Strategy for Protein Sequence Analysis by Fast Atom Bombardment Mass Spectrometry

Toshifumi Takao

1986

ACKNOWLEDGEMENTS

The studies presented in this thesis have been carried out under the direction of Professor Yasutsugu Shimonishi, Institute for Protein Research, Osaka University.

The author wishes to express his deepest gratitude to Professor Yasutsugu Shimonishi for his cordial guidance, helpful criticism and ceaseless encouragements.

The author also wishes to express his gratitude to Professor Toshio Miwatani (Research Institute for Microbial Diseases, Osaka University) and Professor Yoshifumi Takeda (the Institute of Medical Science, the University of Tokyo) for their incitement on biological research of heat-stable enterotoxins and their cordial advices and helpful discussions.

Furthermore, the author wishes to express his thanks to Dr. Tae Takeda (Research Institute for Microbial Diseases, Osaka University) for her advices on biological study, and Dr. Saburo Hara (Department of Chemistry, College of Science, Osaka University) for amino acid analyses.

Finally, the author would like to express his thanks to Dr. Saburo Aimoto for his helpful discussions, and the author is grateful to all the members of Professor Shimonishi's laboratory.

Toshifumi Takao

1986

CONTENTS

GENERAL INTRODUCTION

In recent years, there has been much progress in development of mass spectrometric techniques and these techniques have been used widely for determination of peptide and protein sequences. The main developments have been in soft ionization mass spectrometry, and in particular, field desorption (FD) [1], ^{252}Cf-plasma desorption (PD) [2], and fast atom bombardment (FAB) [3] mass spectrometries. The great advantages of soft ionization mass spectrometry are that it allows observation of quasi-molecular ions with few fragment ions and that it does not require prior derivatization of materials to volatile compounds before ionization, which is necessary for classical mass spectrometric methods such as electron impact (EI) and chemical ionization (CI) mass spectrometry. However, as soft ionization mass spectrometric procedures such as FD mass spectrometry do not provide structural information, they cannot be used to determine the primary structure of peptides and proteins.

In 1979, taking advantage of these characteristics of soft ionization mass spectrometry, Shimonishi and co-workers [4,5] found that when a proteolytic digest of a polypeptide is directly examined by FD mass spectrometry, almost all the signals corresponding to individual peptides in the digest can be observed in the mass spectrum with few signals of fragment ions. This finding provided a clue for the development of a method for determining the amino acid sequences of peptide mixtures by a

combination of FD mass spectrometry and Edman degradation or exopeptidase digestion [5-9]. The great merit of the method is that it does not require prior separation of the individual peptides in the digest, which is necessary for the conventional Edman method [10] in protein sequencing, or derivatization of peptides to volatile compounds, which is required for the EI procedure [11] or gas chromatographic mass spectrometric method [12,13].

However, although FD mass spectrometry is a suitable ionization method for measurement of the molecular ions of non-volatile compounds, it cannot detect high-molecular-weight compounds of over 2000 amu, because of difficulties in mass calibration and ionization. Moreover, the author found that FD mass spectrometry is not an easy method, because the procedures of sample-loading and preparation of the FD ion source are very difficult and an efficient emitter, which consists of whiskers grown on a microneedle [1,14], is also difficult to prepare. Thus for application of this mass spectrometric method to general protein chemistry, a more practical technique is required for ionizing non-volatile, high-molecular-weight compounds of over 3000 amu.

High-molecular-weight compounds can be measured by PD mass spectrometry, which was developed by Macfarlane and Torgerson [2] in 1974. In fact, recently the analyses of very high-molecular-weight compounds such as protected oligonucleotides (ca. 7000 amu) [15] and proteins (over 20,000 amu) [16] by PD mass

spectrometry have been reported. However, when using a PD mass spectrometer equipped with a coincidence time-of-flight mass analyzer [2], data must often be integrated over several hours to obtain sufficiently high signal to noise ratios, because ion fluxes are very low. Furthermore, unit mass resolution above 1000 has not been achieved by this method. Because of these disadvantages, PD mass spectrometry does not seem a practical method for use in protein chemistry.

In 1981, Barber et al. [3] developed an FAB mass spectrometric procedure in which a sample is loaded on a metallic plate with glycerol or α-thioglycerol [17] and ionized by bombardment with an accelerated neutral Ar or Xe atom beam. The remarkable features of this method reported by them [3,17] and others [18-20] including ourselves are as follows: 1) Intense molecular ion signals are mainly observed with few fragment ions as in FD and PD mass spectrometries. 2) A larger amount of ions with a longer life are produced in the ion source than in the FD or PD mass spectrometric procedure, and consequently high mass resolution (over 3000) can be achieved. 3) Molecular ion signals can readily be obtained with small quantities of peptides (ca. 10 pmol). 4) Mass measurement is easy. Because it possesses these advantages, FAB mass spectrometry has proved a very useful ionization method for use in studies on polar and non-volatile compounds such as peptides, antibiotics, vitamins and carbohydrates. Furthermore, with the recent development of a mass spectrometer equipped with a high field magnet [21,22]

high-molecular-weight (over 3000 amu.) compounds can now be measured with unit mass resolution [23].

The author used FAB mass spectrometry with a high field magnet instead of FD mass spectrometry for measurement of a complex peptide mixture. The following advantages were found in measurement of peptide mixtures by this FAB mass method: 1) No previous separation and identification of individual peptides is necessary. 2) FAB mass spectra can be analyzed more easily than FD mass spectra. 3) Almost all the molecular ions of constituent peptides including high-molecular-weight peptides of over 3000 amu in a protein digest can be observed. 4) Spectra can be obtained with microquantities of protein digests (ca. 0.3 nmol) and in a short time. 5) Cystine-containing peptides in a protein digest, which were difficult to detect by FD mass spectrometry, can be observed easily. These features indicate that FAB mass spectrometry is a practical and effective method for analysis of the primary structures of peptides and proteins.

The author considers that the method is useful for 1) peptide mapping of proteins, 2) confirmation of the primary structure of a protein that has already been determined by the conventional Edman method, 3) verification of a protein sequence deduced from the nucleotide sequence, 4) obtaining information on post-translational modifications of a protein, and 5) determination of a primary structure that is difficult to determine by a conventional method.

The first chapter describes a new method for peptide mapping

of proteins by FAB mass spectrometry, with hen and duck egg-white lysozymes as examples. The second chapter describes the application of the method for confirmation of the amino acid sequence of the B subunit of Vibrio cholerae classical biotype Inaba 569B toxin, which has been determined by a conventional method, and also describes determination of the amino acid sequence of Limulus anticoagulant anti-lipopolysaccharide factor by a combination of a conventional method and FAB mass spectrometry. The third chapter describes the application of the method for verification of the amino acid sequence of a development specific protein of Myxococcus xanthus, whose sequence was deduced from the nucleotide sequence. The fourth chapter describes the use of the method for identification of the primary structure of human interleukin-2 produced by a recombinant DNA technique in Escherichia coli. The fifth chapter describes the sequence determinations of several kinds of heat-stable enterotoxins produced by enteric bacteria.

References

1) H. D. Bekey, Int. J. Mass Spectrom. Ion Phys., 2, 500 (1969).

2) R. D. Macfarlane and D. F. Torgerson, Science, 191, 920 (1976).

3) M. Barber, R. S. Bordoli, R. D. Sedgwick, and A. N. Tyler, J. Chem. Soc. Chem. Commun., 325 (1981).

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

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

6) Y. Shimonishi, Y. -M. Hong, T. Takao, S. Aimoto, H. Matsuda, and Y. Izumi, Proc. Japan Acad., 57B, 304 (1981).

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

8) Y. Shimonishi, Methods in Protein Sequence Analysis, (M. Elzinga, ed.), 271, Human Press, Clifton, NJ (1982).

9) Y. -M. Hong, T. Takao, S. Aimoto, and Y. Shimonishi, Biomed. Mass Spectrom., 10, 450 (1983).

10) P. Edman and A. Henschen, Protein Sequence Determination, (S. B. Needleman, ed.), 232, Springer, Berlin (1975).

11) H. R. Morris, Nature, 286, 447 (1980).

12) W. C. Herlihy, R. J. Anderegg, and K. Biemann, Biomed. Mass

Spectrom., 8, 62 (1981).

13) S. A. Carr, P. V. Hauschka, and K. Biemann, J. Biol. Chem., 256, 9944 (1981).

14) T. Matsuo, H. Matsuda, and I. Katakuse, Anal. Chem., 51, 69 (1979).

15) C. J. McNeal, Anal. Chem., 54, 43A (1982).

16) B. Sundqvist, P. Roepstorff, J. Fohlman, A. Hedin, P. Hakansson, I. Kamensky, M. Lindberg, M. Salehpour, and G. Sawe, Science, 226, 696 (1984).

17) M. Barber, R. S. Bordoli, G. J. Elliott, R. D. Sedgwick, A. N. Tyler, and B. N. Green, J. Chem. Soc. Chem. Commun., 936 (1982).

18) S. Aimoto, T. Takao, Y. Shimonishi, S. Hara, T. Takeda, Y. Takeda, and T. Miwatani, Eur. J. Biochem., 129, 257 (1982).

19) T. Takao, Y. Shimonishi, and K. Inouye, Peptide Chemistry, 1984, 187 (1985).

20) T. D. Lee, K. Legesse, D. H. Hawke, and J. E. Shively, Biochem. Biophys. Res. Commun., 132, 520 (1985).

21) A. Dell, R. A. McDowell, H. R. Morris, and G. W. Taylor, Soft Ionization Biological Mass Spectrometry, (H. R. Morris, ed.), 69, Heyden and Son, London (1981).

22) A. Dell and H. R. Morris, Biochem. Biophys. Res. Commun., 106, 1456 (1982).

23) J. S. Cottrell and B. H. Frank, Biochem. Biophys. Res. Commun., 127, 1032 (1985).

CHAPTER I

Peptide Mapping of Proteins by Fast Atom Bombardment Mass Spectrometry

I-1 Introduction

In this chapter, the author describes a new method for peptide mapping of proteins by FAB mass spectrometry. In comparison with peptide mapping by high-performance liquid chromatography, the method is of great advantage in that information on molecular weights of the constituent peptides in a proteolytic digest can be directly obtained without prior separation followed by identification of the individual peptides. To determine the utilities and limitations of the method, the author examined the following problems, using hen egg-white lysozyme as an example: 1) Can all peptides in a protein digest be observed by FAB mass spectrometry? 2) What is the minimum amount of a protein digest necessary for observation of the signals of almost all the peptides in the digest? 3) Can peptides containing disulfide linkages, which the author has found difficult to detect by FD mass spectrometry, be observed by FAB mass spectrometry?

In order to apply the above method to other proteins and to examine its limitations, three kinds of duck egg-white lysozymes have been also measured by FAB mass spectrometry.

I-2 Materials and Methods

Abbreviations. Three-letter and one-letter symbols for amino acids are according to JCBN Recommendations [Eur. J. Biochem., 138, 9-37 (1984)]; m/z denotes the quantity formed by dividing the mass number of an ion (m) by the number of charges carried by the ion (z); amu denotes atomic mass unit.

I-2-1 Materials and sample preparation

All chemicals used were special grade from several suppliers and solvents were redistilled before use. Six times crystallized hen egg-white lysozyme (HEL) was purchased from Seikagaku Kogyo Co., Ltd.. Duck eggs were provided by courtesy of Dr. H. Fujio (Research Institute for Microbial Diseases, Osaka University) and three kinds of egg-white lysozyme were isolated and purified by the procedures described by Kondo et al. [1]. TPCK-treated trypsin was obtained from Worthington Biochemical Corp. (USA). Sample protein was treated with BrCN in 70% formic acid at room temperature for 16 h [2]. The BrCN-treated protein was digested with trypsin at a substrate : enzyme ratio of 50 : 1 (w/w) at pH 8.0 and 37 °C for 4 h.

I-2-2 Fast atom bombardment (FAB) mass spectrometry

FAB mass spectra were obtained with a Jeol JMS HX-100 double-focusing mass spectrometer fitted with a 1.88 or 2.33 tesla magnet and equipped with an FAB ion source and a post-accelerating

system. Sample solution (1-2 μl) containing peptides (5-10 μg) was loaded on a stainless steel sample plate (1.5 X 6 mm) and mixed with glycerol (0.5 μl), 1 M hydrochloric acid (0.5 μl) and in some cases α-thioglycerol (1 μl) [3]. The sample plate was inserted into the FAB ion source. The sample solution was bombarded with a xenon neutral atom beam accelerated by a 7 keV potential. The ion source was a 4 or 5 kV accelerating potential. A data processor (Jeol JMA- 3100 or DA-5000 mass data analysis system) was used for acquisition of mass spectra. The resolving power was 3000. Mass assignment was made using a mixture of CsI and KI (2 : 1, w/w) as a mass reference.

I-2-3 High-performance liquid chromatography (HPLC)

HPLC was performed in a Shimadzu HPLC LC-3A apparatus (Kyoto, Japan) fitted with a solvent programmer GRE-2B and a data processor, chromatopac C-RlA. The tryptic digest of BrCN-treated HEL (200 μg) was separated on a YMC-packed column of ODS A-312 (6 X 150 mm) (Yamamura Chemical Laboratory Co. Ltd, Kyoto, Japan) with a linear gradient of 5-35% acetonitrile in 0.1% trifluoroacetic acid (pH 2.0) at a flow rate of 1.0 ml/min. The absorbance at 220 nm of the eluate was monitored.

I-2-4 Amino acid analysis

The peptides separated were hydrolyzed in 4 M methanesulfonic acid containing 0.2% 3-(2-aminoethyl)indole [4] at 110°C for 24 h in evacuated sealed tubes and the hydrolysates were analyzed in a

Hitachi type 835 amino acid analyzer.

I-2-5 Other methods

The locations of the observed mass values were examined with an ACOS 900S computer in the Crystallographic Research Center of the Institute for Protein Research, Osaka University, using a program (Frag 1-MH) produced in the author's laboratory.

I-3 Results and Discussion

I-3-1 Tryptic digest of hen egg-white lysozyme

In developing a method for measurement of a microquantity of a protein, the following procedures were used: 1) The sample protein was treated with BrCN before digestion with a proteolytic enzyme, because BrCN denatures the sample protein by specific cleavage at Met residues, which are minor components of a protein; 2) the BrCN-treated protein was digested with a proteolytic enzyme in the same minitube to prevent loss of the sample protein. Fig. I-1 shows the FAB mass spectrum of a tryptic digest of BrCN-treated hen egg-white lysozyme. All the observed mass values were collated to the theoretical mass values, calculated from the amino acid sequence of HEL (the mass values of underivatized peptides are observed as $[M + H]^+$ in positive FAB mass spectra, where M denotes molecular weight). For this purpose, a computer program (Frag 1-MH) was designed for selecting

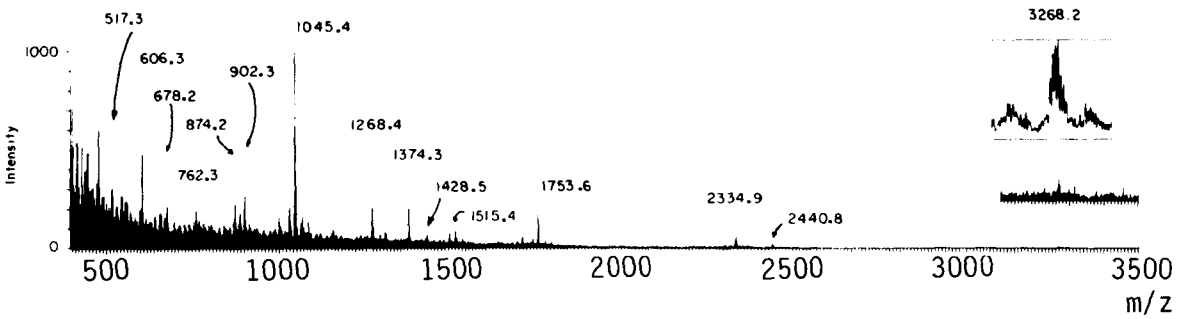

Fig. I-1. FAB mass spectrum of the tryptic digest of BrCN-treated HEL. 15 scans of the spectrum were integrated in the region of 3200 amu and the accumulated spectrum (insert) was compared with the theoretical molecular ion distribution.

