  Bull. Chem. Soc. Jpn., 55, 575—580 (1982)

Computation of Amino Acid Sequences of Polypeptides from Masses of Their Constituent Peptide Fragments and Amino Acid Residues Released in Edman-degradation[1)]

Toyoko KITAGISHI, Yeong-Man HONG, Toshifumi TAKAO, Saburo AIMOTO, and Yasutsugu SHIMONISHI*

Institute for Protein Research, Osaka University, Suita, Osaka 565

(Received July 31, 1981)

A computer program ("PROSEQ2") was designed for determining the amino acid sequence of polypeptides from molecular weights and *N*-terminal amino acid residues of two or more sets of peptide mixtures obtained from polypeptides by two or more specific cleavage methods, together with 3-phenyl-2-thiohydantoin derivatives of amino acids released from two or more sets of peptide mixtures by Edman-degradation. Computation of sequences of several polypeptides using this program is described.

Gas chromatography/mass spectrometry is in use for sequencing a peptide mixture obtained from a polypeptide or protein by partial hydrolysis followed by conversion into volatile derivatives. Numerous experimental data thus obtained were handled with computer programs[2,3)] capable of deducing the sequence of the original polypeptide or protein.

Recently[4–6)] we reported a new method for sequencing a peptide mixture by combination of Edman-degradation and field-desorption mass spectrometry. This method does not involve any extensive chemical processing to convert peptides into volatile derivatives. The principle of this method is to measure mass values of individual "underivatized" peptides in a mixture and their degraded peptide fragments by field-desorption mass spectrometry. *N*-Terminal amino acid residues of individual peptides in a mixture can be determined by calculation of possible mass differences before and after degradation. Amino acid sequences of individual peptides in a mixture can be determined by repeating this operation. If two or more sets of peptide mixtures are to be produced from a polypeptide or protein by two or more specific cleavage methods, the amino acid sequence of the polypeptide or protein can be determined, by use of the computer program "PROSEQ1,"[7)] from the molecular weights and *N*-terminal partial amino acid sequences which have been obtained by the above procedure for the peptides present in the resulting peptide mixtures. This method does not necessarily require determination of complete sequences or major portions of sequences for individual peptides.

After attempting sequence determinations on various polypeptides by this new method, we have succeeded in designing a new program which necessarily requires no amino acid sequences of constituent peptides and which can determine the sequence from molecular weights and *N*-terminal amino acid residues of constituent peptides in two or more sets of peptide mixtures and 3-phenyl-2-thiohydantoin derivatives of amino acids (denoted as PTH-amino acids) released in each cycle of Edman-degradation.

In this paper, we describe some applications of the new program (named "PROSEQ2") to determining amino acid sequences for several polypeptides; sequencing of a polypeptide of unknown structure[8)] is included.

Methods

The computer program "PROSEQ2" described here is written in FORTRAN, its flow chart being outlined in Fig. 1. The single letter abbreviations for amino acid residues presented by Eck and Dayhoff[9)] are adopted as input and output words. The ACOS 700 computer in the Crystallographic Research Center (Institute for Protein Research, Osaka University) was used for calculation.

The analytical data used for the calculation are as follows:

1) The amino acid composition of the sample polypeptide or protein as integer values, as obtained from an amino acid analysis of acid hydrolysate.

2) The *N*- and *C*-terminal amino acid residues of the sample polypeptide or protein, as obtained by well known methods.

