0.4    0.1    0.7    0.4    0.6    0.6    0.3    0.1    0.8    0.1
       ...    ...    ...    ...    ...    ...    ...    ...    ...    ...    ...    ...    ...
       ...    ...    ...    ...    ...    ...    ...    ...    ...    ...    ...    ...    ...
#outut2 1bp2
      ...   ....   ....


#input70 8adh
       ...    ...    ...    ...    ...    ...    ...    ...    ...    ...    ...    ...    ...
       ...    ...    ...    ...    ...    ...    ...    ...    ...    ...    ...    ...    ...
```

Figure 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.

```
SNNS result file V1.4-3D

No. of hidden units : 70
No. of patterns     : 28
No. of input units  : 481
No. of output units : 1
startpattern        : 1
endpattern          : 28
input patterns included
teaching outputs included

#input1
       0.4   0.8   0.5   0.1   0.5   0.7   0.6   0.1   0.6   0.6   0.6   0.6   0.1
       0.1   0.1   0.6   0.4   0.5   0.8   0.5   0.7   0.1   0.1   0.5   0.6   0.1
       0.1   0.1   0.6   0.8   0.8   0.5   0.1   0.8   0.1   0.1   0.7   0.1   0.1
       0.1   0.5   0.6   0.6   0.3   0.6   0.2   0.1   0.2   0.7   0.1   0.1   0.4
       0.1   0.1   0.8   0.4   0.7   0.3   0.1   0.2   0.1   0.8   0.2   0.1   0.5
       0.6   0.6   0.1   0.6   0.1   0.1   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0

#output1
   1
0.7562

#input2
       0.3   0.6   0.2   0.5   0.1   0.1   0.7   0.6   0.7   0.5   0.4   0.1   0.1
       0.4   0.1   0.1   0.4   0.1   0.1   0.1   0.1   0.3   0.7   0.1   0.2   0.7
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
#output2
   ....
   ....

#input28
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
```

Figure 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 structure. The predicted output was the floating number, 0.7562.

Table 4.9 The trained network file, testing result file and percent accuracy of helix structure prediction using 8 groups of amino acid side chain properties.

Hidden units	No. of units	No. of links	Training network file	Testing result file	% accuracy of prediction
7	489	3374	aa8gr1h7.net	teaa8gr1h7.res	100.00
35	517	16870	aa8gr1h35.net	teaa8gr1h35.res	96.42
70	552	33730	aa8gr1h70.net	teaa8gr1h70.res	96.42
100	582	48200	aa8gr1h100.net	teaa8gr1h100.res	96.42
120	602	57840	aa8gr1h120.net	teaa8gr1h120.res	92.85

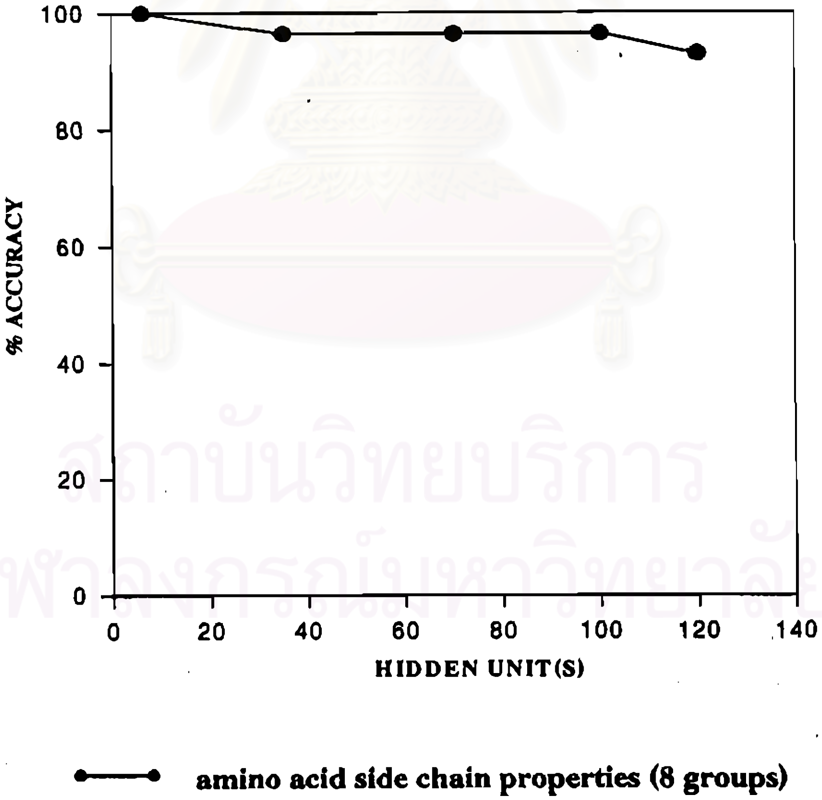

Figure 4.24 Percent accuracy prediction of helix structure using 8 groups of amino acid side chain properties as input vector with various numbers of hidden units (7, 35, 70, 100, 120).

4.2.4 Prediction of helix structure using hydrophobicity (3 groups)

The learn-pattern file, *thydro1h.pat*, and the test-pattern file, *tehydro1h.pat*, were used in training network and the testing respectively. An example of what the training or testing pattern may look like is shown in Figure 4.25. The trained networks and testing result files are shown in Table 4.10 and the example of the testing result patterns is shown in Figure 4.26. From Table 4.10 and Figure 4.27, the prediction from *hydro1h7.net*, *hydro1h35.net* and *hydro1h70.net* are all 96.42% accurate, whereas both *hydro1h100.net* and *hydro1h120.net* gave 92.85% accuracy.

Table 4.10 The trained network file, testing result file and percent accuracy of helix structure prediction using hydrophobicity.

Hidden units	No. of units	No. of links	Training network file	Testing result file	% accuracy of prediction
7	489	3374	hydro1h7.net	tehydro1h7.res	96.42
35	517	16870	hydro1h35.net	tehydro1h35.res	96.42
70	552	33730	hydro1h70.net	tehydro1h70.res	96.42
100	582	48200	hydro1h100.net	tehydro1h100.res	92.86
120	602	57840	hydro1h120.net	tehydro1h120.res	92.86

```
SNNS pattern definition file V1.4

No. of patterns       : 70
No. of input units    : 481
No. of output units   : 1


#input1 1alc
       0.1   0.1   0.3   0.2   0.1   0.3   0.1   0.3   0.2   0.1   0.1   0.3   0.2
       0.1   0.3   0.1   0.2   0.2   0.2   0.1   0.3   0.2   0.3   0.2   0.1   0.3
       0.3   0.3   0.2   0.3   0.3   0.2   0.2   0.2   0.2   0.2   0.1   0.2   0.1
       0.2   0.3   0.3   0.1   0.1   0.1   0.1   0.2   0.2   0.1   0.2   0.2   0.3
       0.3   0.1   0.3   0.2   0.1   0.2   0.3   0.3   0.3   0.1   0.2   0.2   0.1
       0.2   0.2   0.1   0.2   0.1   0.1   0.3   0.3   0.1   0.3   0.2   0.3   0.1
       0.1   0.3   0.3   0.1   0.1   0.1   0.3   0.2   0.1   0.1   0.3   0.3   0.3
       0.2   0.1   0.1   0.3   0.3   0.1   0.3   0.1   0.2   0.3   0.1   0.2   0.3
       0.3   0.2   0.2   0.1   0.2   0.3   0.3   0.2   0.1   0.1   0.3   0.1   0.1
       0.3   0.3   0.3   0.1   0.1   0.1   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
#output1 1alc
  1


#input2 1bp2
       0.2   0.3   0.3   0.1   0.3   0.1   0.2   0.3   0.3   0.1   0.3   0.1   0.3
       0.2   0.2   0.2   0.1   0.2   0.3   0.3   0.1   0.3   0.1   0.1   0.2   0.2
       0.3   0.2   0.3   0.2   0.3   0.2   0.2   0.2   0.2   0.2   0.2   0.3   0.1
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
#outut2 1bp2
      ...


#input70 8adh
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
```

Figure 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 the helix or sheet or turn structure.

```
SNNS result file V1.4-3D

No. of hidden units : 70
No. of patterns     : 28
No. of input units  : 481
No. of output units : 1
startpattern        : 1
endpattern          : 28
input patterns included
teaching outputs included

#input1
       0.2   0.2   0.2   0.3   0.2   0.2   0.2   0.1   0.2   0.1   0.2   0.2   0.1
       0.2   0.3   0.2   0.1   0.1   0.3   0.2   0.3   0.2   0.1   0.2   0.1   0.1
       0.1   0.2   0.2   0.3   0.1   0.3   0.1   0.2   0.2   0.2   0.2   0.1   0.2
       0.1   0.2   0.2   0.3   0.1   0.2   0.3   0.2   0.3   0.2   0.2   0.1   0.1
       0.2   0.2   0.2   0.3   0.1   0.3   0.3   0.1   0.1   0.2   0.2   0.2   0.3
       0.1   0.3   0.1   0.3   0.1   0.1   0.2   0.1   0.1   0.1   0.2   0.3   0.3
       0.1   0.2   0.2   0.2   0.3   0.1   0.2   0.2   0.2   0.1   0.3   0.3   0.1
       0.3   0.2   0.3   0.2   0.3   0.2   0.2   0.3   0.1   0.2   0.1   0.2   0.2
       0.1   0.2   0.2   0.3   0.1   0.1   0.1   0.1   0.2   0.2   0.2   0.1   0.3
       0.2   0.3   0.3   0.2   0.3   0.1   0.2   0.1   0.2   0.2   0.1   0.3   0.2
       0.1   0.2   0.3   0.1   0.1   0.2   0.1   0.2   0.3   0.2   0.3   0.3   0.2
       0.3   0.3   0.3   0.1   0.3   0.2   0.2   0.2   0.1   0.2   0.2   0.3   0.1
       0.1   0.3   0.3   0.1   0.3   0.3   0.1   0.2   0.3   0.1   0.2   0.1   0.2
       0.1   0.2   0.2   0.1   0.3   0.2   0.1   0.3   0.1   0.2   0.1   0.2   0.3
       0.3   0.2   0.1   0.2   0.3   0.1   0.2   0.3   0.2   0.2   0.2   0.2   0.2
       0.3   0.2   0.2   0.2   0.2   0.3   0.3   0.1   0.3   0.3   0.2   0.3   0.3
       0.3   0.3   0.1   0.1   0.2   0.3   0.2   0.3   0.2   0.2   0.1   0.1   0.3
       0.3   0.1   0.3   0.1   0.1   0.3   0.1   0.3   0.1   0.2   0.1   0.2   0.1
       0.2   0.1   0.1   0.3   0.3   0.3   0.1   0.1   0.3   0.1   0.2   0.2   0.1
       0.2   0.3   0.1   0.1   0.1   0.1   0.3   0.1   0.2   0.2   0.3   0.1   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
#output1
     1
  0.8922

#input2
       0.1   0.1   0.3   0.3   0.3   0.2   0.1   0.1   0.2   0.1   0.2   0.3   0.1
       0.3   0.2   0.3   0.2   0.2   0.2   0.1   0.1   0.1   0.2   0.3   0.1   0.3
       0.2   0.3   0.1   0.1   0.2   0.1   0.1   0.3   0.1   0.3   0.3   0.2   0.1
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
#output2
   ....
   ....