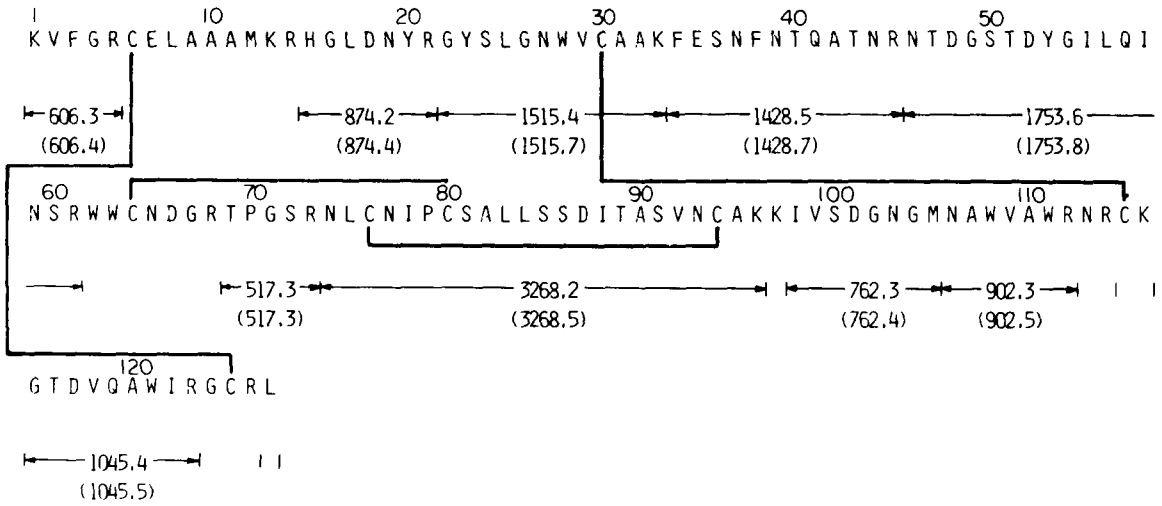

Fig. I-2. Amino acid sequence of HEL. Numbers without and with parentheses indicate observed mass values in Fig. I-1 and theoretical mass values, respectively.

candidate amino acid sequences in a protein from observed mass values. Using this program, all the observed signals could be easily located except those at m/z=1374.3 and 2440.8 in the amino acid sequence of HEL, as shown in Fig.I-2. To verify that these signals observed in the mass spectrum of the tryptic digest were

specific for individual peptides in the digest, the peptide mixture was separated by HPLC, as shown in Fig.I-3, and each peptide was subjected to mass spectrometry, as shown in Fig.I-4, and amino acid analysis (data not shown). Thus, the signals observed in the mass spectrum in Fig.I-1 were unequivocally confirmed to correspond to those of individual peptides.

In this study, it was found that signals of peptides containing a cystine residue(s), which have been difficult to be detected by FD mass spectrometry, were observed at m/z= 1515.4 and 3268.2, respectively, together with signals (m/z=1268.4 and 2334.9) of ions that were probably formed by fragmentation or reductive cleavage of disulfide linkages, although only a fragment ion signal was observed at m/z= 678.2 in the peptide bound between Cys-6 and Cys-129. These observations were confirmed by

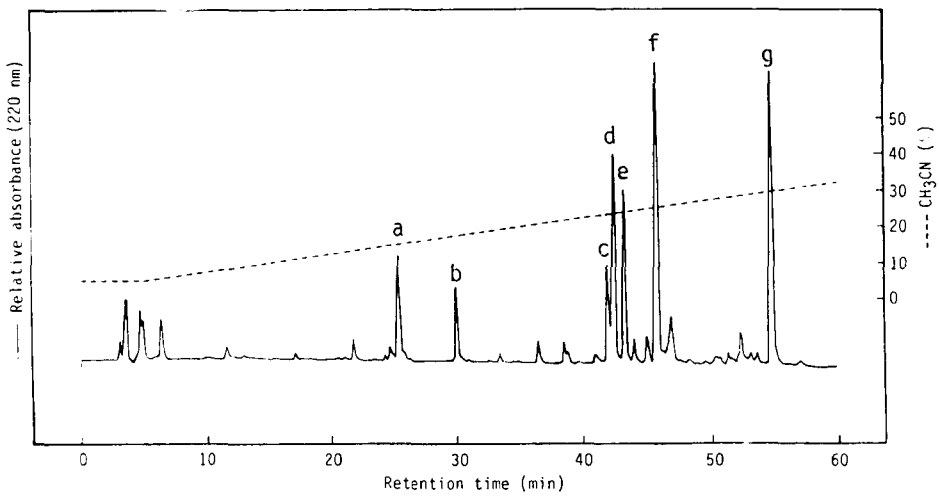

Fig. I-3. High-performance liquid chromatography of the tryptic digest of BrCN-treated HEL. Chromatographic conditions were as described in Materials and Methods.

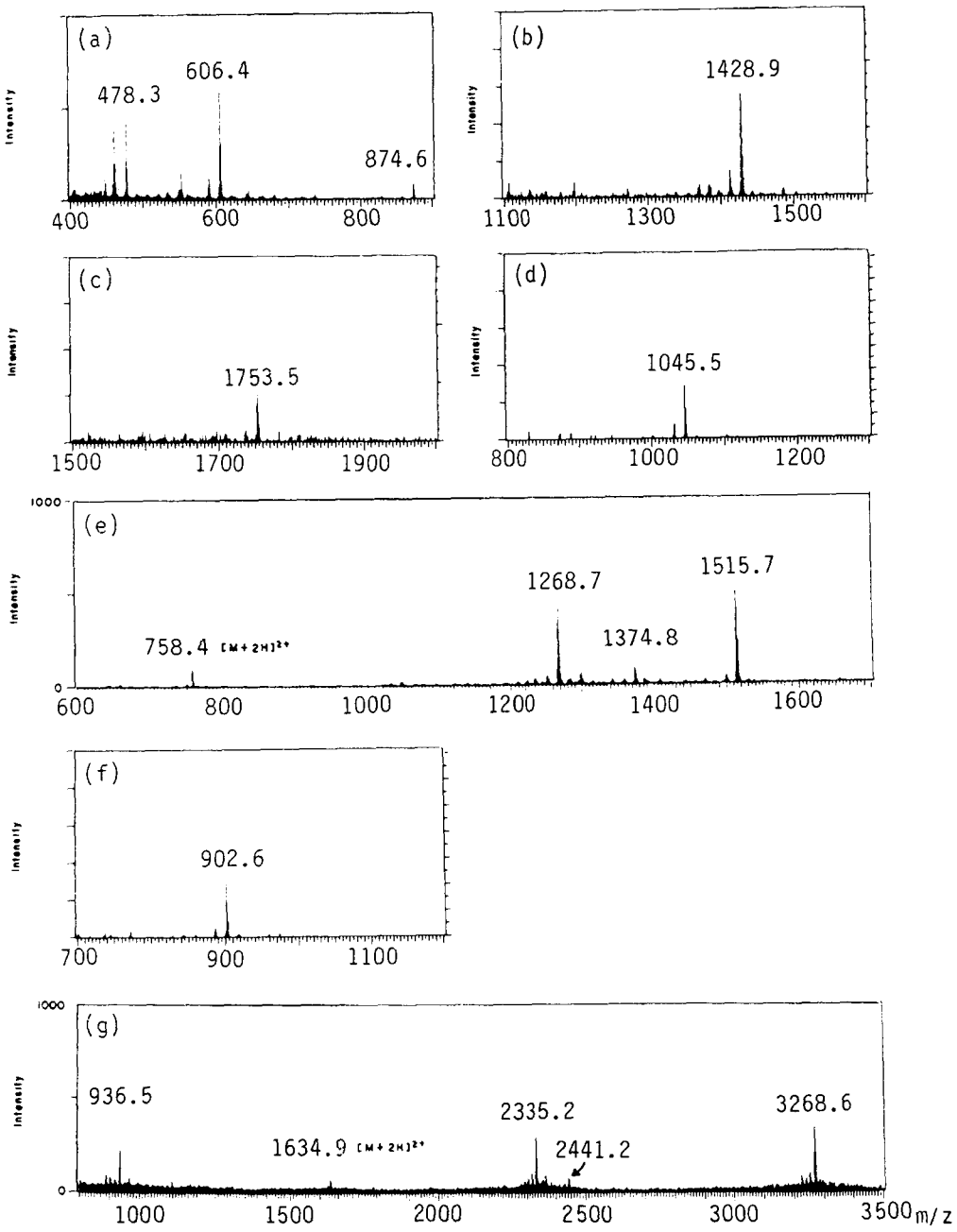

Fig. I-4. FAB mass spectra of the tryptic peptides of BrCN-treated HEL separated by HPLC (Fig. I-3). Figs. (a)-(g) correspond to peak fractions in Fig. I-3 and numbers show observed mass values.

measurements of separated peptides (peaks e and g in Fig.I-3), as shown in Figs.I-4(e) and 4(g), respectively. For example, one peptide separated (peak g) showed signals of a quasi-molecular ion (m/z=3268.6) and both fragment ions (m/z=936.5 and 2335.2) in its spectrum (Fig.I-4(g)), although the peptide gave signals of a quasi-molecular ion (m/z=3268.2) and one (m/z=2334.9) of the two fragment ions in the digest, as shown in Fig.I-1. These findings suggested that the positions of disulfide linkages in a protein could be determined without separation of the components of the protein digest, although the reason why only one of the fragment ions was observed in the digestion mixture could not be elucidated. Another interesting phenomenon was found in the measurement of peptides containing a cystine residue(s). The mass values at m/z=1374.3 and 2440.8 seen in Fig.I-1 did not correspond to any positions in the amino acid sequence of HEL, as shown in Fig.I-2. Similar mass values (m/z=1374.8 and 2441.2) were observed in the mass spectra of the separated peptides, as shown in Figs. I-4(e) and 4(g), respectively. These signals were probably derived from ions at m/z=1268.7 and 2335.2, which were formed from the mother peptides (m/z= 1515.7 and 3268.6) by reductive cleavage or fragmentation and replacement by α-thioglycerol, respectively. This finding indicates that the cystine residue is reduced by α-thioglycerol in a matrix and the signal of the adduct is observed in the spectrum. Thus, the mass spectrum of the tryptic peptides of BrCN-treated HEL covered all the sequence of HEL except those of amino acids and dipeptides and corresponded to 93%

of the sequence.

Next, to examine the minimum amount of sample necessary for detection of signals of as many peptides as possible in a protein digest, the tryptic digest of HEL was examined under various conditions such as with different amounts of sample and different kinds of matrix. A mass spectrum such as that shown in Fig.I-1 was observed with a mixture of glycerol, α-thioglycerol and 1M HCl as a matrix and with a sample of about 5 μg (300-400 pmol). With a sample of less than 5 μg, signals of high-molecular-weight peptides, particularly those of over 3000 amu, were difficult to be observed, although signals of peptides of less than 2000 amu could be clearly seen.

I-3-2 Tryptic digests of duck egg-white lysozyme

To apply the above method to other proteins and to examine the limitations of the method, the author measured the FAB mass spectra of tryptic digests of three kinds of BrCN-treated duck egg-white lysozyme (named DEL-1, DEL-2 and DEL-3 as described in Ref. 1); their spectra are shown in Figs.I-5, 6 and 7, respectively. The observed mass values were collated to the theoretical mass values, calculated from the reported amino acid sequences [1], as for HEL.

Unlike the mass spectrum of the digest of HEL, the mass spectra of the digests of DELs showed that peptides containing two disulfide linkages in the region from Trp-62 to Lys-96 showed

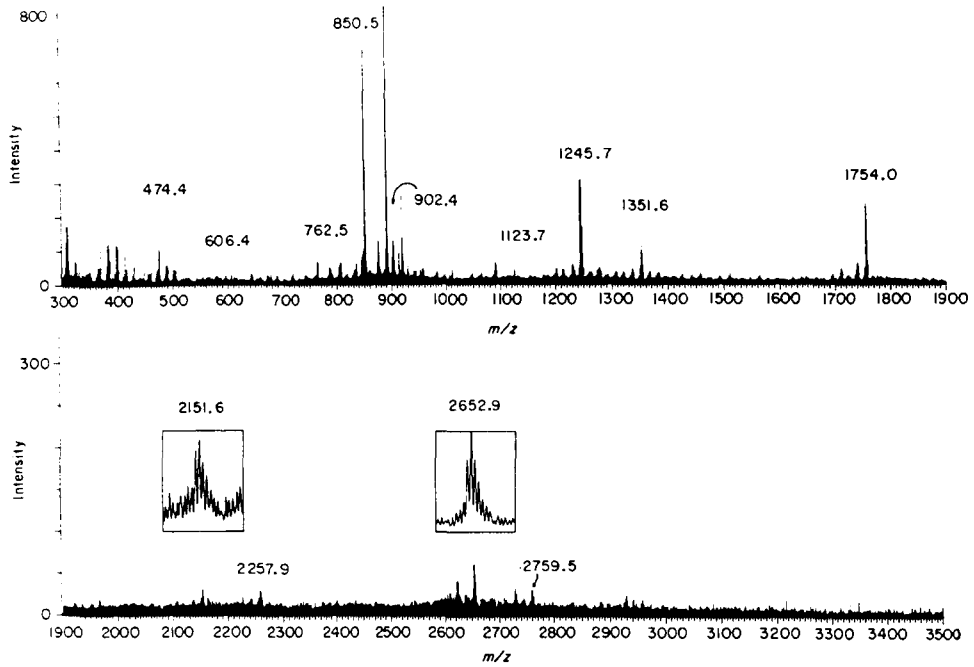

Fig. I-5. FAB mass spectrum of the tryptic digest of BrCN-treated DEL-1.

signals of various ions which were probably produced by reductive cleavage of one or both disulfide bonds. As shown in Fig.I-5, the digest of DEL-1 gave signals at m/z=2151.6 and 1245.7. These mass values were correlated to a peptide linked between Cys-64 and Cys-80 by one disulfide bond and its fragment, respectively, as shown in Fig.I-9. However, no signal of a peptide linked by two disulfide bridges could be observed. In the mass spectrum of the digest of DEL-2 (Fig.I-6), the signal of a peptide containing two disulfide linkages from Trp-62 to Lys-96 could be observed at m/z=2469.0, together with those (m/z=1245.7 and 2151.6) of its fragment, although DEL-2 has the same amino acid sequence in this