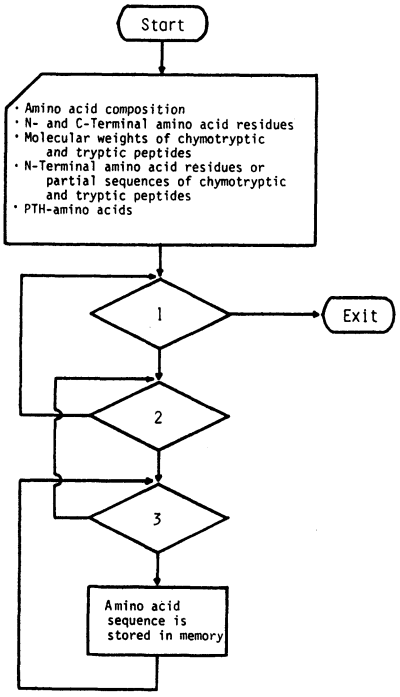


Fig. 1. Flow chart of the computer program "PROSEQ2." The numerals correspond to those given in the explanation of the program in the text.

3) The molecular weights and *N*-terminal amino acid residues of the individual peptides in two or more sets of peptide mixtures prepared from the sample polypeptide or protein by specific cleavage methods, as determined as described in Refs. 4 and 6. Peptide mixtures are degraded by the Edman method[10] with modifications[11] favorable for mass measurements. Mass spectra are obtained with a second-order double focusing mass spectrometer[12] with a mono field-desorption ion source.

4) The kind and compositional ratio, as integer values within some limits between possible maxima and minima, of the PTH–amino acids released from two or more sets of peptide mixtures in each cycle of Edman-degradation and analyzed by high-performance liquid chromatography (HPLC).[13]

In this paper, we use analytical data on peptides obtained from polypeptides by specific cleavage with chymotrypsin and trypsin. Amino acid sequences of polypeptides or proteins are sought by the computer in the following order:

1) Amino acid combinations in chymotryptic peptides are examined in the amino acid composition of the sample polypeptide or protein, and each of the combinations is simultaneously compared with the PTH–amino acids released in each cycle of degradation.

2) Tryptic peptides are aligned in a given order from the *N*-terminus. The tryptic peptide placed at the *N*-terminus has the same *N*-terminal amino acid residue as that of the sample polypeptide or protein. In this case, no amino acid combinations in each tryptic peptide are examined.

3) Chymotryptic peptides are aligned by comparison with the alignment of tryptic peptides from the *N*-terminal position. In this step, the following procedures are carried out simultaneously:

i) Comparison of the molecular weight from the *N*-terminus to the chymotryptic peptide in question with that from the *N*-terminus to the overlapping tryptic peptide.

ii) Collation of the known *N*-terminal amino acid residue of the chymotryptic peptide in question to those of tryptic peptides or the PTH–amino acids released from tryptic peptides.

iii) Collation of the possible amino acid residues in chymotryptic peptides to the PTH–amino acids released from both chymotryptic and tryptic peptides.

iv) Examination of possible similarities of the known *N*-terminal amino acid residues of tryptic peptides with unknown sequences of chymotryptic peptides and also with the PTH–amino acids released from chymotryptic peptides.

In these processes, data on chymotryptic and tryptic peptides are of course interchangeable.