#input28
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
```

Figure 4.26 The example of correct 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. The predicted output was the floating number, 0.8922.

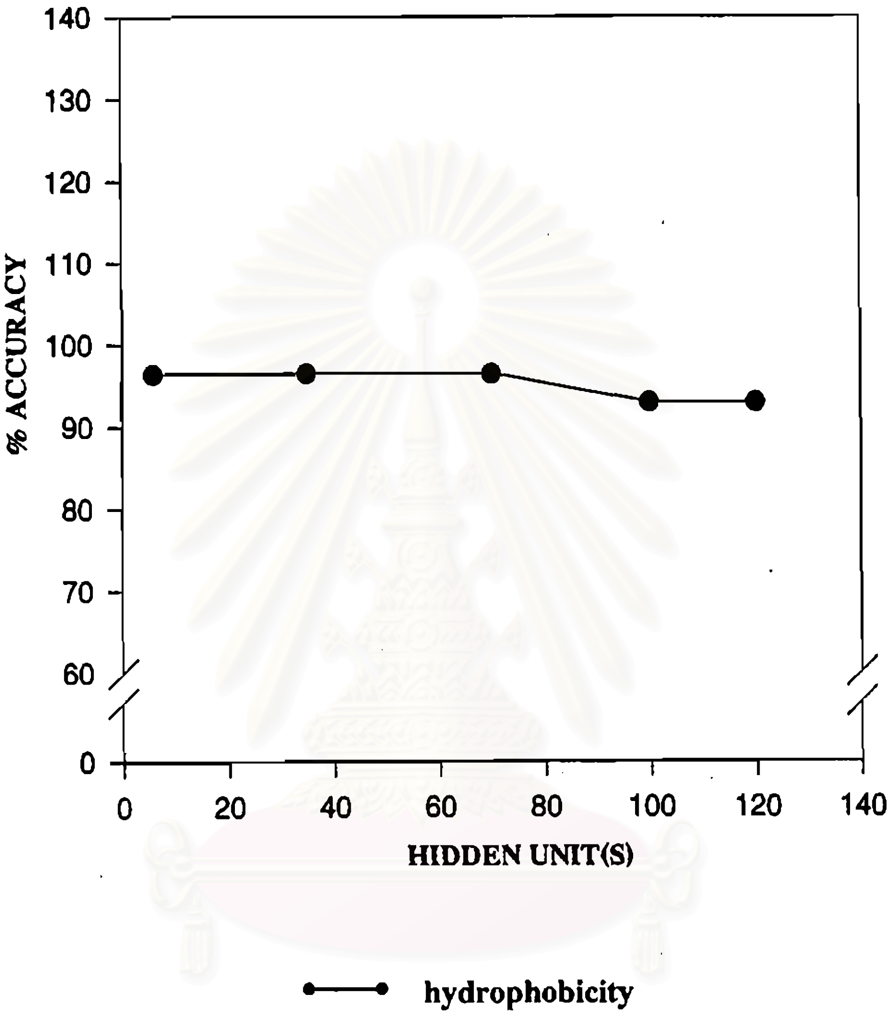

Figure 4.27 Percent accuracy prediction of helix structure using 3 groups of hydrophobicity as input vector with various numbers of hidden units (7, 35, 70, 100, 120).

4.2.5 Comparison of the existence of helical structure 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

Figure 4.28 shows that all group of property gave over 85 percent accuracy for helix structure using a separated helical network. These exact percent accuracy are, however, not significantly different. The network of 2 groups hydropathy with 100 hidden units and the network of amino acid side chain with 7 hidden units gave 100% accuracy which were the best networks for prediction.

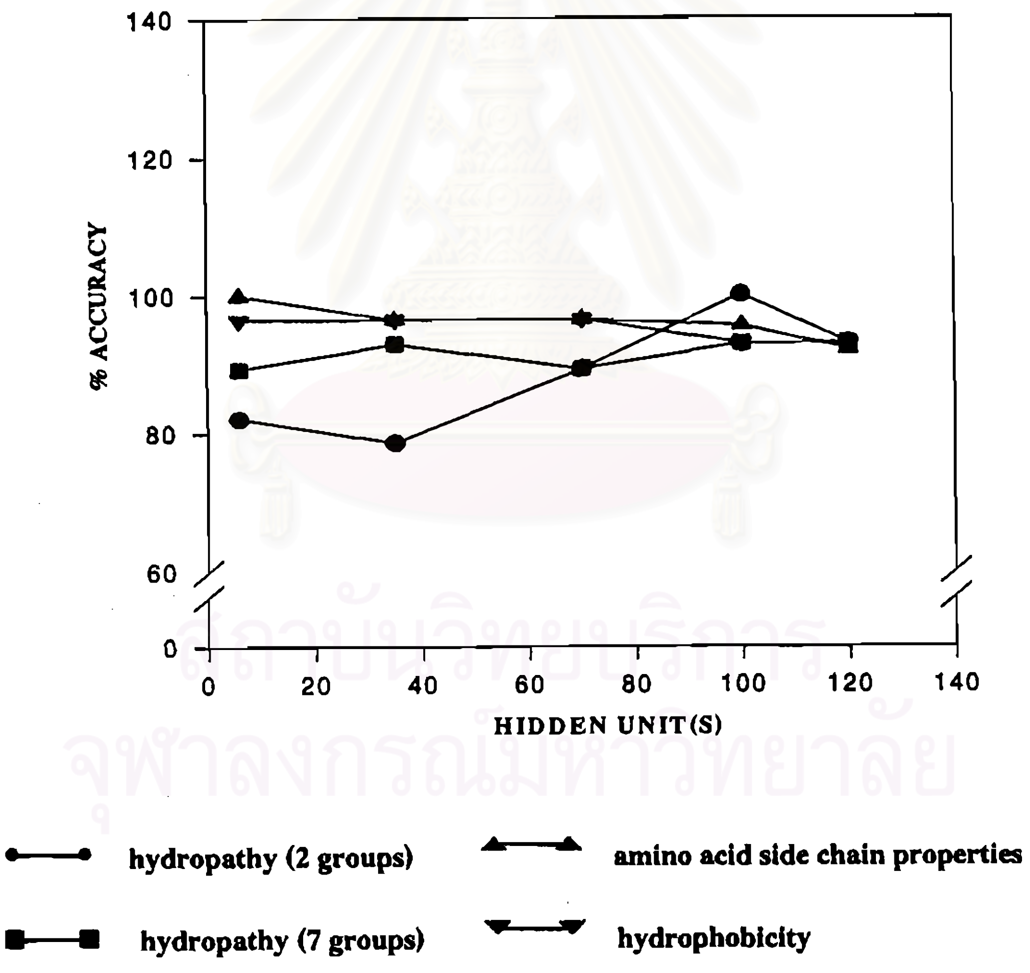

Figure 4.28 The comparison of percent accuracy prediction of helix structure using various types of input properties.

4.3 Sheet structure prediction in amino acid sequence of proteins

Training set.

Seventy patterns of amino acid sequences (Table 4.11) were saved as a train-pattern file for training.

Testing set.

Twenty eight patterns of amino acid sequences (Table 4.11) were saved as a test-pattern file for testing.

Input and output patterns for training and testing.

An input pattern was composed of 481 input units of amino acid which were coded with amino acid properties as previously described in section 3.2.2. An output pattern had one unit of sheet structure. Symbol "1" and "0" represented an existence and non-existence of sheet structure.

A learn-pattern file with 70 patterns was used in training the networks with various hidden units, 7, 35, 70, 100 and 120. The maximum value for learn cycles was set to 100000 and the maximum error to stop was set to 0.01. A test-pattern file with 28 patterns was used for testing the trained networks.

These training and testing sets (Table 4.11) with input and output patterns as describe above were used for all prediction of sheet structure in separate networks (section 4.3.1, 4.3.2, 4.3.3 and 4.3.4)

Table 4.11 The training set, testing set and outputs of sheet structure prediction in separate networks.

Training set (Input No.)*	Testing set (Input No.)*	Output patterns	Structure type existing in amino acid sequence
1, 2, 3, 4, 5, 6, 8, 9, 10, 11, 12, 14, 15, 16, 17, 19, 20, 21, 22, 23, 24, 26, 27, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 41, 42, 43, 44, 45, 46, 47, 49, 50, 51, 52, 54, 55, 56, 57, 58, 60, 62, 63, 64, 65	66, 67, 68, 69, 71, 72, 74, 75, 76, 79, 80, 81, 82, 85, 89, 90, 91, 92, 93, 95, 96, 97, 98	1	sheet
7, 13, 18, 25, 28, 40, 48, 53, 59, 61, 70, 77, 83, 86, 94	73, 78, 84, 87, 88	0	no sheet

* Names of input no. were listed in Table 3.1

4.3.1 Prediction of sheet structure using hydropathy (2 groups)

The learn-pattern file was *thy2grls.pat* and the testing-pattern file was *tehy2grls.pat* (Figure 4.16). The trained network and testing result files are shown in Table 4.12 and the example of result patterns is shown in Figure 4.17. Prediction with highest accuracy (82.14%) were the results of *hy2grls35.net, hy2grls70.net* and *hy2grls100.net* which had 35, 70, and 100 hidden units respectively. The networks with 7 and 120 hidden units gave 78.57 % and 75.00% accurate, respectively (Table 4.12 and Figure 4.29).

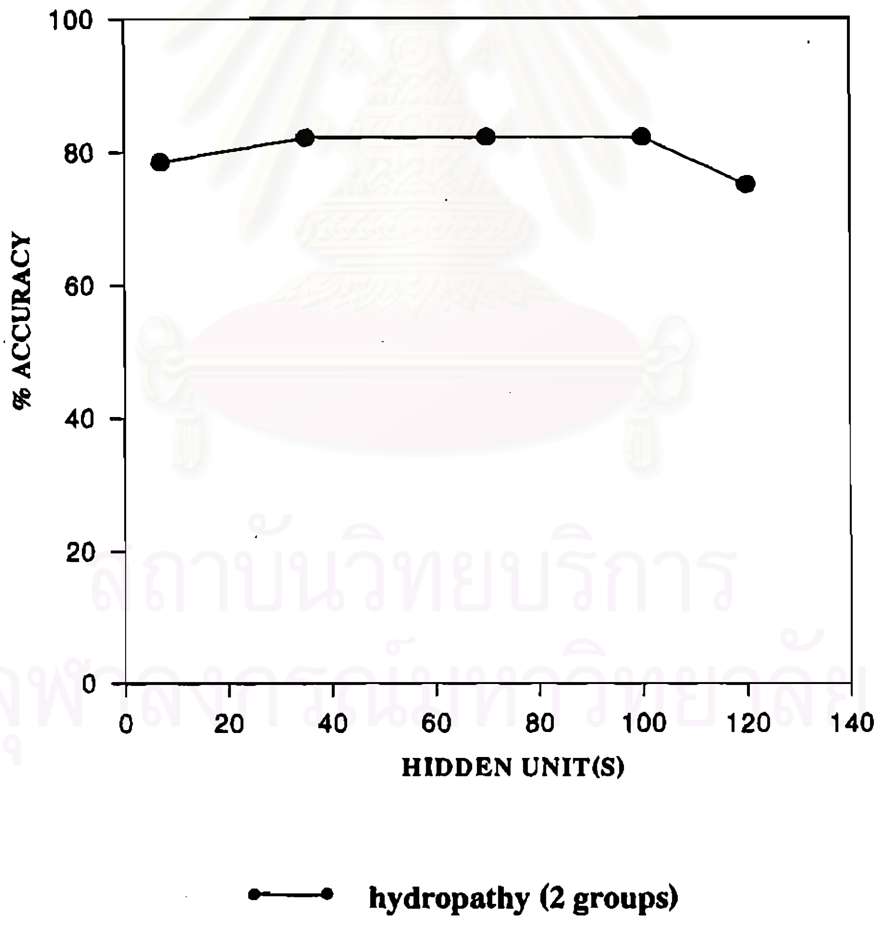

Figure 4.29 Percent accuracy prediction of sheet structure using 2 groups of hydropathy as input vector with various numbers of hidden units (7, 35, 70, 100, 120).

Table 4.12 The trained network file, testing result file and percent accuracy of sheet structure prediction using hydropathy (2 groups).

Hidden units	No. of units	No. of links	Training network file	Testing result file	% accuracy of prediction
7	489	3374	hy2gr1s7.net	tehy2gr1s7.res	78.57
35	517	16870	hy2gr1s35.net	tehy2gr1s35.res	82.14
70	552	33730	hy2gr1s70.net	tehy2gr1s70.res	82.14
100	582	48200	hy2gr1s100.net	tehy2gr1s100.res	82.14
120	602	57840	hy2gr1s120.net	tehy2gr1s120.res	75.00

4.3.2 Prediction of sheet structure using hydropathy (7 groups)

The Learn-pattern file was *thy7gr1s.pat* and the testing-pattern file was *tehy7gr1s.pat* (Figure 4.19). The trained network and testing result files are shown in Table 4.13 and the example of result patterns is shown in Figure 4.20. The predictions with highest accuracy (78.57%) were the results from *hy7gr1s35.net* and *hy7gr1s70.net* which had 35 and 70hidden units respectively. While, others networks with 7 and 100 hidden units gave 75.00% and the network with 120 hidden units gave 71.43% (Table 4.13 and Figure 4.30).

Table 4.13 The trained network file, testing result file and percent accuracy of sheet structure prediction using hydropathy (7 groups).

Hidden units	No. of units	No. of links	Training network file	Testing result file	% accuracy of prediction
7	489	3374	hy7gr1s7.net	tehy7gr1s7.res	75.00
35	517	16870	hy7gr1s35.net	tehy7gr1s35.res	78.57
70	552	33730	hy7gr1s70.net	tehy7gr1s70.res	78.57
100	582	48200	hy7gr1s100.net	tehy7gr1s100.res	75.00
120	602	57840	hy7gr1s120.net	tehy7gr1s120.res	71.43

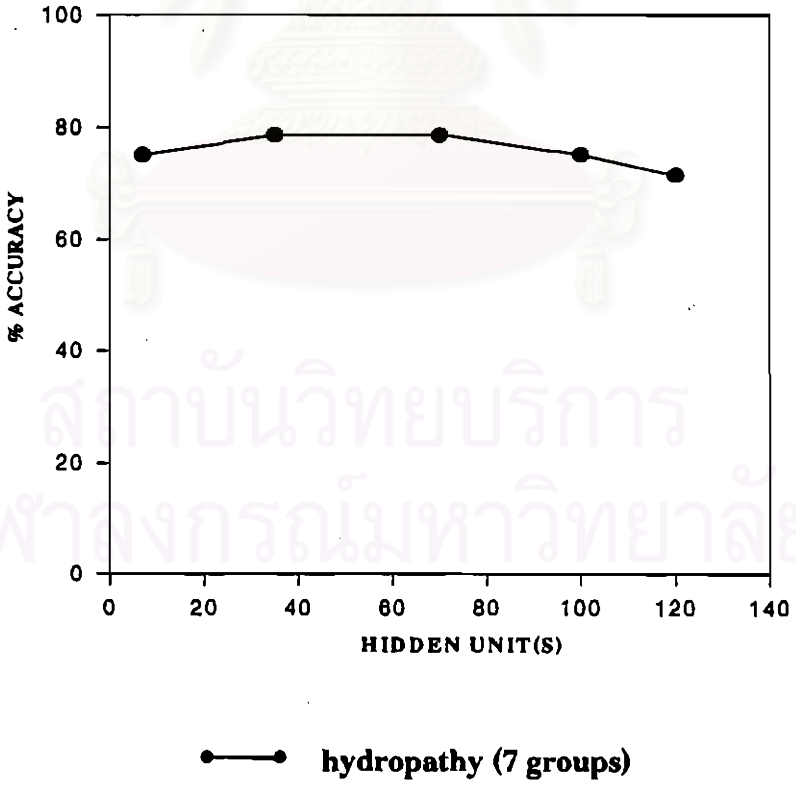

Figure 4.30 Percent accuracy prediction of sheet structure using 7 groups of hydropathy as input vector with various numbers of hidden units (7, 35, 70,100, 120).

4.3.3 Prediction of sheet structure using amino acid side chain properties (8 groups)

The learn-pattern file, *taa8gr1s.pat*, and the testing-pattern file, *aa8gr1s.pat*, were used for training and testing the network respectively. The example of the input and output patterns was shown in Figure 4.22. The trained network and testing result files are shown in Table 4.14 and the example of result patterns is shown in Figure 4.23. From figure 4.31 and Table 4.14, the network with highest prediction accuracy (85.71%) was the result from *aa8gr1s100.net* which had 100 hidden units. Others networks gave lower prediction accuracy of 82.14%.

Table 4.14 The trained network file, testing result file and percent accuracy of sheet structure prediction using amino acid side chain properties (8 groups).

Hidden units	No. of units	No. of links	Training network file	Testing result file	% accuracy of prediction
7	489	3374	aa8gr1s7.net	teaa8gr1s7.res	82.14
35	517	16870	aa8gr1s35.net	teaa8gr1s35.res	82.14
70	552	33730	aa8gr1s70.net	teaa8gr1s70.res	82.14
100	582	48200	aa8gr1s100.net	teaa8gr1s100.res	85.71
120	602	57840	aa8gr1s120.net	teaa8gr1s120.res	82.14

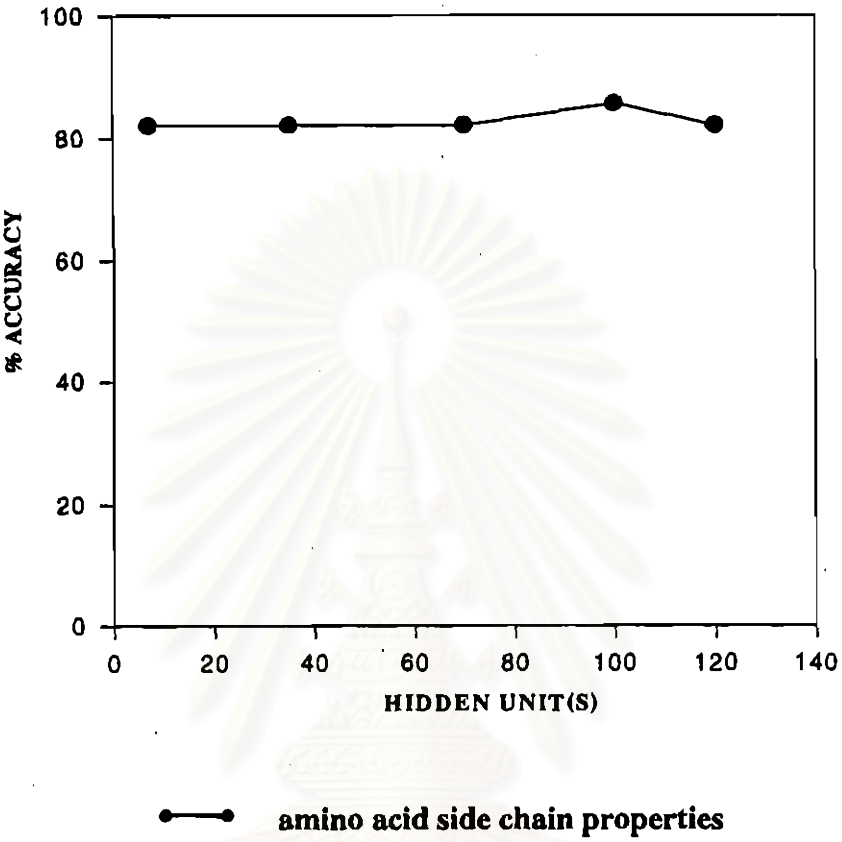

Figure 4.31 Percent accuracy prediction of sheet structure using 8 groups of amino acid side chain properties as input vector with various numbers of hidden units (7, 35, 70, 100, 120).

4.3.4 Prediction of sheet structure using hydrophobicity (3 groups)

The Learn-pattern file, *thydro1s.pat*, and the testing-pattern file, *tehydro1s.pat*, were used for training and testing. The example of the input and output patterns is shown in Figure 4.25. The trained network and testing result files are shown in Table 4.15 and the example of result patterns is shown in Figure 4.26. The predictions with highest accuracy (75.00%) were the result from *hydro1s70.net* and *hydro1s100.net* which had 70 and 100 hidden units respectively. While, *hydro1s7.net, hydro1s35.net* and *hydro1s120.net* were all 71.43 % accurate (Table 4.15 and Figure 4.32).

Table 4.15 The trained network file, testing result file and percent accuracy of sheet structure prediction using hydrophobicity (3 groups).

Hidden units	No. of units	No. of links	Training network file	Testing result file	% accuracy of prediction
7	489	3374	hydro1s7.net	tehydro1s7.res	71.43
35	517	16870	hydro1s35.net	tehydro1s35.res	71.43
70	552	33730	hydro1s70.net	tehydro1s70.res	75.00
100	582	48200	hydro1s100.net	tehydro1s100.res	75.00
120	602	57840	hydro1s120.net	tehydro1s120.res	71.43

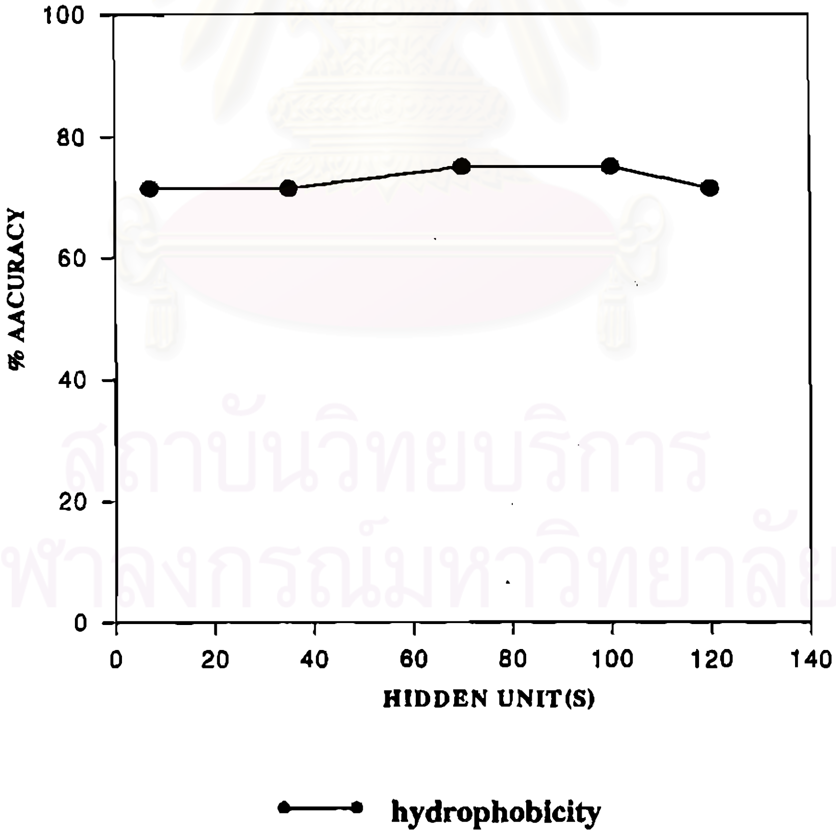

Figure 4.32 Percent accuracy prediction of sheet structure using 3 groups of hydrophobicity as input vector with various numbers of hidden units (7, 35, 70, 100, 120).

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)

Figure 4.33 shows the highest percent accuracy prediction was 85.71% which resulted from *aa8gr1s100.net*. This network with 100 hidden units was trained from the amino acid side chain properties. These properties also gave the over all percent accuracy higher than the others properties. Their percent accuracy predictions were more than 80% in the different numbers of hidden units. Networks with others properties gave the prediction accuracy between 70 - 80 % excepted some networks of hydropathy (2 groups). These networks had 70 and 100 hidden units and gave 82.14 % accuracy prediction.

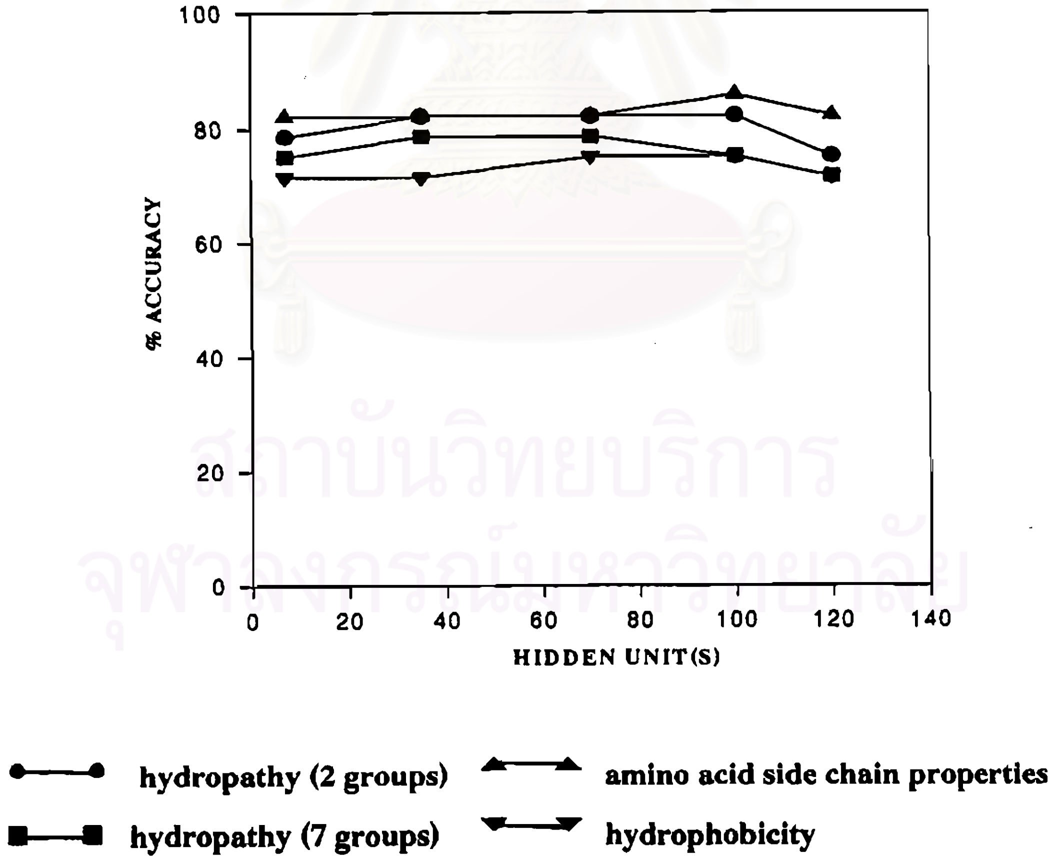

Figure 4.33 The comparison of percent accuracy prediction of sheet structure using various types of input properties.

4.4 Turn structure prediction in amino acid sequence of proteins

Training set.

Seventy patterns of amino acid sequences (Table 4.16) were saved as a train-pattern file for training.

Testing set.

Twenty eight patterns of amino acid sequences (Table 4.16) were saved as a test-pattern file for testing.

Input and output patterns for training and testing.

An input pattern was composed of 481 input units of amino acid which were coded with amino acid properties as previously described in section 3.2.2. An output pattern had one unit of turn structure. Symbol "1" and "0" represented an existence and non-existence of turn structure.

A learn-pattern file with 70 patterns was used in training the networks with various hidden units, 7, 35, 70, 100 and 120. The maximum value for learn cycles was set to 100000 and the maximum error to stop was set to 0.01. A test-pattern file with 28 patterns was used for testing the trained networks.

These training and testing sets (Table 4.16) with input and output patterns as describe above were used for all prediction of turn structure in separate network (section 4.4.1, 4.4.2, 4.4.3 and 4.4.4)

Table 4.16 The training set, testing set and outputs of turn structure prediction in separate networks.

Training set (Input No.)*	Testing set (Input No.)*	Output patterns	Structure type existing in amino acid sequence
2, 3, 6, 10, 11, 12, 14, 16, 17, 19, 20, 21, 22, 24, 26, 27, 29, 30, 31, 32, 35, 36, 40, 42, 44, 47, 49, 49, 50, 53, 54, 56, 58, 60, 61, 62, 63, 64, 65, 66, 68, 70	71, 72, 74, 79, 80, 81, 82, 84, 85, 86, 88, 90, 92, 94, 95, 96, 97, 98	1	turn
1, 4, 5, 7, 8, 9, 13, 15, 18, 23, 25, 28, 33, 34, 37, 38, 39, 41, 43, 45, 46, 48, 50, 51, 52, 55, 57, 59, 67, 69	73, 75, 76, 77, 78, 83, 87, 89, 91, 93	0	no turn

* Names of input no. were listed in Table 3.1

4.4.1 Prediction of turn structure using hydropathy (2 groups)

The learn-pattern file, *thy2gr1t.pat,* and the testing-pattern file, *tehy2gr1t.pat*, were used for training and testing respectively. The example of an input and output patterns is shown in Figure 4.16. The trained network files and testing result files are shown in Table 4.17 and the example of result patterns is shown in Figure 4.17. The prediction with highest accuracy (57.14%) was the result from *hy2gr1t35.net* and *hy2gr1t120.net* which had 35 and 120 hidden units respectively. Whereas the network with 70 hidden units gave 50%. The lowest prediction accuracy resulted form the network with 7 and 100 hidden unit. The lowest prediction accuracy was 46.43% (Table 4.17 and Figure 4.34).

Table 4.17 The trained network, testing result files and percent accuracy of turn structure prediction using hydropathy (2 groups).

Hidden units	No. of units	No. of links	Training network file	Testing result file	% accuracy of prediction
7	489	3374	hy2gr1t7.net	tehy2gr1t7.res	46.43
35	517	16870	hy2gr1t35.net	tehy2gr1t35.res	57.14
70	552	33730	hy2gr1t70.net	tehy2gr1t70.res	50.00
100	582	48200	hy2gr1t100.net	tehy2gr1t100.res	46.43
120	602	57840	hy2gr1t120.net	tehy2gr1t120.res	57.14

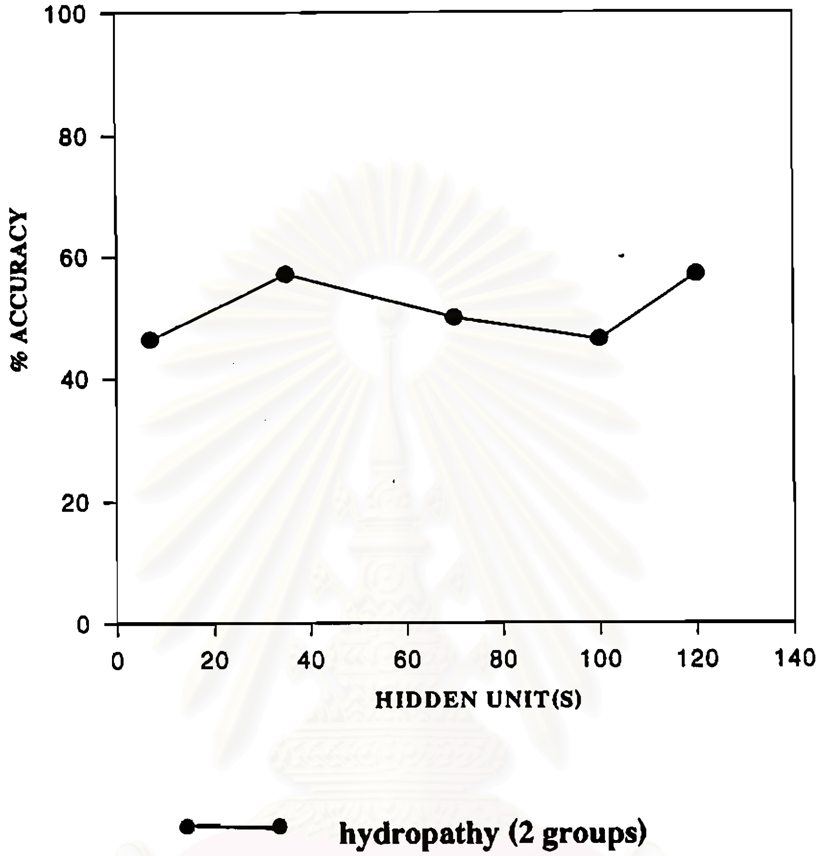

Figure 4.34 Percent accuracy prediction of turn structure using 2 groups of hydropathy as input vector with various numbers of hidden units (7, 35, 70, 100, 120).

4.4.2 Prediction of turn structure using hydropathy (7 groups)

The learn-pattern file, *thy7gr1t.pat*, and the testing-pattern file, *tehy7gr1t.pat*, were used for training and testing respectively. Figure 4.19 shows an example of input and output patterns. The trained network files and testing result files are shown in Table 4.18 and the example of result patterns is shown in Figure 4.20. The most accurate prediction (57.14%) came from *hy7gr1t7.net* and *hy7gr1t100.net* which had 7 and 100 hidden units. Others networks with 35, 70 and 120 hidden units gave 39.29%, 46.43% and 50% percent accuracy respectively (Table 4.18 and Figure 4.35).

Table 4.18 The trained network file, testing result file and percent accuracy of turn structure prediction using hydropathy (7 groups).

Hidden units	No. of units	No. of links	Training network file	Testing result file	% accuracy of prediction
7	489	3374	hy7gr1t7.net	tehy7gr1t7.res	57.14
35	517	16870	hy7gr1t35.net	tehy7gr1t35.res	39.29
70	552	33730	hy7gr1t70.net	tehy7gr1t70.res	46.43
100	582	48200	hy7gr1t100.net	tehy7gr1t100.res	57.14
120	602	57840	hy7gr1t120.net	tehy7gr1t120.res	50.00

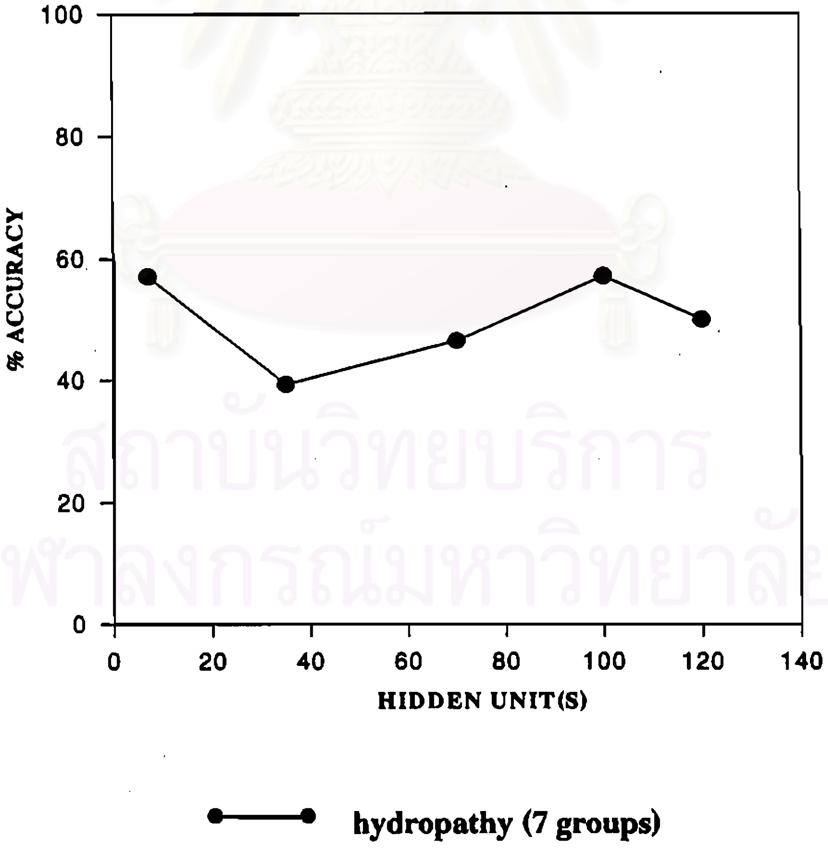

Figure 4.35 Percent accuracy prediction of turn structure using 7 groups of hydropathy as input vector with various numbers of hidden units (7, 35, 70,100, 120).

4.4.3 Prediction of turn structure using amino acid side chain properties (8 groups)

The learn-pattern file, *taa8gr1t.pat*, and the testing-pattern file, *teaa8gr1t.pat*, were used for training and testing respectively. The example of the input and output patterns is shown in Figure 4.22. The trained network files and testing result files are shown in Table 4.19 and the example of result patterns is shown in Figure 4.23.. The most accurate prediction (67.8%) was the result from *aa8gr1t7.net* which had 7 hidden units. While, the network with 35 hidden unit, *aa8gr1t35.net*, gave 64.2% accuracy. The network with 100 and 120 hidden units both gave 60.71% accurate. The lowest accuracy (57.14%) was the result from the network with 70 hidden units.

Table 4.19 The trained network file, testing result file and percent accuracy of turn structure prediction using amino acid side chain properties (8 groups).

Hidden units	No. of units	No. of links	Training network file	Testing result file	% accuracy of prediction
7	489	3374	aa8gr1t7.net	teaa8gr1t7.res	67.80
35	517	16870	aa8gr1t35.net	teaa8gr1t35.res	64.20
70	552	33730	aa8gr1t70.net	teaa8gr1t70.res	57.14
100	582	48200	aa8gr1t100.net	teaa8gr1t100.res	60.71
120	602	57840	aa8gr1t120.net	teaa8gr1t120.res	60.71

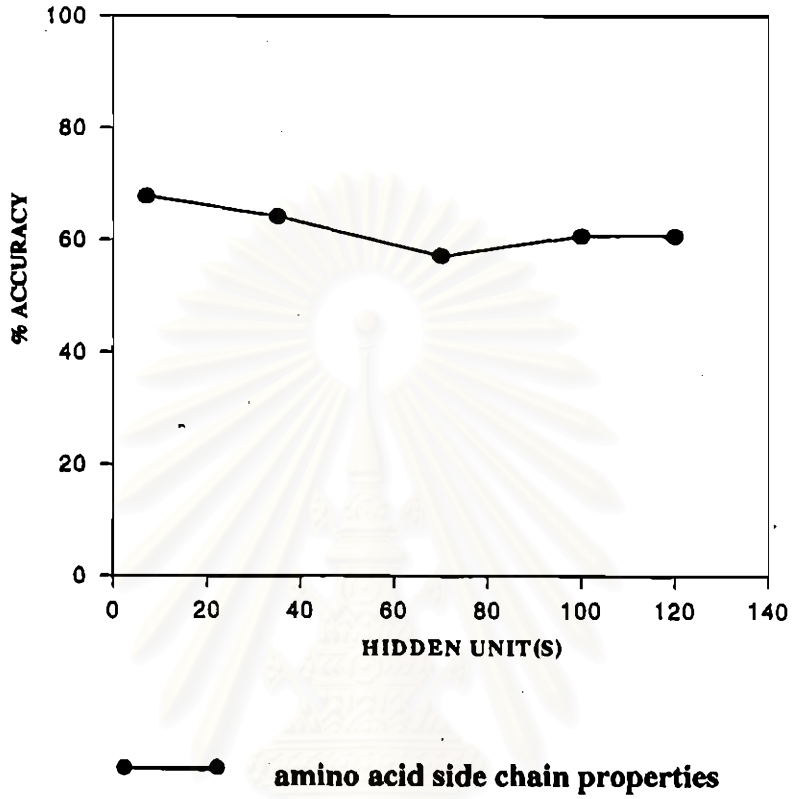

Figure 4.36 Percent accuracy prediction of turn structure using 8 groups of amino acid side chain properties as input vector with various numbers of hidden units (7, 35, 70, 100, 120).

4.4.4 Prediction of turn structure using hydrophobicity (3 groups)

The Learn-pattern file, *thydro1t.pat*, and the testing-pattern file, *tehydro1t.pat*, were used for training and testing. The example of the input and output patterns is shown in Figure 4.25. The trained network files and testing result files are shown in Table 4.20 and the example of result patterns is shown in Figure 4.26. From Figure 4.37 and Table 4.20, the most accurate prediction (60.71%) was the result from *hydro1t7.net* and which had 7 hidden units. Whereas, *hydro1t35.net* and *hydro1t100.net* which had 35 and 100 hidden units respectively gave 53.57 % accurate. The *hydro1t120.net* network with 120 hidden unit gave 57.14% accurate. The lowest prediction (46.43%) was the result from *hydro1t70.net* which had 70 hidden units.

Table 4.20 The trained network file, testing result file and percent accuracy of turn structure prediction using hydrophobicity (3 groups).

Hidden units	No. of units	No. of links	Training network file	Testing result file	% accuracy of prediction
7	489	3374	hydro1t7.net	tehydro1t7.res	60.71
35	517	16870	hydro1t35.net	tehydro1t35.res	53.57
70	552	33730	hydro1t70.net	tehydro1t70.res	46.43
100	582	48200	hydro1t100.net	tehydro1t100.res	53.57
120	602	57840	hydro1t120.net	tehydro1t120.res	57.14

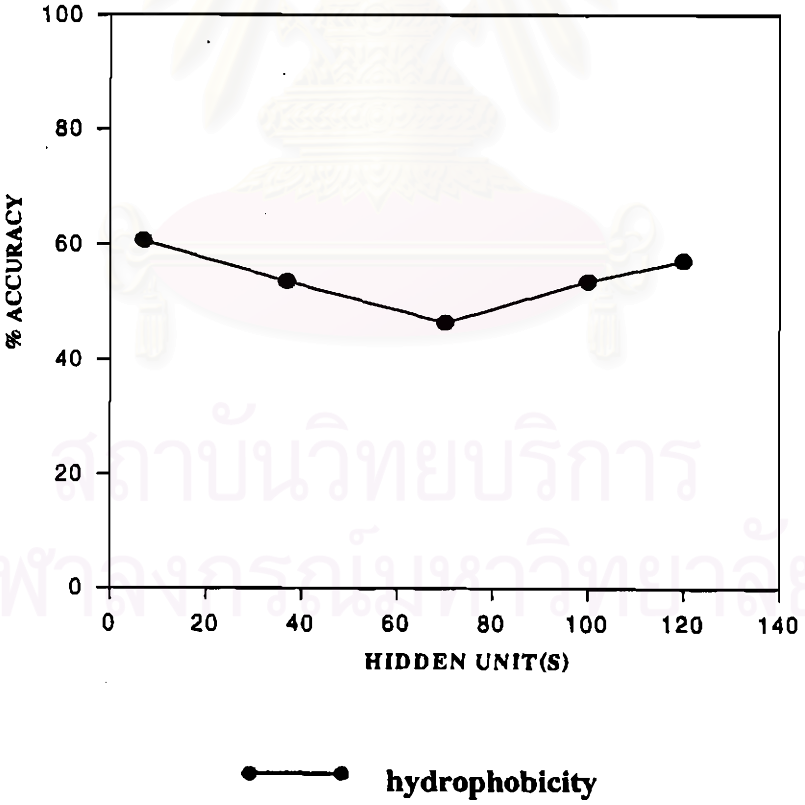

Figure 4.37 Percent accuracy prediction of turn structure using 3 groups of hydrophobicity as input vector with various numbers of hidden units (7, 35, 70, 100, 120).

4.4.5 Comparison of turn structure prediction in separate networks using hydropathy (2 groups), hydropathy (7 groups), amino acid side chain properties (8 groups) and hydrophobicity (3 groups)

Figure 4.38 shows that most of percent accuracy predictions were in the range between 50-60%. The lowest accuracy was 39.29% which resulted from *hy71t35.net.* This network with 35 hidden unit was trained from hydropathy (7 groups). The highest accuracy resulted from *aa8gr1t7.net* which used amino acid side chain properties as input vector and had 7 hidden units. The highest prediction accuracy was 67.80%.

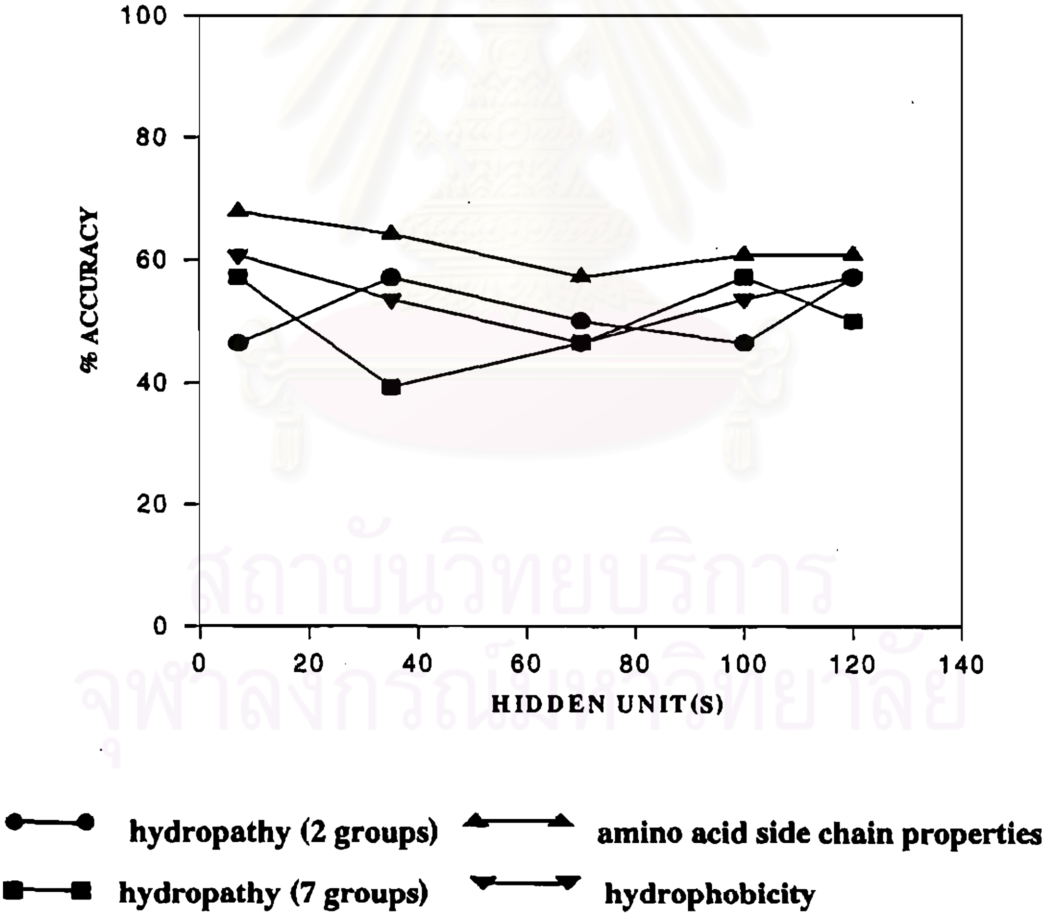

Figure 4.38 The comparison of percent accuracy prediction of turn structure using various types of input properties.

4.5 The existence of helix, sheet and turn structures prediction

4.5.1 The existence of helix, sheet and turn structures prediction using hydropathy (2 groups)

From Figure 4.39, the percent accuracy of helix and sheet structures prediction using the hydropathy (2 groups) were the range between 80-100%. The highest percent accuracy (100%) was the result from the helix structure prediction using the network with 100 hidden units. From the turn structure prediction, the percent accuracy was lower than 60%. The range of percent accuracy for this structure prediction was in the range between 40-60%. In the same way, the range of percent accuracy of helix, sheet and turn which performed in the same network was between 40-60% accurate.

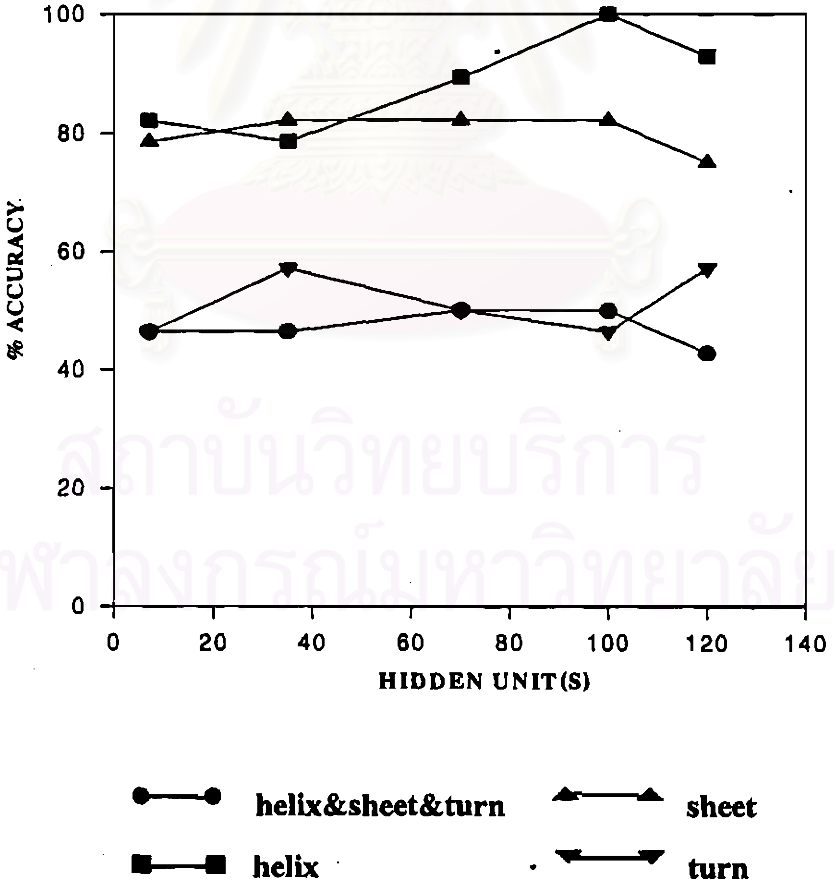

Figure 4.39 The percent accuracy predictions of the existence of helix, sheet, turn in separate networks and in the same network using hydropathy (2 groups).

4.5.2 The existence of helix, sheet, and turn structures prediction using hydropathy (7 groups)

From Figure. 4.40, this hydropathy (7 groups) property gave the overall highest percent accuracy for the helix structure prediction. The percent accuracy of this helix structure prediction was higher than 85%. While, the percent accracy of the sheet and turn structures prediction was in the range between 70-80% and 40 -60% respectively. The percent accuracy of the helix, sheet and turn in the same network was lower than the structure predictions in separate networks.

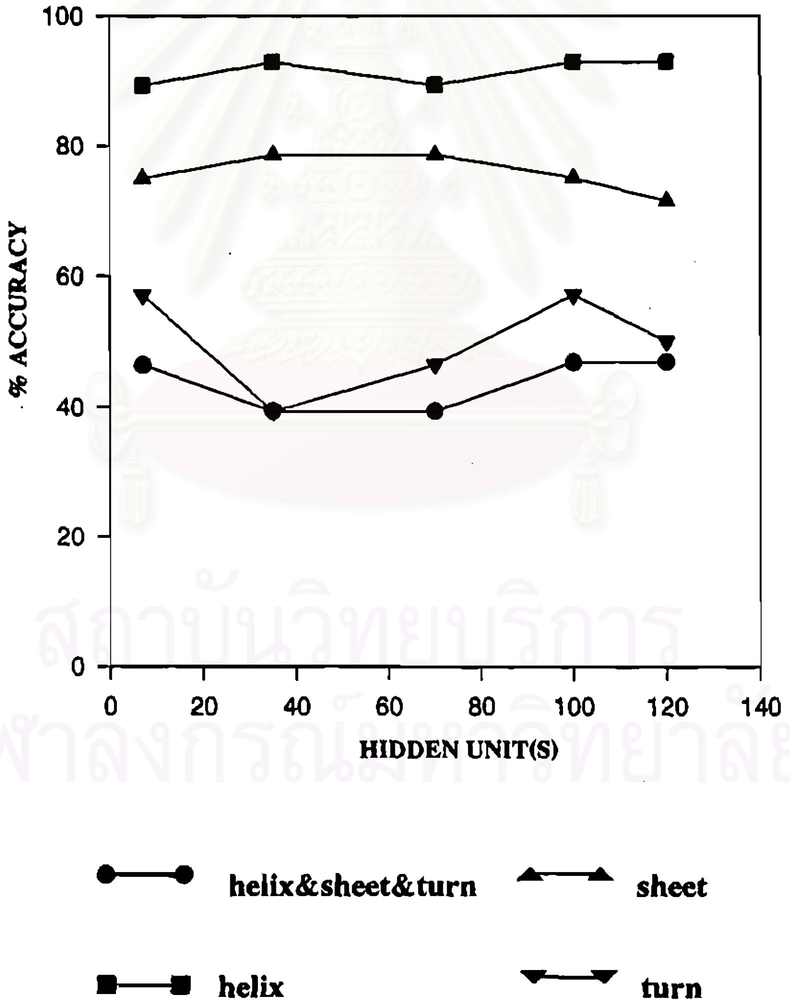

Figure 4.40 The percent accuracy predictions of the existence of helix, sheet, turn in separate networks and in the same network using hydropathy (7 groups).

4.5.3 The existence of helix, sheet and turn structures prediction using amino acid side chain properties (8 groups)

From Figure. 4.41, this amino acid side chain properties also gave the over all higest percent accuracy for helix structure prediction in separate network. The percent of the helix prediction was more more than 90% accurracy. While the range of percent accuracy of sheet and turn structure prediction were in the rage between 80-85% and 55-70% respectively. The percent prediction of helix sheet and turn structure in the same network was lower than the structure prediction in separate network. The range of the structure prediction in the same network was between 40-55% accuracy.

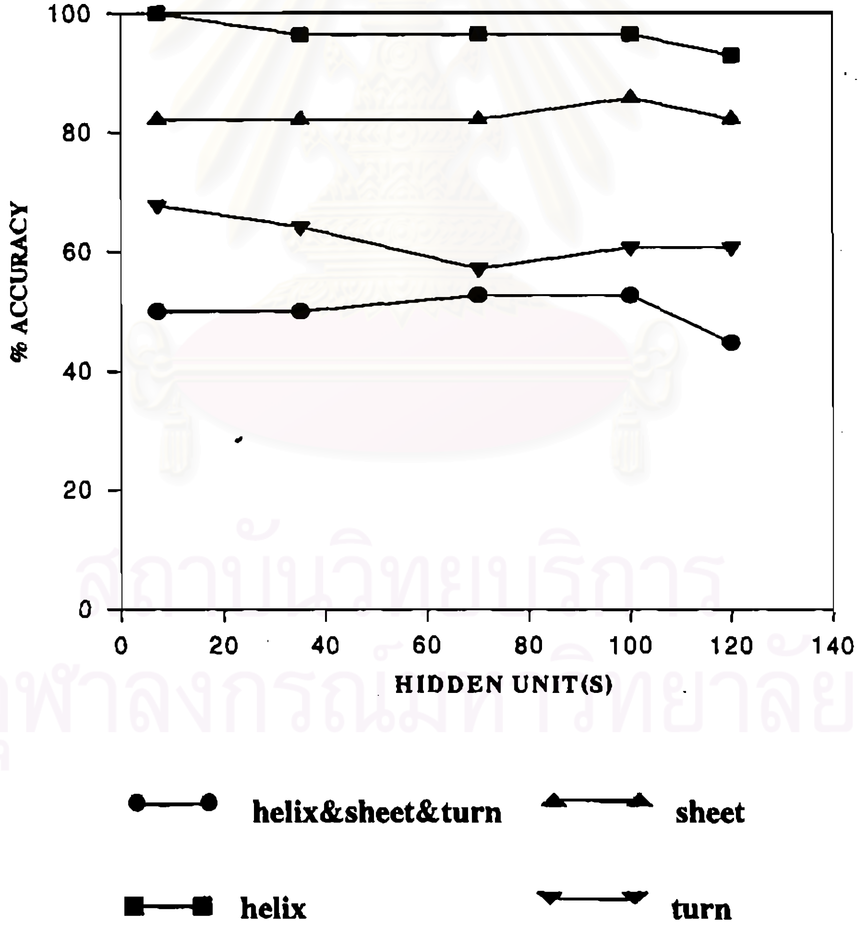

Figure 4.41 The percent accuracy predictions of the existence of helix, sheet, turn in separate networks and in the same network using amino acid side chain properties.

4.5.4 The existence of helix, sheet, and turn structures prediction using hydrophobicity (3 groups)

From Figure. 4.42, the hydrophobicity gave the highest percent accuracy for helix structure prediction. The percent accuracy of this prediction was more than 90%. For the prediction of sheet structure, the range of percent accuracy was between 70-75%. While, percent accuracy of turn structure prediction was between 45-60%. The percent accuracy of the helix sheet and turn structures prediction in the same network was rather low and lower than the structure predictions in separate networks. The range of percent accuracy of this prediction was between 25-35% accurate.

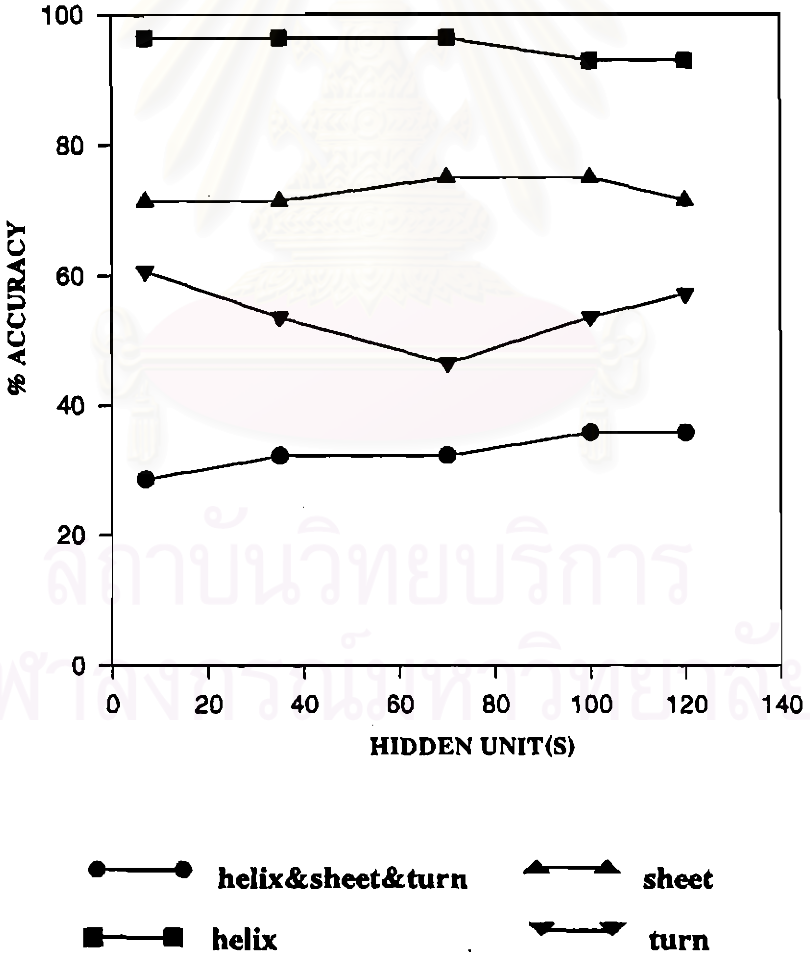

Figure 4.42 The percent accuracy predictions of the existence of helix, sheet, turn in separate networks and in the same network using hydrophobicity (3 groups).

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

Training set.

Seventy patterns of amino acid sequences were saved as a learn pattern file for used in training (Table 4.21). The network were trained for the prediction of six groups of percent helix structure using amino acid properties in the section 4.6.1, 4.6.2, 4.6.3 and 4.6.4

Testing set.

Twenty-eight patterns of amino acid sequences were saved as a test pattern file for used in testing (Table 4.21). The network were tested for the prediction of six groups of percent helix structure using amino acid properties in the section 4.6.1, 4.6.2, 4.6.3 and 4.6.4.

Input and output patterns for training and testing.

The input pattern was composed of 481 input units of amino acid which were coded with properties of amino acid residue as previously described (section 3.2.2).

The output pattern had 3 units of percent helix structure. The percent helix structure in amino acid sequence was divided into 6 groups. There were five possible outputs representing the following ranges of percent values 1-20%, 21-40%, 41-60%, 61-80%, 81-100% and the sixth group represented 0%. The training set, testing set and the possible outputs are shown in Table 4.21.

A learn-pattern file with 70 patterns was used in training the networks with 7, 35, 70, 100 and 120 hidden units. The maximum value for learn cycles was set to 100000 and the maximum error to stop was set to 0.01. A test-pattern file with 28 patterns was used for testing the trained networks.

These training and testing sets (Table 4.21) were the standard sets for prediction of percent helix structure (6 groups) using various types of properties in the next section (section. 4.6.1, 4.6.2, 4.6.3 and 4.6.4).

Table 4.21 The training set, testing set and possible outputs of percent helix structure (6 groups) prediction.

Training set (Input No.)*	**Testing set (Input No.)***	**Output patterns**	**% of helix structure**
4, 6, 10, 11, 43, 64	68	0 0 0	0
3, 5, 9, 16, 17, 20, 21, 30, 32, 34, 36, 38, 45, 47, 52, 54, 55, 57, 58, 62	71, 72, 80, 81, 82, 87, 92, 96	0 0 1	1-20
1, 8, 12, 15, 19, 22, 27, 29, 31, 33, 35, 37, 41, 44, 46, 51, 56, 63, 65, 67, 69	70, 74, 75, 76, 79, 90, 91, 93, 95, 97, 98	0 1 0	21-40
2, 14, 18, 24, 26, 39, 40, 50, 53, 60, 61, 66, 73	84, 85, 88, 89, 94	0 1 1	41-60
7, 13, 23, 25, 28, 42, 48, 49, 59, 77	78, 83, 86	1 0 0	61-80
-	-	1 0 1	81-100

* Names of input no. were listed in Table 3.1

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

The learn-pattern file, *thy2gr%6h.pat*, and testing-pattern file, *tehy2gr%6h.pat*, with hydropathy (2 groups) coded amino acid sequences were used for training and testing respectively. Figure 4.43 shows the example of input and output patterns. The trained network file and testing result files are shown in Table 4.22 and the example of result patterns is shown in Figure 4.44. The best prediction was the result from *hy2gr%6h35.net*. This network with 35 hidden units, gave 70.71% accuracy prediction, the highest in its groups. Files *hy2gr%6h7.net, hy2gr%6h100* and *hy2gr%6h120.net* with 7, 100 and 120 hidden units respectively gave the same accuracy result, 57.14 %. File *hy2gr%6h70.net*, with 7 hidden units gave 53.57% the lowest accuracy (Table 4.22 and Figure 4.45).

Table 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).

Hidden units	No. of units	No. of links	Training network file	Testing result file	% accuracy of prediction
7	491	3388	hy2gr%6h7.net	tehy2gr%6h7.res	57.14
35	519	16940	hy2gr%6h35.net	tehy2gr%6h35.res	70.71
70	554	33880	hy2gr%6h70.net	tehy2gr%6h70.res	53.57
100	584	48400	hy2gr%6h100.net	tehy2gr%6h100.res	57.14
120	604	58080	hy2gr%6h120.net	tehy2gr%6h120.res	57.14

```
SNNS pattern definition file V1.4
No. of patterns        : 70
No. of input units     : 481
No. of output units    : 3

#input1 1alc
       0.2   0.2   0.1   0.2   0.2   0.1   0.2   0.1   0.2   0.2   0.2   0.1   0.2
       0.2   0.1   0.2   0.2   0.2   0.2   0.2   0.1   0.1   0.1   0.2   0.2   0.1
       0.1   0.1   0.2   0.1   0.1   0.2   0.2   0.2   0.2   0.2   0.2   0.2   0.2
       0.1   0.1   0.1   0.2   0.2   0.2   0.2   0.2   0.2   0.2   0.2   0.2   0.1
       0.1   0.2   0.1   0.2   0.2   0.1   0.1   0.2   0.1   0.2   0.2   0.2   0.2
       0.2   0.2   0.2   0.2   0.2   0.2   0.1   0.1   0.2   0.1   0.2   0.1   0.2
       0.2   0.1   0.1   0.2   0.2   0.2   0.1   0.2   0.2   0.2   0.1   0.1   0.1
       0.1   0.2   0.2   0.1   0.1   0.2   0.1   0.2   0.2   0.1   0.2   0.2   0.2
       0.1   0.1   0.2   0.2   0.1   0.1   0.1   0.2   0.2   0.2   0.1   0.2   0.2
       0.2   0.1   0.1   0.2   0.2   0.2   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
#output1 1alc
      0  1  1

#input2 1bp2
       0.1   0.1   0.2   0.2   0.1   0.2   0.2   0.1   0.1   0.2   0.1   0.2   0.1
       0.2   0.2   0.2   0.2   0.2   0.1   0.1   0.2   0.1   0.2   0.2   0.2   0.2
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
#outut2 1bp2
      ...   ....  ....