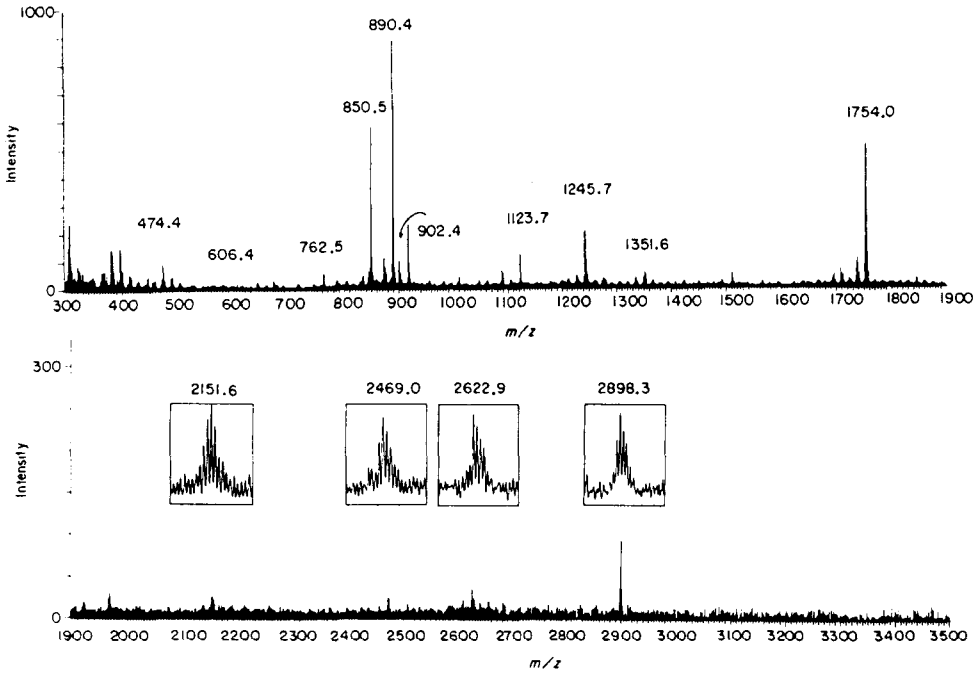

Fig. I-6. FAB mass spectrum of the tryptic digest of BrCN-treated DEL-2.

region as DEL-1. In DEL-3, Pro-79 is replaced by Arg, as shown in Fig.I-8. Therefore, it was expected that in DEL-3 the peptide in both DEL-1 and DEL-2 containing two disulfide bonds from Trp-62 to Lys-96 would be divided into two peptides linked, between Cys-64 and Cys-80 and between Cys-76 and Cys-94, by one disulfide bond each. Only one of these two peptides was detected at m/z=1596.7 together with the signal of its reduced fragment at m/z=690.5, as shown in the mass spectrum of the digest of DEL-3 in Fig.I-7. Similar phenomena were observed in peptides linked by a disulfide bond between Cys-30 and Cys-115; no signal of the peptide in the digest of DEL-1 was observed at m/z=2928.3 (theoretical mass

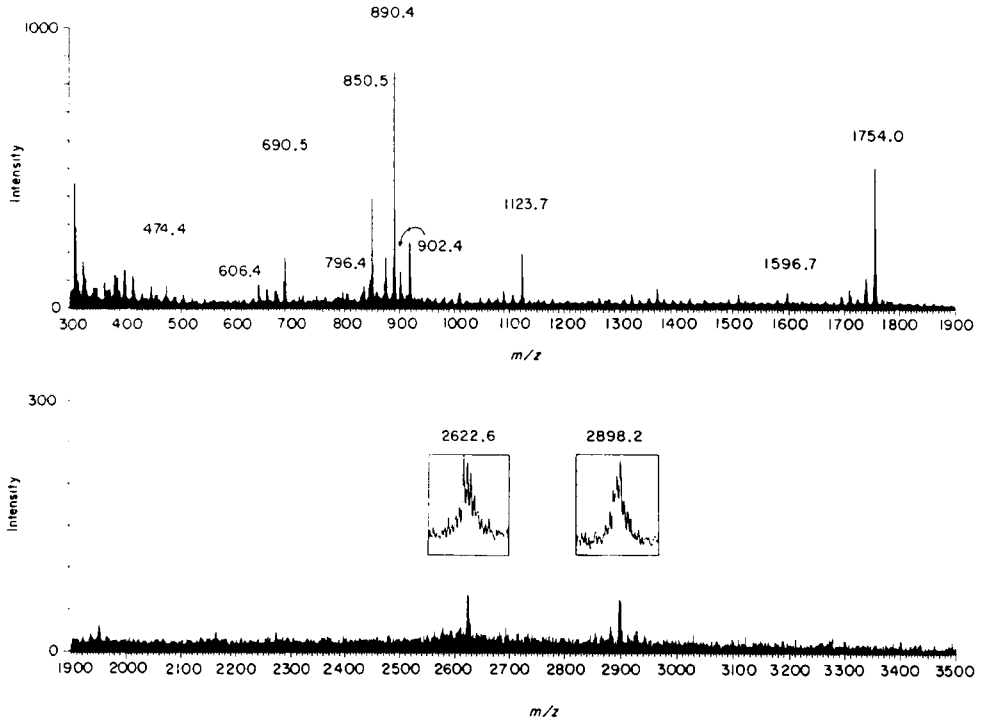

Fig. I-7. FAB mass spectrum of the tryptic digest of BrCN-treated DEL-3.

value), but the fragment ion was observed at m/z=2652.9, as shown in Fig.I-5, while both signals of peptides containing a cystine residue and their fragments were detected at m/z=2898.3 or 2898.2 and 2622.9 or 2622.6 in the mass spectra of the digests of DEL-2 and DEL-3 in Figs.I-6 and 7, respectively.

In the mass spectra of the digests of DEL, signals of the adducts of peptides containing a cystine residue(s) with α-thioglycerol are seen at m/z=1351.6, 2257.9 and 2759.5 in Fig.I-5, at m/z=1351.6 in Fig.I-6 and at m/z=796.4 in Fig.I-7 , like those observed in Fig.I-1. It is unknown why signals of only some adducts could be observed, but these signals should be useful

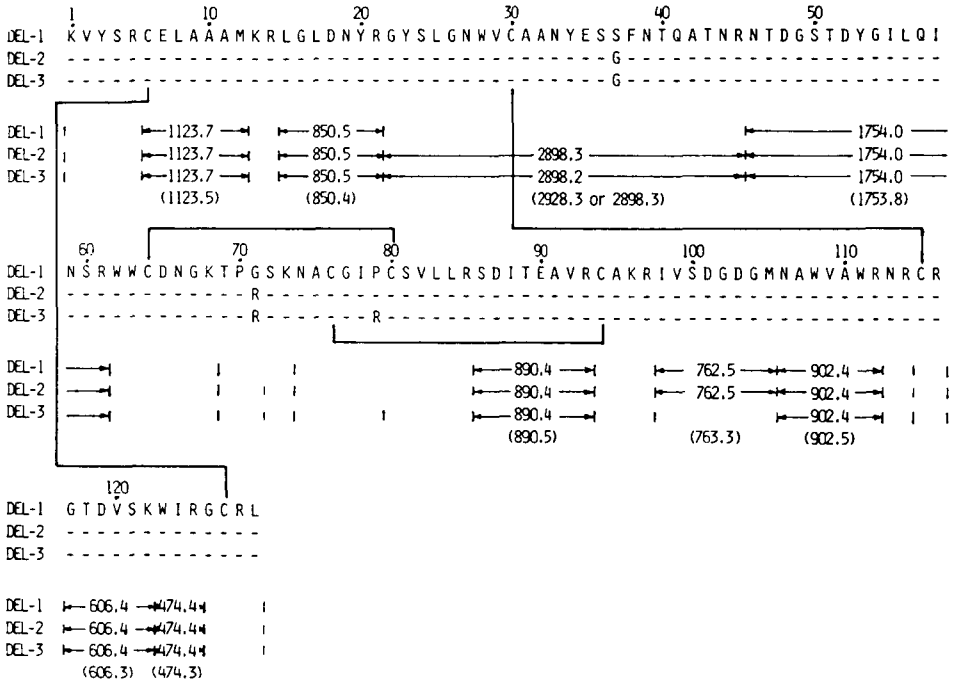

Fig. I-8. Amino acid sequences of three DELs. Bars marked in the sequences of DEL-2 and DEL-3 indicate that the same residues as those in DEL-1 are located in the same positions. Numbers without and with parentheses denote observed and theoretical mass values, respectively.

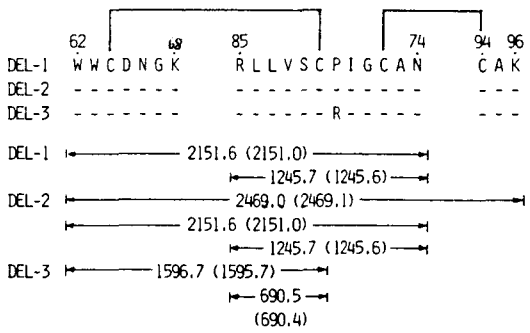

Fig. I-9. Amino acid sequences of three DELs in the region from Trp-62 to Lys-96 linked by two disulfide bonds. Numbers without and with parentheses indicate observed and theoretical mass values, respectively.

for detection of peptides containing a cystine residue(s).

The mass values of the signals at m/z=762.5 in Figs.I-5 and 6 and m/z=918.5 in Fig.I-7 could not be related to any sequence. As Asp-103 was previously revised to Asn in the amino acid sequence of HEL by DNA sequencing [5], Asp-103 of DEL that is phylogenetically related to HEL may be replaced by an Asn residue. If so, the theoretical mass values of the sequences containing Asn-103 need to be revised to 762.4 (98-105) and 918.4 (97-105), which are consistent with the above observed mass values at m/z=762.5 and 918.5, respectively.

No signals of the N-terminal peptides from Lys-1 to Arg-5 (theoretical mass value 652.4) in any DELs or of the peptide from Thr-69 to Lys-73 (theoretical mass value 489.3) in DEL-1 or the peptide linked by one disulfide bond between Cys-76 and Cys-94 in DEL-3 could be observed. Signals of the N-terminal peptide from Lys-1 to Arg-5 and the peptide from Thr-69 to Lys-73 were in fact observed at m/z=652.4 and 489.3 (data not shown) in a separate measurement of a digest of DEL-1, but in this measurement the signals of some other peptides could not be detected. Furthermore, signals of peptides of less than 400 amu and of amino acids were concealed by intense noise signals derived from a matrix (glycerol, 1-thioglycerol, etc.) in the low mass region. These problems may be overcome by use of digests with other enzymes of different specificities or by use of a larger amount of sample peptide or a different technique of mass spectrometry. Signals at m/z=1089.5 was commonly observed in all the digests of DEL and was

not collated to any position in the amino acid sequences of the protein. This signal may be derived from concomitant impurities during purification procedures of the protein. Anyway, almost all of the observed signals could be related to the amino acid sequences of DEL as described for the tryptic digest of HEL.

The present method has the advantages that it does not necessitate chemical modifications, such as reductive alkylation, and hence it saves time and also avoids loss of the sample. In practice, about 5 μg of protein is sufficient for observation of almost all the signals of peptides in a protein digest under the conditions used in this study, although this amount may be reduced by improving the ionizing or detection procedure.

The author experienced that peptides in the mixture gave different intensities of signals in the mass spectrum, those of basic peptides being the more intense [6] even when nearly equimolar amounts of peptide fragments were expected to be present in the digest of a protein. Different intensities of signals of peptides may be caused not only by differences in polarity [7] but also by differences in solubility of peptides, particularly of high-molecular-weight peptides, in a matrix. However, the present results demonstrated that these problems did not prevent measurement of protein digests by mass spectrometry.

It is concluded that over 90% of the amino acid sequence of a protein can be covered by direct measurement of its digest by FAB mass spectrometry. In addition, high-molecular-weight peptides

over 3000 amu and peptides containing cystine residues, which have been difficult to be detected by FD mass spectrometry, were found to be detected easily by FAB mass spectrometry. These findings suggested that direct measurement of a protein digest by FAB mass spectrometry can be a practical and effective procedure for peptide mapping of a protein and that it can be used to determine the positions of disulfide linkages in a protein without separating the peptide fragments present in a protein digest. Furthermore, the present method was found to be very useful for elucidating substitution of amino acid residues of homologous proteins.

References

1) K. Kondo, H. Fujio, and T. Amano, J. Biochem., 91, 571 (1982).

2) E. Gross and B. Witkop, J. Biol. Chem., 237, 1856 (1962).

3) M. Barber, R. S. Bordoli, G. J. Elliott, R. D. Sedgwick, A. N. Tyler, and B. N. Green, J. Chem. Soc. Chem. Commun., 936 (1982).

4) R. J. Shimpson, M. R. Neuberger, and T. -Y. Lin, J. Biol. Chem., 251, 1936 (1976).

5) A. Jung, A. E. Sippel, M. Grez, and G. Schuetz, Proc. Natl. Acad. Sci., 77, 5759 (1980).

6) I. Katakuse, T. Matsuo, H. Matsuda, Y. Shimonishi, and Y. Izumi, Mass Spectroscopy, 27, 127 (1979).

7) A. M. Buko, L. R. Phillips, and B. A. Fraser, Biomed. Mass Spectrom., 10, 324 (1983).

CHAPTER II

Facile Identification of Protein Sequences by Fast Atom Bombardment Mass Spectrometry

II-1 Introduction

In this chapter, the author describes the application of the method described in chapter I to identification of the amino acid sequence of the B subunit of Vibrio cholerae classical biotype Inaba 568B toxin and to determination of the amino acid sequence of limulus anticoagulant anti-lipopolysaccharide (anti-LPS) factor.

Cholera toxin, secreted by V. cholerae when it colonizes in the human small intestine, is a pathogenic substance responsible for watery diarrhea in humans. The toxin is composed of one A subunit (27kDa) and five B subunits (11.6kDa). The B subunits are reported to combine with a receptor (GM1 ganglioside) on the surface of toxin-sensitive cells and facilitate the internalization of the A subunit into the cytosol [1].

Recently, the nucleotide sequences of the genes (ctx) encoding the toxins of V. cholerae El Tor biotype strains 62746 and 2125 have been reported by Lockman and Kaper [2] and Mekalanos et al. [3], respectively. The amino acid sequences of the B-subunit portions translated from these nucleotide sequences of ctx encoding El Tor biotype toxins were very similar, but differed

at several positions from the published amino acid sequences of classical biotype toxin [4, 5]. Information on the correct amino acid sequence of the toxic protein is essential for elucidation of the biological and physicochemical features of its binding to the receptor and internalization into the cytosol, and also for development of an effective preparation of synthetic vaccine [6].

The author used the method to check the amino acid sequence of the B subunit of V. cholerae classical biotype 569B toxin [4, 5] and to cofirm the differences in the amino acid sequences predicted from the nucleotide sequences of the genes of El Tor biotype strains 62746 and 2125 toxins [2, 3].

A potent anticoagulant, anti-LPS factor [7], found in hemocytes from both Tachypleus tridentatus and Limulus polyphemus inhibits the LPS-mediated activation of limulus coagulation cascade. Moreover, it has an antibacterial effect on the growth of Gram-negative bacteria, especially on rough-type ones [8], and a hemolytic activity on red blood cells sensitized with LPS [9]. Recently, Aketagawa and his co-workers [10] determined almost complete amino acid sequence of anti-LPS factor by sequencing the peptides obtained by its selective proteolytic cleavage. Anti-LPS factor was deduced to consist of a single chain of 102 amino acid residues with two half-cystines in disulfide linkage and its N-terminal end to be pyroglutamic acid. Moreover, it has been suggested that anti-LPS factor has two variant residues at position 36 and the C-terminal end, respectively [10]. The author

used FAB mass spectrometry for final determination of the primary structure of anti-LPS factor, especially the location of the variant residue at the C-terminal portion.

II-2 Materials and Methods

All the experimental and analytical procedures and reagents were the same as described in chapter I, except those described below.