Results and Discussion

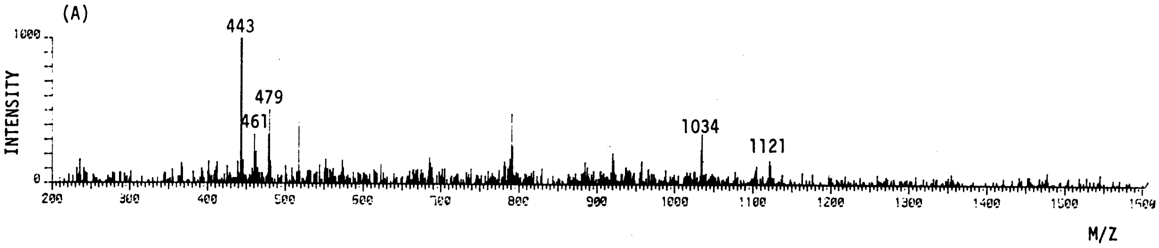
In a previous paper,[7] we showed that amino acid sequences of polypeptides or proteins can be deduced from the molecular weights and *N*-terminal partial sequences of their constituent peptides by the computer program "PROSEQ1." By this program, complete sequences or major portions of amino acid sequences of polypeptides or proteins can be determined from sequence

```
INPUT DATA

 AMINO ACID           MOLE WEIGHTS,N-TERMINAL SEQUENCES AND PTH-AMINO ACIDS IN EACH CYCLE
COMPOSITION
                        CHYMOTRYPTIC PEPTIDES                       TRYPTIC PEPTIDES

                                    2   3   4                                    2   3

   G    1               396  S*  G  0    0  1,0            174 R   G  0    0
   A    1               431  V   A  0    0   0             652 Y   A 1,0   0
   U    0               477  L   U  0    0   0            1351 A   U  0    0
   S    4               484  T   S 2,0 2,0   0            1356 H*  S 1,0   0
   P    0               675  H   P  0    0   0                     P  0    0
   V    1              1106  L   V  0    0   0                     V  0    0
   T    3                        T 1,0   0  1,0                    T  0    0
   L    2                        L  0    0   0                     L 1,0   0
   I    0                        I  0    0   0                     I  0    0
  N+D   4                        N  0   1,0  0                     N  0    0
  Q+E   3                        D 2,0 1,0   0                     D  0   2,0
   K    1                        Q 1,0 2,0   0                     Q 1,0  2,0
   M    1                        K 1,0   0   0                     K  0    0
   H    1                        E  0    0   0                     E  0    0
   F    2                        M 2,0   0   0                     M  0    0
   R    2                        H  0    0   0                     H  0    0
   Y    2                        F  0    0   0                     F  0    0
   W    1                        R  0    0  1,0                    R  0    0
   C    0                        Y  0   1,0 1,0                    Y  0    0
                                 W  0   1,0  0                     W  0    0
TOTAL 29                         C  0    0   0                     C  0    0

 N-TERMINAL AMINO ACID   H
 C-TERMINAL AMINO ACID   T

CANDIDATE SEQUENCES

   1    HSQGTFTSDYSKYLDSRRAQDFVQWLMNT                                          2

   2    HSQGTFTDSYSKYLDSRRAQDFVQWLMNT                                          2

                                                                               3
```

Fig. 2. Computer out-put sequences of glucagon, sought from the molecular weights, the *N*-terminal amino acid residues and the kind and ratio of PTH–amino acids released by Edman-degradation of the chymotryptic and tryptic peptides, respectively. Peptides with asteriscs (*) contain a Lys residue.[4,6] Numerals on the right of sequences show the number of combinations of amino acid residues of chymotryptic peptides and the total number.

```
INPUT DATA

 AMINO ACID          MOLE WEIGHTS,N-TERMINAL SEQUENCES AND PTH-AMINO ACIDS IN EACH CYCLE
COMPOSITION
                      CHYMOTRYPTIC PEPTIDES                          TRYPTIC PEPTIDES

                                   3   4                                        2   3

   G   1              396 SK   G  0   1,0                       174 R    G  0    0
   A   1              431 VQ   A  0    0                        652 Y    A 1,0   0
   U   0              477 LM   U  0    0                       1351 A    U  0    0
   S   4              484 TS   S 2,0   0                       1356 H*   S 1,0   0
   P   0              675 HS   P  0    0                                 P  0    0
   V   1             1106 LD   V  0    0                                 V  0    0
   T   3                       T  0   1,0                                T  0    0
   L   2                       L  0    0                                 L 1,0   0
   I   0                       I  0    0                                 I  0    0
  N+D  4                       N 1,0   0                                 N  0    0
  Q+E  3                       D 1,0   0                                 D  0   2,0
   K   1                       Q 2,0   0                                 Q 1,0  2,0
   M   1                       K  0    0                                 K  0    0
   H   1                       E  0    0                                 E  0    0
   F   2                       M  0    0                                 M  0    0
   R   2                       H  0    0                                 H  0    0
   Y   2                       F  0    0                                 F  0    0
   W   1                       R  0   1,0                                R  0    0
   C   0                       Y 1,0  1,0                                Y  0    0
                               W 1,0   0                                 W  0    0
TOTAL 29                       C  0    0                                 C  0    0

 N-TERMINAL AMINO ACID   H
 C-TERMINAL AMINO ACID   T

CANDIDATE SEQUENCES

   1    HSQGTFTSDYSKYLDSRRAQDFVQWLMNT                                                    2

                                                                                         2
```