#input70 7rsa

       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
```

Figure 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).

```
SNNS result file V1.4-3D
No. of hidden units : 70
No. of patterns     : 28
No. of input units  : 481
No. of output units : 3
startpattern        : 1
endpattern          : 28
input patterns included
teaching outputs included

#input1
       0.2   0.2   0.1   0.1   0.2   0.1   0.2   0.1   0.2   0.2   0.2   0.1   0.2
       0.2   0.2   0.2   0.1   0.2   0.2   0.1   0.2   0.1   0.2   0.2   0.2   0.2
       0.2   0.2   0.1   0.2   0.1   0.1   0.1   0.2   0.2   0.2   0.2   0.2   0.2
       0.1   0.2   0.1   0.2   0.2   0.1   0.2   0.1   0.2   0.2   0.2   0.1   0.1
       0.2   0.2   0.2   0.2   0.2   0.1   0.2   0.2   0.2   0.2   0.2   0.2   0.2
       0.1   0.2   0.2   0.1   0.2   0.2   0.2   0.1   0.2   0.1   0.2   0.2   0.2
       0.2   0.2   0.2   0.1   0.2   0.2   0.1   0.1   0.2   0.2   0.2   0.2   0.1
       0.2   0.1   0.2   0.2   0.1   0.1   0.2   0.1   0.2   0.2   0.2   0.1   0.2
       0.1   0.2   0.2   0.2   0.2   0.2   0.2   0.2   0.1   0.1   0.2   0.2   0.1
       0.2   0.1   0.2   0.2   0.1   0.1   0.2   0.1   0.1   0.2   0.2   0.2   0.1
       0.1   0.2   0.2   0.2   0.2   0.1   0.2   0.1   0.1   0.2   0.2   0.1   0.1
       0.2   0.2   0.2   0.1   0.1   0.1   0.2   0.2   0.1   0.1   0.2   0.1   0.2
       0.1   0.2   0.2   0.2   0.2   0.2   0.2   0.2   0.1   0.1   0.2   0.1   0.2
       0.1   0.1   0.2   0.2   0.2   0.2   0.2   0.2   0.2   0.2   0.1   0.2   0.2
       0.1   0.2   0.1   0.2   0.1   0.2   0.2   0.2   0.2   0.2   0.1   0.2   0.1
       0.2   0.2   0.1   0.2   0.1   0.2   0.2   0.1   0.1   0.1   0.1   0.1   0.2
       0.2   0.2   0.1   0.2   0.1   0.1   0.1   0.2   0.2   0.2   0.2   0.2   0.2
       0.1   0.1   0.2   0.2   0.2   0.2   0.2   0.1   0.1   0.2   0.1   0.2   0.2
       0.1   0.1   0.2   0.1   0.2   0.2   0.2   0.2   0.2   0.2   0.2   0.1   0.2
       0.1   0.2   0.1   0.2   0.2   0.1   0.2   0.2   0.1   0.2   0.2   0.1   0.1
       0.1   0.2   0.1   0.2   0.2   0.2   0.1   0.2   0.2   0.1   0.2   0.2   0.2
       0.1   0.2   0.2   0.2   0.2   0.2   0.2   0.2   0.1   0.1   0.2   0.2   0.2
       0.1   0.2   0.2   0.2   0.2   0.2   0.2   0.2   0.2   0.2   0.1   0.1   0.2
       0.2   0.1   0.1   0.1   0.2   0.2   0.2   0.2   0.2   0.1   0.1   0.2   0.2
       0.1   0.2   0.2   0.1   0.1   0.2   0.1   0.1   0.2   0.1   0.1   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
#output1
    0       0       1
  0.0001  0.8844   0.7771
#input2
       0.1   0.1   0.2   0.2   0.1   0.2   0.2   0.1   0.1   0.2   0.1   0.2   0.1
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
#output
   ....   ....   ....
   ....   ....   ....

#input28
         ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
         ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
```

Figure 4.44 The example of incorrect 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). The predicted outputs were the three floating numbers, 0.001, 0.8854 and 0.7771.

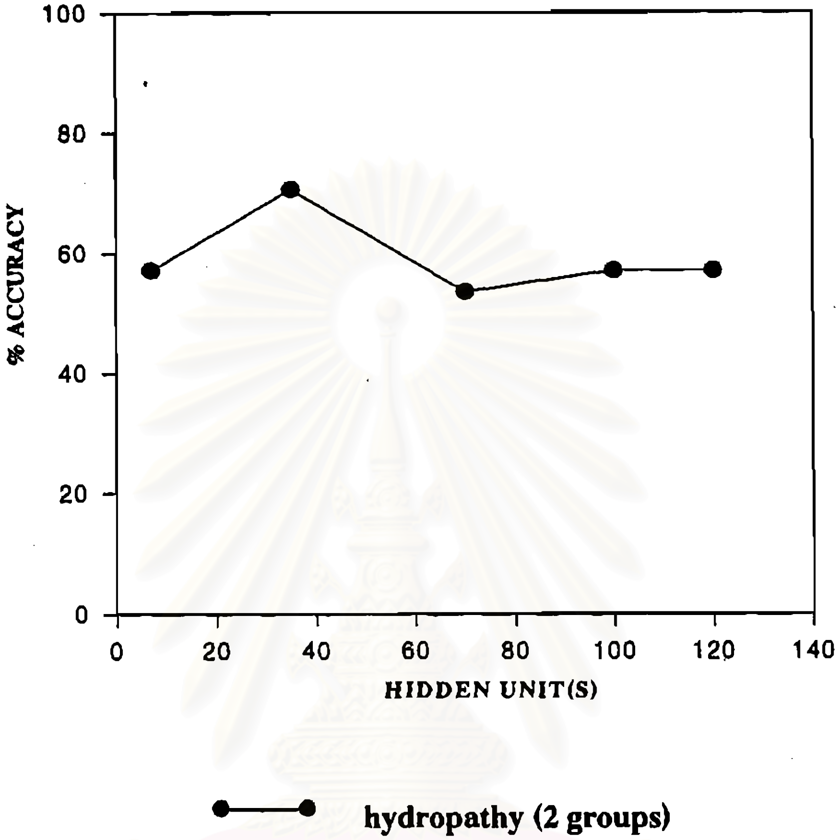

Figure 4.45 Percent accuracy prediction of percent helix structure (6 groups) using 2 groups of hydropathy as input vector with various numbers of hidden units (7, 35, 70, 100, 120).

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

The learn-pattern file, *thy7gr%6h.pat*, and test-pattern file, *tehy7gr%6h.pat*, with hydropathy (7 groups) coded the amino acid sequences were used in training and testing with various hidden units. Figure 4.46 shows the example of input and output patterns. The trained network and testing result file are shown in Table 4.23 and the example of result patterns is in Figure 4.47. Table 4.23 and Figure 4.48 show that, *hy7gr%6h70.net*, which had 70 hidden units gave the highest percent accuracy of 57.14%. Files *hy7gr%6h7.net, hy7gr%6h35.net, hy7gr%6h100.net* and *hy7gr%6h120* gave 50.00%, 42.86%, 46.43% and 50.00% accuracy respectively.

```
SNNS pattern definition file V1.4

No. of patterns       : 70
No. of input units    : 481
No. of output units   : 3

#input1 1alc
        0.7    0.6    0.2    0.4    0.7    0.2    0.6    0.1    0.4    0.6    0.6    0.1    0.5
        0.6    0.1    0.6    0.4    0.5    0.4    0.7    0.1    0.3    0.1    0.5    0.6    0.1
        0.1    0.2    0.4    0.3    0.2    0.6    0.4    0.4    0.4    0.5    0.6    0.4    0.6
        0.3    0.1    0.1    0.6    0.6    0.6    0.6    0.4    0.4    0.6    0.5    0.4    0.1
        0.2    0.6    0.1    0.4    0.6    0.3    0.1    0.4    0.2    0.7    0.4    0.4    0.6
        0.4    0.5    0.6    0.4    0.7    0.6    0.1    0.2    0.6    0.1    0.4    0.2    0.6
        0.7    0.2    0.1    0.6    0.6    0.6    0.1    0.4    0.6    0.6    0.1    0.3    0.2
        0.3    0.7    0.7    0.1    0.1    0.6    0.1    0.7    0.4    0.1    0.6    0.5    0.4
        0.1    0.3    0.6    0.7    0.3    0.1    0.2    0.4    0.6    0.7    0.1    0.6    0.6
        0.4    0.1    0.2    0.6    0.7    0.6    0.0    0.0    0.0    0.0    0.0    0.0    0.0
        0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
        0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
        0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
        0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
        0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
        0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
        0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.C
        0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
        0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
        0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
        0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
        0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
        0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
        0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
        0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
        0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
        0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
        0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
        0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
        0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
        0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
        0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
        0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
        0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
        0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
        0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
        0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
#output1 1alc
      0   1   1

#input2
        0.3    0.1    0.4    0.6    0.2    0.6    0.4    0.3    0.1    0.7    0.2    0.7    0.1
        0.5    0.4    0.4    0.6    0.5    0.1    0.1    0.6    0.2    0.6    0.6    0.5    0.4
        0.2    0.5    0.2    0.4    0.1    0.4    0.4    0.4    0.4    0.4    0.5    0.1    0.6
        ...    ...    ...    ...    ...    ...    ...    ...    ...    ...    ...    ...    ...
        ...    ...    ...    ...    ...    ...    ...    ...    ...    ...    ...    ...    ...
#outut2 1bp2
      ...   ....   ....

#input70 7rsa

        ...    ...    ...    ...    ...    ...    ...    ...    ...    ...    ...    ...    ...
                    ...    ...    ...    ...    ...    ...    ...    ...    ...    ...    ...    ...
...
                    ...    ...    ...    ...    ...    ...    ...    ...    ...    ...    ...    ...
...
```

Figure 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).

```
SNNS result file V1.4-3D

No. of hidden units : 70
No. of patterns     : 28
No. of input units  : 481
No. of output units : 3
startpattern        : 1
endpattern          : 28
input patterns included
teaching outputs included

#input1
       0.4   0.6   0.6   0.4   0.4   0.5   0.4   0.7  .0.6   0.6   0.4   0.5   0.6
       0.6   0.4   0.6   0.7   0.6   0.2   0.5   0.1   0.3   0.7   0.4   0.6   0.7
       0.6   0.4   0.5   0.1   0.6   0.1   0.6   0.4   0.6   0.4   0.3   0.7   0.5
       0.6   0.5   0.4   0.1   0.7   0.5   0.1   0.4   0.1   0.4   0.5   0.6   0.6
       0.3   0.4   0.4   0.1   0.7   0.1   0.1   0.6   0.6   0.4   0.6   0.3   0.2
       0.6   0.1   0.6   0.2   0.6   0.6   0.4   0.6   0.6   0.7   0.3   0.1   0.1
       0.7   0.4   0.4   0.5   0.1   0.6   0.4   0.4   0.5   0.7   0.1   0.1   0.6
       0.2   0.6   0.2   0.6   0.4   0.4   0.4   0.1   0.6   0.4   0.6   0.4   0.4
       0.6   0.6   0.4   0.1   0.6   0.7   0.7   0.7   0.5   0.3   0.3   0.6   0.1
       0.6   0.1   0.1   0.6   0.4   0.6   0.4   0.7   0.5   0.4   0.6   0.2   0.4
       0.7   0.3   0.1   0.6   0.6   0.5   0.6   0.4   0.1   0.3   0.1   0.1   0.4
       0.1   0.2   0.1   0.7   0.1   0.4   0.4   0.3   0.7   0.5   0.4   0.1   0.6
       0.7   0.1   0.1   0.6   0.1   0.1   0.6   0.4   0.1   0.7   0.4   0.7   0.4
       0.7   0.4   0.3   0.6   0.2   0.4   0.6   0.2   0.6   0.5   0.7   0.4   0.1
       0.1   0.5   0.6   0.4   0.1   0.6   0.5   0.4   0.4   0.5   0.5   0.4   0.4
       0.1   0.4   0.4   0.5   0.5   0.1   0.1   0.6   0.2   0.1   0.4   0.4   0.1
       0.1   0.1   0.7   0.6   0.5   0.1   0.4   0.1   0.4   0.4   0.6   0.6   0.1
       0.1   0.7   0.2   0.7   0.7   0.1   0.6   0.2   0.6   0.4   0.6   0.4   0.6
       0.5   0.6   0.6   0.1   0.3   0.1   0.6   0.6   0.4   0.7   0.5   0.3   0.6
       0.5   0.1   0.7   0.6   0.7   0.6   0.1   0.7   0.3   0.4   0.2   0.7   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
#output1
     0       0       1
  0.0002  0.1532  0.7894

#input2
         0.1   0.4   0.2   0.4   0.3   0.4   0.6   0.7   0.7   0.4   0.1   0.6   0.2
         ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
#output2
   ....  ....  ....
   ....  ....  ....

#input28
         ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
         ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
```

Figure 4.47 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 0 0 1 representing percent helix or sheet or turn structures (6 groups). The predicted outputs were the three floating numbers, 0.0002, 0.1532 and 0.7894.

Table 4.23 The trained network file, testing result file and percent accuracy of percent helix structure prediction (6 groups) using hydropathy (7 groups).

Hidden units	No. of units	No. of links	Training network file	Testing result file	% accuracy of prediction
7	491	3388	hy7gr%6h7.net	tehy7gr%6h7.res	50.00
35	519	16940	hy7gr%6h35.net	tehy7gr%6h35.res	42.86
70	554	33880	hy7gr%6h70.net	tehy7gr%6h70.res	57.14
100	584	48400	hy7gr%6h100.net	tehy7gr%6h100.res	46.43
120	604	58080	hy7gr%6h120.net	tehy7gr%6h120.res	50.00

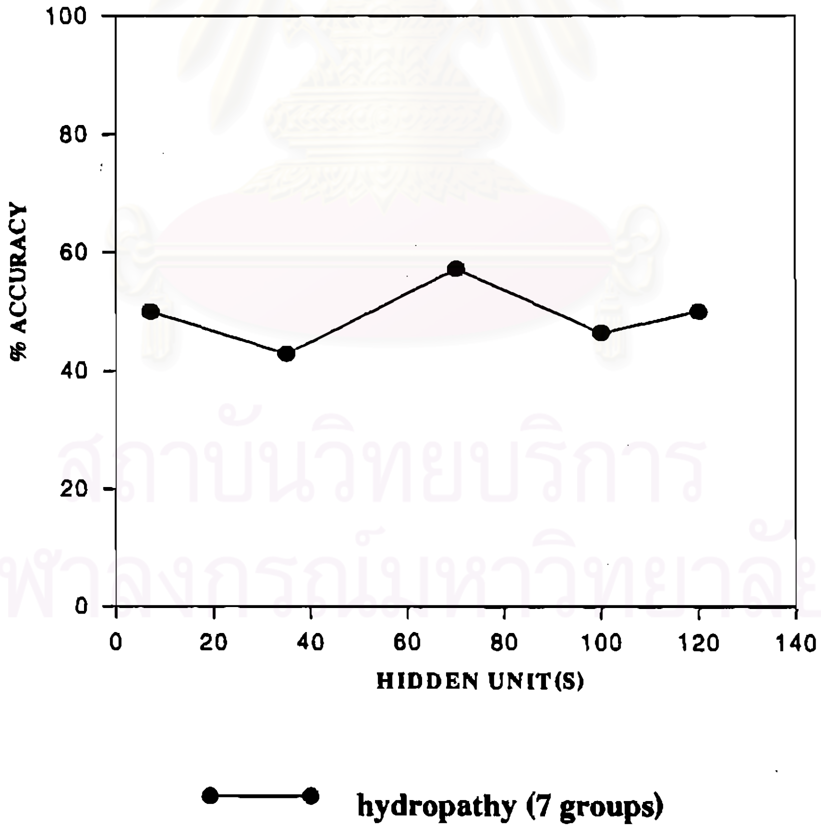

Figure 4.48 Percent accuracy prediction of percent helix structure (6 groups) using 7 groups of hydropathy as input vector with various numbers of hidden units (7, 35, 70, 100, 120).

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

The learn-pattern file, *taa8gr%6h.pat*, and the test-pattern file, *teaa8gr%6h.pat*, with 8 groups of amino acid side chain properties coded amino acid side chains were used in training network and the testing respectively. The example of what training or testing patterns may look like is shown in Figure 4.49. The trained networks files and testing result files are shown in Table 4.24 and the example of the testing result pattern is shown in Figure 4.50. Figure 4.51 and Table 4.24 show that *aa8gr%6h70.net* and *aa8gr%6h120.net* gave the highest percent accuracy of 46.43%. The prediction accuracy from others networks were lower. The trained networks, *aa8gr%6h7.net*, and *aa8gr%6h35.net* both gave 35.71% accuracy prediction. Whereas *aa8gr%6h100.net* gave the lowest accuracy of 28.57%.

Table 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.

Hidden units	No. of units	No. of links	Training network file	Testing result file	% accuracy of prediction
7	491	3388	aa8gr%6h7.net	teaa8gr%6h7.res	35.71
35	519	16940	aa8gr%6h35.net	teaa8gr%6h35.res	35.71
70	554	33880	aa8gr%6h70.net	teaa8gr%6h70.res	46.43
100	584	48400	aa8gr%6h100.net	teaa8gr%6h100.res	28.57
120	604	58080	aa8gr%6h120.net	teaa8gr%6h120.res	46.43

```
SNNS pattern definition file V1.4

No. of patterns        : 70
No. of input units     : 481
No. of output units    : 3

#input1 1alc
        0.1   0.5   0.3   0.7   0.2   0.1   0.2   0.7   0.7   0.6   0.8   0.1   0.0
        0.3   0.7   0.8   0.2   0.5   0.6   0.1   0.2   0.3   0.1   0.1   0.8   0.1
        0.8   0.8   0.8   0.5   0.3   0.1   0.7   0.1   0.6   0.5   0.5   0.7   0.5
        0.6   0.6   0.7   0.5   0.5   0.1   0.6   0.3   0.1   0.5   0.1   0.5   0.2
        0.8   0.3   0.1   0.5   0.1   0.6   0.7   0.1   0.1   0.8   0.1   0.1   0.6
        0.5   0.2   0.6   0.1   0.8   0.2   0.7   0.7   0.8   0.8   0.7   0.6   0.5
        0.1   0.1   0.6   0.1   0.1   0.3   0.2   0.5   0.7   0.5   0.2   0.6   0.1
        0.2   0.8   0.2   0.6   0.2   0.6   0.2   0.1   0.5   0.5   0.8   0.7   0.6
        0.1   0.5   0.7   0.6   0.5   0.1   0.7   0.1   0.1   0.6   0.2   0.5   0.1
        0.7   0.1   0.6   0.1   0.1   0.6   0.2   0.8   0.5   0.1   0.6   0.2   0.5
        0.5   0.1   0.1   0.7   0.1   0.1   0.5   0.6   0.1   0.7   0.1   0.1   0.1
        0.1   0.1   0.1   0.1   0.1   0.4   0.6   0.1   0.1   0.7   0.1   0.8   0.3
        0.6   0.1   0.8   0.6   0.1   0.1   0.7   0.1   0.1   0.8   0.1   0.1   0.6
        0.5   0.6   0.1   0.6   0.6   0.1   0.3   0.2   0.5   0.8   0.2   0.7   0.3
        0.5   0.5   0.1   0.1   0.3   0.5   0.5   0.1   0.7   0.2   0.2   0.5   0.2
        0.3   0.1   0.5   0.1   0.5   0.6   0.3   0.3   0.1   0.2   0.7   0.5   0.1
        0.5   0.2   0.1   0.1   0.4   0.6   0.7   0.5   0.1   0.5   0.1   0.5   0.5
        0.7   0.8   0.1   0.1   0.8   0.2   0.6   0.5   0.1   0.1   0.5   0.8   0.1
        0.7   0.1   0.7   0.8   0.1   0.1   0.3   0.4   0.8   0.6   0.8   0.8   0.6
        0.3   0.5   0.8   0.3   0.1   0.6   0.1   0.6   0.5   0.1   0.6   0.1   0.5
        0.2   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
        0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
        0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
        0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
        0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
        0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
        0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
        0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
        0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
        0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
        0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
        0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
        0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
        0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
        0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
        0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
        0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
#output1 1alc
    0   1   1


#input2 1bp2
        0.1   0.3   0.6   0.1   0.3   0.1   0.7   0.1   0.1   0.6   0.4   0.7   0.6
        0.5   0.6   0.8   0.3   0.1   0.1   0.2   0.8   0.6   0.5   0.5   0.6   0.6
        0.1   0.1   0.6   0.4   0.1   0.7   0.4   0.6   0.6   0.3   0.1   0.8   0.1
        ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
        ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
#outut2 1bp2
      ...   ....   ....


#input70 7rsa
        ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
        ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
```

Figure 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).

```
SNNS result file V1.4-3D
No. of hidden units : 70
No. of patterns     : 28
No. of input units  : 481
No. of output units : 3
startpattern        : 1
endpattern          : 28
input patterns included
teaching outputs included

#input1
       0.4    0.8    0.5    0.1    0.5    0.7    0.6    0.1    0.6    0.6    0.6    0.6    0.1
       0.1    0.1    0.6    0.4    0.5    0.8    0.5    0.7    0.1    0.1    0.5    0.6    0.1
       0.1    0.1    0.6    0.8    0.8    0.5    0.1    0.8    0.1    0.1    0.7    0.1    0.1
       0.1    0.5    0.6    0.6    0.3    0.6    0.2    0.1    0.2    0.7    0.1    0.1    0.4
       0.1    0.1    0.8    0.4    0.7    0.3    0.1    0.2    0.1    0.8    0.2    0.1    0.5
       0.6    0.6    0.1    0.6    0.1    0.1    0.0    0.0    0.0    0.0    0.0    0.0    0.0
       0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
       0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
       0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
       0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
       0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
       0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
       0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
       0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
       0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
       0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
       0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
       0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
       0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
       0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
       0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
       0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
       0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
       0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
       0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
       0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
       0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
       0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
       0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
       0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
       0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
       0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
       0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
       0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
       0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
       0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
       0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0

#output1
   0       0       1
0.0004  0.0025  0.9652

#input2
       0.3    0.6    0.2    0.5    0.1    0.1    0.7    0.6    0.7    0.5    0.4    0.1    0.1
       0.4    0.1    0.1    0.4    0.1    0.1    0.1    0.1    0.3    0.7    0.1    0.2    0.7
       ...    ...    ...    ...    ...    ...    ...    ...    ...    ...    ...    ...    ...
         ...    ...    ...    ...    ...    ...    ...    ...    ...    ...    ...    ...    ...
#output2
   ....    ....    ....
   ....    ....    ....

#input28
       ...    ...    ...    ...    ...    ...    ...    ...    ...    ...    ...    ...    ...
       ...    ...    ...    ...    ...    ...    ...    ...    ...    ...    ...    ...    ...
```

Figure 4.50 The example of correct result pattern after testing with a trained network. The input patterns were amino acid side chain properties (8 groups) coded amino acid residues. The observed outputs were 0 0 1 representing percent helix or sheet or turn structures (6 groups). The predicted outputs were the three floating numbers, 0.0004, 0.0025 and 0.9652.

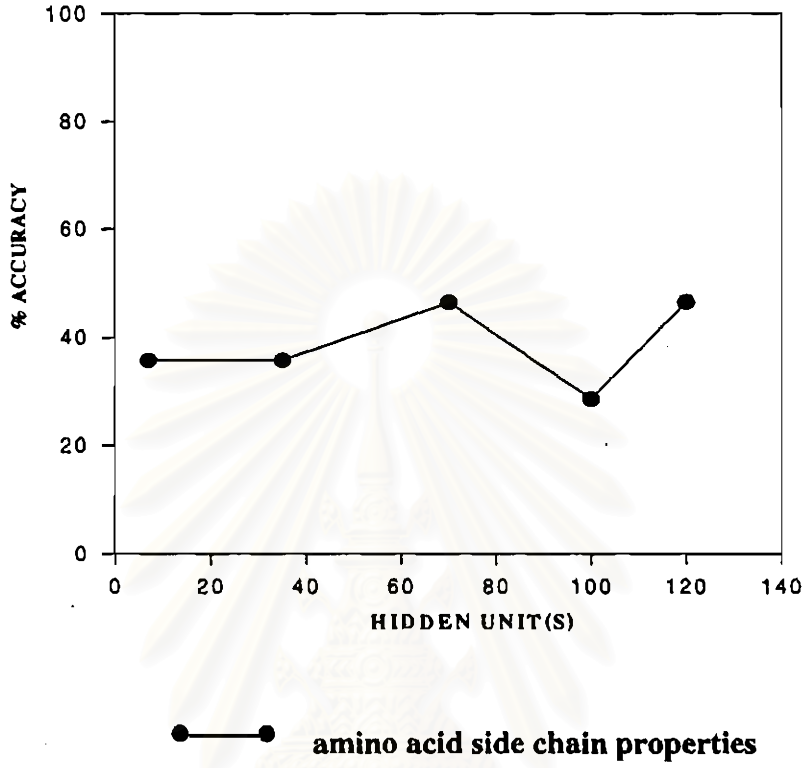

Figure 4.51 Percent accuracy prediction of percent helix structure (6 groups) using 8 groups of amino acid side chain properties as input vector with various numbers of hidden units (7, 35, 70, 100, 120).

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

The learn-pattern file, *thydro%6h.pat*, and a test-pattern file, *tehydro%6h.pat*, with hydrophobicity coded amino acid sequences were used in training network and the testing respectively. The example of what the training or testing pattern may look like is shown in Figure 4.52. The trained networks files and testing result files are shown in Table 4.25 and the example of the testing result patterns is shown in Figure 4.53. Table 4.25 and Figure 4.54, the prediction with highest accuracy (57.14%) was the result from *hydro%6h70.net*, whereas *hydro%6h7.net, hydro%6h35.net* and *hydro%6h100.net* gave 53.57% accuracy. File, *hydro%6h120.net* gave the lowest accuracy of 50.00%.

```
SNNS pattern definition file V1.4

No. of patterns        : 70
No. of input units     : 481
No. of output units    : 3


#input1 lalc
       0.1   0.1   0.3   0.2   0.1   0.3   0.1   0.3   0.2   0.1   0.1   0.3   0.2
       0.1   0.3   0.1   0.2   0.2   0.2   0.1   0.3   0.2   0.3   0.2   0.1   0.3
       0.3   0.3   0.2   0.3   0.3   0.2   0.2   0.2   0.2   0.2   0.1   0.2   0.1
       0.2   0.3   0.3   0.1   0.1   0.1   0.1   0.2   0.2   0.1   0.2   0.2   0.3
       0.3   0.1   0.3   0.2   0.1   0.2   0.3   0.3   0.3   0.1   0.2   0.2   0.1
       0.2   0.2   0.1   0.2   0.1   0.1   0.3   0.3   0.1   0.3   0.2   0.3   0.1
       0.1   0.3   0.3   0.1   0.1   0.1   0.3   0.2   0.1   0.1   0.3   0.3   0.3
       0.2   0.1   0.1   0.3   0.3   0.1   0.3   0.1   0.2   0.3   0.1   0.2   0.3
       0.3   0.2   0.2   0.1   0.2   0.3   0.3   0.2   0.1   0.1   0.3   0.1   0.1
       0.3   0.3   0.3   0.1   0.1   0.1   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
#output1 lalc
  0   1   1


#input2 lbp2
       0.2   0.3   0.3   0.1   0.3   0.1   0.2   0.3   0.3   0.1   0.3   0.1   0.3
       0.2   0.2   0.2   0.1   0.2   0.3   0.3   0.1   0.3   0.1   0.1   0.2   0.2
       0.3   0.2   0.3   0.2   0.3   0.2   0.2   0.2   0.2   0.2   0.2   0.3   0.1
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
#outut2 lbp2
      ...   ....   ....


#input70 7rsa
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
```

Figure 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.

```
SNNS result file V1.4-3D

No. of hidden units : 70
No. of patterns     : 28
No. of input units  : 481
No. of output units : 3
startpattern        : 1
endpattern          : 28
input patterns included
teaching outputs included

#input1
       0.2   0.2   0.2   0.3   0.2   0.2   0.2   0.1   0.2   0.1   0.2   0.2   0.1
       0.2   0.3   0.2   0.1   0.1   0.3   0.2   0.3   0.2   0.1   0.2   0.1   0.1
       0.1   0.2   0.2   0.3   0.1   0.3   0.1   0.2   0.2   0.2   0.2   0.1   0.2
       0.1   0.2   0.2   0.3   0.1   0.2   0.3   0.2   0.3   0.2   0.2   0.1   0.1
       0.2   0.2   0.2   0.3   0.1   0.3   0.3   0.1   0.1   0.2   0.2   0.2   0.3
       0.1   0.3   0.1   0.3   0.1   0.1   0.2   0.1   0.1   0.1   0.2   0.3   0.3
       0.1   0.2   0.2   0.2   0.3   0.1   0.2   0.2   0.2   0.1   0.3   0.3   0.1
       0.3   0.2   0.3   0.2   0.3   0.2   0.2   0.3   0.1   0.2   0.1   0.2   0.2
       0.1   0.2   0.2   0.3   0.1   0.1   0.1   0.1   0.2   0.2   0.2   0.1   0.3
       0.2   0.3   0.3   0.2   0.3   0.1   0.2   0.1   0.2   0.2   0.1   0.3   0.2
       0.1   0.2   0.3   0.1   0.1   0.2   0.1   0.2   0.3   0.2   0.3   0.3   0.2
       0.3   0.3   0.3   0.1   0.3   0.2   0.2   0.2   0.1   0.2   0.2   0.3   0.1
       0.1   0.3   0.3   0.1   0.3   0.3   0.1   0.2   0.3   0.1   0.2   0.1   0.2
       0.1   0.2   0.2   0.1   0.3   0.2   0.1   0.3   0.1   0.2   0.1   0.2   0.3
       0.3   0.2   0.1   0.2   0.3   0.1   0.2   0.3   0.2   0.2   0.2   0.2   0.2
       0.3   0.2   0.2   0.2   0.2   0.3   0.3   0.1   0.3   0.3   0.2   0.3   0.3
       0.3   0.3   0.1   0.1   0.2   0.3   0.2   0.3   0.2   0.2   0.1   0.1   0.3
       0.3   0.1   0.3   0.1   0.1   0.3   0.1   0.3   0.1   0.2   0.1   0.2   0.1
       0.2   0.1   0.1   0.3   0.3   0.3   0.1   0.1   0.3   0.1   0.2   0.2   0.1
       0.2   0.3   0.1   0.1   0.1   0.1   0.3   0.1   0.2   0.2   0.3   0.1   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
#output1
   0       0       1
0.1543  0.0550  0.7641

#input2
       0.1   0.1   0.3   0.3   0.3   0.2   0.1   0.1   0.2   0.1   0.2   0.3   0.1
       0.3   0.2   0.3   0.2   0.2   0.2   0.1   0.1   0.1   0.2   0.3   0.1   0.3
       0.2   0.3   0.1   0.1   0.2   0.1   0.1   0.3   0.1   0.3   0.3   0.2   0.1
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
         ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
#output2
   ....    ....    ....
   ....    ....    ....