II-2-1 Materials and sample preparation

The B subunit of cholera toxin from Vibrio cholerae Inaba 569B was purchased from List Biological Laboratories Inc. (CA, USA), Sigma Chemical Corp. (MS, USA) and Calbiochem-Behring Corp. (CA, USA); the purity of each of preparation was analyzed by HPLC. Samples of each B subunit (200 μg) were treated with BrCN, and then digested with S. aureus protease V8 (Miles, IN, USA). One tenth of the digest was examined directly by FAB mass spectrometry and the remainder was separated by HPLC on a reversed-phase column, as described below. The peptides separated were further digested with TPCK-treated trypsin at 37-40 °C for 4 h or with carboxypeptidase B (Boehringer, Mannheim, FRG) or Y (Oriental Yeast Co., Ltd.) at 37 °C for the indicated times.

Tachypleus tridentatus anti-LPS factor, which was purified in Ref.10, was used. The purified sample of anti-LPS factor was dissolved in 60 mM aqueous ammonium hydrogencarbonate at pH 8.0. The solution was treated with TPCK-treated trypsin or S. aureus

protease V8 at a substrate : enzyme ratio of about 100 : 1 (w/w) at 40 °C for 22 h or 16 h, respectively. The tryptic digest of native protein was degraded manually without washing and extraction by Edman method described below, and directly subjected to FAB mass measurement.

II-2-2 Edman degradation

Edman degradation was performed as described previously [11] with some modifications. A sample was dissolved in a mixture of pyridine and water (1 : 1, v/v) adjusted to pH 9.6 with N-methylmorpholine and allowed to react with phenylisothiocyanate (PITC) (5-10 µl) at 40 °C for 60 min. Then the reaction mixture was washed three times with benzene and lyophilized. The lyophilized material was treated with anhydrous trifluoroacetic acid at 40 °C for 20 min and dissolved in 0.5 M acetic acid, followed by extraction with ethyl acetate. The organic and aqueous solution was dried separately. The former was analyzed by reversed-phase HPLC, as described below, and the latter was directly subjected to FAB mass measurement.

II-2-3 High-performance liquid chromatography

The digest of the BrCN-treated B subunit was separated on a column (4 X 250 mm) of YMC ODS S-5 (Yamamura Chemical Laboratory Co. Ltd., Kyoto, Japan) packed in the author's laboratory with a linear gradient of 5-50% acetonitrile in 0.05% trifluoroacetic acid (pH 2.3) or on a Cosmosil 5C18 column (4.6 X 150 mm, Nakarai

Chemical Ltd, Kyoto, Japan) with a linear gradient of 15-40% acetonitrile in 0.01 M ammonium acetate (pH 5.7). Flow rate was 1.0 ml/min and the absorbance at 220 nm of the eluate was monitored. For identification of phenylthiohydantoin derivatives, released by Edman degradation, the column (Zorbax-ODS, 4.6 X 250 mm) (DuPont, USA) was developed isocratically with a mixture of acetonitrile and 0.01 M sodium acetate (pH 4.5) (42 : 58, v/v) as solvent [12]. HPLC was performed at 55 °C at a flow rate of 1.0 ml/min. The absorbance at 269 nm of the eluate was monitored.

II-3 Results and Discussion

II-3-1 B-subunit of cholera toxin

Fig.II-1 shows an FAB mass spectrum of the protease V8 digest of the BrCN-treated B subunit. The protease V8 digests of the different commercial preparations of the B subunit available all showed the same mass spectrum. All the observed mass values were collated with the theoretical values calculated from the published amino acid sequences [4, 5] of the B subunit of classical biotype Inaba 569 B toxin. All the observed signals, except those at m/z=1322.9, 1539.7 and 2224.4, could be located in the amino acid sequence of the B subunit, as shown in Fig.II-2. The values at m/z=3168.8 corresponding to the sequence positions 1-11 and 84-101 linked between Cys-9 and Cys-86 by a disulfide bond. The signals observed at m/z=1204.8 and 1967.2 corresponded to the same two sequences (1-11 and 84-101, respectively) after

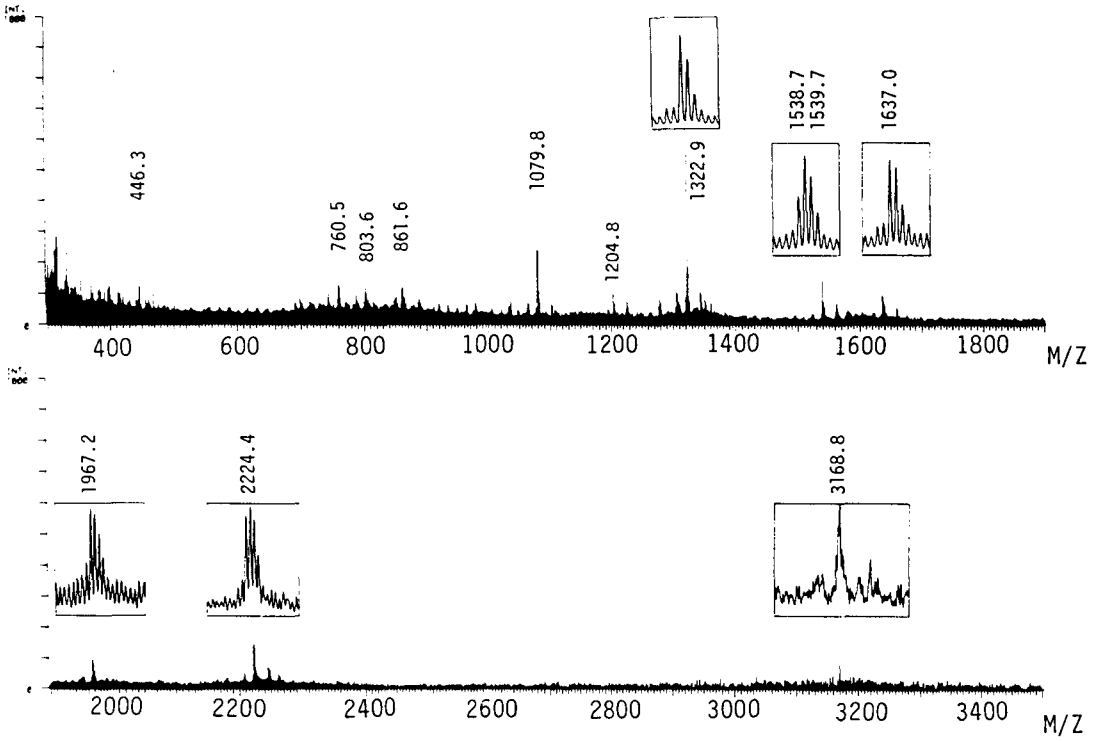

Fig. II-1. FAB mass spectrum of the protease V8 digest of the BrCN-treated B subunit of V. cholerae classical biotype Inaba 569B toxin. Inserts show 15 times accumulated spectra of restricted regions.

reductive cleavage of the disulfide bond, as described in chapter I. The isotopic distribution at about 1540 amu suggested a combination of two signals; that is, the presence of two components with mass values of 1538.7 and 1539.7 . The former coincided with the value calculated from the sequence 38-51. The two dipeptides at positions 67-68 and 102-103 could not be assigned because the signals of low-molecular-weight peptides were difficult to detect under the present experimental conditions. Thus, all the values of signals observed in the mass spectrum in

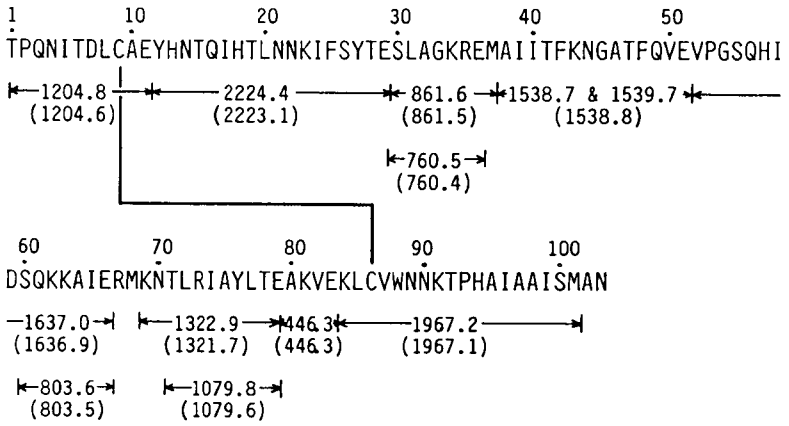

Fig. II-2. Amino acid sequence of the B subunit of V. cholerae classical biotype Inaba 569B toxin. Numbers without parentheses show the mass values observed in Fig. II-1, numbers within parentheses show theoretical mass values calculated from the amino acid sequence.

Fig. II-1 could be located in the amino acid sequence of the B subunit except those of the sequences 12-29 and 69-79.

To determine the location of the mass values of 1322.9, 1539.7 and 2224.4 (Table II-1) in the amino acid sequence of the B subunit and to detect unobserved sequences in the mass spectrum of the protease V8 digest of the BrCN-treated B subunit, the digest was separated by reversed-phase HPLC and the peptides with the mass values of 1322.9, 1539.7 and 2224.4 were isolated (data not shown). The peptide with the mass value of 1322.9 was degraded by the Edman method. The phenythiohydantoin (PTH) derivatives of amino acids released were analyzed by HPLC and PTH-Lys and PTH-Asp were identified in the first and second steps of degradation, respectively. The peptide fragments were submitted to FAB mass

Table II-1. Some mass values observed in the protease V8 digests of V. cholerae classical biotype Inaba 569B toxin from three different suppliers. The sequence is shown in Fig. II-2.

Mass values of peptides from toxin of			Predicted sequence	Calculated mass
List Biological	Sigma	Calbiochem		
2224.4	2224.0	2223.9	12-29	2223.1
1538.7 1539.7	1538.8 1539.8	1538.8 1539.7	38-51	1538.8
1322.9	1322.9	1322.9	69-79	1321.7

spectrometry [13, 14], giving signals at m/z=1194.8 and 1079.7 after the first and second Edman degradations, respectively (data not shown). The mass differences between 1322.9 and 1194.8 amu and between 1194.8 and 1079.7 amu in each step of the degradation were 128.1 (Lys) and 115.1 (Asp) amu, respectively. These results were compatible with those on the PTH derivatives released on Edman degradation. The mass value of 1079.7 of the peptide fragment remaining after the second cycle of Edman degradation was the same as that (1079.8 amu) in the spectrum shown in Fig. II-1 and coincided with the value (1079.6 amu) calculated for the sequence 71-79. These results suggested that the peptide with 1322.9 amu corresponded to the sequence 69-79 with an Asp instead of an Asn residue at position 70.

Peptides with 1538.7 and 1539.7 amu were isolated with

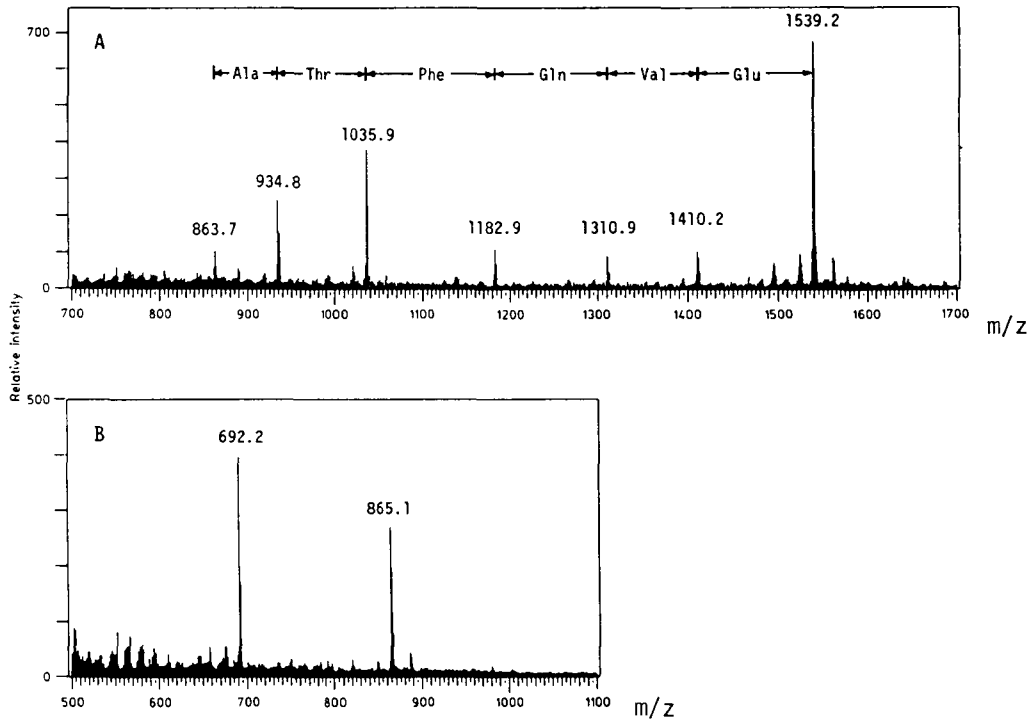

Fig. II-3. FAB mass spectra of a peptide (1538.7 amu) digested with (A) carboxypeptidase Y at 37°C for 45 min and (B) trypsin at 40°C for 4 h.

similar peak heights by HPLC and each was digested with carboxypeptidase Y and submitted to FAB mass spectrometry. As shown in Fig.II-3A, from the values of signals newly observed in the mass spectrum, the C-terminal sequence of the peptide with 1538.7 amu was confirmed to be Ala-Thr-Phe-Gln-Val-Glu, which is consistent with the published amino acid sequence of residues 46-51. A C-terminal sequence Thr-Phe-Gln-Val-Glu was similarly identified in the mass spectrum of the digest of the peptide with 1539.7 amu (data not shown). These results showed that the peptide

with 1539.7 amu has the same C-terminal sequence as the peptide with 1538.7 amu. These peptides were then digested with trypsin and submitted to mass measurement. As shown in Fig.II-3B, the peptide with 1538.7 amu was cleaved into two peptide fragments with 692.2 and 865.1 amu, which corresponded to the sequences 38-43 and 44-51, respectively. The tryptic digest of the peptide with 1539.7 amu gave a signal at m/z=692.2, which corresponded to the sequence 38-43. These results indicated that the peptide with 1539.7 amu should have 866.1 amu for the sequence 44-51. Since this peptide was confirmed to have the sequence Thr-Phe-Gln-Val-Glu at the C-terminus, which corresponded to the sequence 47-51, the mass of the sequence 44-46 of the peptide was calculated as 262.1 amu by subtracting the mass of the sequence 47-51 from that (866.1 amu) of the sequence 44-51. Since this value for the sequence 44-46 is 1 amu larger than that calculated from the published sequence, this sequence must be Asp-Gly-Ala instead of Asn-Gly-Ala in the published amino acid sequence (Fig.II-2). Thus, the peptides with 1538.7 and 1539.7 amu had Asn and Asp residues, respectively, at position 44.