Fig. 3. Computer out-put sequence of glucagon, sought from the molecular weights, the *N*-terminal sequences of two amino acid residues and the kind and ratio of PTH–amino acids released by Edman-degradation of the chymotryptic and tryptic peptides, respectively.



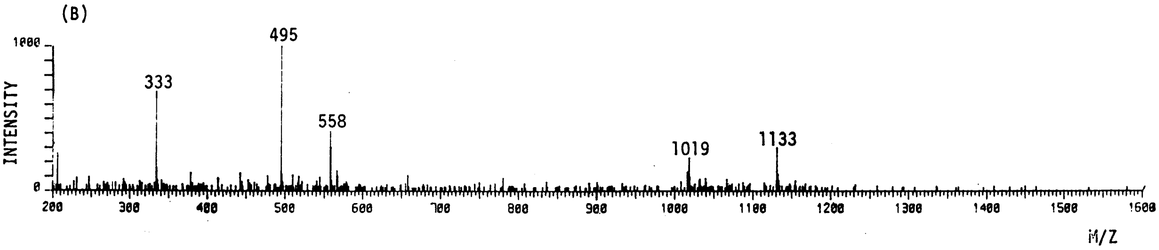


Fig. 4 Field-desorption mass spectra of the chymotryptic (A) and tryptic (B) peptides of human β-endorphin. The spectra were measured with a JMS-DX300 double focusing mass spectrometer equipped with a JMA-2000 mass data analysis system.

information on constituent peptides, but with increasing difficulty with increase in the number of amino acid residues. Therefore, increasing numbers of mass measurements on degraded peptide fragments are necessary to obtain sufficient information on the amino acid sequence of constituent peptides for deducing almost complete amino acid sequences. The present program uses information obtained on PTH–amino acids liberated from peptide mixtures by Edman-degradation in place of that on amino acid sequences of constituent peptides, because using HPLC makes it easy to determine PTH–amino acids liberated from peptide mixtures by successive Edman-degradation.

Figures 2, 3, 5, and 6 show computer output sequences of some polypeptides. The first example (Fig. 2) illustrates the sequence of glucagon, which has been deduced from the molecular weights and *N*-terminal amino acid residues of the chymotryptic and tryptic peptides (Table 1) and the PTH–amino acids released in each cycle of Edman-degradation (Table 2). The output sequences have exchangeable residues (Ser and Asp) at the 8th and 9th positions from the *N*-terminus. This is because PTH–Ser (S) and PTH–Asp (D) were liberated together in the 2nd and 3rd cycles of degradation from the chymotryptic peptides, and the ratio of serine to other amino acids was inaccurate, because

```
INPUT DATA

 AMINO ACID        MOLE WEIGHTS,N-TERMINAL SEQUENCES AND PTH-AMINO ACIDS IN EACH CYCLE
COMPOSITION
                        TRYPTIC PEPTIDES                      CHYMOTRYPTIC PEPTIDES

                                   2   3                                  2   3   4

   G    3              332 K    G 3,0 3,0                 442 Y    G 2,0 2,0 1,0
   A    2              494 N*   A 2,0  0                  460 K*   A 2,0 1,0  0
   U    0              557 N*   U  0   0                  478 V    U  0   0   0
   S    2             1018 Y*   S  0  1,0                1033 K*   S  0  1,0  0
   P    1             1132 S*   P  0   0                 1120 M*   P  0   0  1,0
   V    1                       V  0   0                           V  0   0   0
   T    3                       T 1,0 1,0                          T 2,0 1,0  0
   L    2                       L  0   0                           L  0  1,0  0
   I    2                       I  0  1,0                          I  0   0  1,0
  N+D   2                       N  0   0                           N 2,0  0   0
  Q+E   3                       D  0   0                           D  0   0   0
   K    5                       Q 2,0  0                           Q 1,0  0   0
   M    1                       K 1,0  0                           K 2,0  0  1,0
   H    0                       E 1,0 2,0                          E 1,0 1,0 4,0
   F    2                       M  0   0                           M  0   0   0
   R    0                       H  0   0                           H  0   0   0
   Y    2                       F  0   0                           F  0   0  2,0
   W    0                       R  0   0                           R  0   0   0
   C    0                       Y  0  1,0                          Y  0  1,0  0
                                W  0   0                           W  0   0   0
TOTAL 31                        C  0   0                           C  0   0   0

 N-TERMINAL AMINO ACID   Y
 C-TERMINAL AMINO ACID   E

CANDIDATE SEQUENCES

  1   YGGFMTSEKSQTPLVTLFKNAIIKNAYKKGE                                          22

                                                                               24
```