#input28
         ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
         ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
```

Figure 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). The predicted outputs were the three floating numbers, 0.1543, 0.0550 and 0.7641.

Table 4.25 The trained network file, testing result file and percent accuracy of percent helix structure prediction (6 groups) using hydrophobicity.

Hidden units	No. of units	No. of links	Training network file	Testing result file	% accuracy of prediction
7	491	3388	hydro%6h7.net	tehydro%6h7.res	53.57
35	519	16940	hydro%6h35.net	tehydro%6h35.res	53.57
70	554	33880	hydro%6h70.net	tehydro%6h70.res	57.14
100	584	48400	hydro%6h100.net	tehydro%6h100.res	53.57
120	604	58080	hydro%6h120.net	tehydro%6h120.res	50.00

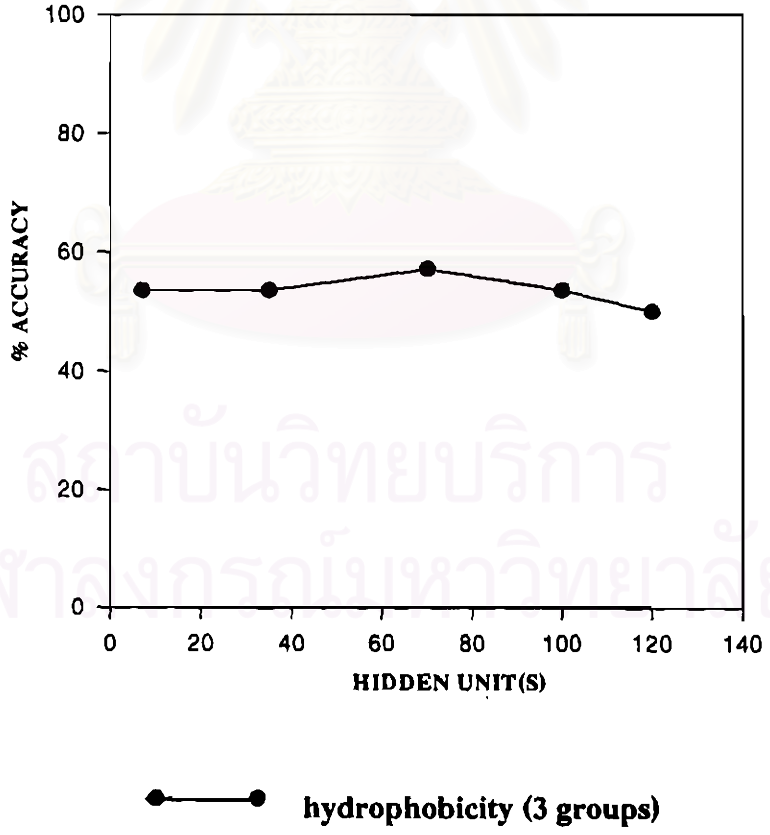

Figure 4.54 Percent accuracy prediction of percent helix structure (6 groups) using hydrophobicity (3 groups) as input vector with various numbers of hidden units (7, 35, 70, 100, 120).

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

Figure 4.55 shows that most of percent accuracy of 6 groups of percent helix structure were in the range between 35-60%. The highest percent accuracy was 70.71% (hy2gr%6h35.net). This network with 35 hidden units was trained from the hydropathy (2 groups). Apparently, this property also gave overall percent accuracy higher than those resulting from the others properties. The lowest percent accuracy was 28.57% (aa8gr%6h100.net). This network with 100 hidden units was trained from the amino acid side chain properties (8 groups). The eight properties also gave overall percent accuracy lower than those resulting from the others properties.

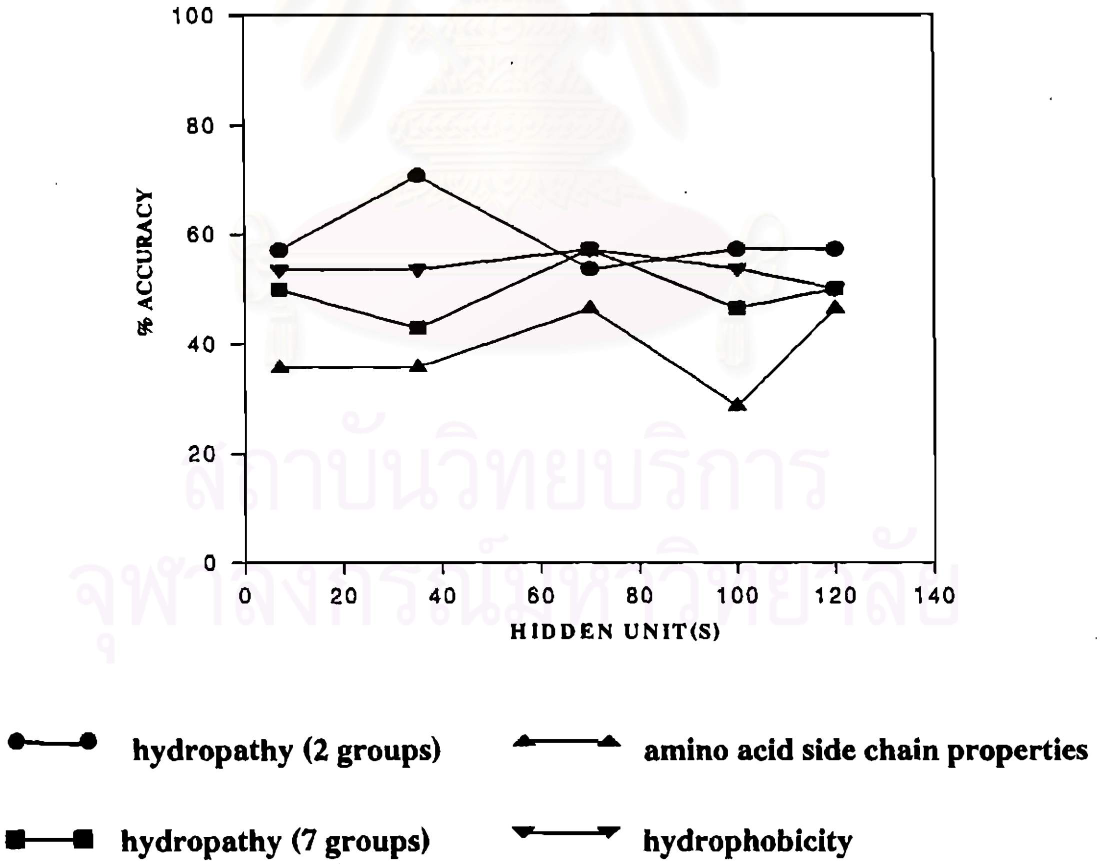

Figure 4.55 The comparison of percent accuracy prediction of percent helix structure (6 groups) using various types of input properties.

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

Training set.

Seventy patterns of amino acid sequences were saved as a learn pattern file for used in training (Table 4.26). The networks were trained for the prediction of six groups of percent sheet structure using amino acid properties in section 4.7.1, 4.7.2, 4.7.3 and 4.7.4

Testing set.

Twenty-eight patterns of amino acid sequences were saved as a test pattern file for used in testing (Table 4.26). The networks were tested for the prediction of six groups of percent sheet structure using amino acid properties in the section 4.7.1, 4.7.2, 4.7.3 and 4.7.4.

Input and output patterns for training and testing.

The input pattern was composed of 481 input units of amino acid sequences with properties coded amino acid residue as previously described (section 3.2.2).

The output pattern had 3 unit of percent sheet structure. The percent sheet structure in amino acid sequence was divided into 6 groups. There were five possible outputs representing the following ranges of percent values between 1-20%, 21-40%, 41-60%, 61-70%, 71-80%, 81-100% and the sixth group represent the percent value as 0%. The training set, testing set and the possible outputs were shown in Table 4.26.

A learn-pattern file with 70 patterns was used in training the networks with 7, 35, 70, 100 and 120 hidden units. The maximum value for learn cycles was set to 100000 and the maximum error to stop was set to 0.01. A test-pattern file with 28 patterns was used for testing the trained networks.

These training and testing set (Table 4.26) were the standard sets for prediction of percent sheet structure (6 groups) using various types of properties in the next section (section. 4.7.1, 4.7.2, 4.7.3 and 4.7.4).

Table 4.26 The training set, testing set and possible outputs of percent sheet structure (6 groups) prediction.

Training set (Input No.)*	**Testing set (Input No.)***	**Output patterns**	**% of sheet structure**
7, 13, 18, 23, 25, 28, 40, 48, 53, 59, 61, 70, 73, 77, 83	78, 84, 86, 87, 88, 94	0 0 0	0
1, 8, 9, 12, 14, 19, 24, 27, 29, 33, 37, 39, 49, 50, 51, 57, 60	65, 72, 74, 79, 90, 91, 95, 98	0 0 1	1-20
3, 6, 15, 20, 21, 26, 30, 31, 35, 36, 41, 42, 44, 46, 52, 55, 56, 63, 66, 67	68, 71, 75, 76, 80, 82, 85, 89, 92, 93	0 1 0	21-40
2, 5, 16, 17, 22, 32, 38, 43, 45, 47, 58	64, 81, 97	0 1 1	41-60
4, 10, 11, 34, 54, 69	96	1 0 0	61-80
62	-	1 0 1	81-100

* Names of input no. were listed in Table 3.1

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

The learn-pattern file, *thy2gr%6s.pat*, and the testing-pattern file, *tehy2gr%6s.pat*, which the amino acid sequence coded by hydropathy (2 groups) were used for training and testing respectively. Figure 4.43 shows the example of input and output patterns in these files. The trained network and testing result files are shown in Table 4.27 and the example of result patterns is shown in Figure 4.44. From Table 4.27 and Figure 4.56, the network with highest predictions accuracy (50.00%) was *hy2gr%6s35.net* which had 35 hidden units. While file *hy2gr%6s70.net*, with 70 hidden units gave only 46.43% accuracy. The networks with 100 and 120 hidden units were all 42.85%. The lowest percent accuracy (39.28%) resulted from the network with 7 hidden units.

Table 4.27 The trained network file, testing result files and percent accuracy of percent sheet structure (6 groups) prediction using hydropathy (2 groups).

Hidden units	No. of units	No. of links	Training network file	Testing result file	% accuracy of prediction
7	491	3388	hy2gr%6s7.net	tehy2gr%6s7.res	39.28
35	519	16940	hy2gr%6s35.net	tehy2gr%6s35.res	50.00
70	554	33880	hy2gr%6s70.net	tehy2gr%6s70.res	46.43
100	584	48400	hy2gr%6s100.net	tehy2gr%6s100.res	42.85
120	604	58080	hy2gr%6s120.net	tehy2gr%6s120.res	42.85

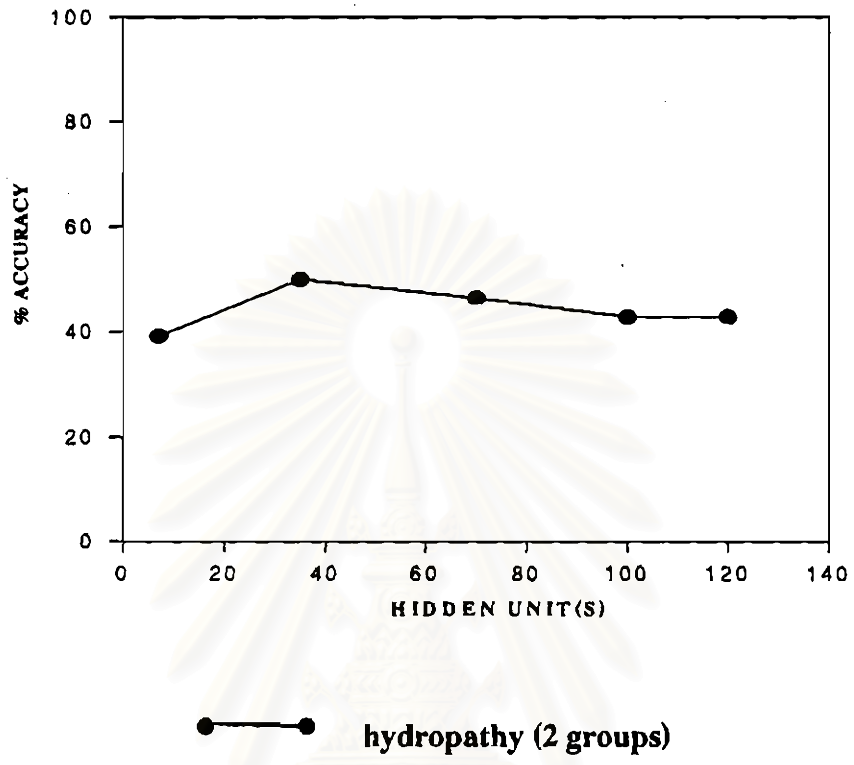

Figure 4.56 Percent accuracy prediction of percent sheet structure (6 groups) using 2 groups of hydropathy as input vector with various numbers of hidden units (7, 35, 70, 100, 120).

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

The learn-pattern file, *thy7gr%6s.pat*, and the testing-pattern file, *tehy7gr%6s.pat*, with hydropathy (7 groups) coded the amino acid sequences were used for training and testing respectively. Figure 4.46 shows the example of input and output patterns. The trained network and testing result files are shown in Table 4.28 and the example of result patterns is shown in Figure 4.47. From Table 4.28 and Figure 4.57, the network with highest predictions accuracy (50.00%) was *hy7gr%6s70.net* which had 70 hidden units. While, the networks with 7, 35 and 100 hidden unit were all 46.43%. The lowest percent accuracy (42.85%) was the result from the network with 120 hidden units (*hy7gr%6s120.net*).

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

Hidden units	No. of units	No. of links	Training network file	Testing result file	% accuracy of prediction
7	491	3388	hy7gr%6s7.net	tehy7gr%6s7.res	46.43
35	519	16940	hy7gr%6s35.net	tehy7gr%6s35.res	46.43
70	554	33880	hy7gr%6s70.net	tehy7gr%6s70.res	50.00
100	584	48400	hy7gr%6s100.net	tehy7gr%6s100.res	46.43
120	604	58080	hy7gr%6s120.net	tehy7gr%6s120.res	42.85

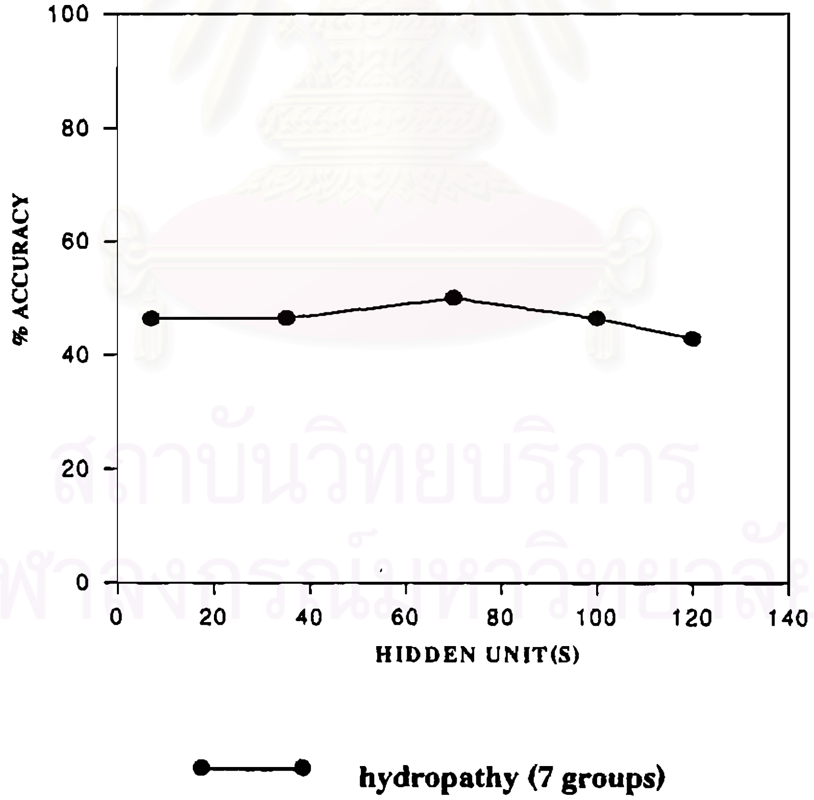

Figure 4.57 Percent accuracy prediction of percent sheet structure (6 groups) using 7 groups of hydropathy as input vector with various numbers of hidden units (7, 35, 70, 100, 120).

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

The learn-pattern file, *taa8gr%6s.pat*, and the testing-pattern file, *teaa8gr%6s.pat*, with amino acid side chain properties (8 groups) coded the amino acid sequences were used for training and testing respectively. Figure 4.49 shows the example of input and output patterns. The trained network files and testing result files are shown in Table 4.29 and the example of result patterns is shown in Figure 4.50. From Table 4.29 and Figure 4.58, the network with highest predictions accuracy (39.29%) is *aa8gr%6s120.net* (120 hidden units). Others networks with 7, 35 and 70 hidden units were all 35.71% accuracy. The lowest percent accuracy (28.57%) was the result from the network with 100 hidden units.

Table 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).

Hidden units	No. of units	No. of links	Training network file	Testing result file	% accuracy of prediction
7	491	3388	aa8gr%6s7.net	teaa8gr%6s7.res	35.71
35	519	16940	aa8gr%6s35.net	teaa8gr%6s35.res	35.71
70	554	33880	aa8gr%6s70.net	teaa8gr%6s70.res	35.71
100	584	48400	aa8gr%6s100.net	teaa8gr%6s100.res	28.57
120	604	58080	aa8gr%6s120.net	teaa8gr%6s120.res	39.28

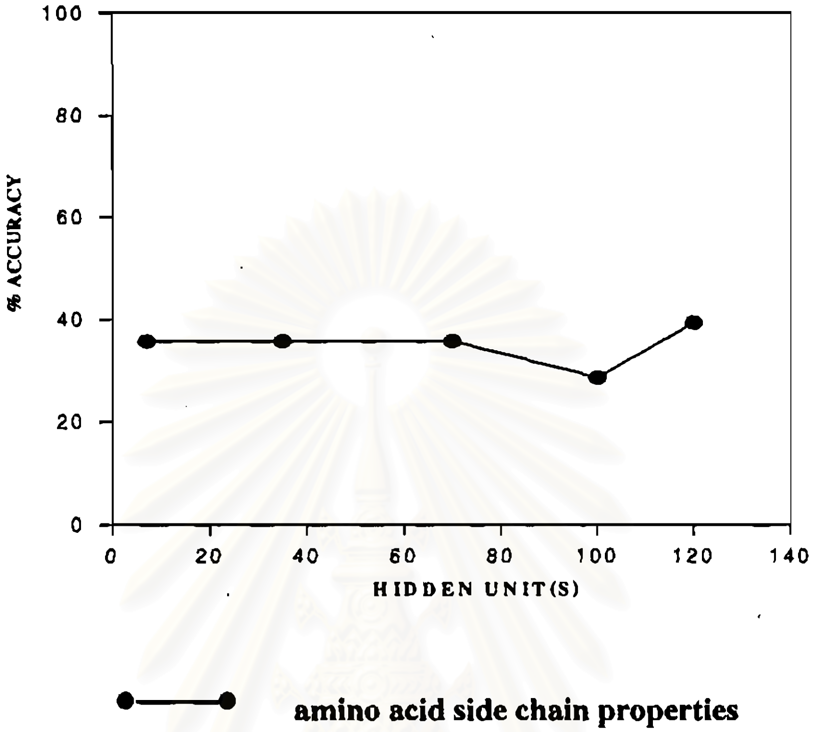

Figure 4.58 Percent accuracy prediction of percent sheet structure (6 groups) using 8 groups of amino acid side chain properties as input vector with various numbers of hidden units (7, 35, 70, 100, 120).

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

The learn-pattern file, *thydrogr%6s.pat*, and the testing-pattern file, *tehydrogr%6s.pat*, with hydrophobicity (3 groups) coded the amino acid sequences were used for training and testing respectively. Figure 4.52 shows the example of input and output patterns. The trained network and testing result files are shown in Table 4.30 and the example of result patterns is shown in Figure 4.53. From Table 4.30 and Figure 4.59, predictions with highest accuracy (35.71%) were the results of *hydrogr%6s35.net* and *hydro%6s70.net* which had 35 and 70 hidden units respectively. Whereas the networks with 100 and 120 hidden unit were all 32.14%. The lowest percent accuracy (28.57%) was the result from the network with 7 hidden units.

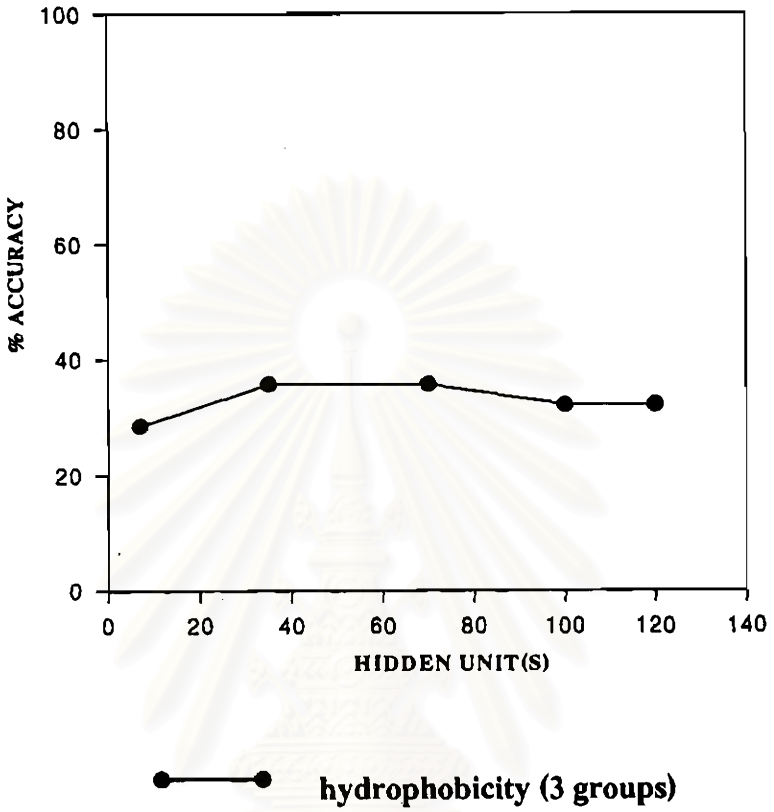

Figure 4.59 Percent accuracy prediction of percent sheet structure (6 groups) using hydrophobicity (3 groups) as input vector with various numbers of hidden units (7, 35, 70, 100, 120).

Table 4.30 The trained network file, testing result file and percent accuracy of percent sheet structure (6 groups) prediction using hydrophobicity (3 groups).

Hidden units	No. of units	No. of links	Training network file	Testing result file	% accuracy of prediction
7	491	3388	hydrogr%6s7.net	tehydrogr%6s7.res	28.57
35	519	16940	hydrogr%6s35.net	tehydrogr%6s35.res	35.71
70	554	33880	hydrogr%6s70.net	tehydrogr%6s70.res	35.71
100	584	48400	hydrogr%6s100.net	tehydrogr%6s100.res	32.14
120	604	58080	hydrogr%6s120.net	tehydrogr%6s120.res	32.14

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

Figure 4.60 shows that the range of percent accurate for percent sheet structure prediction are between 28 -50%. The highest percent accuracy was 50.00% which result from *hy2gr%6s35.net* and *hy7gr%6s70.net*. These networks with 35 and 70 hidden units were trained from the hydropathy (2 groups) and hydropathy (7 groups) respectively. The lowest percent accuracy was 28.57% which resulted from *aa8gr%6s100.net* and *hydro%6s7.net*. These networks with 100 and 7 hidden units were trained from the amino acid side chains properties (8 groups) and hydrophobicity.

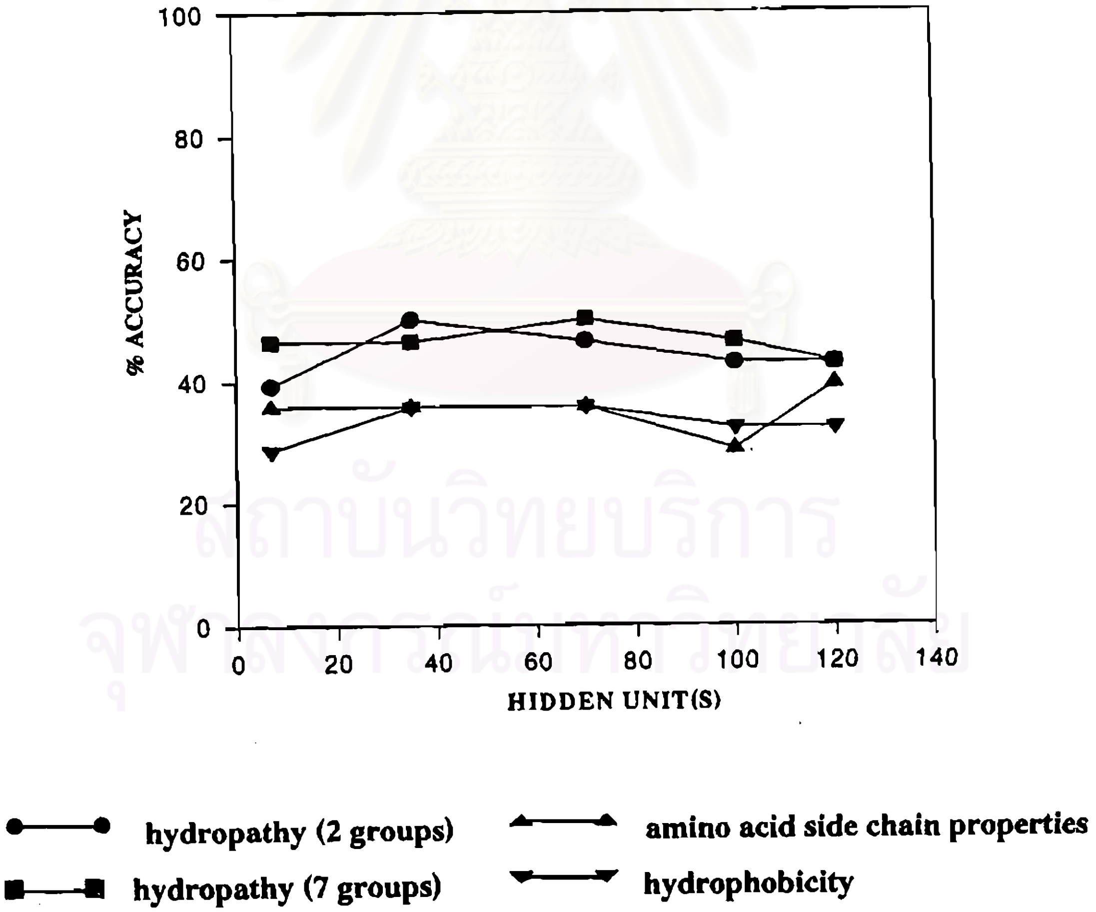

Figure 4.60 The comparison of percent accuracy prediction of percent sheet structure (6 groups) using various types of input properties.

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

Training set.

Seventy patterns of amino acid sequences were saved as a learn pattern file for used in training (Table 4.31). The networks were trained for the prediction of six groups of percent turn structure using amino acid properties in the section 4.8.1, 4.8.2, 4.8.3 and 4.8.4

Testing set.

Twenty-eight patterns of amino acid sequences were saved as a test pattern file for used in testing (Table 4.31). The network were tested for the prediction of six group of percent turn structure using amino acid properties in the section 4.8.1, 4.8.2, 4.8.3 and 4.8.4.

Input and output patterns for training and testing.

The input pattern was composed of 481 input units of amino acid which were coded with properties of amino acid residue as previously described (section 3.2.2).

The output pattern had 3 unit of percent turn structure. The percent turn structure in amino acid sequence was divided into 6 groups. There were five possible outputs represented the percent value between 1-20%, 21-40%, 41-60%, 61-70%, 71-80%, 81-100% and the sixth group represent the percent value as 0%. The training set, testing set and the possible outputs were shown in Table 4.26.

A learn-pattern file with 70 patterns was used in training the networks with various hidden units, 7, 35, 70, 100 and 120. The maximum value for learn cycles was set to 100000 and the maximum error to stop was set to 0.01. A test-pattern file with 28 patterns was used for testing the trained networks.

These training and testing set (Table 4.31) were the standard sets for prediction of percent turn structure (6 groups) using various types of properties in the next section (section. 4.8.1, 4.8.2, 4.8.3 and 4.8.4).

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

Training set (Input No.)*	**Testing set (Input No.)***	**Output patterns**	**% of turn structure**
1, 4, 5, 7, 8, 9, 13, 15, 18, 25, 26, 27, 33, 34, 37, 38, 39, 41, 43, 45, 46, 48, 50, 52, 55, 57	59, 67, 69, 73, 75, 76, 77, 78, 83, 87, 88, 89, 91, 93	0 0 0	0
3, 10, 11, 24, 30, 35, 40, 42, 47, 49, 51, 54, 56, 60, 71, 74, 81	82, 84, 86, 94, 95, 97	0 0 1	1-20
2, 6, 12, 14, 17, 19, 20, 21, 28, 29, 31, 32, 36, 44, 53, 58, 61, 63, 64	66, 70, 72, 80, 92, 96, 98	0 1 0	21-40
16, 22, 23, 62, 65, 79, 85	90	0 1 1	41-60
68	-	1 0 0	61-80
-	-	1 0 1	81-100

* Names of input no. were listed in Table 3.1

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

The learn-pattern file, *thy2gr%6t.pat*, and the testing-pattern file, *tehy2gr%6t.pat*, with hydropathy (2 groups) coded the amino acid sequences were used for training and testing respectively. Figure 4.43 shows the example of input and output patterns. The trained network and testing result files are shown in Table 4.32 and the example of result patterns is shown in Figure 4.44. From Table 4.32 and Figure 4.61, the network with highest predictions accuracy (32.14%) was *hy2gr%6t35.net* which had 35 hidden units. Whereas the *hy2gr%6t7.net* with 7 hidden units gave 28.57% accuracy. Others networks with 70, 100 and 120 hidden units were all 25.00% accuracy.

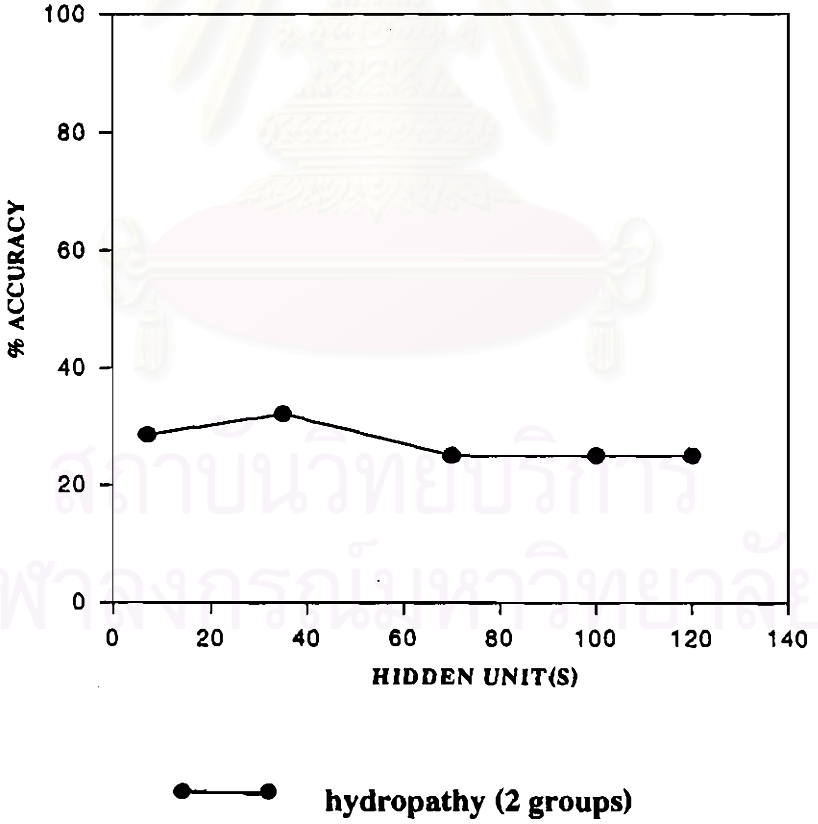

Figure 4.61 Percent accuracy prediction of percent turn structure (6 groups) using 2 groups of hydropathy as input vector with various numbers of hidden units (7, 35, 70, 100, 120).

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

Hidden units	No. of units	No. of links	Training network file	Testing result file	% accuracy of prediction
7	491	3388	hy2gr%6t7.net	tehy2gr%6t7.res	28.57
35	519	16940	hy2gr%6t35.net	tehy2gr%6t35.res	32.14
70	554	33880	hy2gr%6t70.net	tehy2gr%6t70.res	25.00
100	584	48400	hy2gr%6t100.net	tehy2gr%6t100.res	25.00
120	604	58080	hy2gr%6t120.net	tehy2gr%6t120.res	25.00

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

The learn-pattern file, *thy7gr%6t.pat,* and testing-pattern file, *tehy7gr%6t.pat*, with hydropathy (7 groups) coded the amino acid sequences were used for training and testing respectively. Figure 4.46 shows the example of input and output patterns. The trained network and testing result files are shown in Table 4.33 and the example of result patterns is shown in Figure 4.47. From Table 4.33 and Figure 4.62, predictions with highest accuracy (35.71%) were the results of *hy7gr%6t7.net hy7gr%6t35.net* and *hy7gr%6t70.net* which had 7, 35 and 70 hidden units respectively. Whereas, others networks with 100 and 120 hidden units were all 28.57%.

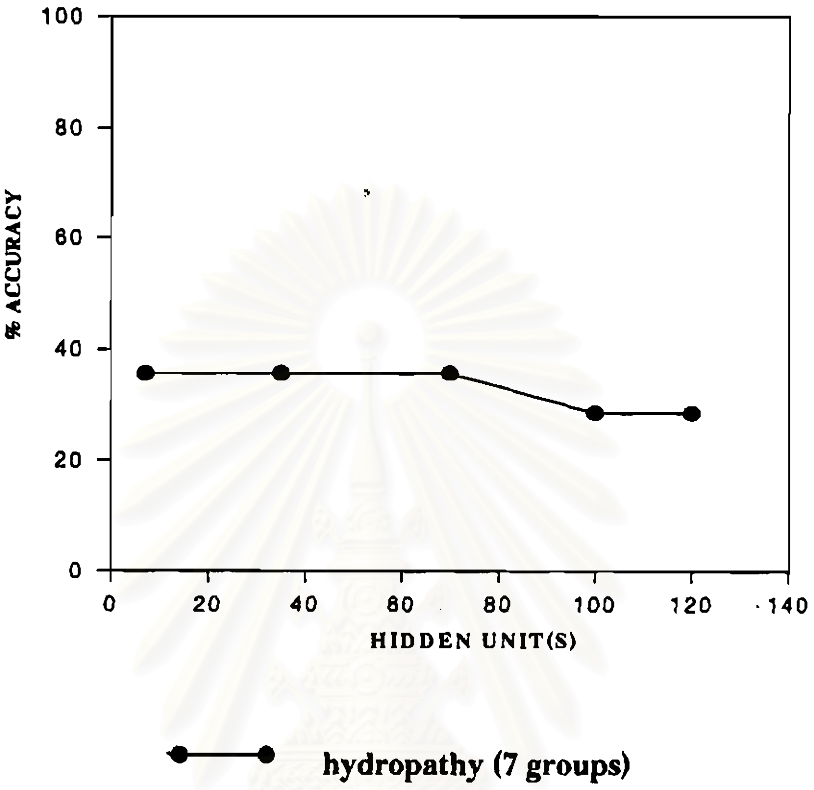

Figure 4.62 Percent accuracy prediction of percent turn structure (6 groups) using 7 groups of hydropathy as input vector with various numbers of hidden units (7, 35, 70, 100, 120).

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

Hidden units	No. of units	No. of links	Training network file	Testing result file	% accuracy of prediction
7	491	3388	hy7gr%6t7.net	tehy7gr%6t7.res	35.71
35	519	16940	hy7gr%6t35.net	tehy7gr%6t35.res	35.71
70	554	33880	hy7gr%6t70.net	tehy7gr%6t70.res	35.71
100	584	48400	hy7gr%6t100.net	tehy7gr%6t100.res	28.57
120	604	58080	hy7gr%6t120.net	tehy7gr%6t120.res	28.57

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

The learn-pattern file, *taa8gr%6t.pat*, and the testing-pattern file, *teaa8gr%6t.pat*, with amino acid side chain properties (8 groups) coded the amino acid sequences were used for training and testing respectively. Figure 4.49 shows the example of input and output patterns. The trained network and testing result files are shown in Table 4.34 and the example of result patterns is shown in Figure 4.50. From Table 4.34 and Figure 4.63, predictions with highest accuracy (46.43%) were the results of *aa8gr%6t35.net* and *aa8gr%6t100.net* which had 35 and 100 hidden units respectively. Whereas, the networks with 70 and 120 hidden units were all 39.29% accuracy. The lowest percent accuracy (32.14%) was the result from the network with 7 hidden units.

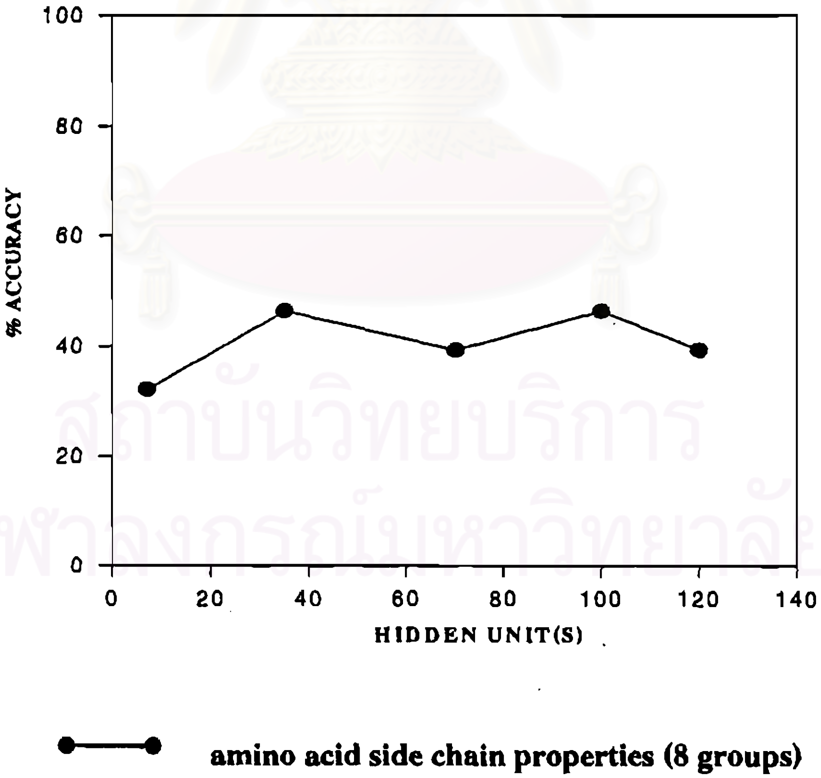

Figure 4.63 Percent accuracy prediction of percent turn structure (6 groups) using 8 groups of amino acid side chain properties as input vector with various numbers of hidden units (7, 35, 70, 100, 120).

Table 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).

Hidden units	No. of units	No. of links	Training network file	Testing result file	% accuracy of prediction
7	491	3388	aa8gr%6t7.net	teaa8gr%6t7.res	32.14
35	519	16940	aa8gr%6t35.net	teaa8gr%6t35.res	46.43
70	554	33880	aa8gr%6t70.net	teaa8gr%6t70.res	39.29
100	584	48400	aa8gr%6t100.net	teaa8gr%6t100.res	46.43
120	604	58080	aa8gr%6t120.net	teaa8gr%6t120.res	39.29

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

The learn-pattern file, *thydrogr%6t.pat*, and the testing-pattern file, *tehydrogr%6t.pat*, with hydrophobicity (3 groups) coded the amino acid sequence were used for training and testing respectively. Figure 4.52 shows the example of input and output patterns in these files. The trained network and testing result files are shown in Table 4.35 and the example of result patterns is shown in Figure 4.53. From Table 4.35 and Figure 4.64, predictions with highest accuracy (42.86%) were the results of *hydrogr%6t7.net* and *hydro%6t35.net* which had 7 and 35 hidden units respectively. The networks with 70 and 120 hidden units were all 35.71% accurate. Whereas the network with 100 hidden units was 39.71% accurate.

Table 4.35 The trained network file, testing result file and percent accuracy of percent turn structure (6 groups) prediction using hydrophobicity (3 groups).

Hidden units	No. of units	No. of links	Training network file	Testing result file	% accuracy of prediction
7	491	3388	hydrogr%6t7.net	tehydrogr%6t7.res	42.86
35	519	16940	hydrogr%6t35.net	tehydrogr%6t35.res	42.86
70	554	33880	hydrogr%6t70.net	tehydrogr%6t70.res	35.71
100	584	48400	hydrogr%6t100.net	tehydrogr%6t100.res	39.29
120	604	58080	hydrogr%6t120.net	tehydrogr%6t120.res	35.71

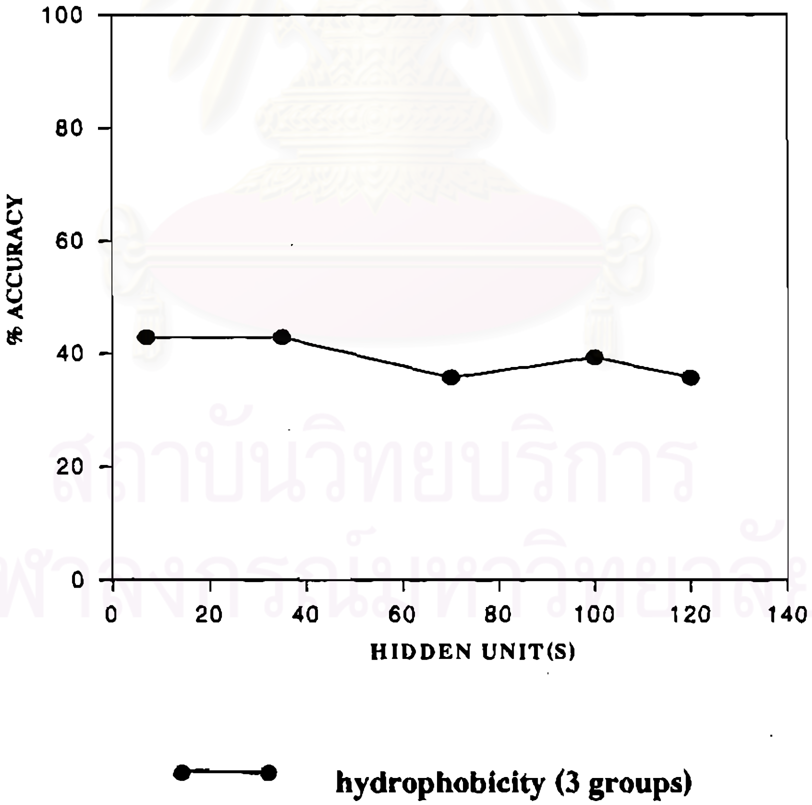

Figure 4.64 Percent accuracy prediction of percent turn structure (6 groups) using 3 groups of hydrophobicity as input vector with various numbers of hidden units (7, 35, 70, 100, 120)

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

Figure 4.65 shows that the range of percent accurate for percent turn structure prediction were between 25-50%. The highest percent accuracy was 46.43% which resulted from *aa8gr%6t35.net* and *aa8grgr%6t100.net*. These networks with 35 and 70 hidden units were trained from the amino acid side chain properties. The lowest percent accuracy was 25% which result from *hy2gr%6t70.net*, *hy2gr%6t100.net* and *hy2gr%6t120.net*. These networks, with 70, 100 and 120 hidden units, were trained from hydrophobicity.

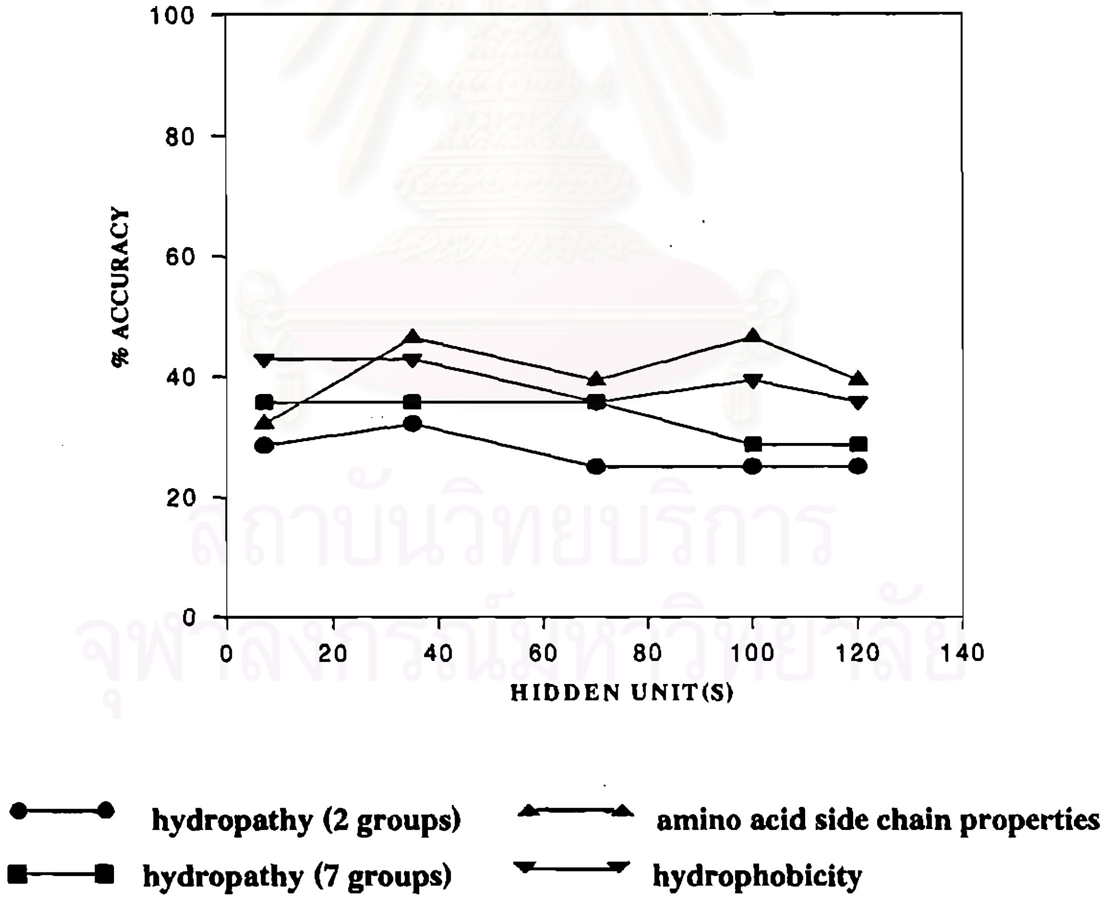

Figure 4.65 The comparison of percent accuracy prediction of percent turn structure (6 groups) using various types of input properties.

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

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

From Figure 4.66, the hydropathy (2 groups) gave the overall percent accuracy for the prediction of percent helix structure higher than the others structures.The range of percent accuracy for percent helix prediction was between 50-70%. While, the range of percent accuracy for the prediction of sheet structure was between 40-50%. The range of percent accuracy for turn structure prediction was between 25-35%.

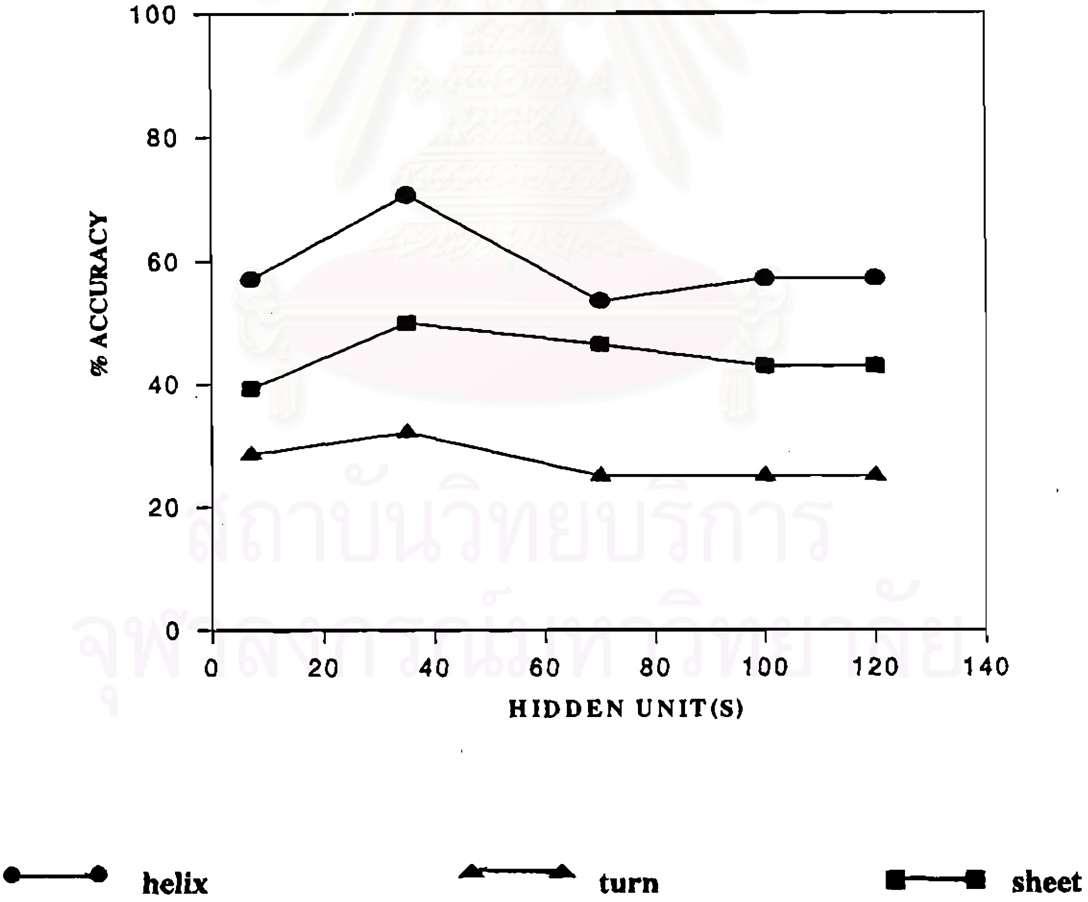

Figure 4.66 The percent accuracy predictions of the percent of helix, sheet, turn (6 groups) in separate networks using hydropathy (2 groups).

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

From Figure 4.67, the hydropathy (7 groups) gave the percent accuracy for helix and sheet structures in the same range, 40-50%. While the range of percent accuracy for turn structure prediction was between 25-35% when using this property.

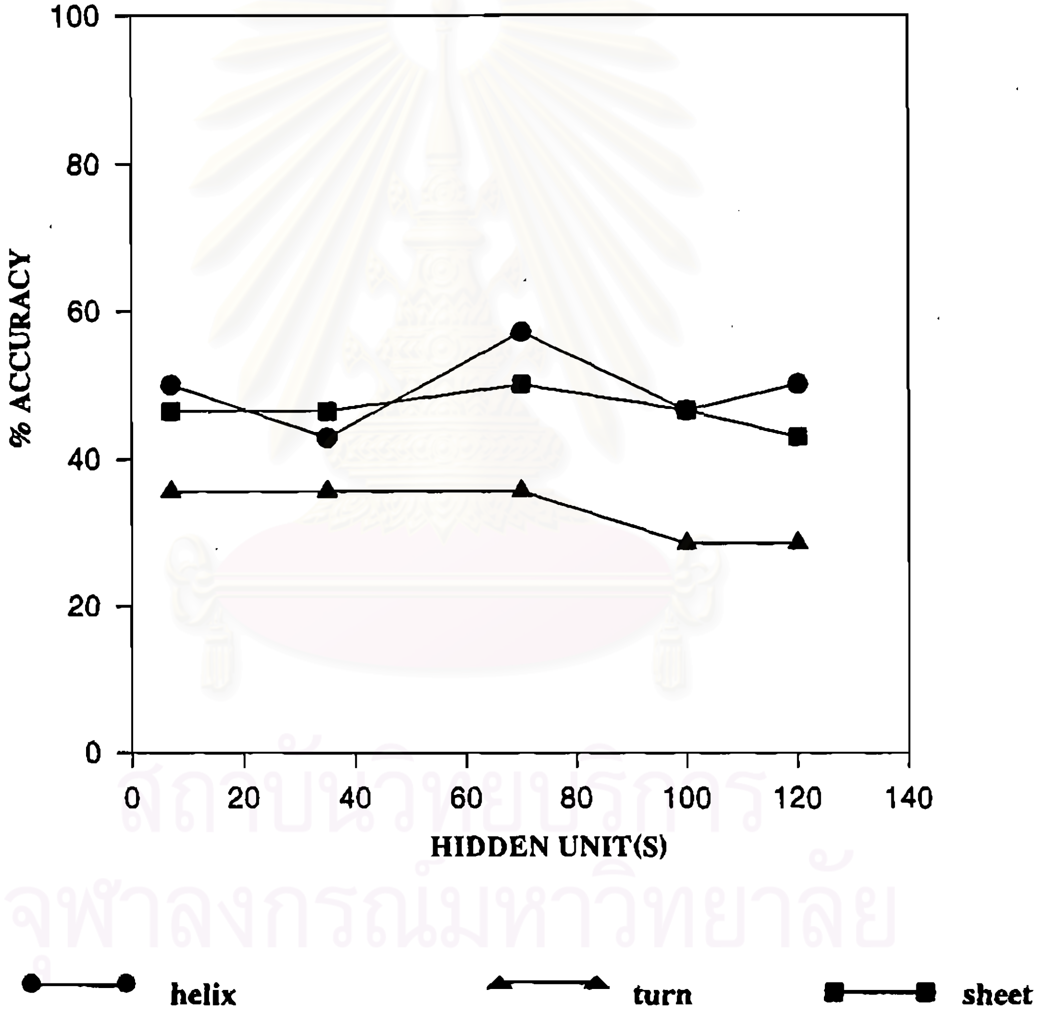

Figure 4.67 The percent accuracy predictions of the percent of helix, sheet, turn (6 groups) in separate networks using hydropathy (7 groups).

4.9.3 Percent helix, sheet and turn structures predictions using amino acid side chain properties

From Figure 4.68, the range of percent accuracy for prediction of helix, sheet and turn structures was the same when using this amino acid side chain properties as input vector. The range was between 25-50%

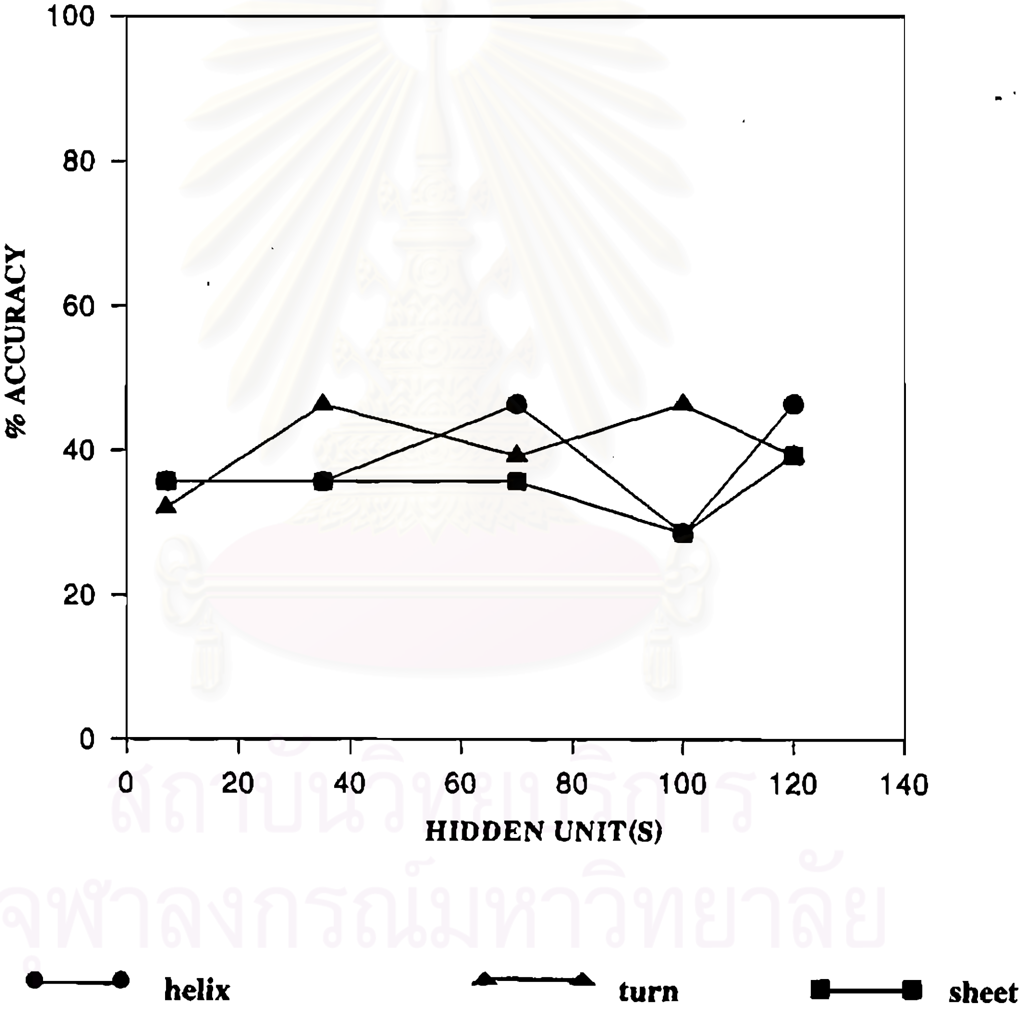

Figure 4.68 The percent accuracy predictions of the percent of helix, sheet, turn (6 groups) in separate networks using amino acid side chain properties (8 groups).

4.9.4 Percent helix, sheet and turn structure prediction using hydrophobicity

From Figure 4.69, the hydrphobicity gave highest percent accuracy for percent helix structure prediction. The range of percent accuracy of this helix prediction was between 50-55%. While the range of percent accuracy fot percent sheet and turn structures predictions was between 25-45%.

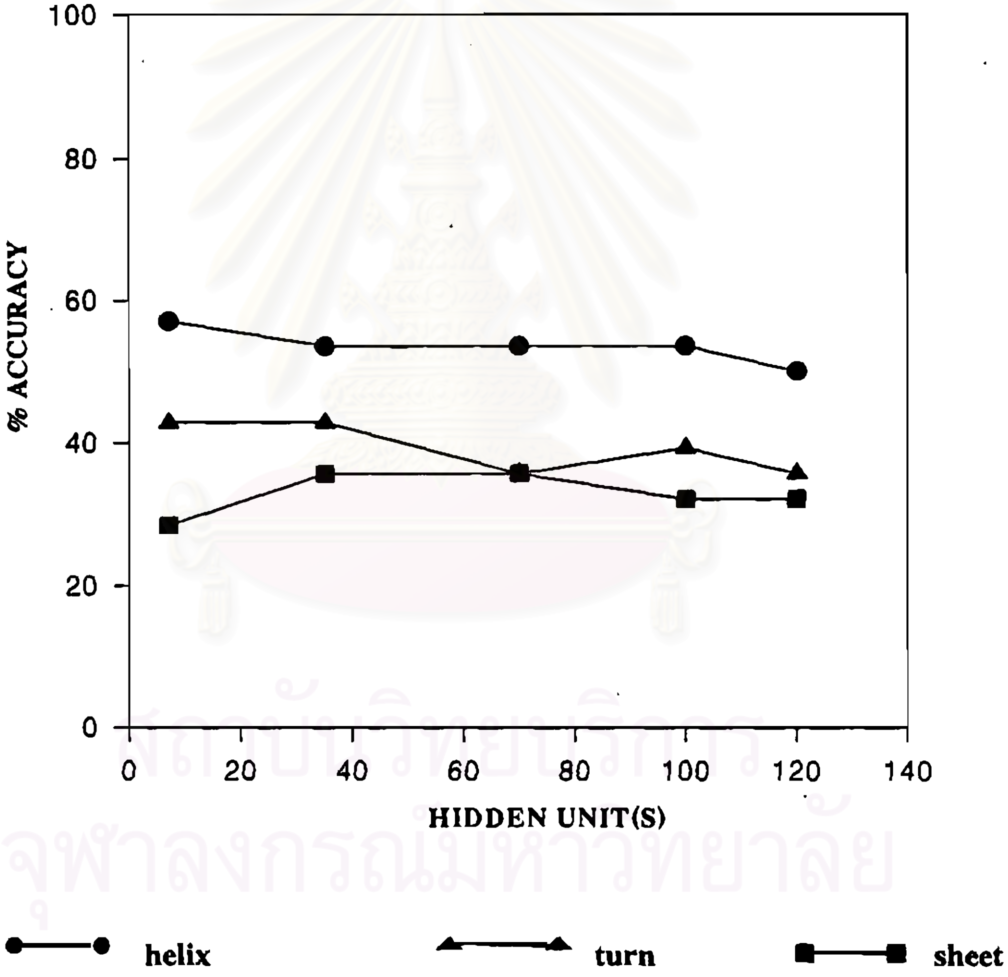

Figure 4.69 The percent accuracy predictions of the percent of helix, sheet, turn (6 groups) in separate networks using hydrophobicity.

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

Training set.

Seventy patterns of amino acid sequences were saved as a learn pattern file for used in training (Table 4.36). The network were trained for the prediction of three groups of percent helix structure using amino acid properties in the section 4.10.1, 4.10.2, 4.10.3 and 4.10.4

Testing set.

Twenty-eight patterns of amino acid sequences were saved as a test pattern file for used in testing (Table 4.36). The network were tested for the prediction of three groups of percent helix structure using amino acid properties in the section 4.10.1, 4.10.2, 4.10.3 and 4.10.4.

Input and output patterns for training and testing.

The input pattern was composed of 481 input units of amino acid which were coded with properties of amino acid residue as previously described (section 3.2.2).

The output pattern had 3 units of percent helix structure. The percent helix structure in amino acid sequence was divided into 3 groups. There were 2 possible outputs representing the following ranges of percent values 1-50%, 51-100%, and the third group represented 0%. The training set, testing set and the possible outputs are shown in Table 4.36.

A learn-pattern file with 70 patterns was used in training the networks with 7, 35, 70, 100 and 120 hidden units. The maximum value for learn cycles was set to 100000 and the maximum error to stop was set to 0.01. A test-pattern file with 28 patterns was used for testing the trained networks.

These training and testing sets (Table 4.36) were the standard sets for prediction of percent helix structure (3 groups) using various types of properties in the next section (section. 4.10.1, 4.10.2, 4.10.3 and 4.10.4).

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

Training set (Input No.)*	Testing set (Input No.)*	Output patterns	% of helix structure
4, 6, 11, 43, 64	68	0 0	0
1, 3, 5, 89, 12, 15, 16, 17, 19, 20, 21, 22, 24, 26, 27, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 41, 44, 45, 46, 47, 50, 51, 52, 54, 55, 56, 57, 58, 62, 63, 65, 66, 67, 69, 70, 71	72, 74, 75, 76, 79, 80, 81, 82, 84, 85, 87, 88, 89, 90, 91, 92, 93, 95, 96, 97, 98	0 1	1-50
2, 7, 10, 13, 14, 18, 23, 25, 28, 40, 42, 48, 49, 53, 59, 60, 61	73, 77, 78, 83, 86, 94	1 0	51-100

Names of input no. were listed in Table 3.1