Then the peptide with 2224.4 amu was digested with trypsin. The digest examined by FAB mass spectrometry without separation gave new signals at m/z=759.5 and 1483.8, as shown in Fig.II-4A, together with the signal of the undigested peptide at m/z=2224.4. The 759.5 amu corresponded to the theoretical mass value of 759.4 amu calculated for the sequence 24-29, which was formed by cleavage at Lys-23 from the sequence 12-29. However, the other

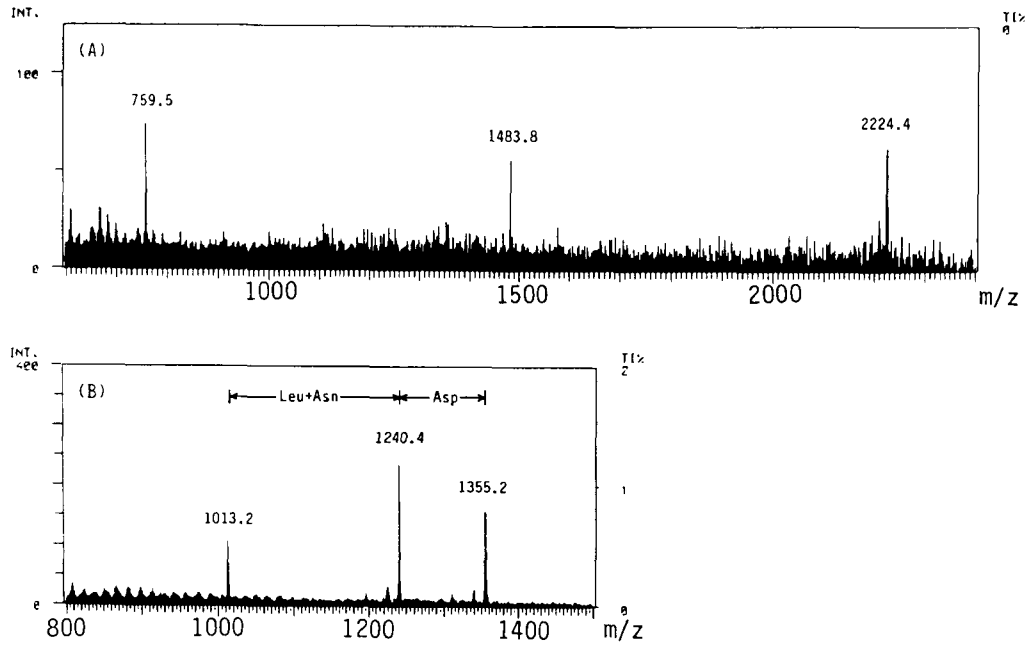

Fig. II-4. FAB mass spectra of (A) a peptide (2224.4 amu) digested with trypsin at 37°C for 4 h and (B) a peptide (1483.8) digested with carboxypeptidase B at 37°C for 2 h.

mass value of 1483.8 amu did not correspond to the mass value (1482.7 amu) calculated from the published amino acid sequence 12-23, suggesting that the observed values were incompatible with the published amino acid sequence 12-23. Then, the peptide with 1483.8 amu was isolated from the tryptic digest by HPLC. This peptide was treated with carboxypeptidase B and examined by FAB mass spectrometry. As shown in Fig.II-4B, new signals were observed at m/z=1355.2, 1240.4 and 1013.2, indicating that this peptide released Lys, Asp and Asp + Leu (or Ile) in that order from its C-terminus, as recognized from the mass differences between 1483.8 and 1355.2, 1355.2 and 1240.4, and 1240.4 and

1013.2, respectively (Leu and Ile cannot be distinguished in this method, because they have the same residue weights). These results indicated that the peptide with 1483.8 amu has the sequence (Asp, Leu or Ile)-Asp-Lys at 20-23; that is, the amino acid residue at position 22 is Asp instead of Asn in Fig.II-2.

Thus, the mass values observed at 1322.9, 1539.7 and 2224.4 could be located in the sequence of the B subunit and the unobserved sequences 12-29 and 69-79 were confirmed. The present data showed that commercial preparations of the B subunit of classical biotype Inaba 569B toxin have Asp residues at positions 22 and 70 in place of Asn in the published amino acid sequences [4, 5]. The results agreed well with the amino acid sequences

```
                -20       -10       -1 1       10        20
569B                                 TPQNITDLCAEYHNTQIHTLNNKIF
El Tor 62746    MIKLKFGVFFTVLLSSAYAHG----------------Y---D---
El Tor 2125     ---------------------------------------Y---D---
Present data                         ----------------H---D---

                30        40        50        60        70
SYTESLAGKREMAIITFKNGATFQVEVPGSQHIDSQKKAIERMKNT
--------------------I------G--------------D-
--------------------I------S--------------D-
--------------------T------G--------------D-

                80        90        100
LRIAYLTEAKVEKLCVWNNKTPHAIAAISMAN
--------------------------------
--------------------------------
--------------------------------
```

Fig. II-5. Sequence comparison of the B subunit of V. cholerae classical and El Tor biotype toxins. The top three sequences were obtained from other studies: 569B [4,5], El Tor 62746 [2] and El Tor 2125 [3].

translated from the nucleotide sequences of genes (*ctx*) encoding El Tor biotype toxins, as shown in Fig.II-5. Peptides with 1538.7 and 1539.7 amu were isolated in equal amounts from the digest of the B subunit and were verified to have Asn and Asp residues, respectively, at position 44. This fact suggests that Asn at position 44 was partially deamidated to Asp during preparation or purification of the holotoxin or separation of the subunits, because it is known that side-chain amide groups tend to deamidation in the presence of vicinal basic amino acid residues [15]. The present data revealed that the amino acid sequence of the classical biotype Inaba 569B toxin is different only at positions 18 (His-Tyr), 47 (Thr-Ile) and 54 (Gly-Ser) from that of El Tor biotype toxins (Fig.II-5).

II-3-2 Limulus anticoagulant anti-lipopolysaccharide factor

Fig.II-6 shows FAB mass spectra in the range of 400-4050 amu of the tryptic peptides derived from native anti-LPS factor before and after Edman degradation (upper and lower, respectively). The tryptic peptides gave 11 intense signals, T1, T1', T2, T3a, T3b, T4, T2-T4, T5, T6, T7, and T8. Out of these signals, T2, T3a, or T3b, T4, T5, T6, T7, and T8 shifted to T2#, T3#, T4#, T5#, T6#, T7#, and T8# after one cycle of Edman degradation by losing their N-terminal amino acid residues, Asn, Val, or Ile (or Leu), Phe, Ser, Asp, Ser, and Glu, respectively. It should be noted that the mass value containing Lys residue(s) is observed with an increase

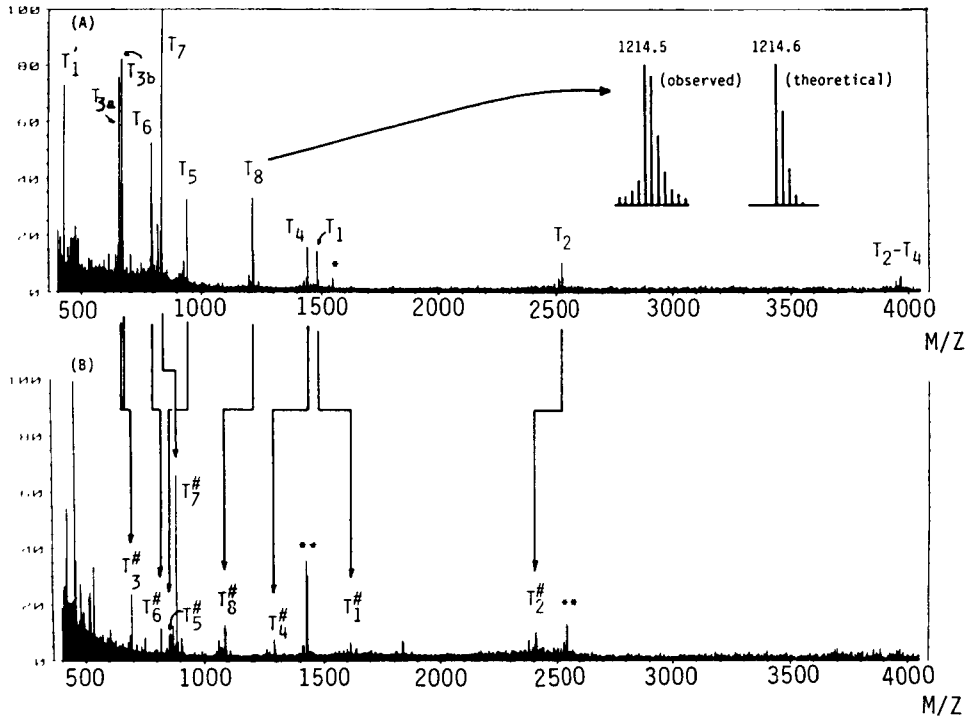

Fig. II-6. FAB mass spectra of tryptic digest of native anti-LPS factor before (upper) and after (lower) one cycle-Edman degradation. The upper inset shows the observed and theoretical isotopic molecular ion distribution of the peptide (Glu-93 to Gln-102). The signal with an asterisk mark may be derived from the peptide T4 with α-thioglycerol adducted to its Cys residue. The signals with two asterisks may be derived from the peptides T2# and T4# with PITC adducted to their Cys residues, respectively.

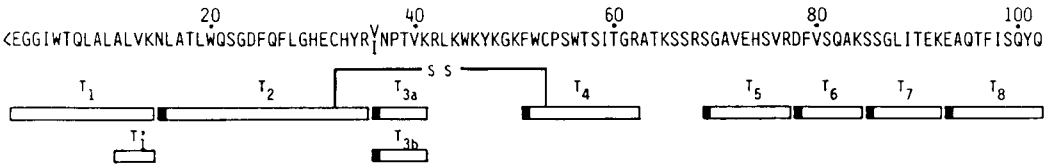

Fig. II-7. Amino acid sequence of anti-LPS factor. ▭ show the peptides observed in the upper spectrum (Fig. II-6), and ■ show the amino acid residues released by one cycle-Edman degradation.

of the mass value of PITC after Edman degradation, because the ε-amino group of Lys residue is modified with this reagent [13]. The observed mass values were in good agreement with the theoretical values calculated from the amino acid sequence (Fig.II-7), as listed in Table II-2.

The peptide T1 was not susceptible to Edman degradation to give T1# by adding PITC. This confirmed that the N-terminal residue of anti-LPS factor is a pyroglutamic acid. The mass values of T3a and T3b corresponded to those calculated for the peptides containing Val and Ile at the first residue, respectively. The mass value of T8 was consistent with that calculated for the

Table II-2. Mass values and locations of tryptic peptides of native anti-LPS factor.

Before Edman degradation				After Edman degradation				NH_2-terminal
Peptide	Observed values	Theoretical values	Sequences	Peptide	Observed values	Theoretical values	Sequences	
T_1	1480.7	1480.9	1-14	$T_1^{\#}$ a)	1615.5	1615.9	1-14	
T_1'	430.3	430.3	11-14					
T_2	2522.1	2522.2	15-35	$T_2^{\#}$	2408.3	2408.1	16-35	Asn
T_{3a}	657.4	657.4	36-41	$T_3^{\#}$ a)	693.2	693.3	37-41	Val
T_{3b}	671.4	671.3	36-41					Ile or Leu
T_4	1440.6	1440.7	51-62	$T_4^{\#}$	1293.3	1293.6	52-62	Phe
T_2-T_4	3959.8	3959.9	15-35 51-62					
T_5	941.4	941.5	69-77	$T_5^{\#}$	854.3	854.3	70-77	Ser
T_6	794.4	794.4	78-84	$T_6^{\#}$ a)	814.3	814.4	79-84	Asp
T_7	834.4	834.5	85-92	$T_7^{\#}$ a)	882.3	882.4	86-92	Ser
T_8	1214.5	1214.6	93-102	$T_8^{\#}$	1085.4	1085.5	94-102	Glu

a) Mass values of peptides containing the Lys residue(s) are observed with increase of 135 x n (n is number of Lys residues) after Edman degradation.

C-terminal peptide consisting of the sequence from Glu-93 to Gln-102. However, the isotopic molecular ion distribution was little different from the theoretical one calculated from the sequence from Glu-93 to Gln-102 and showed the contamination of the peptide (1215.4 amu) having one more atomic mass unit than the expected one, as shown in the upper inset of Fig.II-6. This mass value corresponded to the sequence from Glu-93 to Glu-102 instead of Gln-102. In fact, the two C-terminal peptides consisting of the same amino acid composition have been isolated by reversed-phase HPLC (data not shown). To further confirm the presence of two kinds of C-terminal residues, native anti-LPS factor was digested with protease V8, and the digest was measured by FAB mass spectrometry (Fig.II-8). The signal corresponding to the C-terminal peptide from Ala-94 to Glu-102 showed similar isotopic molecular ion distribution, with mass values at m/z=1085.3 and 1086.3, to that of the tryptic peptide as shown in the inset of

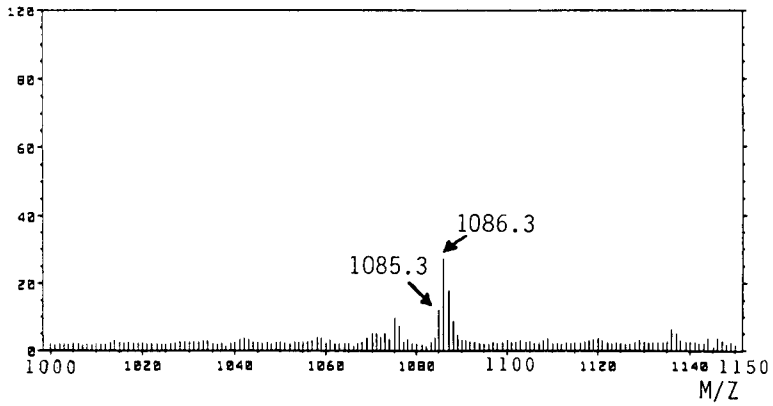

Fig. II-8. FAB mass spectrum of the protease V8 digest of native anti-LPS factor in the mass range of 1000 to 1150 amu.

Fig.II-6. These mass values were consistent with the sequence from Ala-94 to Gln-102 and Ala-94 to Glu-102, respectively. Thus, the author concluded that there exist two molecular species of anti-LPS factor; one has Gln and another Glu at the C-terminal residue. The mass value of T2-T4 was consistent with that calculated for a disulfide-containing peptide which consists of T2 and T4. Moreover, the signals of T2 and T4 corresponded to the peptides produced by reductive cleavage of the disulfide bond during the mass measurement, as described in chapter I. These results indicate that two half-cystine residues at positions 32 and 53 are linked by a disulfide bond.

Based on the above results, the complete amino acid sequence of anti-LPS factor isolated from the hemocyte lysate of Tachypleus tridentatus has been identified, as shown in Fig.II-7. The protein consists of a total of 102 amino acid residues with 2 half-cystines in disulfide linkage [10]. Its N-terminal is a pyroglutamic acid. The isolated anti-LPS factor contains the variants of Val and Ile at position 36 and of Gln and Glu at the C-terminus. However, it is not yet clear whether these two molecular species with variants of Gln/Glu at the C-terminus are present natively or the species with Glu at the C-terminal is derived from deamidation of the Gln residue during purification or enzymatic digestion of the native molecule. Moreover, the precise combination of these variant residues along the polypeptide chain remains to be clarified.

It was demonstrated that the direct measurement of protein digests by FAB mass spectrometry is a very easy and useful procedure for confirmation or determination of protein sequences without separation of the individual peptides in the digests. In particular, it is important to confirm or determine the exact amino acid sequences of proteins, if they are used in the studies on the structure-function relationships, physicochemical and immunochemical properties, etc. of proteins, even if their amino acid sequences have been reported. The present method should be useful for these purposes. Furthermore, a combination of FAB mass spectrometry and carboxypeptidase digestion or Edman degradation provides an easy and reliable method for determining the C- or N-terminal amino acid sequences of peptides or peptide mixtures, similarly to the method by FD mass spectrometry [13,16-18].

References

1) H. L. Vine and P. Cuatrecasas, Pharmacol. Ther., 12, 167 (1981).

2) H. Lockman and J. B. Kaper, J. Biol. Chem., 258, 13722 (1983).

3) J. J. Mekalanos, D. J. Swartz, G. D. N. Pearson, N. Harford, F. Groyne, and M. de Wilde, Nature, 306, 551 (1983).

4) C. -Y. Lai, J. Biol. Chem., 252, 7249 (1977).

5) A. Kurosky, D. E. Markel, and J. W. Peterson, J. Biol. Chem., 252, 7257 (1977).