Fig. 5. Computer out-put sequence of human β-endorphin, sought from the molecular weights, the *N*-terminal amino acid residues and the kind and ratio of PTH-amino acids released by Edman-degradation of the tryptic and chymotryptic peptides, respectively.

```
INPUT DATA

 AMINO ACID        MOLE WEIGHTS,N-TERMINAL SEQUENCES AND PTH-AMINO ACIDS IN EACH CYCLE
COMPOSITION
                    CHYMOTRYPTIC PEPTIDES                         TRYPTIC PEPTIDES

                          2   3   4   5   6   7                        2   3   4   5   6   7

   G    7        252 A   G 1,0 1,0 1,0 1,0 1,0  0          174 R    G  0   0  1,0 2,0 2,0 1,0
   A    5        385 N   A 2,0 1,0 2,0 1,0  0   0          626 F*   A 1,0  0  1,0 1,0  0   0
   U    1        439 N   U 1,0 1,0 1,0 1,0 1,0 1,0         825 A*   U  0   0   0  1,0 1,0 1,0
   S    2        495 Q   S  0  1,0  0   0   0  1,0        1309 A    S 1,0  0   0   0   0   0
   P    1        520 K   P  0  1,0  0   0   0   0         1622 N*   P 1,0  0   0   0   0   0
   V    4        532 N   V 1,0 2,0  0   0   0   0         1817 N    V 1,0 1,0 1,0  0   0   0
   T    3        655 F   T 2,0  0  1,0  0  1,0 1,0                  T 2,0  0   0  1,0 1,0  0
   L    0        924 V*  L  0   0   0   0   0   0                   L  0   0   0   0   0   0
   I    2        934 W*  I 1,0  0   0   0  1,0  0                   I  0   0   0   0  2,0  0
  N+D   7       1309 A   N 2,0  0   0   0   0   0                   N  0   0   0   0   0   0
  Q+E   6                D  0  1,0  0   0   0  1,0                  D  0  1,0  0   0   0  2,0
   K    3                Q 1,0 2,0 1,0  0   0   0                   Q  0  1,0 1,0 1,0 1,0 1,0
   M    0                K 1,0  0  1,0 1,0  0   0                   K  0   0   0  1,0  0   0
   H    1                E 1,0 1,0 1,0 1,0 1,0 1,0                  E  0  2,0 1,0 1,0 1,0  0
   F    2                M  0   0   0   0   0   0                   M  0   0   0   0   0   0
   R    2                H  0   0   0   0   0   0                   H  0   0   0   0   0   0
   Y    6                F  0   0   0   0   0  1,0                  F  0   0  1,0  0   0   0
   W    3                R  0   0   0  1,0 1,0  0                   R  0   0   0   0   0   0
   C    0                Y 2,0 1,0 2,0 2,0  0  1,0                  Y  0  2,0  0  1,0 1,0 1,0
                         W  0  1,0 1,0 1,0 1,0 1,0                  W  0   0  1,0 1,0 1,0 1,0
TOTAL 55                 C  0   0   0   0   0   0                   C  0   0   0   0   0   0

 N-TERMINAL AMINO ACID   A
 C-TERMINAL AMINO ACID   U

CANDIDATE SEQUENCES

  1   ATVAGID(G,S,V,Q,H)RNVDWQYWWNQGKRFAYVKATEGTGYKNPYFAQQYNGSYNIGU            3242

                                                                              9276
```