```
SNNS pattern definition file V1.4
No. of patterns         : 70
No. of input units      : 481
No. of output units     : 2

#input1 1alc
       0.2   0.2   0.1   0.2   0.2   0.1   0.2   0.1   0.2   0.2   0.2   0.1   0.2
       0.2   0.1   0.2   0.2   0.2   0.2   0.2   0.1   0.1   0.1   0.2   0.2   0.1
       0.1   0.1   0.2   0.1   0.1   0.2   0.2   0.2   0.2   0.2   0.2   0.2   0.2
       0.1   0.1   0.1   0.2   0.2   0.2   0.2   0.2   0.2   0.2   0.2   0.2   0.1
       0.1   0.2   0.1   0.2   0.2   0.1   0.1   0.2   0.1   0.2   0.2   0.2   0.2
       0.2   0.2   0.2   0.2   0.2   0.2   0.1   0.1   0.2   0.1   0.2   0.1   0.2
       0.2   0.1   0.1   0.2   0.2   0.2   0.1   0.2   0.2   0.2   0.1   0.1   0.1
       0.1   0.2   0.2   0.1   0.1   0.2   0.1   0.2   0.2   0.1   0.2   0.2   0.2
       0.1   0.1   0.2   0.2   0.1   0.1   0.1   0.2   0.2   0.2   0.1   0.2   0.2
       0.2   0.1   0.1   0.2   0.2   0.2   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
#output1 1alc
     0     1

#input2 1bp2
       0.1   0.1   0.2   0.2   0.1   0.2   0.2   0.1   0.1   0.2   0.1   0.2   0.1
       0.2   0.2   0.2   0.2   0.2   0.1   0.1   0.2   0.1   0.2   0.2   0.2   0.2
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
#outut2 1bp2
       ...   ....  ....

#input70 8adh

       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
```

Figure 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).

```
SNNS result file V1.4-3D
No. of hidden units : 70
No. of patterns     : 28
No. of input units  : 481
No. of output units : 2
startpattern        : 1
endpattern          : 28
input patterns included
teaching outputs included

#input1
       0.2   0.2   0.1   0.2   0.2   0.1   0.2   0.1   0.2   0.2   0.2   0.1   0.2
       0.2   0.1   0.2   0.2   0.2   0.2   0.2   0.1   0.1   0.1   0.2   0.2   0.1
       0.1   0.1   0.2   0.1   0.1   0.2   0.2   0.2   0.2   0.2   0.2   0.2   0.2
       0.1   0.1   0.1   0.2   0.2   0.2   0.2   0.2   0.2   0.2   0.2   0.2   0.1
       0.1   0.2   0.1   0.2   0.2   0.1   0.1   0.2   0.1   0.2   0.2   0.2   0.2
       0.2   0.2   0.2   0.2   0.2   0.2   0.1   0.1   0.2   0.1   0.2   0.1   0.2
       0.2   0.1   0.1   0.2   0.2   0.2   0.1   0.2   0.2   0.2   0.1   0.1   0.1
       0.1   0.2   0.2   0.1   0.1   0.2   0.1   0.2   0.2   0.1   0.2   0.2   0.2
       0.1   0.1   0.2   0.2   0.1   0.1   0.1   0.2   0.2   0.2   0.1   0.2   0.2
       0.2   0.1   0.1   0.2   0.2   0.2   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
#output1
   0         1
 0.0002    0.9554

#input2
       0.1   0.1   0.2   0.2   0.1   0.2   0.2   0.1   0.1   0.2   0.1   0.2   0.1
       0.2   0.2   0.2   0.2   0.2   0.1   0.1   0.2   0.1   0.2   0.2   0.2   0.2
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
#output
  ....   ....   ....
  ....   ....   ....

#input70
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
```

Figure 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). The predicted output was the floating numbers, 0.0002 and 0.9554

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

The learn-pattern file, *thy2gr%3h.pat*, and testing-pattern file, *tehy2gr%3h.pat*, with hydropathy (2 groups) coded amino acid sequences were used for training and testing respectively. Figure 4.70 shows the example of input and output patterns. The trained network and testing result files are shown in Table 4.37 and the example of result patterns is shown in Figure 4.71. Table 4.37 and Figure 4.72, the predictions from all network in this group are all 71.43% accuracy.

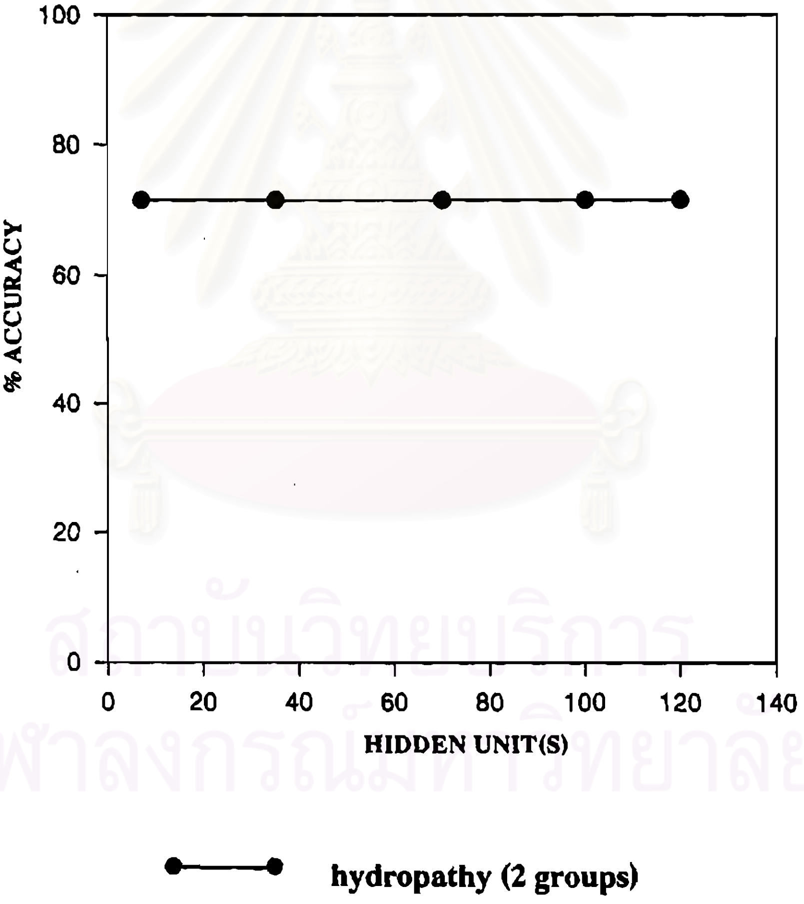

Figure 4.72 Percent accuracy prediction of percent helix structure (3 groups) using 2 groups of hydropathy as input vector with various numbers of hidden units (7, 35, 70, 100, 120).

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

Hidden units	No. of units	No. of links	Training network file	Testing result file	% accuracy of prediction
7	490	3381	hy2gr%3h7.net	tehy2gr%3h7.res	71.43
35	518	16950	hy2gr%3h35.net	tehy2gr%3h35.res	71.43
70	553	33810	hy2gr%3h70.net	tehy2gr%3h70.res	71.43
100	583	48300	hy2gr%3h100.net	tehy2gr%3h100.res	71.43
120	603	57960	hy2gr%3h120.net	tehy2gr%3h120.res	71.43

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

The learn-pattern file, *thy7gr%3h.pat*, and test-pattern file, *tehy7gr%3h.pat*, with hydropathy (7 groups) coded the amino acid sequences were used in training and testing with various hidden units. Figure 4.73 shows the example of input and output patterns. The trained network and testing result file are shown in Table 4.38 and the example of result patterns is in Figure 4.74. From Table 4.38 and Figure 4.75 show that *hy7gr%3h100.net*, which had 100 hidden units, gave highest percent accuracy of 71.43%. While files, *hy7gr%3h7.net, hy7gr%3h35.net* and *hy7gr%3h70.net* were all 67.8% accurate. The lowest percent accuracy (64.29%) resulted from the network with 120 hidden units.

```
SNNS pattern definition file V1.4

No. of patterns          : 70
No. of input units       : 481
No. of output units      : 2

#input1 1a1c
              0.7   0.6   0.2   0.4   0.7   0.2   0.6   0.1   0.4   0.6   0.6   0.1   0.5
              0.6   0.1   0.6   0.4   0.5   0.4   0.7   0.1   0.3   0.1   0.5   0.6   0.1
              0.1   0.2   0.4   0.3   0.2   0.6   0.4   0.4   0.4   0.5   0.6   0.4   0.6
              0.3   0.1   0.1   0.6   0.6   0.6   0.6   0.4   0.4   0.6   0.5   0.4   0.1
              0.2   0.6   0.1   0.4   0.6   0.3   0.1   0.4   0.2   0.7   0.4   0.4   0.6
              0.4   0.5   0.6   0.4   0.7   0.6   0.1   0.2   0.6   0.1   0.4   0.2   0.6
              0.7   0.2   0.1   0.6   0.6   0.6   0.1   0.4   0.6   0.6   0.1   0.3   0.2
              0.3   0.7   0.7   0.1   0.1   0.6   0.1   0.7   0.4   0.1   0.6   0.5   0.4
              0.1   0.3   0.6   0.7   0.3   0.1   0.2   0.4   0.6   0.7   0.1   0.6   0.6
              0.4   0.1   0.2   0.6   0.7   0.6   0.0   0.0   0.0   0.0   0.0   0.0   0.0
              0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
              0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
              0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
              0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
              0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
              0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
              0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
              0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
              0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
              0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
              0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
              0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
              0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
              0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
              0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
              0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
              0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
              0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
              0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
              0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
              0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
              0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
              0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
              0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
              0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
              0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
              0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
#output1 1a1c
      0   1

#input2
              0.3   0.1   0.4   0.6   0.2   0.6   0.4   0.3   0.1   0.7   0.2   0.7   0.1
              0.5   0.4   0.4   0.6   0.5   0.1   0.1   0.6   0.2   0.6   0.6   0.5   0.4
              0.2   0.5   0.2   0.4   0.1   0.4   0.4   0.4   0.4   0.4   0.5   0.1   0.6
              ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
              ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
#outut2 1bp2
      ...   ....   ....

#input70 8adh

              ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
              ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
              ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
```

Figure 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).

```
SNNS result file V1.4-3D

No. of hidden units : 70
No. of patterns     : 28
No. of input units  : 481
No. of output units : 2
startpattern        : 1
endpattern          : 28
input patterns included
teaching outputs included

#input1
       0.4   0.6   0.6   0.4   0.4   0.5   0.4   0.7   0.6   0.6   0.4   0.5   0.6
       0.6   0.4   0.6   0.7   0.6   0.2   0.5   0.1   0.3   0.7   0.4   0.6   0.7
       0.6   0.4   0.5   0.1   0.6   0.1   0.6   0.4   0.6   0.4   0.3   0.7   0.5
       0.6   0.5   0.4   0.1   0.7   0.5   0.1   0.4   0.1   0.4   0.5   0.6   0.6
       0.3   0.4   0.4   0.1   0.7   0.1   0.1   0.6   0.6   0.4   0.6   0.3   0.2
       0.6   0.1   0.6   0.2   0.6   0.6   0.4   0.6   0.6   0.7   0.3   0.1   0.1
       0.7   0.4   0.4   0.5   0.1   0.6   0.4   0.4   0.5   0.7   0.1   0.1   0.6
       0.2   0.6   0.2   0.6   0.4   0.4   0.4   0.1   0.6   0.4   0.6   0.4   0.4
       0.6   0.6   0.4   0.1   0.6   0.7   0.7   0.7   0.5   0.3   0.3   0.6   0.1
       0.6   0.1   0.1   0.6   0.4   0.6   0.4   0.7   0.5   0.4   0.6   0.2   0.4
       0.7   0.3   0.1   0.6   0.6   0.5   0.6   0.4   0.1   0.3   0.1   0.1   0.4
       0.1   0.2   0.1   0.7   0.1   0.4   0.4   0.3   0.7   0.5   0.4   0.1   0.6
       0.7   0.1   0.1   0.6   0.1   0.1   0.6   0.4   0.1   0.7   0.4   0.7   0.4
       0.7   0.4   0.3   0.6   0.2   0.4   0.6   0.2   0.6   0.5   0.7   0.4   0.1
       0.1   0.5   0.6   0.4   0.1   0.6   0.5   0.4   0.4   0.5   0.5   0.4   0.4
       0.1   0.4   0.4   0.5   0.5   0.1   0.1   0.6   0.2   0.1   0.4   0.4   0.1
       0.1   0.1   0.7   0.6   0.5   0.1   0.4   0.1   0.4   0.4   0.6   0.6   0.1
       0.1   0.7   0.2   0.7   0.7   0.1   0.6   0.2   0.6   0.4   0.6   0.4   0.6
       0.5   0.6   0.6   0.1   0.3   0.1   0.6   0.6   0.4   0.7   0.5   0.3   0.6
       0.5   0.1   0.7   0.6   0.7   0.6   0.1   0.7   0.3   0.4   0.2   0.7   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
#output1
    0        1
  0.0542   0.9840

#input2
       0.7   0.7   0.1   0.1   0.1   0.4   0.7   0.7   0.4   0.6   0.4   0.1   0.6
       0.1   0.4   0.2   0.4   0.3   0.4   0.6   0.7   0.7   0.4   0.1   0.6   0.2
       0.6   0.4   0.7   0.6   0.4   0.6   0.6   0.1   0.7   0.1   0.1   0.4   0.6
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
#output2
    ....    ....    ....
    ....    ....    ....

#input28
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
```

Figure 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. The predicted outputs were the two floating numbers, 0.0542 and 0.9840.

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

Hidden units	No. of units	No. of links	Training network file	Testing result file	% accuracy of prediction
7	490	3381	hy7gr%3h7.net	tehy7gr%3h7.res	67.80
35	518	16905	hy7gr%3h35.net	tehy7gr%3h35.res	67.80
70	553	33810	hy7gr%3h70.net	tehy7gr%3h70.res	67.80
100	583	48300	hy7gr%3h100.net	tehy7gr%3h100.res	71.43
120	603	57960	hy7gr%3h120.net	tehy7gr%3h120.res	64.29

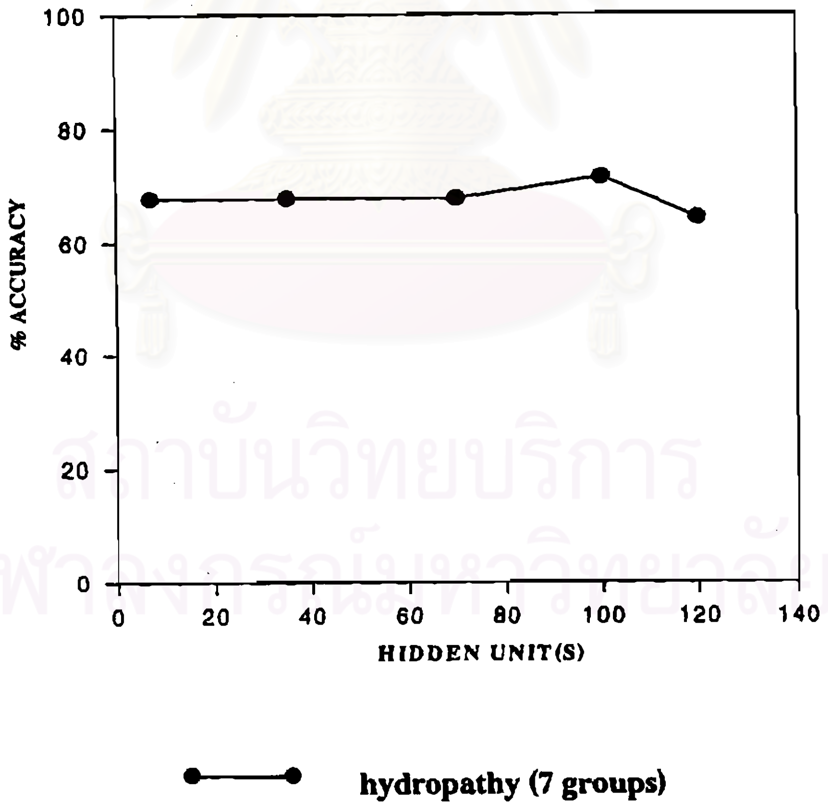

Figure 4.75 Percent accuracy prediction of percent helix structure (3 groups) using 7 groups of hydropathy as input vector with various numbers of hidden units (7, 35, 70, 100, 120).

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

The learn-pattern file, *taa8gr%3h.pat*, and a test-pattern file, *teaa8gr%3h.pat*, with 8 groups of amino acid side chain properties coded amino acid side chains were used in training network and the testing respectively. The example of what training or testing patterns may look like is shown in Figure 4.76. The trained networks and testing result files are shown in Table 4.39 and the example of the testing result pattern is shown in Figure 4.77. Figure 4.78 and Table 4.39 show that *aa8gr%3h70.net* and *aa8gr%3h100.net* gave the highest percent accuracy of 85.70%. The prediction accuracy from others networks were lower. The trained networks, *aa8gr%3h35.net* and *aa8gr%3h120.net* both gave 78.57% accuracy. Whereas *aa8gr%3h7.net* gave the lowest accuracy of 75.00%.

Table 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.

Hidden units	No. of units	No. of links	Training network file	Testing result file	% accuracy of prediction
7	490	3381	aa8gr%3h7.net	teaa8gr%3h7.res	75.00
35	518	16905	aa8gr%3h35.net	teaa8gr%3h35.res	78.57
70	553	33810	aa8gr%3h70.net	teaa8gr%3h70.res	85.70
100	583	48300	aa8gr%3h100.net	teaa8gr%3h100.res	85.70
120	603	57960	aa8gr%3h120.net	teaa8gr%3h120.res	78.57

```
SNNS pattern definition file V1.4

No. of patterns          : 70
No. of input units       : 481
No. of output units      : 2

#input1 lalc
       0.1   0.5   0.3   0.7   0.2   0.1   0.2   0.7   0.7   0.6   0.8   0.1   0.0
       0.3   0.7   0.8   0.2   0.5   0.6   0.1   0.2   0.3   0.1   0.1   0.8   0.1
       0.8   0.8   0.8   0.5   0.3   0.1   0.7   0.1   0.6   0.5   0.5   0.7   0.5
       0.6   0.6   0.7   0.5   0.5   0.1   0.6   0.3   0.1   0.5   0.1   0.5   0.2
       0.8   0.3   0.1   0.5   0.1   0.6   0.7   0.1   0.1   0.8   0.1   0.1   0.6
       0.5   0.2   0.6   0.1   0.8   0.2   0.7   0.7   0.8   0.8   0.7   0.6   0.5
       0.1   0.1   0.6   0.1   0.1   0.3   0.2   0.5   0.7   0.5   0.2   0.6   0.1
       0.2   0.8   0.2   0.6   0.2   0.6   0.2   0.1   0.5   0.5   0.8   0.7   0.6
       0.1   0.5   0.7   0.6   0.5   0.1   0.7   0.1   0.1   0.6   0.2   0.5   0.1
       0.7   0.1   0.6   0.1   0.1   0.6   0.2   0.8   0.5   0.1   0.6   0.2   0.5
       0.5   0.1   0.1   0.7   0.1   0.1   0.5   0.6   0.1   0.7   0.1   0.1   0.1
       0.1   0.1   0.1   0.1   0.1   0.4   0.6   0.1   0.1   0.7   0.1   0.8   0.3
       0.6   0.1   0.8   0.6   0.1   0.1   0.7   0.1   0.1   0.8   0.1   0.1   0.6
       0.5   0.6   0.1   0.6   0.6   0.1   0.3   0.2   0.5   0.8   0.2   0.7   0.3
       0.5   0.5   0.1   0.1   0.3   0.5   0.5   0.1   0.7   0.2   0.2   0.5   0.2
       0.3   0.1   0.5   0.1   0.5   0.6   0.3   0.3   0.1   0.2   0.7   0.5   0.1
       0.5   0.2   0.1   0.1   0.4   0.6   0.7   0.5   0.1   0.5   0.1   0.5   0.5
       0.7   0.8   0.1   0.1   0.8   0.2   0.6   0.5   0.1   0.1   0.5   0.8   0.1
       0.7   0.1   0.7   0.8   0.1   0.1   0.3   0.4   0.8   0.6   0.8   0.8   0.6
       0.3   0.5   0.8   0.3   0.1   0.6   0.1   0.6   0.5   0.1   0.6   0.1   0.5
       0.2   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
#output1 lalc
    0   1


#input2 1bp2
       0.1   0.3   0.6   0.1   0.3   0.1   0.7   0.1   0.1   0.6   0.4   0.7   0.6
       0.5   0.6   0.8   0.3   0.1   0.1   0.2   0.8   0.6   0.5   0.5   0.6   0.6
       0.1   0.1   0.6   0.4   0.1   0.7   0.4   0.6   0.6   0.3   0.1   0.8   0.1
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
#outut2 1bp2
      ...   ....  ....