6) C. D. Jacob, M. Sela, and R. Arnon, Proc. Natl. Acad. Sci., 80, 7611 (1983).

7) S. Tanaka, T. Nakamura, T. Morita, and S. Iwanaga, Biochem. Biophys. Res. Commun., 105, 717 (1982).

8) T. Morita, S. Ohtsubo, T. Nakamura, S. Tanaka, S. Iwanaga, K. Ohashi, and M. Niwa, J. Biochem., 97, 1611 (1985).

9) K. Ohashi, M. Niwa, T. Nakamura, T. Morita, and S. Iwanaga, FEBS Lett., 176, 207 (1984).

10) J. Aketagawa, T. Miyata, S. Ohtsubo, T. Nakamura, T. Morita, H. Hayashida, T. Miyata, S. Iwanaga, T. Takao, and Y. Shimonishi, J. Biol. Chem., 261, 7357 (1986).

11) P. Edman and A. Henschen, Protein Sequence Determination, (S. B. Needleman, ed.), 232, Springer, Berlin (1975).

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

13) Y. Shimonishi, Y. -M. Hong, T. Kitagishi, T. Matsuo, H.

Matsuda, and I. Katakuse, Eur. J. Biochem., 112, 251 (1980).

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

15) A. B. Robinson, J. H. McKerrow, and P. Cary, Proc. Natl. Acad. Sci., 66, 753 (1970).

16) Y. Shimonishi, Methods in Protein Sequence Analysis, (M. Elzinga, ed.), 271, Humana Press, Clifton, NJ (1982).

17) Y. Shimonishi, Y. -M. Hong, T. Takao, S. Aimoto, H. Matsuda, and Y. Izumi, Proc. Jap. Acad., 57B, 304 (1981).

18) Y. -M. Hong, T. Takao, S. Aimoto, and Y. Shimonishi, Biomed. Mass Spectrom., 10, 450 (1983).

CHAPTER III

Verification of Protein Sequences by Fast Atom Bombardment Mass Spectrometry Deduced from Nucleotide Sequences

III-1 Introduction

The determination of protein sequences has recently become fast and facile by the development of techniques for the nucleotide sequence analysis of the DNA of the genes encoding proteins [1-4]. Even though the results are most likely complete and accurate, it seems indispensable to verify protein sequences deduced from nucleotide sequences because errors such as point mutations and post-translational modifications are still possible. Biemann and his co-workers [5, 6] have used gas-chromatographic mass spectrometry, by which sequences of small peptides derived from enzymic or chemical cleavage of proteins can be analyzed, for verification of protein sequences deduced from cDNA sequences. The sequences of peptides determined by this method are aligned along to protein sequences deduced from cDNA sequences, providing the information on the DNA reading frame. However, the method seemed difficult in sample volatilization and in analysis of very complex mixture of fairly small peptides. The author considered that the method using FAB mass spectrometry described in chapter I is much suitable for verifying the amino acid sequences deduced from nucleotide sequences and searching for the DNA reading frames in

terms of accuracy, easiness and rapidity. In this chapter, the author describes the application of the method to verification of the amino acid sequence of protein S, a development-specific protein of Myxococcus xanthus whose sequence was deduced from the nucleotide sequence of its gene [7].

Protein S is a protein which is produced only during differentiation of M. xanthus, accumulates in very large amounts, and assembles on the surface of myxospores in the presence of Ca^{2+} [8]. Protein S has been purified and its C-terminal amino acid sequence has been determined using a C-terminal peptide fragment obtained by BrCN cleavage [9]. The gene encoding protein S has been identified and cloned using synthetic oligonucleotides corresponding to the C-terminal amino acid sequence of protein S as a probe [10]. It was found that two homologous genes (gene 1 and gene 2) are tandemly repeated in the same direction within a short distance [7]. DNA sequencing revealed that the gene 1 and gene 2 products have 88% homology, and protein S was deduced to be derived from gene 2 on the basis of its partial amino acid sequence [7]. In this study, in order to confirm this conclusion and develop a new method for verifying protein sequences deduced from the DNA sequences, the author applied the method to verification of the amino acid sequence of protein S.

III-2 Materials and Methods

All the experimental and analytical procedures and reagents were

the same as described in the preceding chapters, except those described below.

Sample preparation

Protein S was treated with BrCN, and a large fragment lacking 10 amino acid residues from the C-terminal end was isolated by gel filtration as described previously [9]. This large fragment is referred to as BrCN-treated protein S. BrCN-treated protein S [9] (300 μg) was digested with TPCK-treated trypsin at a substrate : enzyme ratio of 50 : 1 (w/w) at pH 8.0 and 37 °C for 4 h. 100 μg and 80 μg of the tryptic digest were submitted to Edman degradation [11] and digestion with *S. aureus* protease V8 [12], respectively. BrCN-treated protein S (200 μg) was also digested with protease V8. Then, 100 μg of the digest was submitted to Edman degradation.

III-3 Results and Discussion

Fig.III-1 shows FAB mass spectra in the range from 300 to 4100 amu of the tryptic digest of BrCN-treated protein S before and after Edman degradation (middle and top, respectively) and of the digest with protease V8 (bottom). The tryptic peptides gave 12 intense signals, T1, T2, T3, T4, T5, T6, T8, T10, T11, T11(a), T11(b), and T12, as shown in the middle spectrum in Fig.III-1. These signals are accompanied by weak signals which occurred 15amu lower than the individual parent signals. These weak signals were

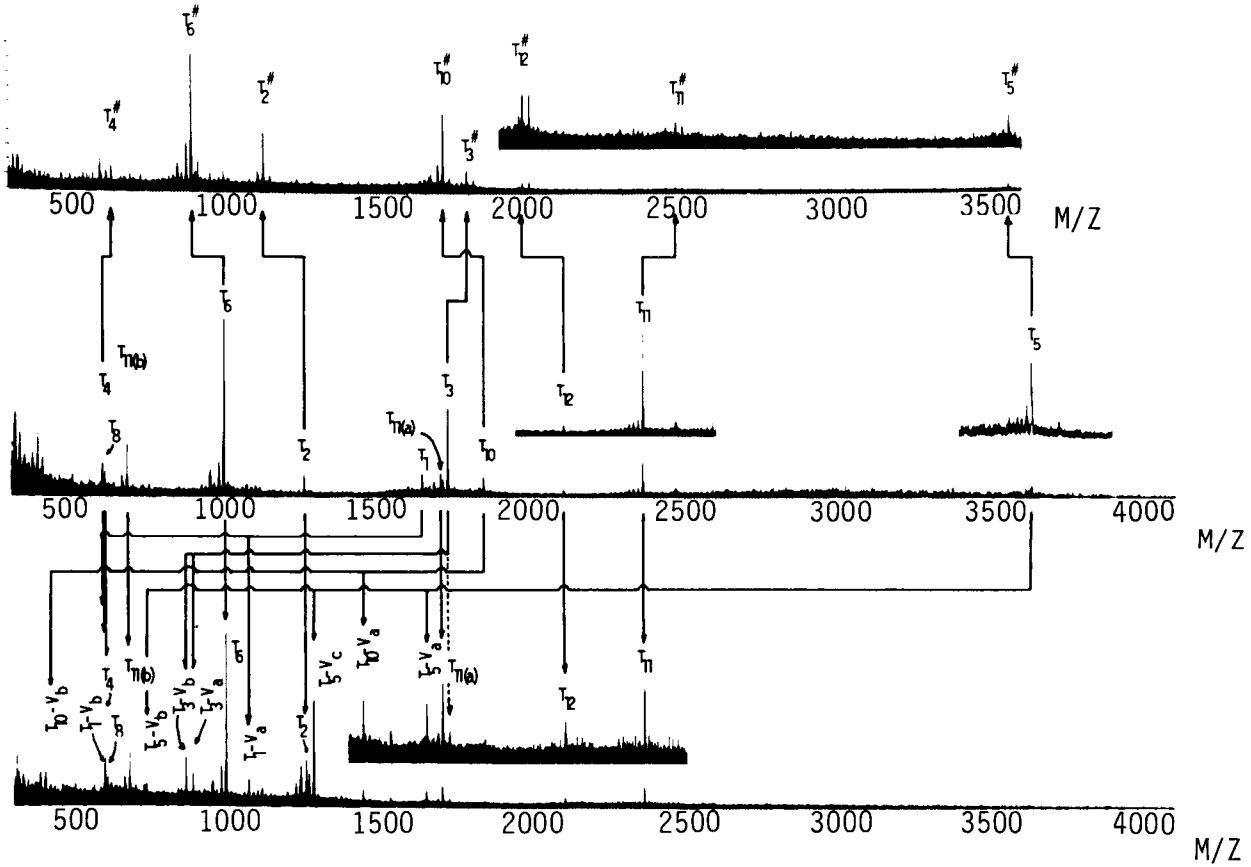

Fig. III-1. FAB mass spectra of BrCN-treated protein S. Tryptic peptides after one cycle-Edman degradation (top), tryptic peptides (middle), and tryptic peptides after digestion with protease V8 (bottom) are shown.

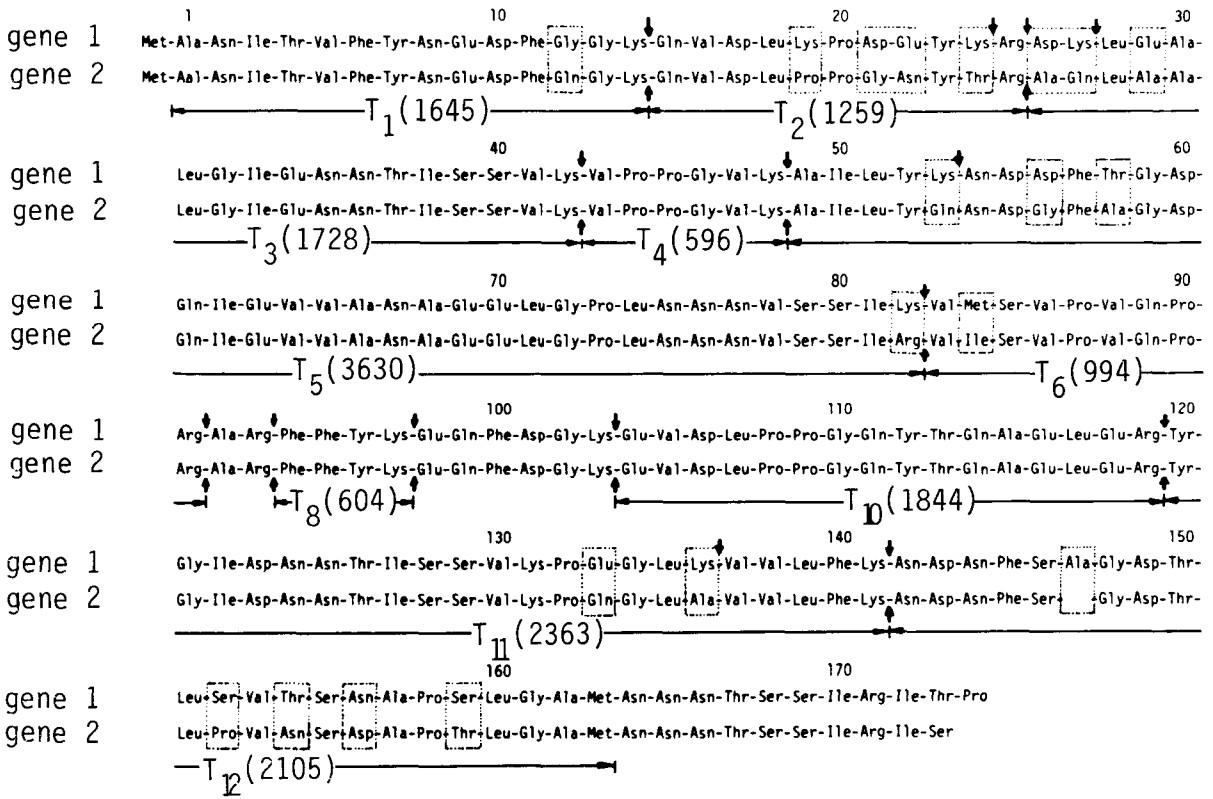

Fig. III-2. Amino acid sequences of gene 1 and gene 2 products determined by DNA sequencing. Arrows denote cleavage sites by trypsin. Residues enclosed by dashed lines indicate differences between the two sequences. Tn and numbers with parentheses show tryptic peptides and observed mass values, respectively.

Table III-1. Mass values and locations of tryptic peptides of BrCN-treated protein S.

Before Edman degradation				After Edman degradation				NH_2-terminal
Peptide	Observed mass value	Theoretical mass value	Sequence	Peptide	Observed mass value	Theoretical mass value	Sequence	
T1	1645.4	1645.79	1-14			1709.76[a]	2-14	
T2	1259.6	1259.65	15-25	T'2	1131.6	1131.60	16-25	Gln
T3	1728.7	1728.95	26-42	T'3	1793.3[a]	1792.93	27-42	Ala
T4	596.3	596.38	43-48	T'4	632.2[a]	632.32	44-48	Val
T5	3630.5	3630.80	49-82	T'5	3559	3559.77	50-82	Ala
T6	994.4	994.60	83-91	T'6	895.3	895.54	84-91	Val
T8	604.3	604.31	94-97			592.26	95-97	
T10	1844.7	1844.90	104-119	T'10	1715.9	1715.86	105-119	Glu
T11	2363.1	2363.30	120-141	T'11	2470[a]	2470.26	121-141	Tyr
T11(a)	1705.6	1705.88	120-135			1677.83	121-135	
T11(b)	676.4	676.44	136-141			740.42[a]	137-141	
T12	2105.9	2105.96	142-162	T'12	1974.0[b]	1973.91	143-162	Asn

[a] Mass values of peptides containing the Lys residue(s) are observed with increase of $135 \times n$ (n is number of Lys residues) after Edman degradation.
[b] The mass value of T'12 was observed as that of a peptide containing COOH-terminal homoserine lactone.

presumably caused by the loss of amino or methyl groups. The observed mass values were correlated to all the theoretical mass values calculated for the two possible amino acid sequences for protein S as shown in Fig.III-2. It should be pointed out that the tryptic peptides T1, T2, T3, T5, T6, T11, and T12, all of which can be assigned to the signals in Fig.III-1, are unique for the gene 2 product (see Fig.III-2). The theoretical and observed values for the tryptic peptides from the gene 2 product are in a very good agreement with each other, as listed in Table III-1. The observed mass value 1259.6 of signal T2 could correspond to any of the following three possible peptides: (a) from residue 77 (Asn) to residue 88 (Val) of the gene 1 product, (b) from residue 15 (Gln) to residue 25 (Arg) of the gene 2 product, or (c) from residue 123 (Asp) to residue 134 (Gly) of the gene 2 product (see Fig.III-2). Of these three possible sequences, however, the

sequence from residue 15 (Gln) to residue 25 (Arg) of the gene 2 product was assigned to signal T2 because of the arginine residue at the C-terminal end. The observed mass values for the signals T11(a) and T11(b) corresponded to the sequence from 120 (Tyr) to 135 (Leu) and that from 136 (Ala) to 141 (Lys) of the gene 2 product, respectively, which most likely resulted from nonspecific cleavage between Leu (135) and Ala (136). The middle spectrum in Fig.III-1 could not identify the positions from 92 (Ala) to 93 (Arg) and from 98 (Glu) to 103 (Lys), because their molecular weights were too small to be detected under the given conditions. These results clearly demonstrate that protein S is derived from gene 2. In order to further confirm the conclusion, the tryptic digest was subjected to Edman degradation without separating individual peptides [11]. The mixture was then subjected to mass measurement (top spectrum in Fig.III-1), yielding eight signals assigned as T#2, T#3, T#4, T#5, T#6, T#10, T#11, and T#12, the mass values of which are listed in Table III-1. It should be noted that mass values for T3, T4, and T11 increased after Edman degradation. This is because the ε-amino group of Lys residues in these peptides were modified with phenylisothiocyanate after Edman degradation. The phenylthiohydantoin (PTH) derivatives released by Edman degradation were quantitatively identified by HPLC (data not shown) to be the following seven PTH derivatives: Glu, Asn, Gln, Ala, Tyr, Val, and Phe. Using a computer program designed in the author's laboratory [11], all the mass shifts before and after Edman degradation were examined by correlation to these PTH

derivatives. For example, signal T2 with mass value 1259.6 could change to signal T#2 with mass value 1131.6 by releasing Gln (residue molecular weight, 128) during the degradation. Thus, peptide T2 was concluded to have Gln at its N-terminal end. This result confirmed the predicted sequence of peptide T2 from residue 15 (Gln) to residue 25 (Arg) of the gene 2 product as discussed earlier. Similarly, signals were concluded to shift from T3, T4, T5, T6, T10, T11, and T12 to T#3, T#4, T#5, T#6, T#10, T#11, and T#12, respectively, by losing Ala, Val, Ala, Val, Glu, Tyr, and Asn, respectively. These results completely agree with the predicted structures shown in Fig.III-2.