Fig. 6. Computer out-put sequence of the *N*-terminal BrCN fragment of *Streptomyces erythraeus* lysozyme,[14] sought from the molecular weights, the *N*-terminal amino acid residues and the kind and ratio of PTH-amino acids released by Edman-degradation of the chymotryptic and tryptic peptides, respectively.

the recovery of PTH–Ser is generally poor in Edman-degradation. The output sequences suggest that if sequences of two amino acid residues from the *N*-terminus of each peptide are determined as described previously,[4–6] either of the two sequences output is determined. In fact, one sequence was output as illustrated in Fig. 3.

The second example is a sequence determination for human β-endorphin. The mass spectra of the tryptic and chymotryptic peptides are shown in Fig. 4. While the tryptic peptides gave five intense mass peaks of peptides, as expected from the content of basic amino acid residues (4 Lys residues), the chymotryptic peptides gave a number of mass peaks as $[M+H]^+$ and/or $[M+2H]^{2+}$ of expected peptides together with weak mass peaks of peptides, resulting from unspecific cleavage

Table 1. Molecular weights and *N*-terminal amino acid residues of proteolytic fragments of several polypeptides, determined by field-desorption mass spectrometry and Edman-degradation

Glucagon[a]		Human β-endorphin[b]		BrCN *N*-terminal fragment of SE lysozyme[c]	
Chymotryptic	Tryptic	Chymotryptic	Tryptic	Chymotryptic	Tryptic
1106 L	1356 H	1120 M	1132 S	1309 A	1817 N
675 H	1351 A	1033 K	1018 Y	934 W	1622 N
484 T	652 Y	478 V	557 N	924 V	1309 A
477 L	174 R	460 K	494 N	655 F	825 A
431 V		442 Y	332 K	532 N	626 F
396 S				520 K	174 R
				495 Q	
				439 N	
				385 N	
				252 A	

a) Data are cited from Ref. 4. b) Data are based on the mass values in Fig. 4 and PTH–amino acids in Table 3. c) Data are cited from Ref. 8.

Table 2. Recoveries (nmol × 10) of PTH–amino acids released from chymotryptic (1.20 mg) and tryptic (0.66 mg) peptides of glucagon by Edman-degradation

Thr was estimated in addition to *Δ*-Thr. The 3-phenyl-2-thiohydantoin derivatives of Asn and Gln and Ser and of Ala and Tyr were not satisfactorily separated in some cases. In these cases (*), recoveries were estimated as those of the main component.

	Chymotryptic peptides				Tryptic peptides		
	1	2	3	4	1	2	3
Asp	34	74	44	2	3	4	78
Asn	5	—	41*	1	4		—
Glu		10	19			18	31
Gln		12	30*	2		25	45
Ser	—	—	—		—	—	
Thr	66	12	3	34	6	2	
Gly	8	6	3	18	2	1	5
Tyr	2*	—*	2*	11*	59*	—*	—*
Ala	7*	4*	—*	—*	62*	44*	6
Met	6	60	—		2	10	1
Val	78	5	1	1	2	1	1
Trp		1	1	—	1		
Pro							
Lys	1	1					
Phe	2	1			1		
Ile							
Leu	157	12	3	1	18	58	2
His	64				39		
Arg				26	27		

under the conditions used. The recoveries of the PTH–amino acids liberated from the tryptic and chymotryptic peptides by Edman-degradation are summarized in Table 3. In the method described previously,[4,6] the molecular weights and *N*-terminal amino acid residues of the tryptic and chymotryptic peptides were determined from their mass values and the PTH–amino acids released in the first-cycle of Edman-degradation, as described in Table 1. Figure 5 shows the computer output sequence of human β-endorphin, based on the data described above. The results suggest that the