#input70 8adh
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
```

Figure 4.76 The example of input and output patterns for training and testing. The inputs were 8 groups of amino acid side chain properties coded amino acid sequences. The outputs consisted of 2 units representing the helix or sheet or turn structures.

```
SNNS result file V1.4-3D
No. of hidden units : 70
No. of patterns     : 28
No. of input units  : 481
No. of output units : 2
startpattern        : 1
endpattern          : 28
input patterns included
teaching outputs included

#input1
      0.4   0.8   0.5   0.1   0.5   0.7   0.6   0.1   0.6   0.6   0.6   0.6   0.1
      0.1   0.1   0.6   0.4   0.5   0.8   0.5   0.7   0.1   0.1   0.5   0.6   0.1
      0.1   0.1   0.6   0.8   0.8   0.5   0.1   0.8   0.1   0.1   0.7   0.1   0.1
      0.1   0.5   0.6   0.6   0.3   0.6   0.2   0.1   0.2   0.7   0.1   0.1   0.4
      0.1   0.1   0.8   0.4   0.7   0.3   0.1   0.2   0.1   0.8   0.2   0.1   0.5
      0.6   0.6   0.1   0.6   0.1   0.1   0.0   0.0   0.0   0.0   0.0   0.0   0.0
      0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
      0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
      0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
      0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
      0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
      0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
      0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
      0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
      0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
      0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
      0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
      0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
      0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
      0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
      0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
      0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
      0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
      0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
      0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
      0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
      0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
      0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
      0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
      0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
      0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
      0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
      0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
      0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
      0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
      0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
      0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0

#output1
 0        1
0.3252   0.8745

#input2
      0.3   0.6   0.2   0.5   0.1   0.1   0.7   0.6   0.7   0.5   0.4   0.1   0.1
      0.4   0.1   0.1   0.4   0.1   0.1   0.1   0.1   0.3   0.7   0.1   0.2   0.7
      ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
      ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
#output2
   ....   ....   ....
   ....   ....   ....

#input28
      ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
      ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
```

Figure 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. The predicted outputs were the two floating numbers, 0.3252 and 0.8745.

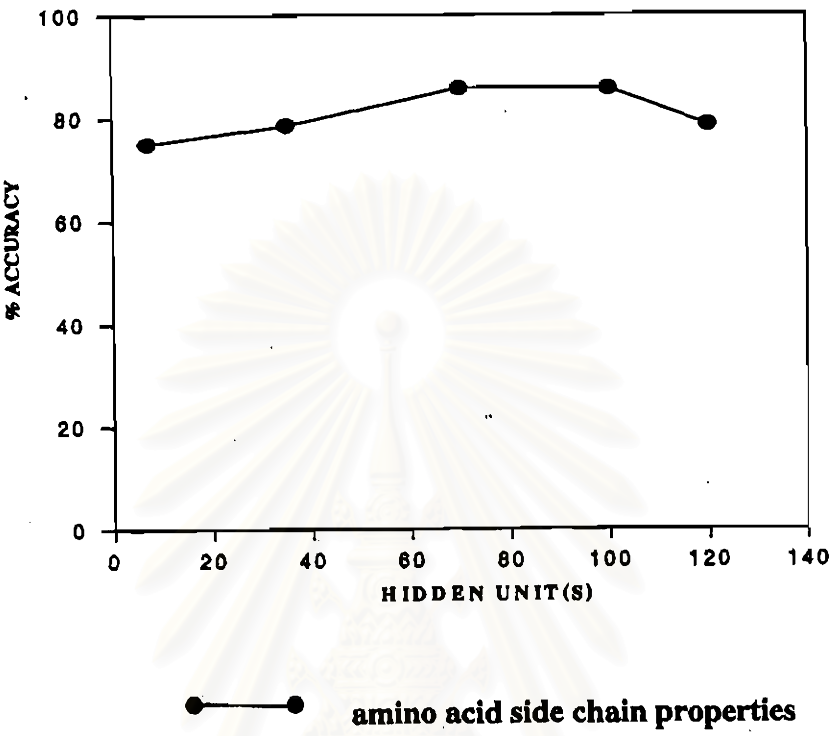

Figure 4.78 Percent accuracy prediction of percent helix structure (3 groups) using 8 groups of amino acid side chain properties as input vector with various numbers of hidden units (7, 35, 70, 100, 120).

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

The learn-pattern file, *thydro%3h.pat*, and a test-pattern file, *tehydro%3h.pat*, with hydrophobicity coded amino acid sequences were used in training network and the testing respectively. The example of what the training or testing pattern may look like is shown in Figure 4.79. The trained networks and testing result files are shown in Table 4.40 and the example of the testing result patterns is shown in Figure 4.80. Table 4.40 and Figure 4.81, the predictions with highest accuracy (71.43%) are the results from *hydro%3h70.net* and *hydro%3h100.net* which had 70 and 100 hidden units, respectively. The trained networks, *hydro%3h7.net* and *hydro%3h35.net* both gave 64.29% accurate, whereas, *hydro%3h120.net* gave 67.86%.

```
SNNS pattern definition file V1.4

No. of patterns        : 70
No. of input units     : 481
No. of output units    : 2


#input1 1a1c
        0.1   0.1   0.3   0.2   0.1   0.3   0.1   0.3   0.2   0.1   0.1   0.3   0.2
        0.1   0.3   0.1   0.2   0.2   0.2   0.1   0.3   0.2   0.3   0.2   0.1   0.3
        0.3   0.3   0.2   0.3   0.3   0.2   0.2   0.2   0.2   0.2   0.1   0.2   0.1
        0.2   0.3   0.3   0.1   0.1   0.1   0.1   0.2   0.2   0.1   0.2   0.2   0.3
        0.3   0.1   0.3   0.2   0.1   0.2   0.3   0.3   0.3   0.1   0.2   0.2   0.1
        0.2   0.2   0.1   0.2   0.1   0.1   0.3   0.3   0.1   0.3   0.2   0.3   0.1
        0.1   0.3   0.3   0.1   0.1   0.1   0.3   0.2   0.1   0.1   0.3   0.3   0.3
        0.2   0.1   0.1   0.3   0.3   0.1   0.3   0.1   0.2   0.3   0.1   0.2   0.3
        0.3   0.2   0.2   0.1   0.2   0.3   0.3   0.2   0.1   0.1   0.3   0.1   0.1
        0.3   0.3   0.3   0.1   0.1   0.1   0.0   0.0   0.0   0.0   0.0   0.0   0.0
        0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
        0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
        0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
        0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
        0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
        0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
        0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
        0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
        0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
        0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
        0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
        0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
        0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
        0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
        0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
        0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
        0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
        0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
        0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
        0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
        0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
        0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
        0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
        0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
        0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
        0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
        0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
#output1 1a1c
  0   1


#input2 1bp2
        0.2   0.3   0.3   0.1   0.3   0.1   0.2   0.3   0.3   0.1   0.3   0.1   0.3
        0.2   0.2   0.2   0.1   0.2   0.3   0.3   0.1   0.3   0.1   0.1   0.2   0.2
        0.3   0.2   0.3   0.2   0.3   0.2   0.2   0.2   0.2   0.2   0.2   0.3   0.1
        ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
        ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
#outut2 1bp2
       ...   ....  ....


#input70 8adh
        ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
        ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
```

Figure 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).

```
SNNS result file V1.4-3D
No. of hidden units : 70
No. of patterns     : 28
No. of input units  : 481
No. of output units : 2
startpattern        : 1
endpattern          : 28
input patterns included
teaching outputs included

#input1
       0.2   0.2   0.2   0.3   0.2   0.2   0.2   0.1   0.2   0.1   0.2   0.2   0.1
       0.2   0.3   0.2   0.1   0.1   0.3   0.2   0.3   0.2   0.1   0.2   0.1   0.1
       0.1   0.2   0.2   0.3   0.1   0.3   0.1   0.2   0.2   0.2   0.2   0.1   0.2
       0.1   0.2   0.2   0.3   0.1   0.2   0.3   0.2   0.3   0.2   0.2   0.1   0.1
       0.2   0.2   0.2   0.3   0.1   0.3   0.3   0.1   0.1   0.2   0.2   0.2   0.3
       0.1   0.3   0.1   0.3   0.1   0.1   0.2   0.1   0.1   0.1   0.2   0.3   0.3
       0.1   0.2   0.2   0.2   0.3   0.1   0.2   0.2   0.2   0.1   0.3   0.3   0.1
       0.3   0.2   0.3   0.2   0.3   0.2   0.2   0.3   0.1   0.2   0.1   0.2   0.2
       0.1   0.2   0.2   0.3   0.1   0.1   0.1   0.1   0.2   0.2   0.2   0.1   0.3
       0.2   0.3   0.3   0.2   0.3   0.1   0.2   0.1   0.2   0.2   0.1   0.3   0.2
       0.1   0.2   0.3   0.1   0.1   0.2   0.1   0.2   0.3   0.2   0.3   0.3   0.2
       0.3   0.3   0.3   0.1   0.3   0.2   0.2   0.2   0.1   0.2   0.2   0.3   0.1
       0.1   0.3   0.3   0.1   0.3   0.3   0.1   0.2   0.3   0.1   0.2   0.1   0.2
       0.1   0.2   0.2   0.1   0.3   0.2   0.1   0.3   0.1   0.2   0.1   0.2   0.3
       0.3   0.2   0.1   0.2   0.3   0.1   0.2   0.3   0.2   0.2   0.2   0.2   0.2
       0.3   0.2   0.2   0.2   0.2   0.3   0.3   0.1   0.3   0.3   0.2   0.3   0.3
       0.3   0.3   0.1   0.1   0.2   0.3   0.2   0.3   0.2   0.2   0.1   0.1   0.3
       0.3   0.1   0.3   0.1   0.1   0.3   0.1   0.3   0.1   0.2   0.1   0.2   0.1
       0.2   0.1   0.1   0.3   0.3   0.3   0.1   0.1   0.3   0.1   0.2   0.2   0.1
       0.2   0.3   0.1   0.1   0.1   0.1   0.3   0.1   0.2   0.2   0.3   0.1   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
       0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
#output1
 0        1
0.2768   0.9654

#input2
       0.1   0.1   0.3   0.3   0.3   0.2   0.1   0.1   0.2   0.1   0.2   0.3   0.1
       0.3   0.2   0.3   0.2   0.2   0.2   0.1   0.1   0.1   0.2   0.3   0.1   0.3
       0.2   0.3   0.1   0.1   0.2   0.1   0.1   0.3   0.1   0.3   0.3   0.2   0.1
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
#output2
  ....   ....   ....
  ....   ....   ....