The tryptic digest of BrCN-treated protein S was further treated by protease V8 and subjected to mass measurement by FAB mass spectrometry. The result is shown in the bottom spectrum in Fig.III-1. By comparing it with the mass spectra before the digestion (middle spectrum), signals T1, T3, T5, and T10 were found to be shifted to T1-Va and T1-Vb, T3-Va and T3-Vb, T5-Va and T5-Vb and T5-Vc, and T10-Va and T10-Vb, respectively. On the other hand, peptides T2, T4, T6, T8, T11(a), T11(b), and T12 remained undigested by the enzyme. Table III-2 shows the identification of the newly produced peptides with their observed and theoretical mass values. The results are consistent with the fact that protease V8 mainly cleaves peptides at the carboxyl side of Glu residues [12].

In order to align the tryptic peptides, BrCN-treated protein S was digested with protease V8. One-half of the digest was

Table III-2. Mass values of BrCN-treated protein S successively digested with trypsin and protease V8.

Peptide	Observed mass value	Theoretical mass value	Sequence
T1-V(a)	1070.3	1070.52	1-9
T1-V(b)	594.2	594.29	10-14
T2	1259.6	1259.64	15-25
T3	1728.7	1728.95	26-42
T3-V(a)	885.4	885.50	26-34
T3-V(b)	862.3	862.46	35-42
T4	596.3	596.38	43-48
T5-V(a)	1653.6	1653.78	49-63
T5-V(b)	731.3	731.36	64-70
T5-V(c)	1283.7	1283.71	71-82
T6	994.4	994.60	83-91
T8	604.3	604.31	94-97
T10-V(a)	1446.6	1446.68	104-116
T10-V(b)	417.2	417.25	117-119
T11	2363.1	2363.30	120-141
T11(a)	1705.6	1705.88	120-135
T11(b)	676.4	676.44	136-141
T12	2105.9	2105.96	142-162

subjected to mass measurement and the other half to Edman degradation followed by FAB mass spectrometry, as in the case of tryptic digestion. The mass spectrum of the digest before Edman degradation gave 11 signals with moderately high intensity in the range from 300 to 3600amu (upper spectrum in Fig.III-3). Comparison of the mass spectra before and after Edman degradation leads to the conclusion that signals V1, V2, V2(b), V3(b), V4, V5, V5(a), and V8 were shifted to V#1, V#2, V#2(b), V#3(b), V#4, V5#, V#5(a), and V#8 by releasing Ala, Asp, Leu (or Ile), Gly, Val, Leu (or Ile), Leu (or Ile), and Val, respectively (Leu and Ile cannot be distinguished in this method, because they have the same residue weights). These PTH derivatives were identified by HPLC (data not shown). The observed mass values were compared with the theoretical mass values calculated for the gene 1 and the gene 2 products. It was concluded that signals V1, V2, V2(b), V3, V3(a), V3(b), V4, V5, V5(a), V6, and V8 correspond to the sequences from

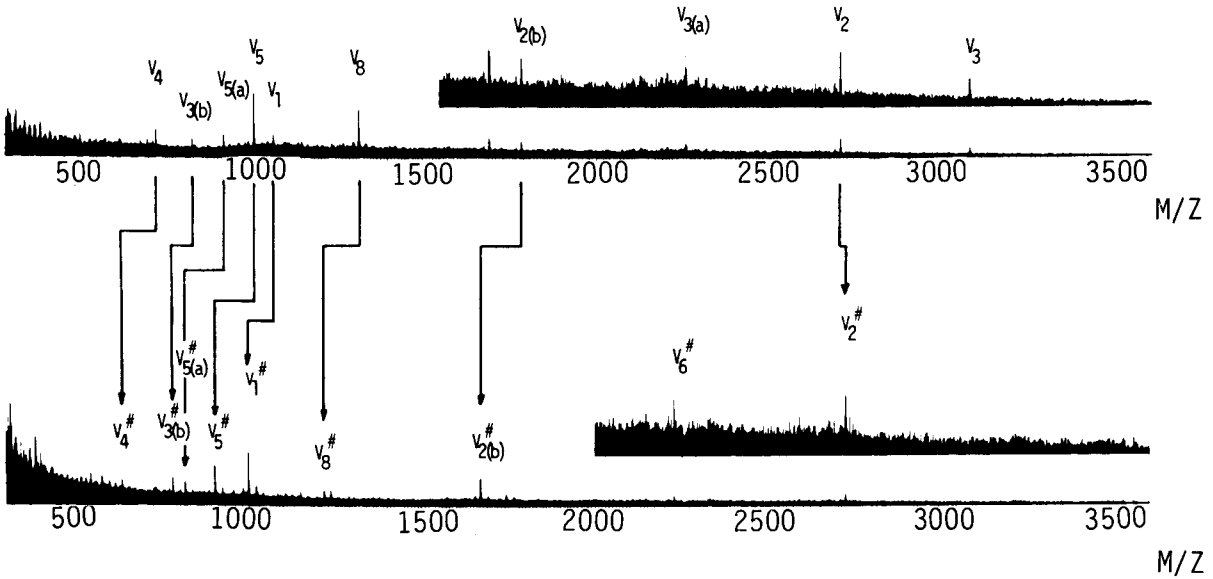

Fig. III-3. FAB mass spectra of the protease V8 digest of BrCN-treated protein S before (upper) and after (lower) Edman degradation.

Table III-3. Mass values and locations of the protease V8 peptides of BrCN-treated protein S.

Before Edman degradation				After Edman degradation				NH_2-terminal
Peptide	Observed mass value	Theoretical mass value	Sequence	Peptide	Observed mass value	Theoretical mass value	Sequence	
V1	1070.6	1070.52	1-9	V'1	999.2	999.48	2-9	Ala
V2	2701.3	2701.39	10-34	V'2	2720.9[a]	2721.38	11-34	Asp
V2(b)	1784.1	1783.97	18-34	V'2(b)	1670.4	1670.89	19-34	Leu (or Ile)
V3	3075.4	3074.58	35-63			3230.57[a]	36-63	
V3(a)	2257.1	2257.22	35-55			2413.20[a]	36-55	
V3(b)	836.4	836.38	56-63	V'3(b)	779.0	779.36	57-63	Gly
V4	731.3	731.36	64-70	V'4	632.0	632.29	65-70	Val
V5	1014.6	1014.52	71-80	V'5	901.0	901.44	72-80	Leu (or Ile)
V5(a)	927.5	927.49	71-79	V'5(a)	814.1	814.41	72-79	Leu (or Ile)
V6[b]	2205.2	2205.27	81-98	V'6	2227.2[a]	2227.20	82-98	Ile (or Leu)
V8	1317.8	1317.63	105-116	V'8	1218.3	1218.56	106-116	Val

[a] Mass values of peptides containing the Lys residue(s) are observed with increase of $135 \times n$ (n is the number of Lys residues) after Edman degradation.

[b] The signal at $m/z = 2205.2$ was obtained in another mass spectrum, although this signal was not observed in the spectrum in Fig. III-3.

1 (Ala) to 9, 10 (Asp) to 34, 18 (Leu) to 34, 35 (Asn) to 63, 35 (Asn) to 55, 56 (Gly) to 63, 64 (Val) to 70, 71 (Leu) to 80, 71 (Leu) to 79, 81 (Ile) to 98, and 105 (Val) to 116 of the gene 2 product (Fig.III-4, and also see Table III-3). Signals at m/z=1692.0 in the upper spectrum in Fig.III-3 and from 500 to 600amu in the lower spectrum could not be assigned to any sequences in protein S. Thus, the mass spectra of the protease V8 digest could account for the sequence from the N-terminus to residue 116 of protein S. The signal corresponding to the sequence from residue 117 to the C-terminus was not detected because the C-terminal peptide has a molecular weight of over 4000, which is beyond the mass range in the present experiment. The sequence from 92 to 93, which could not be identified in the tryptic digest, was now identified in the protease V8 digest. Table III-3 summarizes the observed and theoretical mass values for the peptides generated by the protease V8 digestion before and after Edman

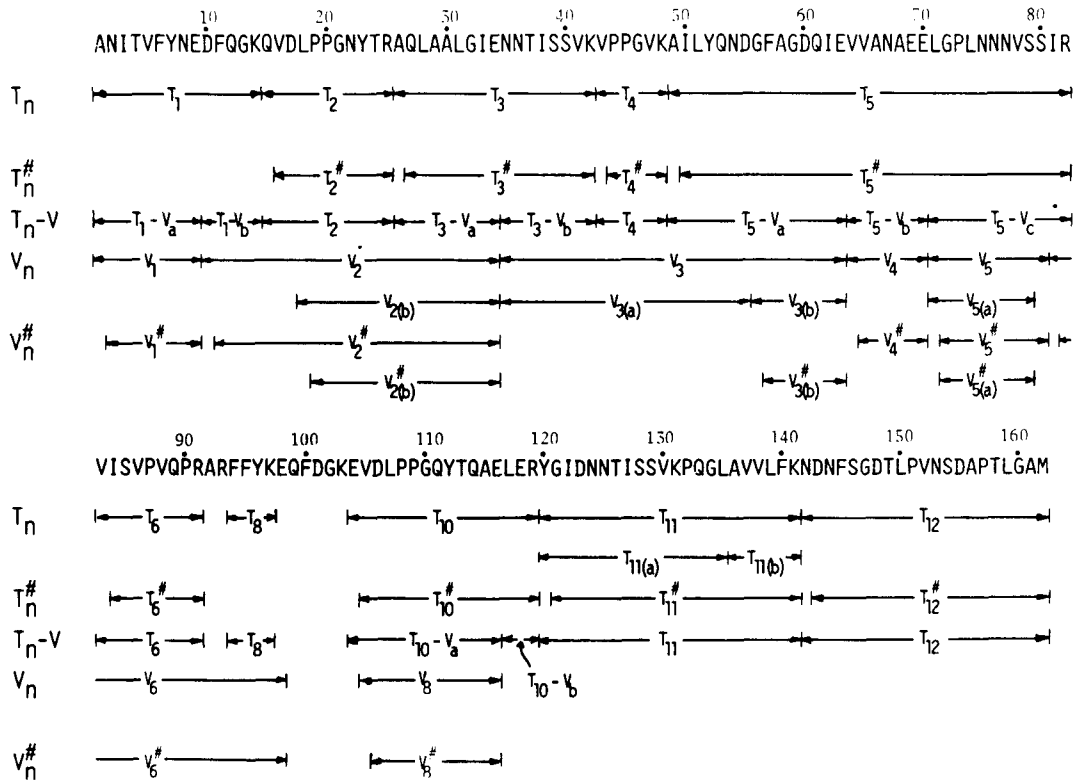

Fig. III-4. Summary of locations of observed mass values in the amino acid sequence of the BrCN-treated protein S.

degradation and their locations in the amino acid sequence of protein S. With use of the protease V8 peptides, the tryptic peptides could now be aligned along the amino acid sequence of protein S, as shown in Fig.III-4. The entire amino acid sequence of BrCN-treated protein S except for the sequence from 99 to 103 (see Fig.III-4) is covered by the mass spectra on the basis of the amino acid sequence deduced from the DNA sequence of gene 2, but not gene 1. Thus, it is unambiguously concluded that protein S is produced exclusively from gene 2 during differentiation of M. xanthus.

These results demonstrate that the direct measurement of protein digests by FAB mass spectrometry is a very useful and easy method for verifying protein sequences deduced from the DNA sequences and searching for the DNA reading frames.

References

1) A. M. Maxam and W. Gilbert, Proc. Natl. Acad. Sci., 74, 560 (1977).

2) F. Sanger and A. R. Coulson, J. Mol. Biol., 94, 441 (1975).

3) F. Sanger, A. R. Coulson, B. G. Barrel, A. J. H. Smith, and B. A. Roe, J. Mol. Biol., 143, 161 (1980).

4) J. Messing, R. Crea, and P. H. Seenburg, Nucleic Acids Res., 9, 309 (1981).

5) S. D. Putney, N. J. Royal, H. N. DeVegvar, W. C. Herlihy, K. Biemann, and P. Schimmel, Science, 213, 1497 (1981).

6) P. Hoben, N. J. Royal, A. Cheung, F. Yamao, K. Biemann, and D. Soll, J. Biol. Chem., 257, 11644 (1982).

7) S. Inouye, T. Franceschini, and M. Inouye, Proc. Natl. Acad. Sci., 80, 6829 (1983).

8) M. Inouye, S. Inouye, and D. R. Zusman, Proc. Natl. Acad. Sci., 76, 209 (1979).

9) S. Inouye, W. Harada, D. R. Zusman, and M. Inouye, J. Bacteriol., 148, 678 (1981).

10) S. Inouye, Y. Ike, and M. Inouye, J. Biol. Chem., 258, 38 (1983).

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

12) J. Houmard and G. R. Drapeau, Proc. Natl. Acad. Sci., 69, 3506 (1972).

CHAPTER IV

Identification of Primary Structures of Proteins by Fast Atom Bombardment Mass Spectrometry Produced by a Recombinant DNA Technique in Escherichia coli

IV-1 Introduction

Recently, advances in recombinant DNA technology has made it possible to produce many biologically interesting and clinically valuable proteins in large quantities by using bacterial cells such as Escherichia coli and Bacillus subtilis. However, it has been indicated in some cases that structurally modified proteins are synthesized endogenously in bacterial cells or produced during purification [1-3]. Therefore, it seemed necessary to identify the amino acid sequences of the proteins produced by a recombinant DNA technique in bacteria, even if they are confirmed by sequencing cDNA.

As described in chapter III, direct measurement of protein digests by FAB mass spectrometry is very useful for verification of protein sequences deduced from the nucleotide sequences. The author considered that the method can also be used for the above purpose. In this chapter, the author applied the method to identification of the primary structure of human interleukin-2 produced by a recombinant DNA technique in E. coli as well as that isolated from cultured human leukemic T-cells (Jurkat-III).

Interleukin-2 (IL-2), formerly named T-cell growth factor, is a lymphokine produced by activated T-cells [4, 5]. The important biological activities of this protein are reported to be promotion of long-term *in vitro* proliferation of antigen-specific effector T lymphocytes and induction of cytotoxic T lymphocyte reactivity [6-9], suggesting the clinical value of this protein in treatment of neoplastic diseases. Recently, mRNAs for human IL-2 from various sources, such as a leukemic T-cell line [10], peripheral blood lymphocytes [11], and tonsillar mononuclear cells [12], were isolated, and their cDNAs were cloned and sequenced. The amino acid sequences of mature human IL-2 predicted from these nucleotide sequences were identical, and hence their human IL-2s appeared to be homologous. Moreover, cloned cDNA to mRNA of a leukemic T-cell line was expressed in *E. coli*, and the expressed protein was purified by successive chromatographies, including HPLC [13]. For further investigation of the biological properties of the expressed protein and its use for various purposes, including clinical studies, it seemed important to analyze the primary structure of the protein.