Table 3. Recoveries (nmol) of PTH–amino acids released from tryptic (0.55 mg) and chymotryptic (0.55 mg) peptides of human β-endorphin by Edman-degradation

See the legend to Table 3 for further details. t means the presence of a trace of material.

	Tryptic peptides			Chymotryptic peptides			
	1	2	3	1	2	3	4
Asp	33	2	1	13	11	2	1
Asn	84	1	t	26	26	1	1
Glu		26	15		12	13	23
Gln		13			t		
Ser	36		2	7	t	3	1
Thr		5	4	13	26	6	
Gly	42	50	24	15	34	45	3
Ala		30		t	30*	10	1
Tyr	55		7	55*	t	12	1
Met	10*	1		33	2*		1
Val	7	t		40	1*	t	
Pro					t	t	6
Phe			1			1	7
Lys	36	3	1	92	35	3	7
Ile			4	t	2	2	4
Leu					1	4	t
His							
Arg							

sequence can be completed by the molecular weights and *N*-terminal amino acid residues of the tryptic and chymotryptic peptides, respectively, and also that the PTH–amino acids released from the second to third cycle of Edman-degradation of the tryptic peptides and from the second to fourth cycle of Edman-degradation of the chymotryptic peptides, although these were rather complicated as shown in Table 3.

Figure 6 shows the amino acid sequence of the *N*-terminal BrCN fragment of *Streptomyces erythraeus* lysozyme,[14] which consists of 55 amino acid residues. The sequence was deduced from the molecular weights, the *N*-terminal amino acid residues of the tryptic and chymotryptic peptides, and the ratios of PTH–amino acids liberated until the 7th step of Edman-degradation. U and W were put in as many data as possible, because

TABLE 4. RECOVERIES (nmol) OF PTH–AMINO ACIDS RELEASED FROM CHYMOTRYPTIC (1.0 mg) AND TRYPTIC (0.7 mg) PEPTIDES OF THE *N*-TERMINAL BrCN FRAGMENT OF *Streptomyces erythraeus* LYSOZYME BY EDMAN-DEGRADATION

See the legend to Table 3 for further details.

	Chymotryptic peptides							Tryptic peptides						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Asp	62	30	46	8	4	3	33	24	3	26	4	2	3	20
Asn	277*	113*	11	2	1	1		96	3*			2	2	
Glu	59	16*	75	37	45	9	12			62	11	6	7	1
Gln	—*	11*	65*	31	5	1	1		—*	11*	8*	5	6	5
Ser			—*				1		4*	—*	—*			
Thr		64	13	25	7	7	10	1	80	8	1	12	3	1
Gly	28	51	33	39	43	34	10	6	12	3	25	28	16	7
Tyr		23	—*	25	17*	4	3			22*	—*	6	4	9
Ala	211	102	48*	64	22*	5	1	121	55	—*	32*	17	2	
Met														
Hse														
Val	67	47	75	15	3	2	2	3	58	29	8	3	1	1
Trp	69		—*	1				2			2			
Pro		2	50*	3					40	2				
Lys	113	71	5	29	21	1		1	1			1		
Phe	97	3	1				t	44	2		9	1		
Ile	3	15	2	1	1	39	5	1	4	1		1	20	4
Leu														
His														
Arg	8	8			3	5		26	4					

the former derivative (PTH–Homoser) was not identified and the latter residue (PTH–Trp) was not recovered in good yield under our experimental conditions.[11] The results suggest that the whole sequence was determined except for 4 amino acid residues from the 8th to 11th position from the *N*-terminus and was completed by 4 further cycles of Edman-degradation of the tryptic peptides, as described elsewhere.[8,11] In the calculation reported in this paper, peptides containing a Lys residue[4,6] were used for sequence calculation, and these are shown by asterisks (*) in Figs. 2—5, although the data were not used in the previous paper.[8] This improvement greatly reduced the number of possible combinations of amino acid residues in the peptides with consequent shortening of calculation time.