#input28
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
       ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
```

Figure 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). The predicted outputs were the three floating numbers, 0.2768 and 0.9654.

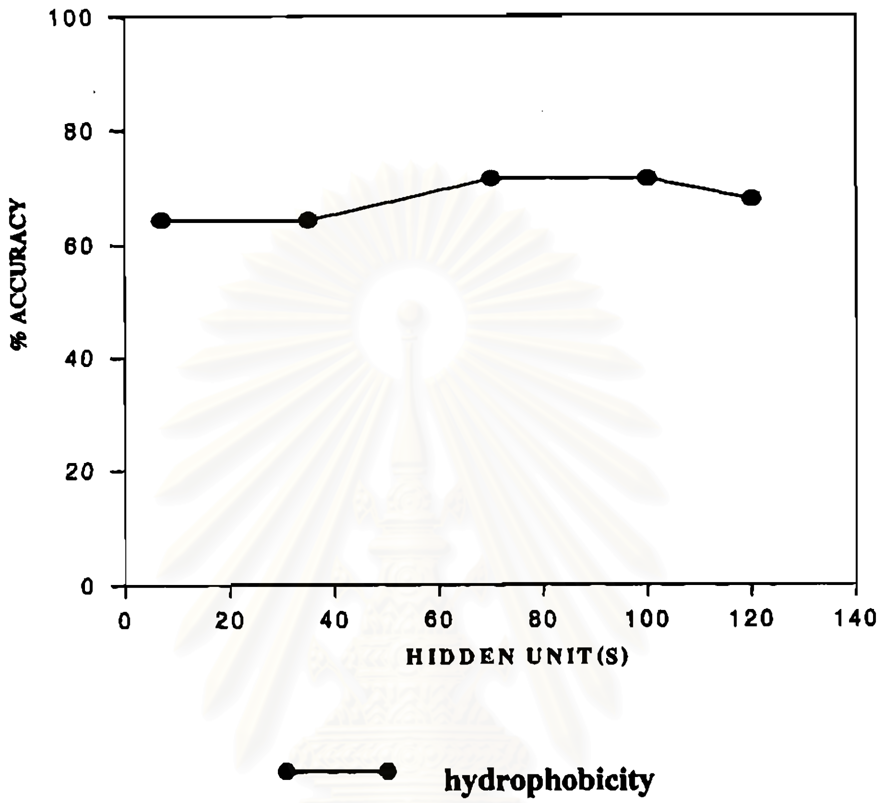

Figure 4.81 Percent accuracy prediction of percent helix structure (3 groups) using hydrophobicity as input vector with various numbers of hidden units (7, 35, 70, 100, 120).

Table 4.40 The trained network file, testing result file and percent accuracy of percent helix structure prediction (3 groups) using hydrophobicity.

Hidden units	No. of units	No. of links	Training network file	Testing result file	% accuracy of prediction
7	490	3381	hydro%3h7.net	tehydro%3h7.res	64.29
35	518	16905	hydro%3h35.net	tehydro%3h35.res	64.29
70	553	33810	hydro%3h70.net	tehydro%3h70.res	71.43
100	583	48300	hydro%3h100.net	tehydro%3h100.res	71.43
120	603	57960	hydro%3h120.net	tehydro%3h120.res	67.86

4.10.5 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

Figure 4.82 shows that most of percent accuracy prediction of 5 groups of percent sheet structure were in the range between 67-85%. The highest percent accuracy was 78.7% (*aa8gr%3h70.net* and *aa8gr%3h100.net*). This network with 70 and 100 hidden units were trained from the amino acid side chain properties (8 groups). Apparently this property also gave overall percent accuracy higher than those resulting from the others properties. The lowest percent accuracy was 64.29% (*hy7gr%3h120.net, hydro%3h7.net* and *hydro%3h35.net*).

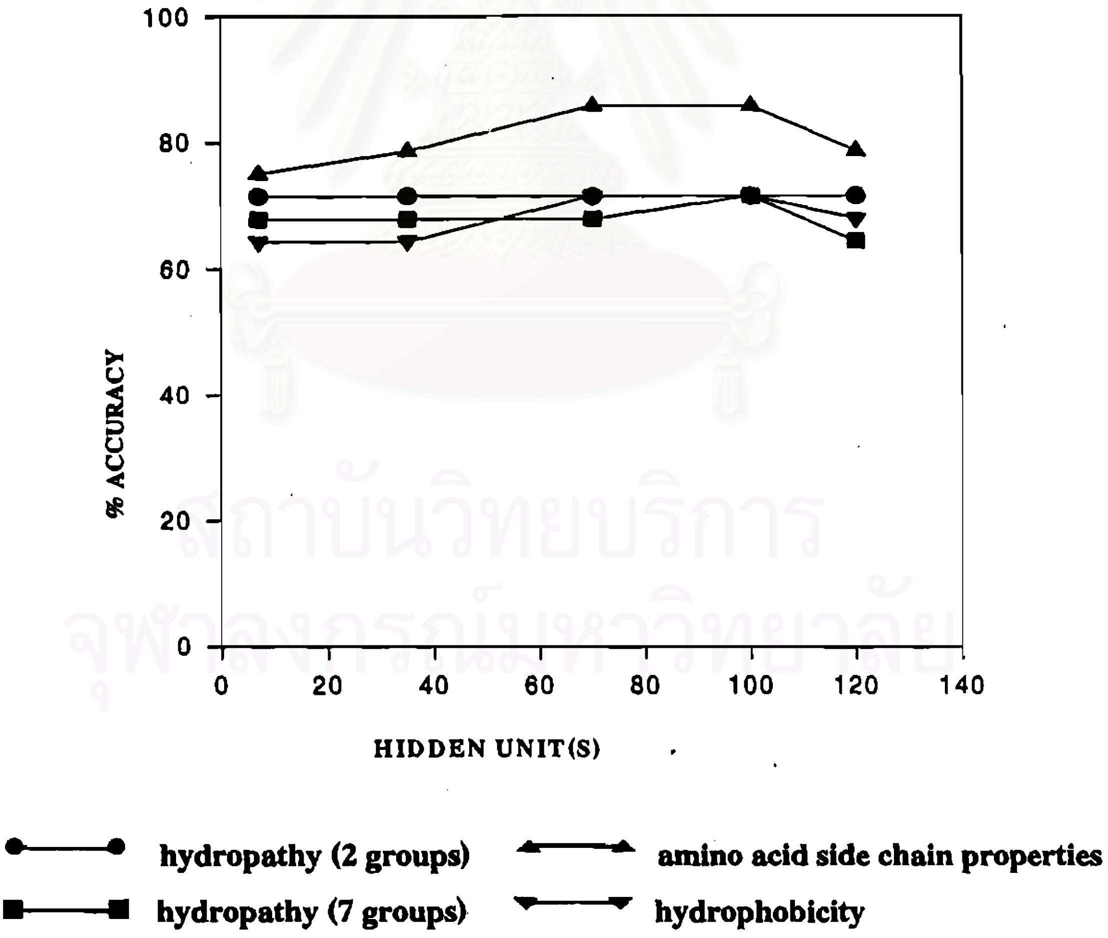

Figure 4.82 The comparison of percent accuracy prediction of percent turn structure (3 groups) using various types of input properties.

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

Training set.

Seventy patterns of amino acid sequences were saved as a learn pattern file for used in training (Table 4.41). The network were trained for the prediction of three groups of percent sheet structure using amino acid properties in the section 4.11.1, 4.11.2, 4.11.3 and 4.11.4

Testing set.

Twenty-eight patterns of amino acid sequences were saved as a test pattern file for used in testing (Table 4.41). The network were tested for the prediction of three group of percent sheet structure using amino acid properties in the section 4.11.1, 4.11.2, 4.11.3 and 4.11.4.

Input and output patterns for training and testing.

The input pattern was composed of 481 input units of amino acid which were coded with properties of amino acid residue as previously described (section 3.2.2).

The output pattern had 3 units of percent sheet structure. The percent sheet structure in amino acid sequence was divided into 3 groups. There were 2 possible outputs representing the following ranges of percent values 1-50%, 51-100%, and the seventh group represented 0%. The training set, testing set and the possible outputs are shown in Table 4.41.

A learn-pattern file with 70 patterns was used in training the networks with 7, 35, 70, 100 and 120 hidden units. The maximum value for learn cycles was set to 100000 and the maximum error to stop was set to 0.01. A test-pattern file with 28 patterns was used for testing the trained networks.

These training and testing sets (Table 4.41) were the standard sets for prediction of percent sheet structure (3 groups) using various types of properties in the next section (section. 4.11.1, 4.11.2, 4.11.3 and 4.11.4).

Table 4.41 The training set, testing set and possible outputs of percent sheet structure (3 groups) prediction.

Training set (Input No.)*	**Testing set (Input No.)***	**Output patterns**	**% of sheet structure**
7, 13, 18, 23, 25, 28, 40, 48, 53, 59, 61, 70, 78, 83, 86, 88	73, 77, 84, 87, 94	0 0	0
1, 2, 3, 4, 5, 6, 8, 9, 12, 14, 15, 16, 17, 19, 20, 21, 22, 24, 26, 27, 29, 30, 31, 32, 33, 35, 36, 37, 39, 41, 42, 43, 44, 45, 46, 49, 50, 51, 52, 54, 55, 56, 57, 58, 60, 63	64, 65, 66, 67, 68, 71, 72, 74, 75, 76, 79, 80, 81, 82, 85, 89, 90, 91, 92, 93, 95, 98	0 1	1-50
10, 11, 34, 38, 47, 62, 69, 96	97	1 0	51-100

Names of input no. were listed in Table 3.1

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

The learn-pattern file, *thy2gr%3s.pat,* and testing-pattern file, *tehy2gr%3s.pat,* with hydropathy (3 groups) coded amino acid sequences were used for training and testing respectively. Figure 4.70 shows the example of input and output patterns in these files. The trained network and testing result files are shown in Table 4.42 and the example of result patterns is shown in Figure 4.71. The most accurate predictions were the results from *hy2gr%3s7.net* and *hy2gr%3s35.net*. These networks with 7 and 35 hidden units, gave 75.00% accuracy, the highest in their groups. The prediction accuracy from others networks were lower. Files, *hy2gr%3s70.net, hy2gr%3s100* and *hy2gr%3s120.net* gave 71.43%, 71.43%, and 67.8% respectively.

Table 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).

Hidden units	No. of units	No. of links	Training network file	Testing result file	% accuracy of prediction
7	490	3381	hy2gr%3s7.net	tehy2gr%3s7.res	75.00
35	518	16905	hy2gr%3s35.net	tehy2gr%3s35.res	75.00
70	553	33810	hy2gr%3s70.net	tehy2gr%3s70.res	71.43
100	583	48300	hy2gr%3s100.net	tehy2gr%3s100.res	71.43
120	603	57960	hy2gr%3s120.net	tehy2gr%3s120.res	67.80

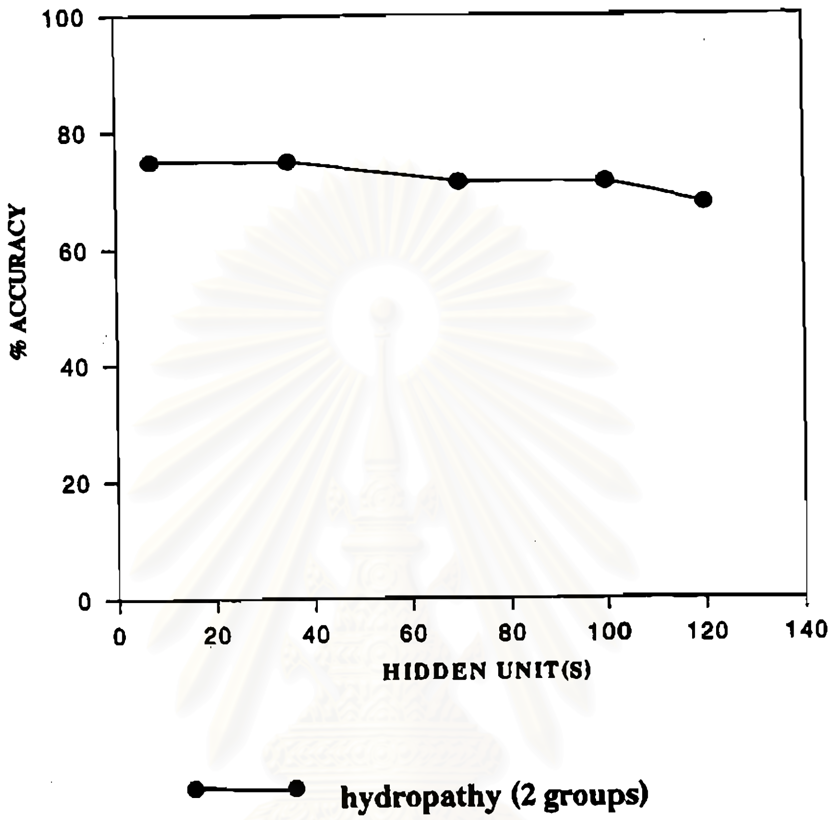

Figure 4.83 Percent accuracy prediction of percent sheet structure (3 groups) using 2 groups of hydropathy as input vector with various numbers of hidden units (7, 35, 70, 100, 120).

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

The learn-pattern file, *thy7gr%3s.pat*, and test-pattern file, *tehy7gr%3s.pat*, with hydropathy (7 groups) coded the amino acid sequences were used in training and testing with various hidden units. Figure 4.73 shows the example of input and output patterns. The trained network and testing result file are shown in Table 4.43 and the example of result patterns is in Figure 4.74. Table 4.43 and Figure 4.84 show that, *hy7gr%3s7.net*, which had 7 hidden units gave the highest percent accuracy of 78.57%. Files *hy7gr%3s35.net*, *hy7gr%3s70.net*, *hy7gr%3s100.net* and *hy7gr%3s120* gave 75.00%, 64.29%, 60.71% and 60.71% accuracy respectively.

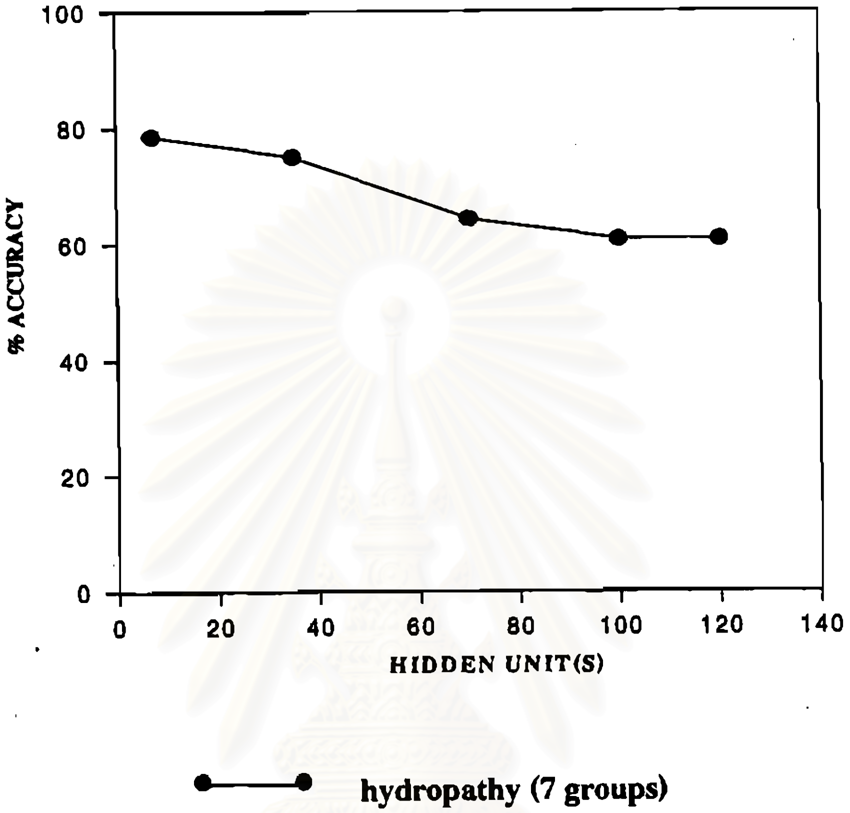

Figure 4.84 Percent accuracy prediction of percent sheet structure (3 groups) using 7 groups of hydropathy as input vector with various numbers of hidden units (7, 35, 70, 100, 120).

Table 4.43 The trained network file, testing result file and percent accuracy of percent sheet structure prediction (3 groups) using hydropathy (7 groups).

Hidden units	No. of units	No. of links	Training network file	Testing result file	% accuracy of prediction
7	490	3381	hy7gr%3s7.net	tehy7gr%3s7.res	78.57
35	518	16905	hy7gr%3s35.net	tehy7gr%3s35.res	75.00
70	553	33810	hy7gr%3s70.net	tehy7gr%3s70.res	64.24
100	583	48300	hy7gr%3s100.net	tehy7gr%3s100.res	60.71
120	603	57960	hy7gr%3s120.net	tehy7gr%3s120.res	60.71

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

The learn-pattern file, *taa8gr%3s.pat*, and a test-pattern file, *teaa8gr%3s.pat*, with 8 groups of amino acid side chain properties coded amino acid side chains were used in training network and the testing respectively. The example of what training or testing patterns may look like is shown in Figure 4.76. The trained networks and testing result files are shown in Table 4.44 and the example of the testing result pattern is shown in Figure 4.77. Figure 4.85 and Table 4.44 show that *aa8gr%3s100.net* gave the highest percent accuracy of 78.57%. The prediction accuracy from others networks were lower. The trained networks, *aa8gr%3s7.net*, and *aa8gr%3s35.net* and *aa8gr%3s70.net* all gave 75.00% accuracy. Whereas *aa8gr%3s120.net* gave the lowest accuracy of 64.28%.

Table 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.

Hidden units	No. of units	No. of links	Training network file	Testing result file	% accuracy of prediction
7	490	3381	aa8gr%3s7.net	teaa8gr%3s7.res	75.00
35	518	16905	aa8gr%3s35.net	teaa8gr%3s35.res	75.00
70	553	33810	aa8gr%3s70.net	teaa8gr%3s70.res	75.00
100	583	48300	aa8gr%3s100.net	teaa8gr%3s100.res	78.57
120	603	57960	aa8gr%3s120.net	teaa8gr%3s120.res	64.28

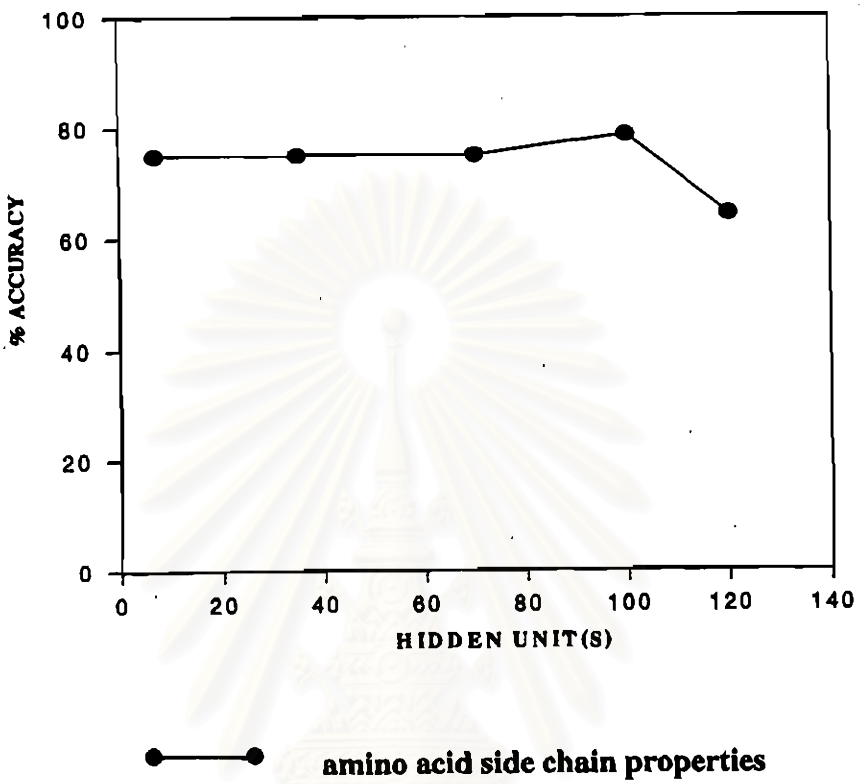

Figure 4.85 Percent accuracy prediction of percent sheet structure (3 groups) using 8 groups of amino acid side chain properties as input vector with various numbers of hidden units (7, 35, 70, 100, 120).

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

The learn-pattern file, *thydro%3s.pat*, and a test-pattern file, *tehydro%3s.pat*, with hydrophobicity coded amino acid sequences were used in training network and the testing respectively. The example of what the training or testing pattern may look like is shown in Figure 4.79. The trained networks and testing result files are shown in Table 4.45 and the example of the testing result patterns is shown in Figure 4.80. Table 4.45 and Figure 4.86, the prediction from *hydro%3s35.net* is 67.86% accurate, whereas *hydro%3s7.net*, *hydro%3s100.net* and *hydro%3s120.net* gave 64.29% accuracy. File, *hydro%3s70.net* gave the lowest accuracy of 60.81%.

Table 4.45 The trained network file, testing result file and percent accuracy of percent sheet structure prediction (3 groups) using hydrophobicity.

Hidden units	No. of units	No. of links	Training network file	Testing result file	% accuracy of prediction
7	490	3381	hydro%3s7.net	tehydro%3s7.res	64.29
35	518	16905	hydro%3s35.net	tehydro%3s35.res	67.86
70	553	33810	hydro%3s70.net	tehydro%3s70.res	60.81
100	583	48300	hydro%3s100.net	tehydro%3s100.res	64.29
120	603	57960	hydro%3s120.net	tehydro%3s120.res	64.29

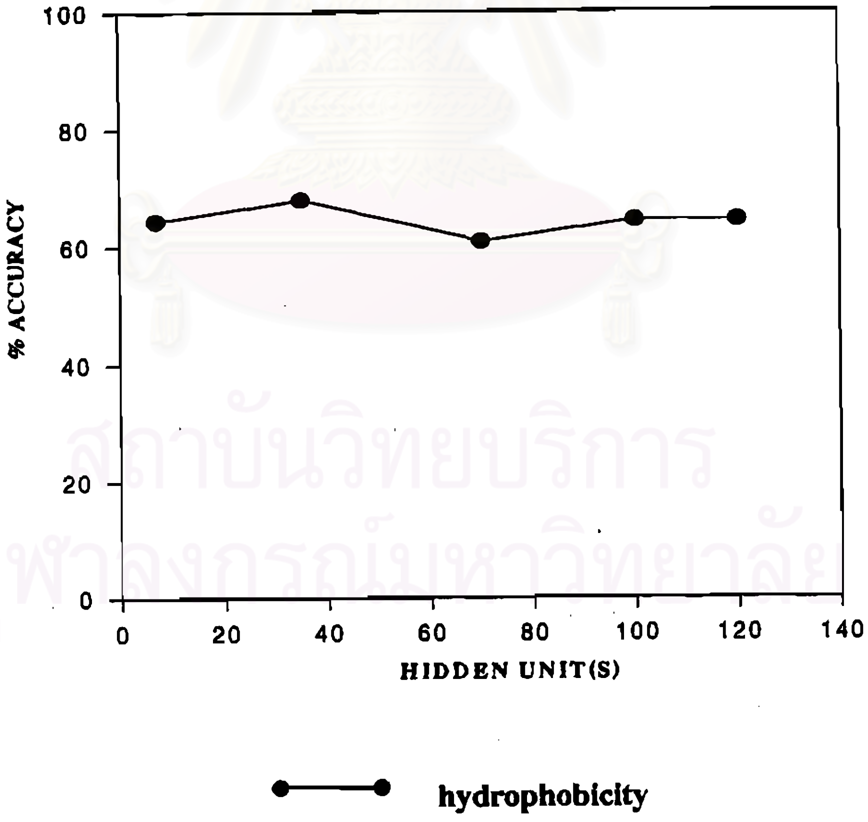

Figure 4.86 Percent accuracy prediction of percent sheet structure (3 groups) using hydrophobicity as input vector with various numbers of hidden units (7, 35, 70, 100, 120).

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

Figure 4.87 shows that most of percent accuracy prediction of 3 groups of percent sheet structure were in the range between 60-80%. The highest percent accuracy was 78.57% (*hy7gr%3s7.net* and *aa8gr%3s100.net*). The lowest percent accuracy was 60.71% (*hy7gr%3s100.net and hydro%3s70.net*).

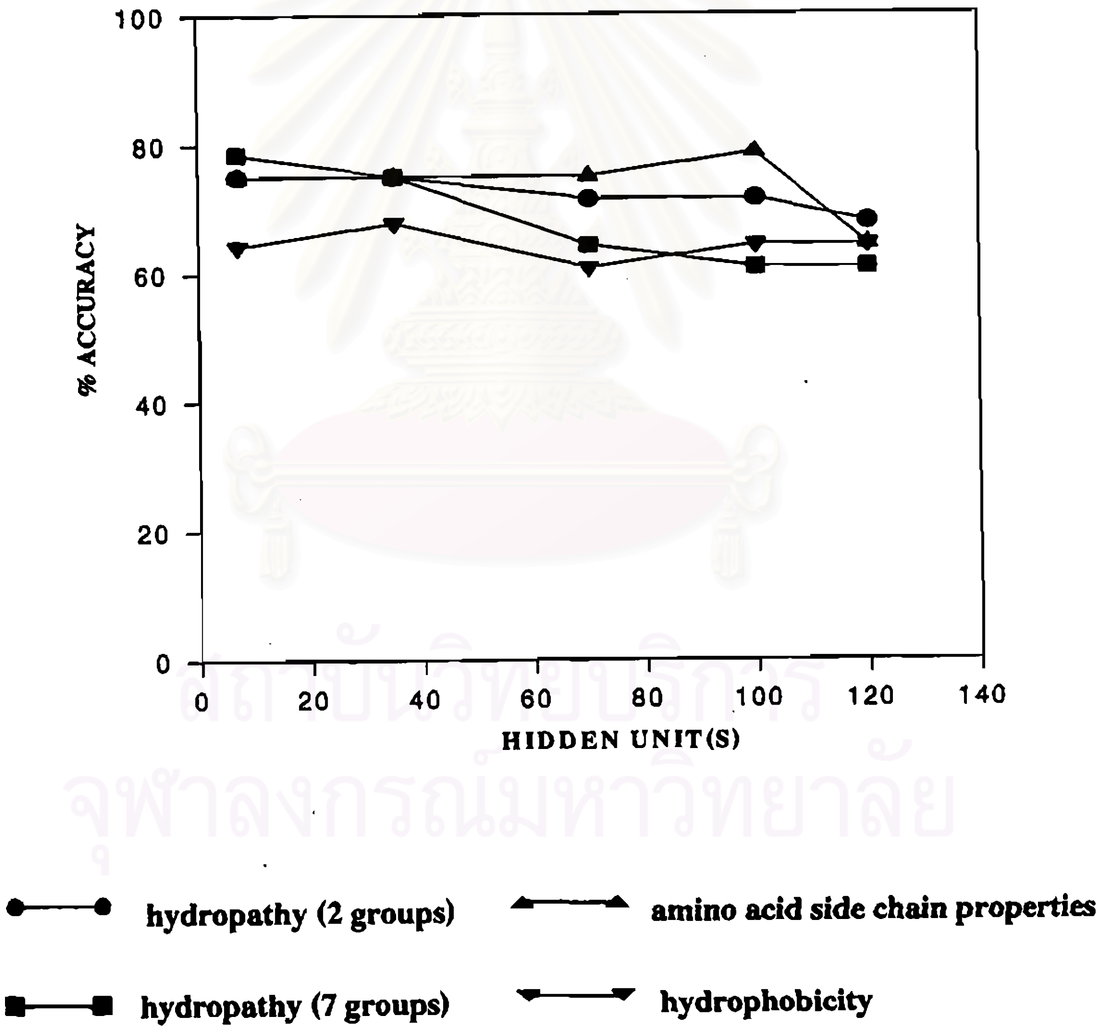

Figure 4.87 The comparison of percent accuracy prediction of percent sheet structure (3 groups) using various types of input properties.

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

Training set.

Seventy patterns of amino acid sequences were saved as a learn pattern file for used in training (Table 4.46). The network were trained for the prediction of 3 groups of percent turn structure using amino acid properties in the section 4.12.1, 4.12.2, 4.12.3 and 4.12.4

Testing set.

Twenty-eight patterns of amino acid sequences were saved as a test pattern file for used in testing (Table 4.46). The network were tested for the prediction of 3 groups of percent turn structure using amino acid properties in the section 4.12.1, 4.12.2, 4.12.3 and 4.12.4.

Input and output patterns for training and testing.

The input pattern was composed of 481 input units of amino acid which were coded with properties of amino acid residue as previously described (section 3.2.2).

The output pattern had 2 units of percent turn structure. The percent turn structure in amino acid sequence was divided into 3 groups. There were 2 outputs representing the following ranges of percent values 1-50%, 51-100% and the third group represented 0%. The training set, testing set and the possible outputs are shown in Table 4.46.

A learn-pattern file with 70 patterns was used in training the networks with 7, 35, 70, 100 and 120 hidden units. The maximum value for learn cycles was set to 100000 and the maximum error to stop was set to 0.01. A test-pattern file with 28 patterns was used for testing the trained networks.

These training and testing sets (Table 4.46) were the standard sets for prediction of percent turn structure (3 groups) using various types of properties in the next section (section. 4.12.1, 4.12.2, 4.12.3 and 4.12.4).

Table 4.46 The training set, testing set and possible outputs of percent turn structure (3 groups) prediction.

Training set (Input No.)*	**Testing set (Input No.)***	**Output patterns**	**% of turn structure**
1, 4, 5, 7, 8, 9, 13, 15, 18, 24, 25, 26, 33, 34, 37, 38, 39, 41, 43, 45, 46, 48, 50, 52, 55, 57, 59, 67, 69, 73	75, 76, 77, 78, 83, 87, 89, 91, 93	0 0	0
2, 3, 6, 10, 11, 12, 14, 16, 17, 19, 20, 21, 22, 23, 27, 28, 29, 30, 31, 32, 35, 36, 40, 42, 44, 47, 49, 51, 53, 54, 56, 58, 60, 61, 63, 64, 65	66, 70, 71, 72, 74, 79, 80, 81, 82, 84, 85, 86, 88, 92, 94, 95, 96, 97, 98	0 1	1-50
62, 68	90	1 0	51-100

Names of input no. were listed in Table 3.1

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

The learn-pattern file, *thy2gr%3t.pat*, and testing-pattern file, *tehy2gr%3t.pat*, with hydropathy (2 groups) coded amino acid sequences were used for training and testing respectively. Figure 4.70 shows the example of input and output patterns. The trained network and testing result files are shown in Table 4.47 and the example of result patterns is shown in Figure 4.71. The best prediction was the result from *hy2gr%3t7.net, hy2gr%3t70.net* and *hy2gr%3t100.net*. These networks with 7, 70 and 100 hidden units, gave 64.29% accuracy, the highest in their groups. Files, *hy2gr%3t100* and *hy2gr%3t120.net* with 35 and 120 hidden units respectively were both 60.71 accurate (Table 4.47 and Figure 4.88).

Table 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).

Hidden units	No. of units	No. of links	Training network file	Testing result file	% accuracy of prediction
7	490	3381	hy2gr%3t7.net	tehy2gr%3t7.res	64.29
35	518	16905	hy2gr%3t35.net	tehy2gr%3t35.res	60.17
70	553	33810	hy2gr%3t70.net	tehy2gr%3t70.res	64.29
100	583	48300	hy2gr%3t100.net	tehy2gr%3t100.res	64.29
120	603	57960	hy2gr%3t120.net	tehy2gr%3t120.res	60.17

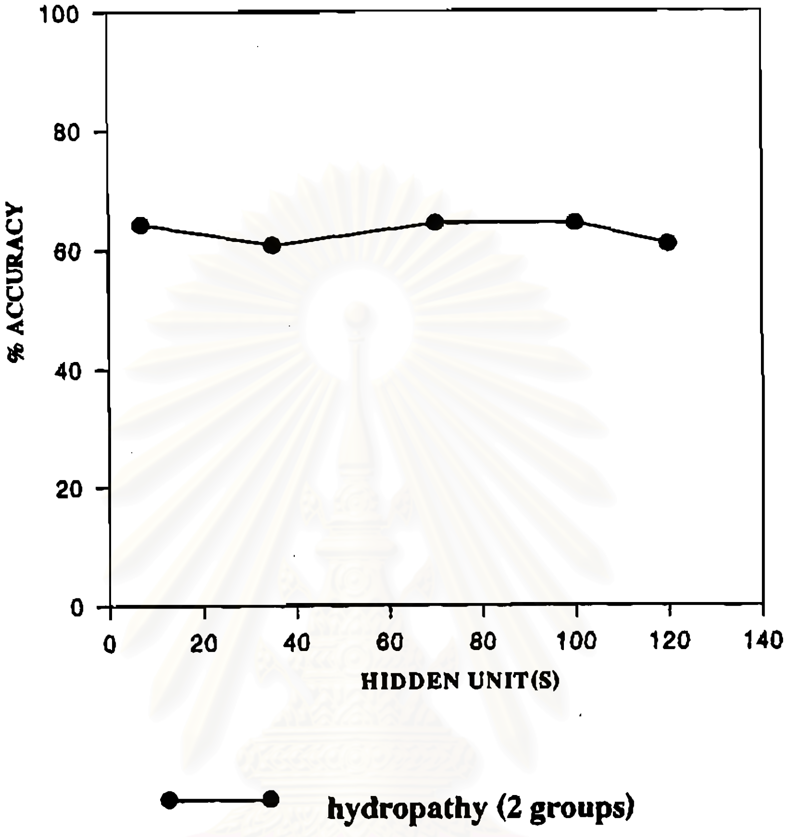

Figure 4.88 Percent accuracy prediction of percent turn structure (3 groups) using 2 groups of hydropathy as input vector with various numbers of hidden units (7, 35, 70, 100, 120).

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

The learn-pattern file, *thy7gr%3t.pat*, and test-pattern file, *tehy7gr%3t.pat*, with hydropathy (7 groups) coded the amino acid sequences were used in training and testing with various hidden units. Figure 4.73 shows the example of input and output patterns. The trained network and testing result file are shown in Table 4.48 and the example of result patterns is in Figure 4.74. Table 4.48 and Figure 4.89 show that, *hy7gr%3t70.net* and *hy7gr%3t120*, which had 70 and 120 hidden units gave the highest percent accuracy of 57.14%. Files *hy7gr%3t7.net, hy7gr%3t35.net* and *hy7gr%3t100.net* gave 53.57% accuracy respectively.

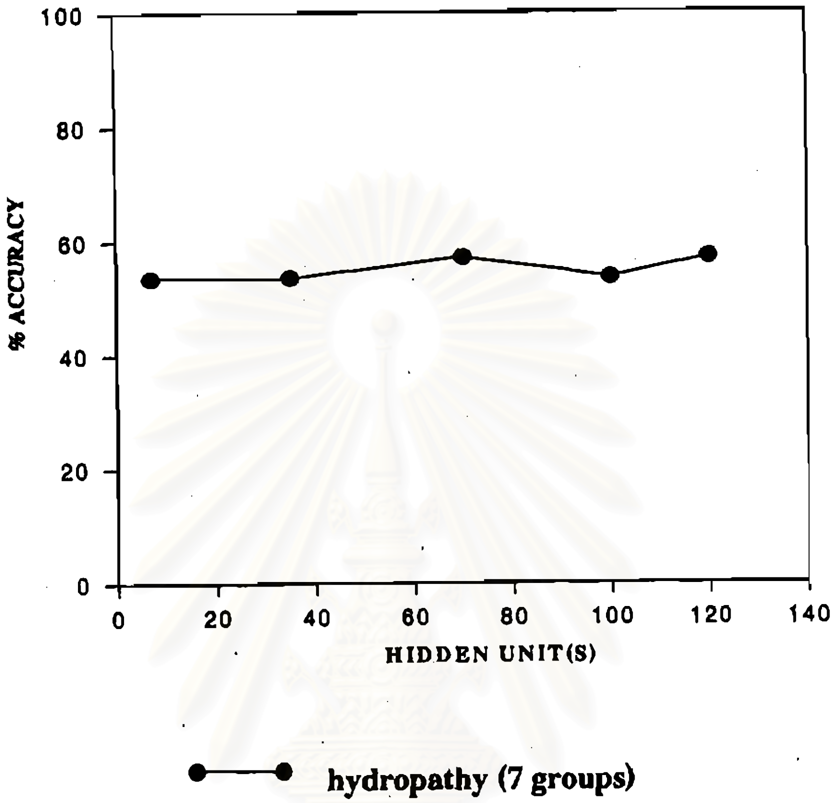

Figure 4.89 Percent accuracy prediction of percent turn structure (3 groups) using 7 groups of hydropathy as input vector with various numbers of hidden units (7, 35, 70, 100, 120).

Table 4.48 The trained network file, testing result file and percent accuracy of percent turn structure prediction (3 groups) using hydropathy (7 groups).

Hidden units	No. of units	No. of links	Training network file	Testing result file	% accuracy of prediction
7	490	3381	hy7gr%3t7.net	tehy7gr%3t7.res	53.57
35	518	16905	hy7gr%3t35.net	tehy7gr%3t35.res	53.57
70	553	33810	hy7gr%3t70.net	tehy7gr%3t70.res	57.14
100	583	48300	hy7gr%3t100.net	tehy7gr%3t100.res	53.57
120	603	57960	hy7gr%3t120.net	tehy7gr%3t120.res	57.14

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

The learn-pattern file, *taa8gr%3t.pat*, and a test-pattern file, *teaa8gr%3t.pat*, with 8 groups of amino acid side chain properties coded amino acid side chains were used in training network and the testing respectively. The example of what training or testing patterns may look like is shown in Figure 4.76. The trained networks and testing result files are shown in Table 4.49 and the example of the testing result pattern is shown in Figure 4.77. Figure 4.90 and Table 4.49 show that *aa8gr%3t70.net* gave the highest percent accuracy of 64.29%. The prediction accuracy from others networks were lower. The trained networks, *aa8gr%3t35.net*, *aa8gr%3t100.net* and *aa8gr%3t120.net* all gave 60.71% accuracy. Whereas, *aa8gr%3t7.net* gave the lowest accuracy of 57.14%.

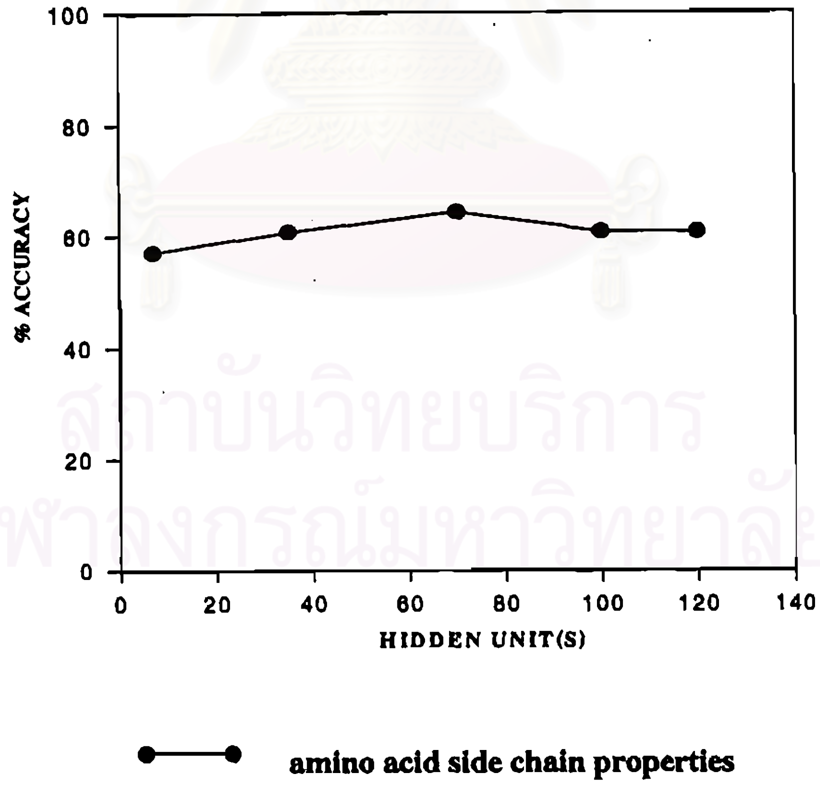

Figure 4.90 Percent accuracy prediction of percent turn structure (3 groups) using 8 groups of amino acid side chain properties as input vector with various numbers of hidden units (7, 35, 70, 100, 120).

Table 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.

Hidden units	No. of units	No. of links	Training network file	Testing result file	% accuracy of prediction
7	490	3381	aa8gr%3t7.net	teaa8gr%3t7.res	57.14
35	518	16905	aa8gr%3t35.net	teaa8gr%3t35.res	60.71
70	553	33810	aa8gr%3t70.net	teaa8gr%3t70.res	64.29
100	583	48300	aa8gr%3t100.net	teaa8gr%3t100.res	60.71
120	603	57960	aa8gr%3t120.net	teaa8gr%3t120.res	60.71

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

The learn-pattern file, *thydro%3t.pat*, and a test-pattern file, *tehydro%3t.pat*, with hydrophobicity coded amino acid sequences were used in training network and the testing respectively. The example of what the training or testing pattern may look like is shown in Figure 4.79. The trained networks and testing result files are shown in Table 4.50 and the example of the testing result patterns is shown in Figure 4.80. Table 4.50 and Figure 4.91, the predictions from *hydro%3t7.net* and *hydro%3t70.net* are 60.71% accurate, whereas, *hydro%2t100.net* and *hydro%2t120.net* give 57.14% accuracy. File, *hydro%2t35.net* gave the lowest accuracy of 50.00%.

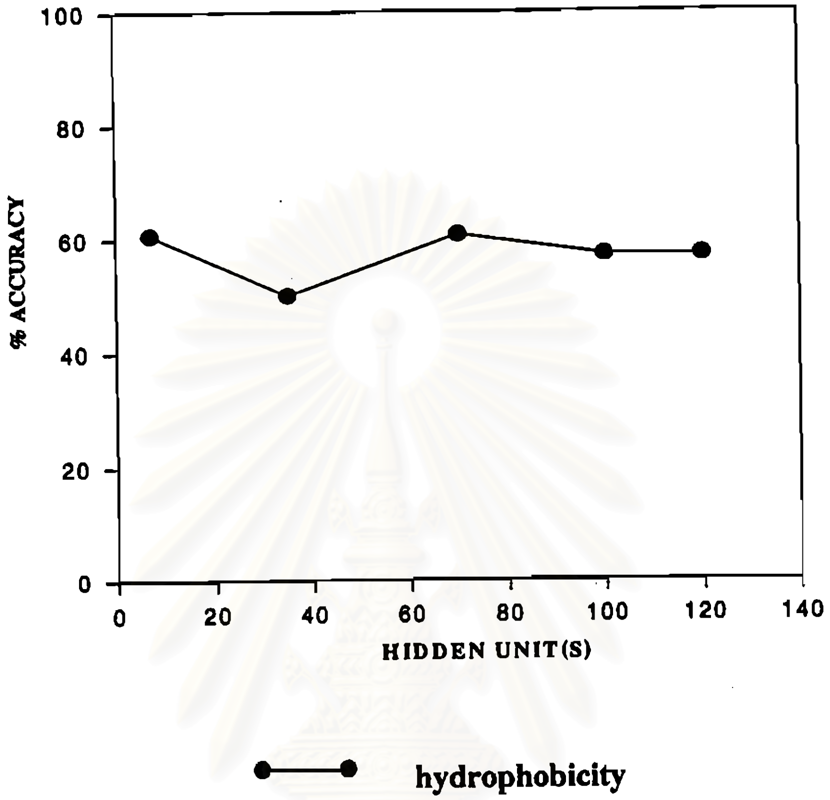

Figure 4.91 Percent accuracy prediction of percent turn structure (3 groups) using hydrophobicity as input vector with various numbers of hidden units (7, 35, 70, 100, 120).

Table 4.50 The trained network file, testing result file and percent accuracy of percent turn structure prediction (3 groups) using hydrophobicity.

Hidden units	No. of units	No. of links	Training network file	Testing result file	% accuracy of prediction
7	490	3381	hydro%3t7.net	tehydro%3t7.res	60.71
35	518	16905	hydro%3t35.net	tehydro%3t35.res	50.00
70	553	33810	hydro%3t70.net	tehydro%3t70.res	60.71
100	583	48300	hydro%3t100.net	tehydro%3t100.res	57.14
120	603	57960	hydro%3t120.net	tehydro%3t120.res	57.14

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