Thus, the computation by the program described here is very effective for determining amino acid sequences, because all possible candidates of sequences can be searched from a little, simple information. In the current experimental procedure, the recovery of PTH–amino acids from a peptide mixture becomes rather more complicated with increase in the number of peptides in the mixture, because the recovery of PTH–amino acids from smaller or hydrophobic peptides is generally poorer than that from larger or hydrophilic peptides, *etc.* However, the improvement of Edman-degradation increases the usefulness of the present procedure.

References

1) A part of this work was presented at the 18th Symposium on Peptide Chemistry: T. Kitagishi, Y.-M. Hong, and Y. Shimonishi, "Peptide Chemistry 1980," Proceedings of the 18th Symposium on Peptide Chemistry, ed by K. Okawa, Protein Research Foundation, Minoh, Osaka (1981), p. 135.

2) J. A. Kelley, H. Nau, H.-J. Forster, and K. Biemann, *Biomed. Mass Spectrom.*, **2**, 313 (1975).

3) H. Nau, H.-J. Forster, J. A. Kelley, and K. Biemann, *Biomed. Mass Spectrom.*, **2**, 326 (1975).

4) Y. Shimonishi, Y.-M. Hong, T. Matsuo, I. Katakuse, and H. Matsuda, *Chem. Lett.*, **1979**, 1369.

5) T. Matsuo, I. Katakuse, H. Matsuda, Y. Shimonishi, Y.-M. Hong, and Y. Izumi, *Mass Spectroscopy* (*Japan*), **28**, 169 (1980).

6) Y. Shimonishi, Y.-M. Hong, T. Kitagishi, T. Matsuo, H. Matsuda, and I. Katakuse, *Eur. J. Biochem.*, **112**, 251 (1980).

7) T. Kitagishi, Y.-M. Hong, and Y. Shimonishi, *Int. J. Pept. Protein Res.*, **17**, 436 (1981).

8) Y. Shimonishi, Y.-M. Hong, T. Kitagishi, I. Katakuse, T. Matsuo, H. Matsuda, S. Hara, T. Ikenaka, and Y. Izumi, *Chem. Lett.*, **1981**, 499.

9) "Atlas of Protein Sequence and Structure," ed by R. V. Eck and M.O. Dayhoff, National Biomedical Research Foundation Publishers, Silver Spring, Md. (1966): A=Ala, D=Asp, E=Glu, F=Phe, G=Gly, H=His, I=Ile, K=Lys, L=Leu, M=Met, N=Asn, P=Pro, Q=Gln, R=Arg, S=Ser, T=Thr, V=Val, W=Trp, Y=Tyr.

10) P. Edman and A. Henschen, "Protein Sequence Determination," ed by S. B. Needleman, Springer-Verlag, Berlin-Heidelberg- New York (1975), pp. 232—271.

11) Y. Shimonishi, Y.-M. Hong, I. Katakuse, and S. Hara, *Bull. Chem. Soc. Jpn.*, **54**, 3069 (1981).

12) H. Matsuda, *Atomic Masses Fundam. Constants*, **5**, 185 (1976).

13) C. L. Zimmerman, E. Appella, and J. J. Pisano, *Anal. Biochem.*, **77**, 569 (1977).

14) T. Morita, S. Hara, and Y. Matsushima, *J. Biochem.* (*Tokyo*), **83**, 893 (1978).