Figure 4.92 shows that most of percent accuracy prediction of 3 groups of percent sheet structure were in the range between 50-65%. The highest percent accuracy was 64.29% (*hy2gr%3t70.net, hy2gr%3t100.net* and *aa8gr%3t70.net*). These network with 70, 100 and 70 hidden units were trained from the hydropathy (2 groups) and amino acid side chain properties. This properties also gave over all percent accuracy higher than the others properties. The lowest percent accuracy was 53.57% (*hy7gr%3t7.net, hy7gr%3t35.net and hy7gr%3t100.net*). These networks were trained from the hydropathy (7 groups). This property also gave over all percent accuracy lower than those results from the others properties.

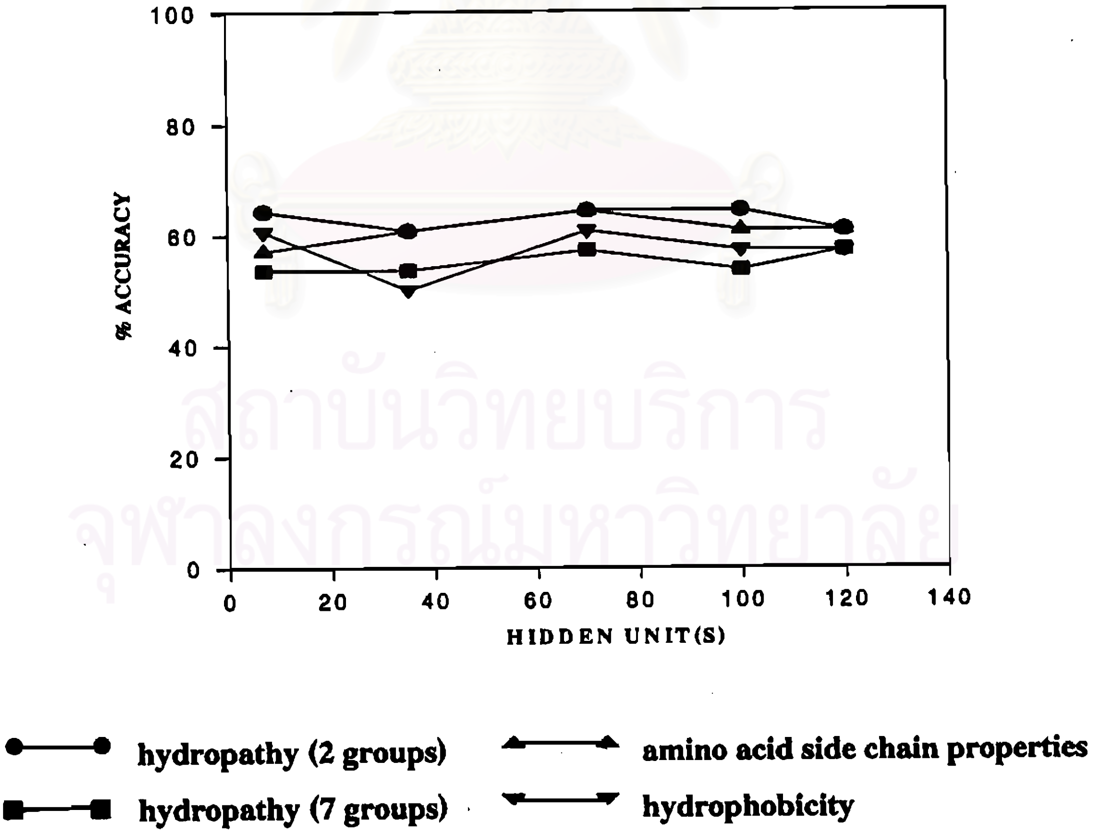

Figure 4.92 The comparison of percent accuracy prediction of percent turn structure (3 groups) using various types of input properties.

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

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

From Figure 4.93, all networks which different hidden units gave the same prediction accuracy (71.43%) for percent (3 groups) helix structure prediction. While the prediction of percent sheet structure, this hydropathy (2 groups) property could give the highest percent accuray (75%) when using the network with 7 and 35 hidden units. The others network for sheet prediction gave the same rang of percent accuracy as helix prediction. This property gave the overall percent accuracy for the prediction of turn structure lower than the others structures. The range of turn structure prediction was between 60-65%.

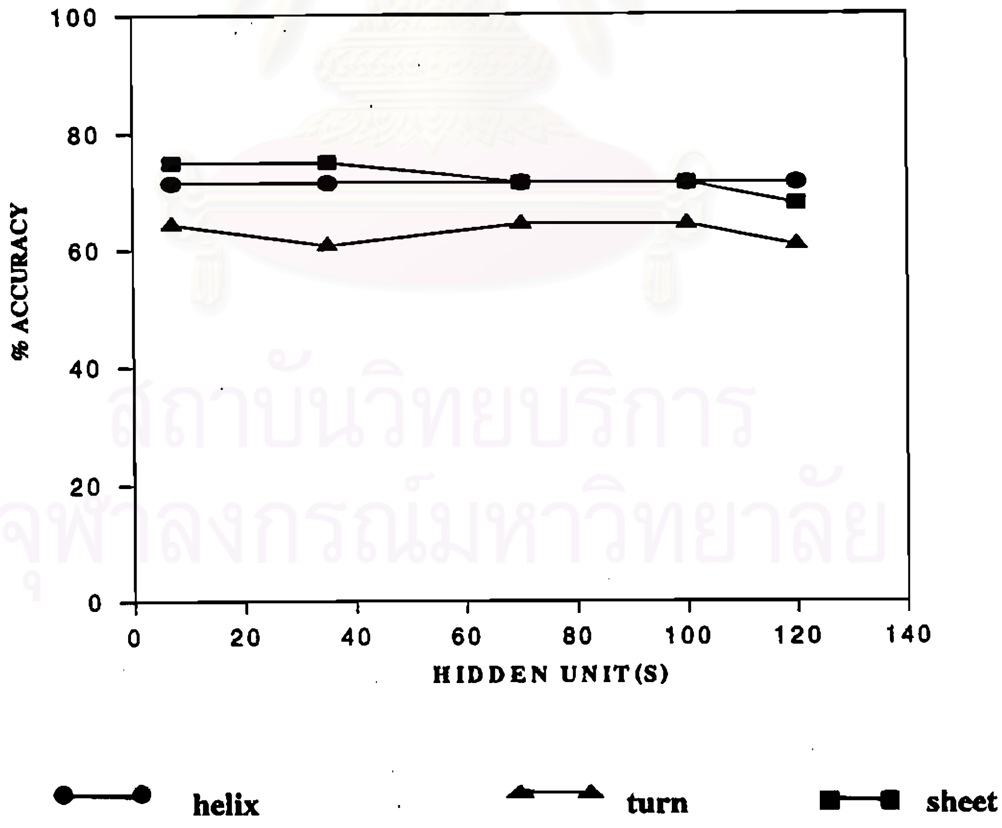

Figure 4.93 The percent accuracy predictions of the percent of helix, sheet, turn (3 groups) in separate networks using hydropathy (2 groups).

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

From Figure 4.94, the hydropathy (7 groups) gave the higest (78% and 75%) percent accuracy for sheet prediction when using the networks with 7 and 35 hidden units. The others networks of sheet prediction gave the same range (60-70%) of percent accuracy same as the networks of helix prediction. This 7 groups property also gave the lowest accuray when used for prediction of turn structure. The lowest range of accuracy was 50-60%.

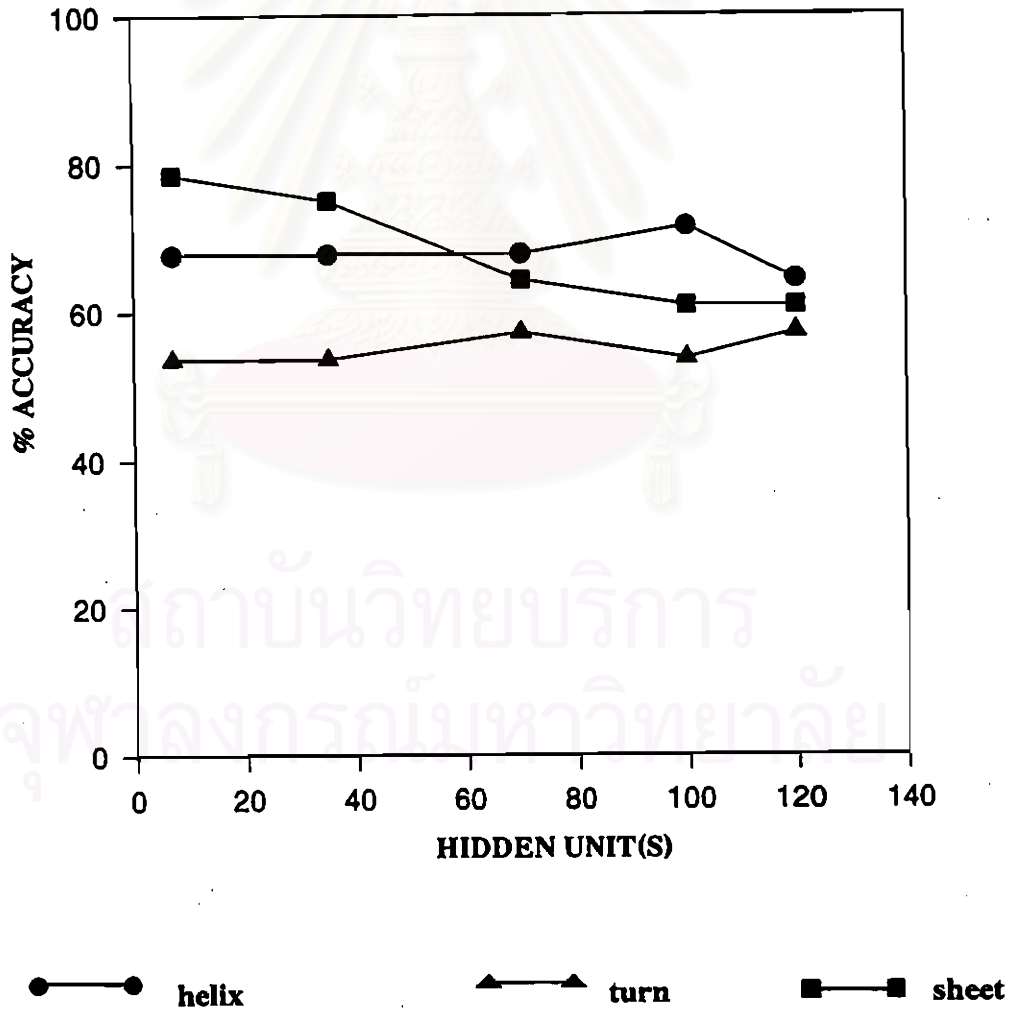

Figure 4.94 The percent accuracy predictions of the percent of helix, sheet, turn (3 groups) in separate networks using hydropathy (7 groups).

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

From Figure 4.95, the amino acid side chain properties gave the highest (85%) accuracy for helix prediction. The range of this predicition was between 75-85%. For percent sheet prediction, these properties gave the percent accuracy more than 75% except the result from the network with 120 hidden units. The lowest accuracy (57%) was the result from the network for turn prediction. The accuracy for this turn prediction was lower than another structures prediction.

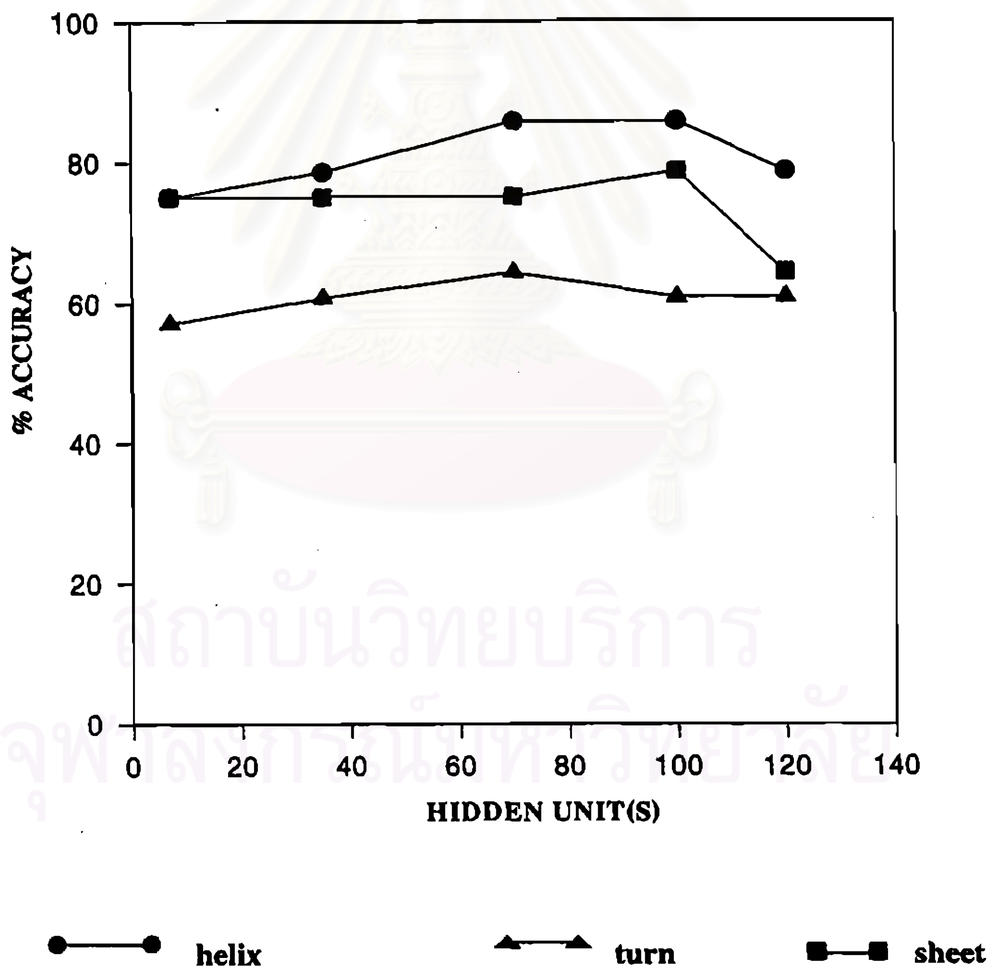

Figure 4.95 The percent accuracy predictions of the percent of helix, sheet, turn (3 groups) in separate networks using amino acid side chain properties.

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

From Figure 4.96, the hydrophobicity could gave highest accuracy (71%) for the helix prediction when using the network with 70 and 100 hidden unit. While another networks of helix prediction gave the accuracy lower than 70% same as sheet prediction. The lowest accuracy was the result from the turn prediction. This hydrophocity gave accuracy for turn prediction lower than 60%. The range of percent helix, sheet and turn (3 groups) prediction using hydrophobicity was between 50-70%.

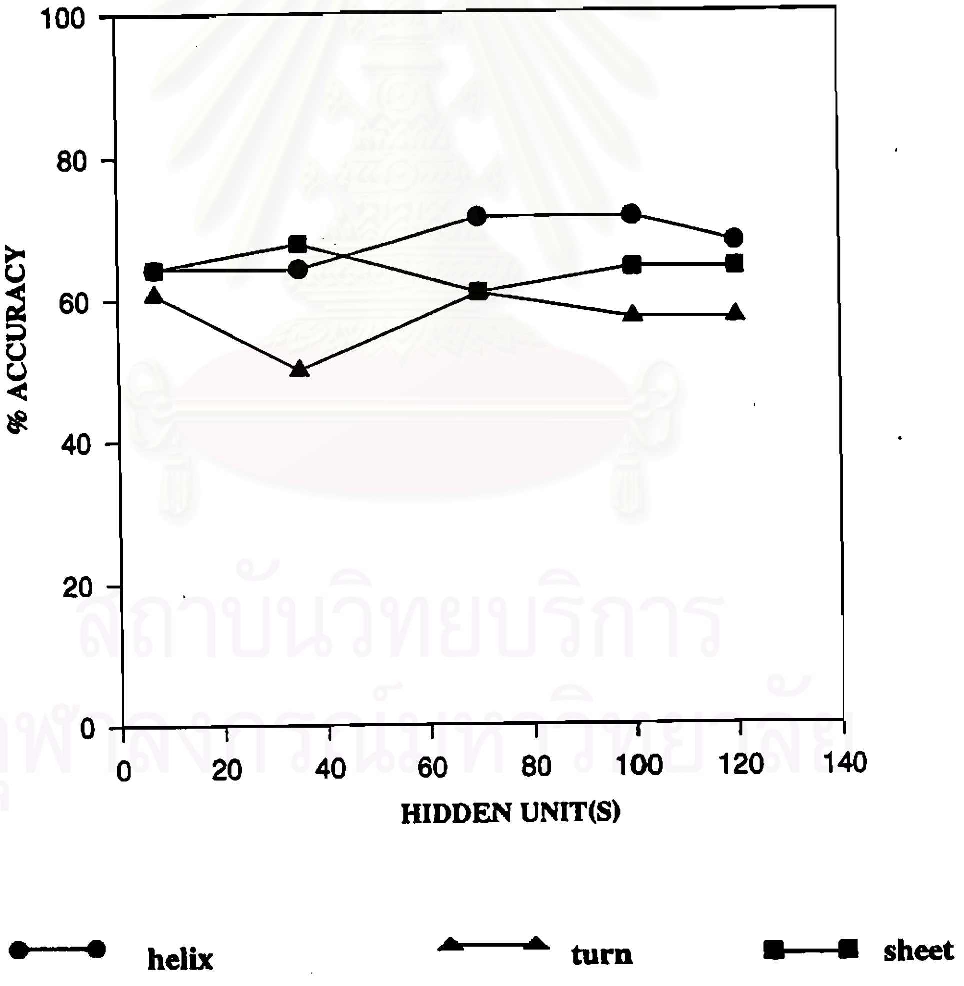

Figure 4.96 The percent accuracy predictions of the percent of helix, sheet, turn (3 groups) in separate networks using hydrophobicity.

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

Training set.

Seventy patterns of amino acid sequences were saved as a learn pattern file for used in training (Table 4.51). The network were trained for the prediction of five groups of percent helix structure using amino acid properties in the section 4.14.1, 4.14.2, 4.14.3, 4.14.4 and 4.14.5

Testing set.

Twenty-eight patterns of amino acid sequences were saved as a test pattern file for used in testing (Table 4.51). The network were tested for the prediction of two groups of percent helix structure using amino acid properties in the section 4.14.1, 4.14.2, 4.14.3, and 4.14.4

Input and output patterns for training and testing.

The input pattern was composed of 481 input units of amino acid which were coded with properties of amino acid residue as previously described (section 3.2.2).

The output pattern had 3 units of percent helix structure. The percent helix structure in amino acid sequence was divided into 2 groups. There were 2 outputs representing the following ranges of percent values 0-15%, and 16-100%. The training set, testing set and the possible outputs are shown in Table 4.51.

A learn-pattern file with 70 patterns was used in training the networks with 7, 35, 70, 100 and 120 hidden units. The maximum value for learn cycles was set to 100000 and the maximum error to stop was set to 0.01. A test-pattern file with 28 patterns was used for testing the trained networks.

These training and testing sets (Table 4.51) were the standard sets for prediction of percent helix structure (2 groups) using various types of properties in the next section (section. 4.14.1, 4.14.2, 4.14.3, and 4.14.4).

Table 4.51 The training set, testing set and possible outputs of percent helix structure (2 groups) prediction.

Training set (Input No.)*	**Testing set (Input No.)***	**Output patterns**	**% of helix structure**
4, 5, 6, 9, 10, 11, 16, 20, 21, 32, 34, 36, 38, 43, 45, 47, 54, 55, 57	58, 64, 68, 81, 82, 87, 96	0 1	1-15
1, 2, 3, 7, 8, 12, 13, 14, 15, 17, 18, 19, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 33, 35, 37, 39, 40, 41, 42, 44, 46, 48, 49, 50, 51, 52, 53, 56, 59, 60, 61, 62, 63, 65, 66, 67, 69, 70, 78, 79, 80	71, 72, 73, 74, 75, 76, 77, 83, 84, 85, 86, 88, 89, 90, 91, 92, 93, 94, 95, 97, 98	1 0	16-100

* Names of input no. were listed in Table 3.1

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

The learn-pattern file, *thy2gr%2h.pat*, and testing-pattern file, *tehy2gr%2h.pat*, with hydropathy (2 groups) coded amino acid sequences were used for training and testing respectively. The trained network and testing result files are shown in Table 4.52. From Table 4.52 and Figure 4.97, predictions with highest accuracy (71.43%) are the results of *hy2gr%2h35.net*, *hy2gr%2h70.net*, *hy2gr%2h100.net* and *hy2gr%2h120.net* which had 35, 70, 100 and 120 hidden units, respectively. While *hy2gr%2h7.net*, with 7 hidden units gave 67.86%, the lowest accuracy.

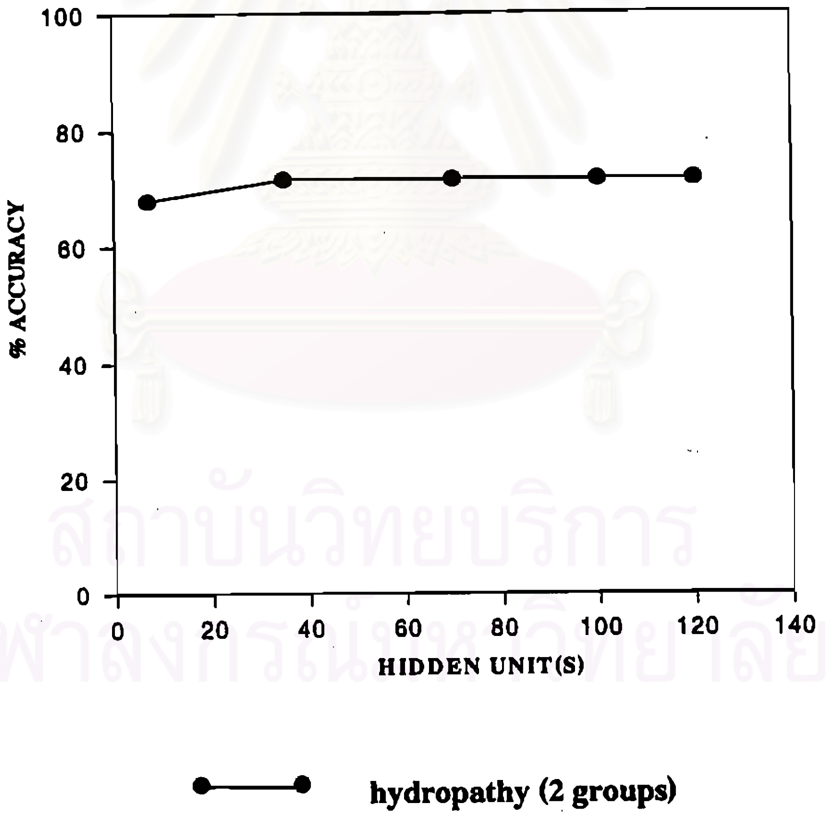

Figure 4.97 Percent accuracy prediction of percent helix structure (2 groups) using 2 groups of hydropathy as input vector with various numbers of hidden units (7, 35, 70, 100, 120).

Table 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).

Hidden units	No. of units	No. of links	Training network file	Testing result file	% accuracy of prediction
7	490	3381	hy2gr%2h7.net	tehy2gr%2h7.res	67.86
35	518	16905	hy2gr%2h35.net	tehy2gr%2h35.res	71.43
70	553	33810	hy2gr%2h70.net	tehy2gr%2h70.res	71.43
100	583	48300	hy2gr%2h100.net	tehy2gr%2h100.res	71.43
120	603	57960	hy2gr%2h120.net	tehy2gr%2h120.res	71.43

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

The learn-pattern file, *thtgr%2h.pat*, and test-pattern file, *tehtgr%2h.pat*, with helical tendencies coded the amino acid sequences were used in training and testing with various hidden units. From Table 4.53 and Figure 4.98 show that *htgr%2h70.net*, which had 70 hidden units, gave highest percent accuracy of 71.43%. While files, *htgr%2h7.net* and *ht%2h35.net* were both 67.8% accurate. The lowest percent accuracy (64.29%) resulted from both *htgr%2h100.net* and *htgr%2h120.net* which had 100 and 120 hidden units respectively.

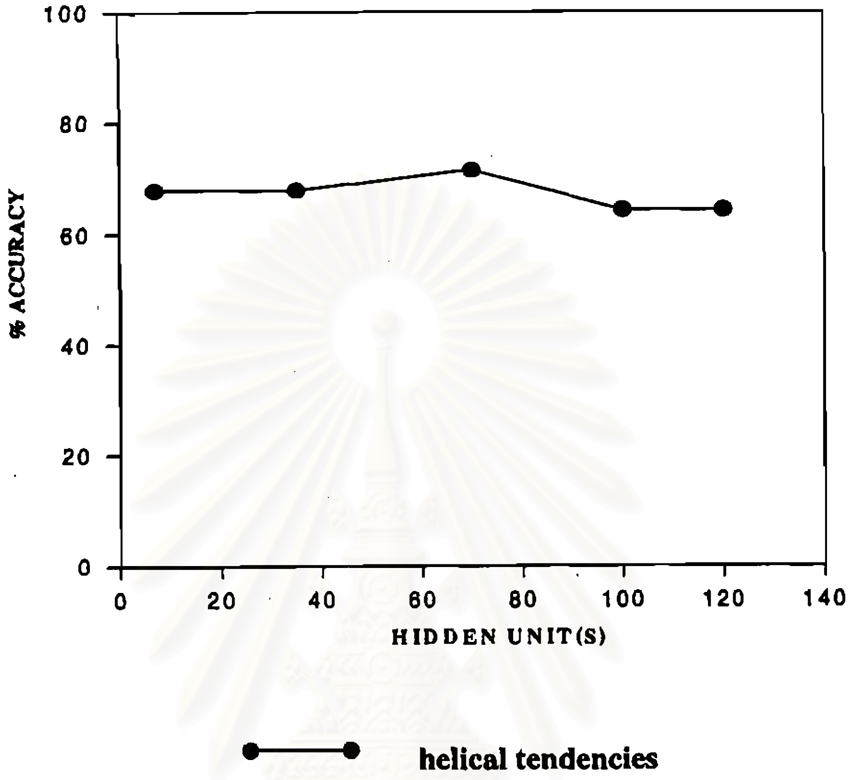

Figure 4.98 Percent accuracy prediction of percent helix structure (2 groups) using helical tendencies as input vector with various numbers of hidden units (7, 35, 70, 100, 120).

Table 4.53 The trained network file, testing result file and percent accuracy of percent helix structure prediction (2 groups) using helical tendencies (5 groups).

Hidden units	No. of units	No. of links	Training network file	Testing result file	% accuracy of prediction
7	490	3381	htgr%2h7.net	tehtgr%2h7.res	67.80
35	518	16905	htgr%2h35.net	tehtgr%2h35.res	67.80
70	553	33810	htgr%2h70.net	tehtgr%2h70.res	71.43
100	583	48300	htgr%2h100.net	tehtgr%2h100.res	64.29
120	603	57960	htgr%h120.net	tehtgr%2h120.res	64.29

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

The learn-pattern file, *taa8gr%2h.pat*, and a test-pattern file, *teaa8gr%2h.pat*, with 8 groups of amino acid side chain properties coded amino acid side chains were used in training network and the testing respectively. Figure 4.99 and Table 4.54 show that *aa8gr%2h100.net* gave the highest percent accuracy of 85.70%, whereas the prediction accuracy from others networks gave lower prediction accuracy with 71.43%.

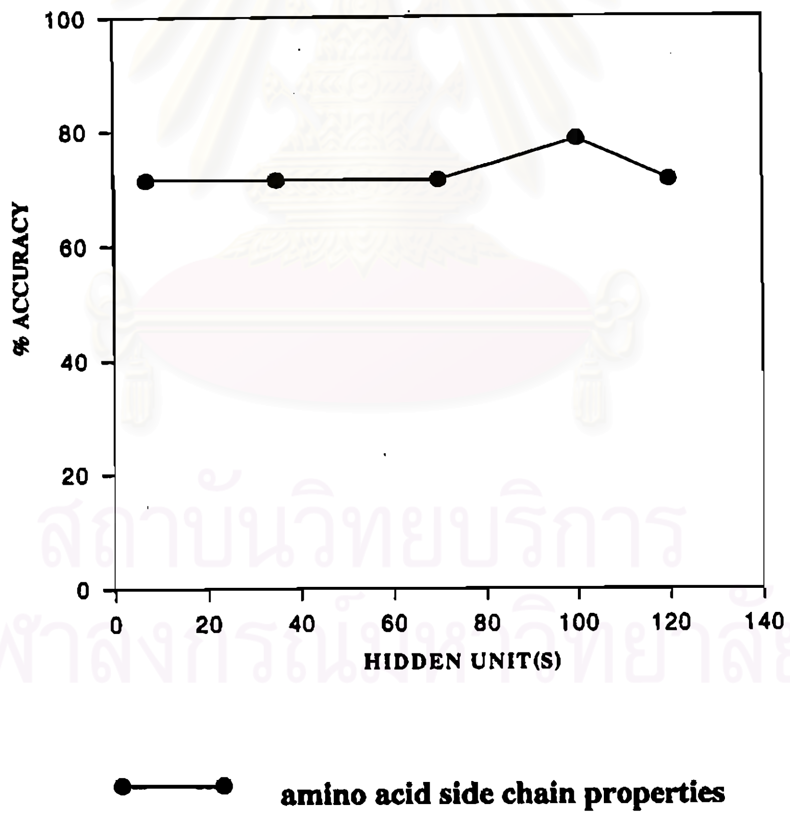

Figure 4.99 Percent accuracy prediction of percent helix structure (2 groups) using 8 groups of amino acid side chain properties as input vector with various numbers of hidden units (7, 35, 70, 100, 120).

Table 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.

Hidden units	No. of units	No. of links	Training network file	Testing result file	% accuracy of prediction
7	490	3381	aa8gr%2h7.net	teaa8gr%2h7.res	71.43
35	518	16905	aa8gr%2h35.net	teaa8gr%2h35.res	71.43
70	553	33810	aa8gr%2h70.net	teaa8gr%2h70.res	71.43
100	583	48300	aa8gr%2h100.net	teaa8gr%2h100.res	78.57
120	603	57960	aa8gr%2h120.net	teaa8gr%2h120.res	71.43

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

The learn-pattern file, *thydro%2h.pat*, and a test-pattern file, *tehydro%2h.pat*, with hydrophobicity coded amino acid sequences were used in training network and the testing respectively. Table 4.55 and Figure 4.100, the predictions with highest accuracy (78.57%) is the results from *hydro%2h120.net* which had 120 hidden units. File, *hydro%2h70.net* with 70 hidden units gave 75% accurate whereas, *hydro%2h7.net*, *hydro%2h35.net* and *hydro%2h100.net* all gave 71.43%.

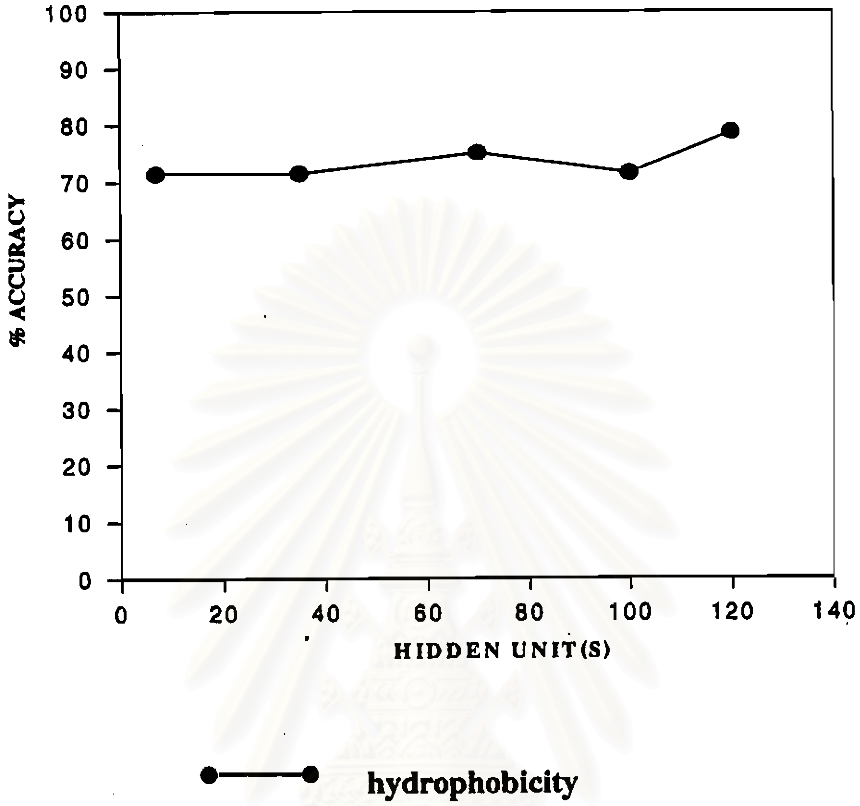

Figure 4.100 Percent accuracy prediction of percent helix structure (2 groups) using hydrophobicity as input vector with various numbers of hidden units (7, 35, 70, 100, 120).

Table 4.55 The trained network file, testing result file and percent accuracy of percent helix structure prediction (2 groups) using hydrophobicity.

Hidden units	No. of units	No. of links	Training network file	Testing result file	% accuracy of prediction
7	490	3381	hydro%2h7.net	tehydro%2h7.res	71.43
35	518	16905	hydro%2h35.net	tehydro%2h35.res	71.43
70	553	33810	hydro%2h70.net	tehydro%2h70.res	75.00
100	583	48100	hydro%2h100.net	tehydro%2h100.res	71.43
120	603	57960	hydro%2h120.net	tehydro%2h120.res	78.57

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

Figure 4.101 shows that most of percent accuracy prediction of 2 groups of percent helix structure were in the range between 60-80%. The highest percent accuracy was 78.7% (*hydro%2h120.net* and *aa8gr%2h100.net*). These networks were trained from hydrophobicity and the amino acid side chain properties respectively. Apparently these properties also gave overall percent accuracy more than 70%. The lowest percent accuracy was 64.29% (*ht%2h100.net* and *ht%2h120.net*). These networks were trained from helical tendencies properties.

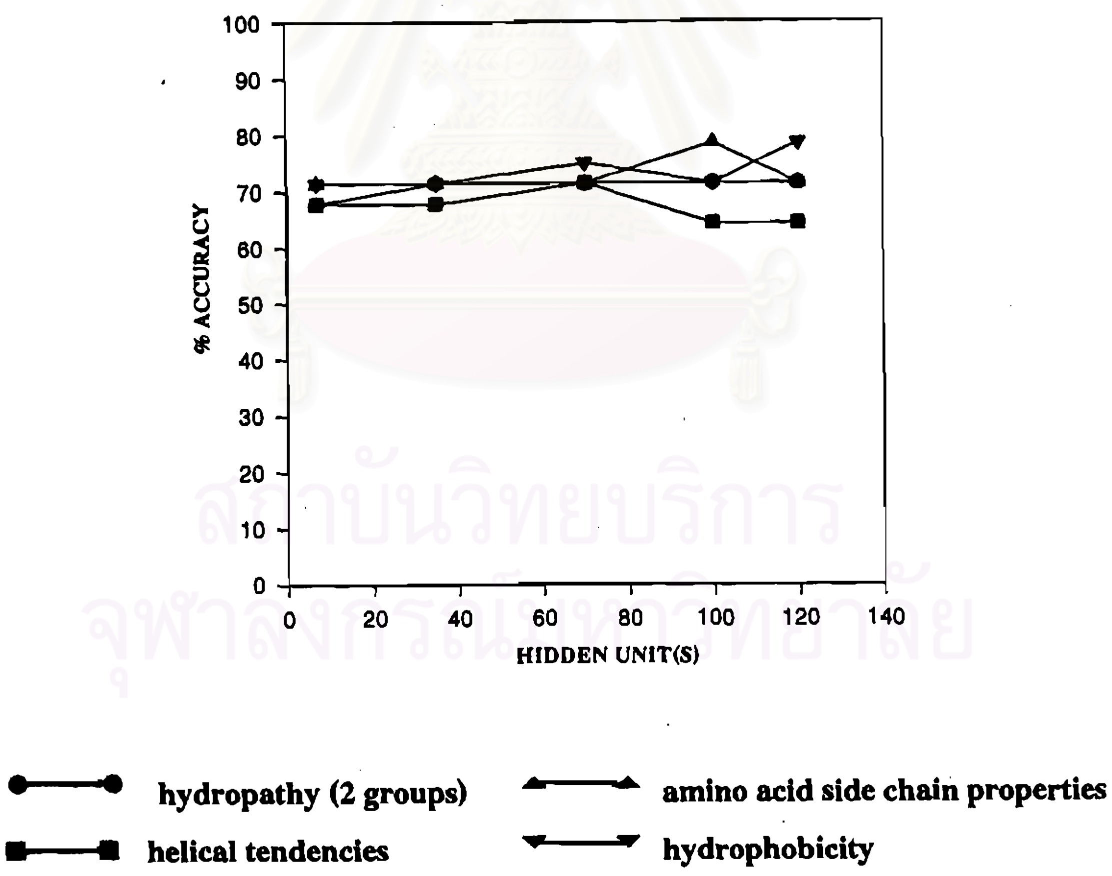

Figure 4.101 The comparison of percent accuracy prediction of percent helix structure (2 groups) using various types of input properties.

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

Training set.

Seventy patterns of amino acid sequences were saved as a learn pattern file for used in training (Table 4.56). The network were trained for the prediction of five groups of percent sheet structure using amino acid properties in the section 4.15.1, 4.15.2 and 4.15.3.

Testing set.

Twenty-eight patterns of amino acid sequences were saved as a test pattern file for used in testing (Table 4.56). The network were tested for the prediction of two groups of percent sheet structure using amino acid properties in the section 4.15.1, 4.15.2, 4.15.3 and 4.15.4.

Input and output patterns for training and testing.

The input pattern was composed of 481 input units of amino acid which were coded with properties of amino acid residue as previously described (section 3.2.2).

The output pattern had 2 units of percent sheet structure. The percent sheet structure in amino acid sequence was divided into 2 groups. There were 2 possible outputs representing the following ranges of percent values 0-15% and 15-100%. The training set, testing set and the possible outputs are shown in Table 4.56.

A learn-pattern file with 70 patterns was used in training the networks with 7, 35, 70, 100 and 120 hidden units. The maximum value for learn cycles was set to 100000 and the maximum error to stop was set to 0.01. A test-pattern file with 28 patterns was used for testing the trained networks.

These training and testing sets (Table 4.56) were the standard sets for prediction of percent sheet structure (2 groups) using various types of properties in the next section (section. 4.15.1, 4.15.2 and 4.15.3.)

Table 4.56 The training set, testing set and possible outputs of percent sheet structure (2 groups) prediction.

Training set (Input No.)*	**Testing set (Input No.)***	**Output patterns**	**% of helix structure**
1, 7, 13, 14, 18, 19, 23, 25, 28, 33, 37, 40, 48, 49, 51, 53, 59, 60, 61, 65, 70, 78, 79, 83, 84	72, 73, 74, 77, 86, 87, 88, 90, 91, 94, 95	0 1	1-15
2, 3, 4, 5, 6, 8, 9, 10, 11, 12, 15, 16, 17, 20, 21, 22, 24, 26, 27, 29, 30, 31, 32, 34, 35, 36, 38, 39, 41, 42, 43, 44, 45, 46, 47, 50, 52, 54, 55, 56, 57, 58, 62, 63, 64	66, 67, 68, 69, 71, 75, 76, 80, 81, 82, 85, 92, 93, 96, 97, 98	1 0	16-100

* Names of input no. were listed in Table 3.1

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

The learn-pattern file, *thy2gr%2s.pat*, and testing-pattern file, *tehy2gr%2s.pat*, with hydropathy (2 groups) coded amino acid sequences were used for training and testing respectively. Table 4.57 and Figure 4.102, shows the best prediction was the result from *hy2gr%2s35.net* and *hy2gr%2s120.net*. These networks with 35 and 120 hidden units, gave 75.00% accuracy, the highest in its groups. The prediction accuracy from others networks were lower. Both *hy2gr%2s7.net*, and *hy2gr%2s120.net* with 7 and 120 hidden units respectively, gave 71.43% accurate. Whereas, *hy2gr%2s70.net* gave the lowest accuracy of 64.29%.

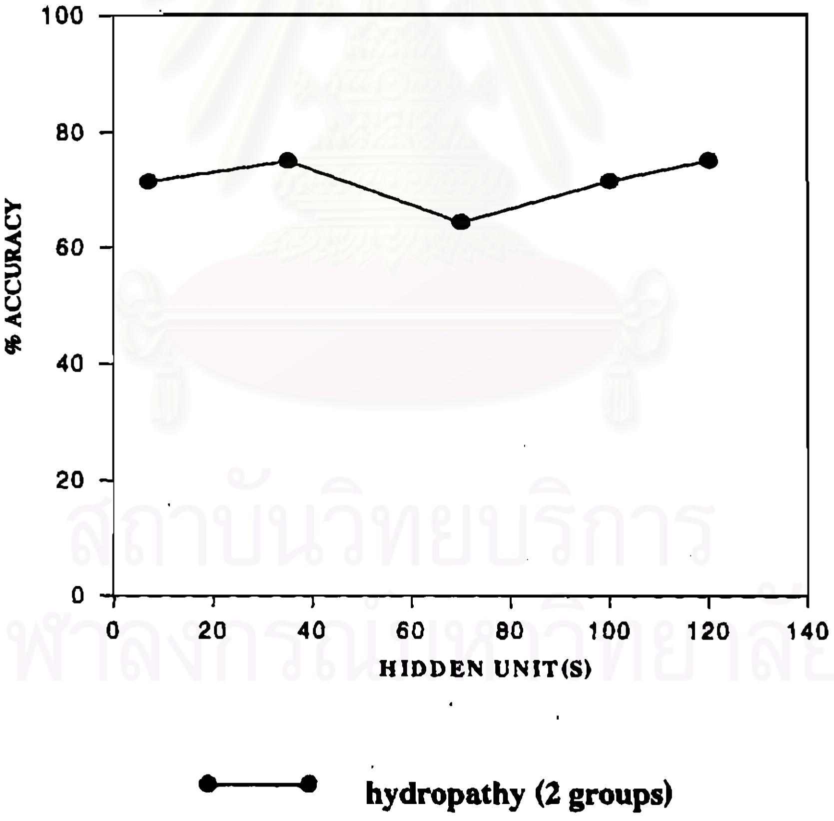

Figure 4.102 Percent accuracy prediction of percent sheet structure (2 groups) using 2 groups of hydrophathy as input vector with various numbers of hidden units (7, 35, 70, 100, 120).

Table 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).

Hidden units	No. of units	No. of links	Training network file	Testing result file	% accuracy of prediction
7	490	3381	hy2gr%2s7.net	tehy2gr%2s7.res	71.43
35	518	16905	hy2gr%2s35.net	tehy2gr%2s35.res	75.00
70	553	33810	hy2gr%2s70.net	tehy2gr%2s70.res	64.29
100	583	48300	hy2gr%2s100.net	tehy2gr%2s100.res	71.43
120	603	57960	hy2gr%2s120.net	tehy2gr%2s120.res	75.00

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

The learn-pattern file, *taa8gr%2s.pat*, and a test-pattern file, *teaa8gr%2s.pat*, with 8 groups of amino acid side chain properties coded amino acid side chains were used in training network and the testing respectively. Figure 4.103 and Table 4.58 show that *aa8gr%2s7.net* gave the highest percent accuracy of 67.86%. The prediction accuracy from others networks were lower. The trained networks, *aa8gr%2s35.net*, and *aa8gr%2s70.net* *aa8gr%2s100.net* and *aa8gr%2s120* all gave 60.71% accuracy.

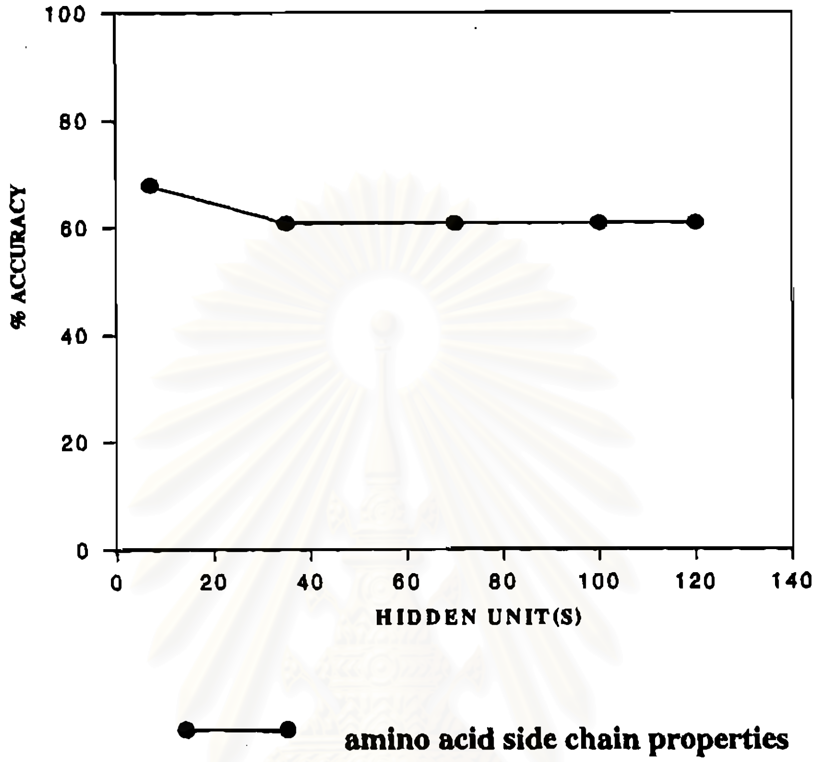

Figure 4.103 Percent accuracy prediction of percent sheet structure (2 groups) using 8 groups of amino acid side chain properties as input vector with various numbers of hidden units (7, 35, 70, 100, 120).

Table 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.

Hidden units	No. of units	No. of links	Training network file	Testing result file	% accuracy of prediction
7	490	3381	aa8gr%2s7.net	teaa8gr%2s7.res	67.86
35	518	16905	aa8gr%2s35.net	teaa8gr%2s35.res	60.71
70	553	33810	aa8gr%2s70.net	teaa8gr%2s70.res	60.71
100	583	48300	aa8gr%2s100.net	teaa8gr%2s100.res	60.71
120	603	57960	aa8gr%2s120.net	teaa8gr%2s120.res	60.71

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

The learn-pattern file, *thydro%2s.pat*, and a test-pattern file, *tehydro%2s.pat*, with hydrophobicity coded amino acid sequences were used in training network and the testing respectively. Table 4.59 and Figure 4.104, the prediction from *hydro%2s7.net* is 64.29% accurate, whereas *hydro%2s35.net* gave 60.71% accuracy. File, *hydro%2s70.net hydro%2s100.net* and *hydro%2s120.net* gave the lowest accuracy of 57.14%.

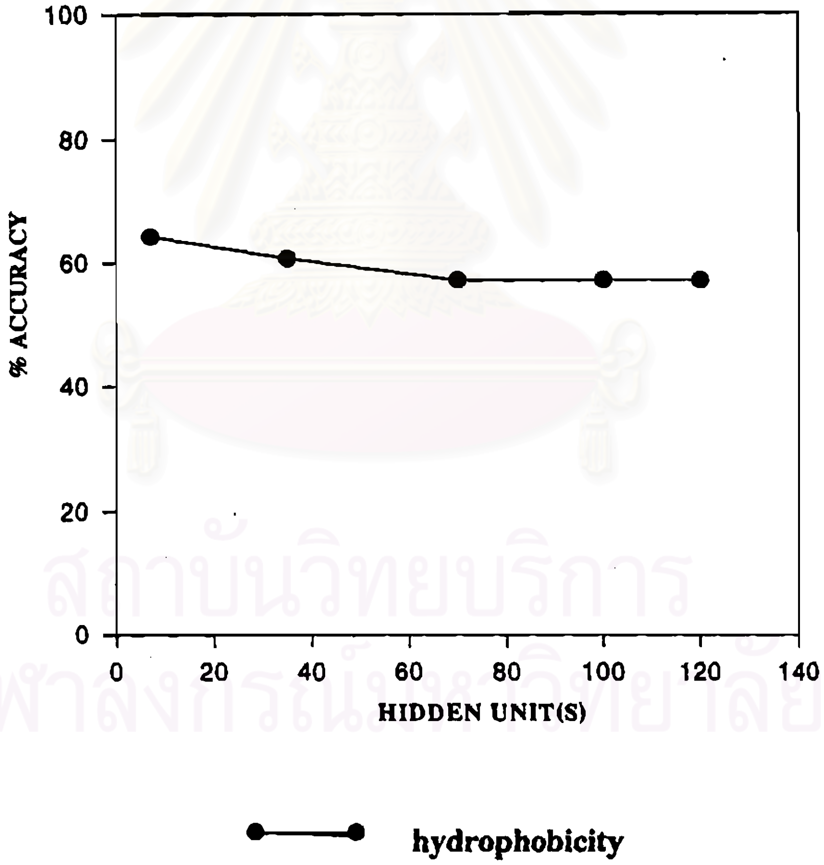

Figure 4.104 Percent accuracy prediction of percent sheet structure (2 groups) using hydrophobicity as input vector with various numbers of hidden units (7, 35, 70, 100, 120).

Table 4.59 The trained network file, testing result file and percent accuracy of percent sheet structure prediction (2 groups) using hydrophobicity.

Hidden units	No. of units	No. of links	Training network file	Testing result file	% accuracy of prediction
7	491	3381	hydro%2s7.net	tehydro%2s7.res	64.29
35	519	16905	hydro%2s35.net	tehydro%2s35.res	60.71
70	554	33810	hydro%2s70.net	tehydro%2s70.res	57.14
100	584	48300	hydro%2s100.net	tehydro%2s100.res	57.14
120	604	57960	hydro%2s120.net	tehydro%2s120.res	57.14

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

Figure 4.105 shows that most of percent accuracy prediction of 5 groups of percent sheet structure were in the range between 57-75%. The highest percent accuracy was 75% (*hy2gr%2s35.net* and *hy2gr%2s120*). The lowest percent accuracy was 57.14% (*hydro%2s70.net, hydrogr%2s100.net* and *hydro%2s120.net*).

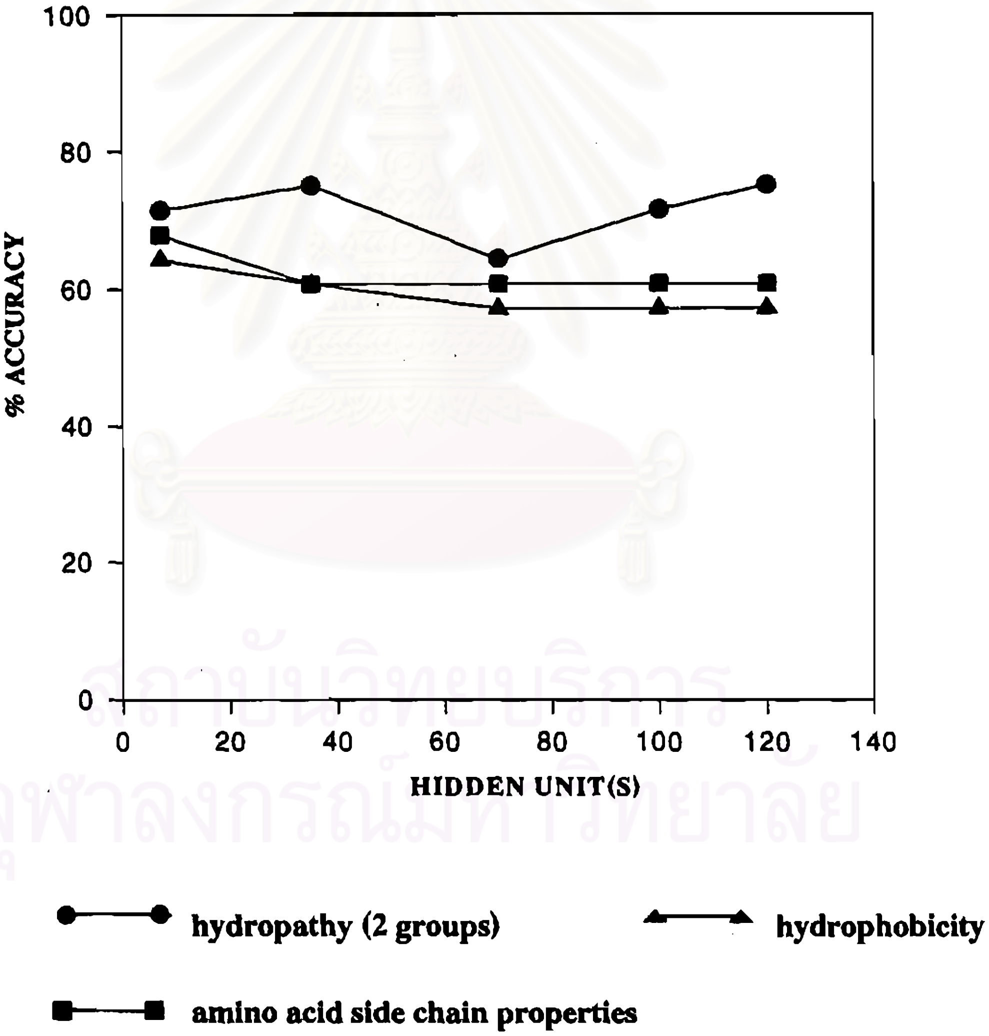

Figure 4.105 The comparison of percent accuracy prediction of percent sheet structure (2 groups) using various types of input properties.