IV-2 Materials and Methods

All the experimental and analytical procedures and reagents were the same as described in the preceding chapters, except those described below.

IV-2-1 Sample preparation

The purified sample (ca. 20 μg) of human IL-2 expressed in *E. coli* or Jurkat cell was treated with BrCN. The BrCN-treated protein was digested with *S. aureus* protease V8 and TPCK-treated trypsin for 3 h and 6 h, respectively, at a substrate : enzyme ratio of 50 : 1 (w/w) at pH 8.0 and 37 °C. Digestion with carboxypeptidase Y was carried out in pyridine/acetate buffer at pH 7.0 and 37 °C for the given period.

IV-2-2 Peptide synthesis

Two peptides (Phe-Leu-Asn-Arg-Trp-Ile-Thr-Phe-Cys-Gln-Ser-Ile and Trp-Ile-Thr-Phe-Cys-Gln-Ser-Ile) with sequences from position 117 to 128 and from 121 to 128, respectively, predicted from the nucleotide sequence of the human IL-2 gene were synthesized manually by the solid-phase method [14]. Portions of the synthesized peptides were oxidized to disulfides by treatment with $K_3[Fe(CN)_6]$ in 50% acetonitrile and purified by reversed-phase HPLC under the conditions described below.

IV-2-3 High-performance liquid chromatography

HPLC apparatus consisted of a Hitachi 655 liquid chromatograph equipped with a multiwave ultraviolet monitor, type 635M (Tokyo, Japan). For purification of synthetic peptides, a reversed-phase column of μBondapack C-18 (4.6 X 300 mm, Waters) was developed with a linear gradient of 15-40% acetonitrile containing 0.1% trifluoroacetic acid (pH 2.0) at a flow rate of 1 ml/min. For separation of the protease V8 digest of

carboxymethylated and BrCN-treated IL-2, a reversed-phase column (YMC-ODS, S-5, 4 X 250 mm) packed in the author's laboratory was developed with a linear gradient of 10-60% acetonitrile containing 0.05% trifluoroacetic acid (pH 2.3) at a flow rate of 1 ml/min. The absorbances at 220 and 280 nm of the eluate were monitored.

IV-3 Results and Discussion

Fig.IV-1 shows the FAB mass spectrum in the range from 300 to 2800 amu of the protease V8 digest of BrCN-treated IL-2 expressed in *E. coli* with a recombinant plasmid. Since the amount of the digest was limited, the intensities of signals of over 3000 amu were low. Therefore, peptides with mass values of over 3000 amu could not easily be detected. Low-molecular-weight peptides could also not be observed because they were buried under intense noise signals in the low mass region below 500 amu. However, the protease V8 digest gave 10-15 prominent signals of over 600 amu. The mass values of these signals were collated to the theoretical mass values calculated from the predicted amino acid sequence of Jurkat IL-2 as shown in Fig.IV-3. The mass values of signals V1, V2, V2', V3, V4, V6, V7, V7', and V13 agreed well with the theoretical mass values of peptide fragments prepared by specific cleavage with protease V8, as shown in Table IV-1. The peptide from position 53 to 62 containing a Cys residue was observed as the reduced form (signal V6). The V6 signal was accompanied by a signal at m/z=1347.8 (Fig.IV-1, asterisk) of an adduct with

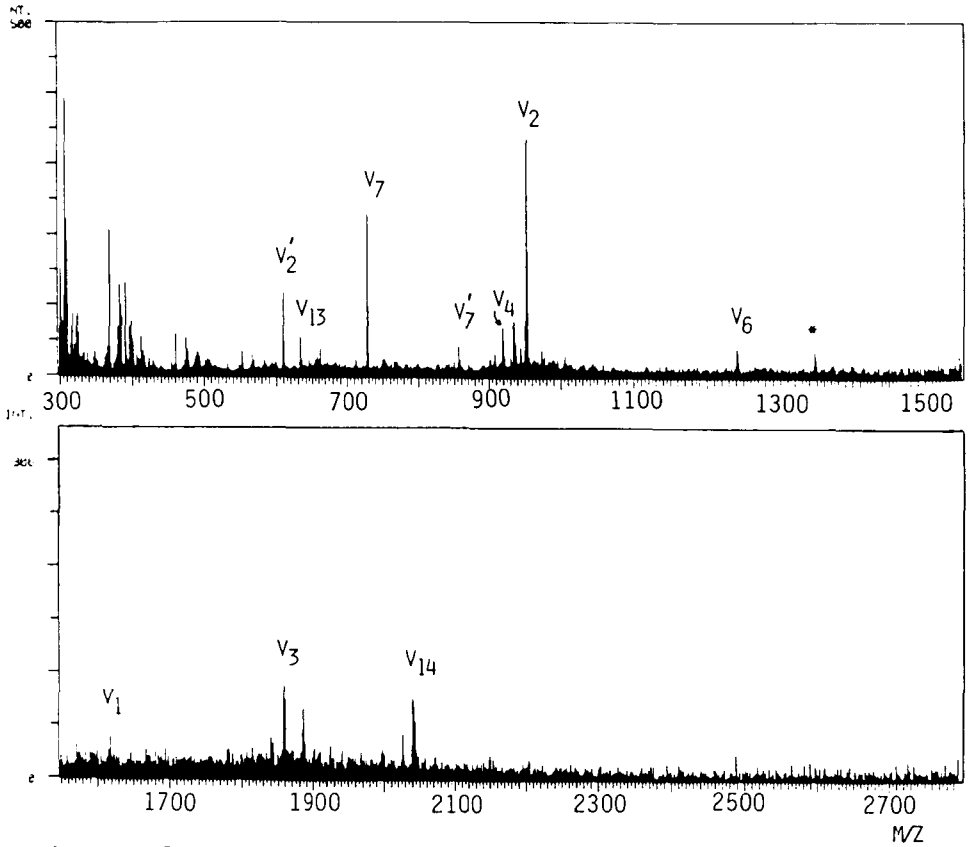

Fig. IV-1. FAB mass spectrum of the protease V8 digest of BrCN-treated IL-2.

α-thioglycerol, as described in chapter I. The mass value 2041.1 of signal V14 could not be related to any sequence predicted from the nucleotide sequence. However, as described below, this value was concluded to be due to a region from position 117 to 133 located in the C-terminal sequence, although the observed mass value was 2 amu smaller than that (2043.1) calculated from the amino acid sequence. Signals of sequences from residues 47 to 52 and from 69 to 110 could not be detected in the mass spectrum of the protease V8 digest for the reason mentioned above.

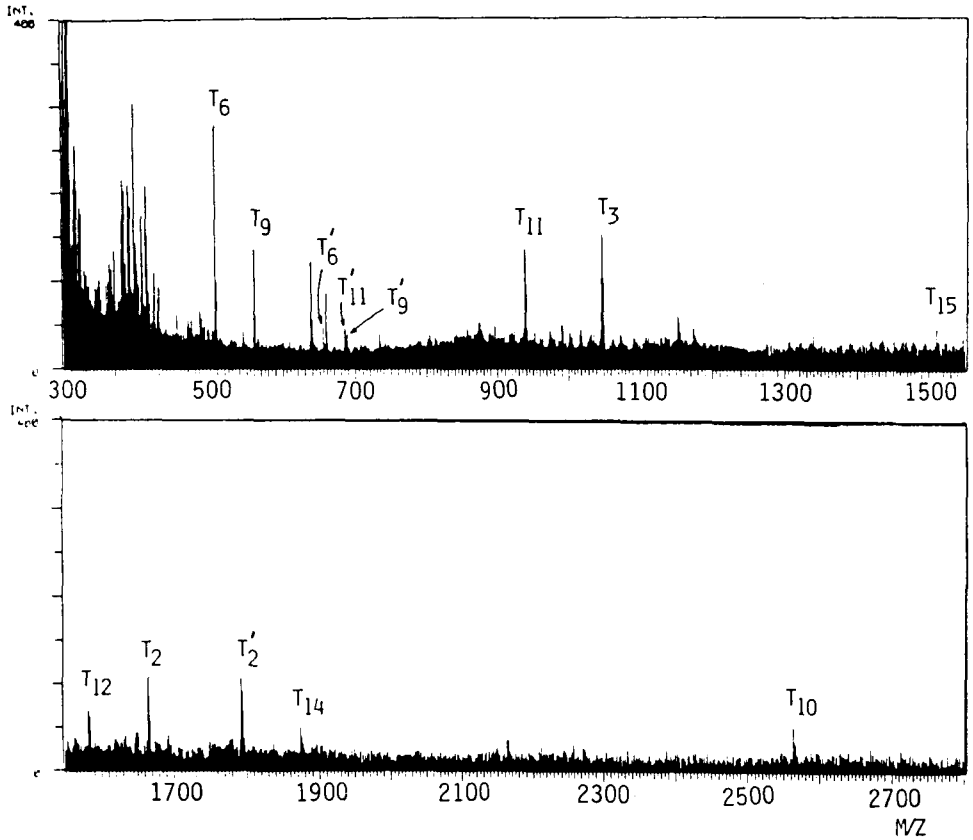

Fig. IV-2. FAB mass spectrum of the tryptic digest of BrCN-treated IL-2.

To obtain further information on the sequence of IL-2, the FAB mass spectrum of the tryptic digest was measured. The mass values of the intense signals T2, T2', T3, T6, T6', T9, T9', T10, T11, T11', T12, and T14 observed in the region of over 500 amu in Fig.IV-2 could be collated with the amino acid sequence of IL-2 as shown in Fig.IV-3 and Table IV-1. A few signals did not correspond to any sequence and might be derived from nonspecific cleavage or impurities. The sequence from position 69 to 97, which could not be identified in the mass spectrum of the protease V8

Table IV-1. Mass values and locations of protease V8 and tryptic peptides of BrCN-treated IL-2.

Protease V8 peptides				Tryptic peptides			
Peptide	Observed mass values	Theoretical mass values	Sequence	Peptide	Observed mass values	Theoretical mass values	Sequence
V_1	1618.8	1618.9	1-15	T_2	1665.1	1664.9	10-23
V_2	952.6	952.5	16-23	T_2'	1793.3	1793.0	9-23
V_2'	610.4	610.4	16-20	T_3	1048.7	1048.6	24-32
V_3	1859.0	1859.1	24-39	T_6	508.3	508.3	40-43
V_4	919.5	919.5	40-46	T_8'	655.3	655.4	40-44
V_6	1241.6	1241.6	53-62	T_9	561.3	561.3	50-54
V_7	728.4	728.4	63-68 (62-67)	T_9'	689.4	689.4	49-54
V_7'	857.5	857.5	62-68 (61-67)	T_{10}	2563.6	2563.3	55-76
V_{13}	633.4	633.3	111-116	T_{11}	939.4	939.5	77-83
V_{14}	2041.1	2043.1	117-133	T_{11}'	686.4	686.4	77-81
				T_{12}	1583.2	1582.9	84-97
				T_{14}	1874.3	1873.9	105-120
				T_{15}	1510.9	1512.8	121-133

digest, was seen in the mass spectrum of the tryptic peptides. The peptide containing the Cys residue at position 105 was observed in the reduced form (signal T14). The mass value (1510.9) of a signal T15 was considered to be that of the C-terminal fragment from position 121 to 133, as in the case of the protease V8 digest, although the value was 2 amu smaller than that deduced from the sequence. Thus, almost the entire amino acid sequence of IL-2 except for the Lys at position 48 and the sequence from position 98 to 104 was present in the mass spectra of the protease V8 and tryptic peptides as predicted from the DNA sequence.

To elucidate the difference between the observed and calculated mass values in the C-terminal sequence, BrCN-treated IL-2 was separated into four fractions (A-D) by reversed-phase HPLC (data not shown) and the FAB mass spectrum of each was measured. Fractions A-C showed signals at m/z=1840.9, 901.3, and 2534.1, corresponding to sequences with C-terminal homoserine lactone from positions 24 to 39, 40 to 46, and 1 to 23,

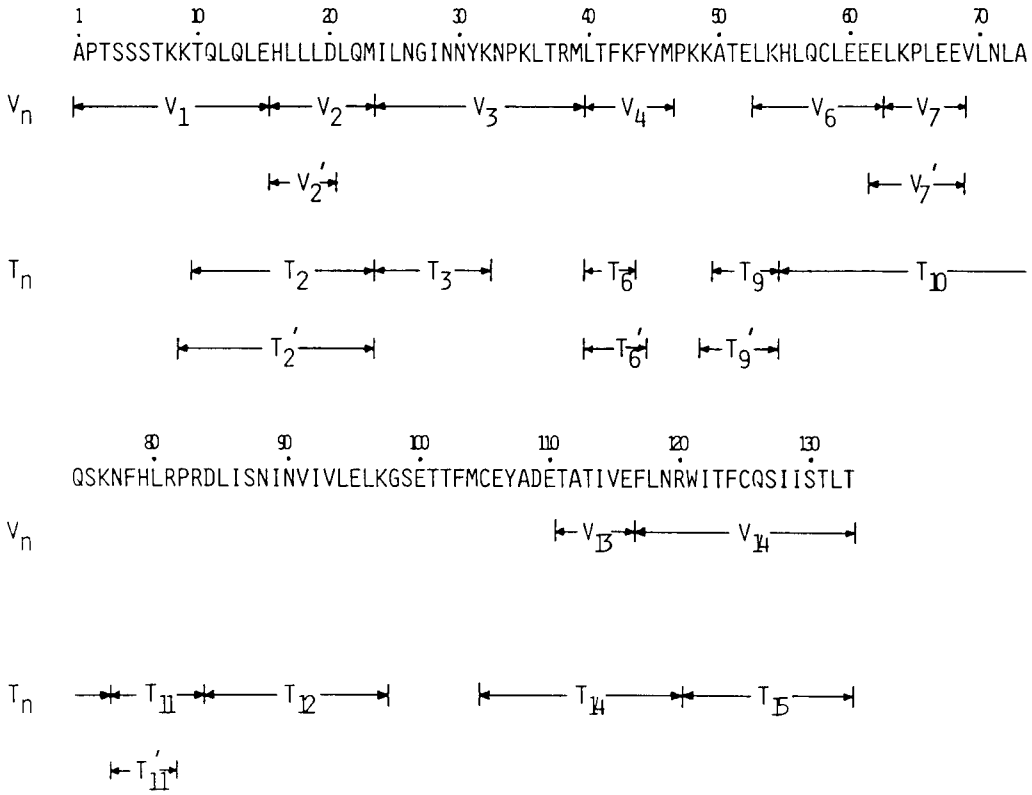

Fig. IV-3. Amino acid sequence of human IL-2 predicted from the nucleotide sequence of the cDNA. Vn and Tn indicate protease V8 and tryptic peptides, respectively, and correspond to the symbols described in Figs. IV-1 and 2.

respectively. These data support the results obtained in mass measurements of the protease V8 and tryptic peptides of IL-2 in Figs.IV-1 and 2, respectively. The protease V8 digest of fraction D gave five signals at m/z=633.5, 728.5, 983.8, 1241.8, and 2041.0, as shown in Fig.IV-4. The first four of these mass values were attributed to the sequences from position 111 to 116, 63 to 68 (or 62 to 67), 53 to 60, and 53 to 62, respectively. The mass value observed at m/z=2041.0 was identical with that in Fig.IV-1.

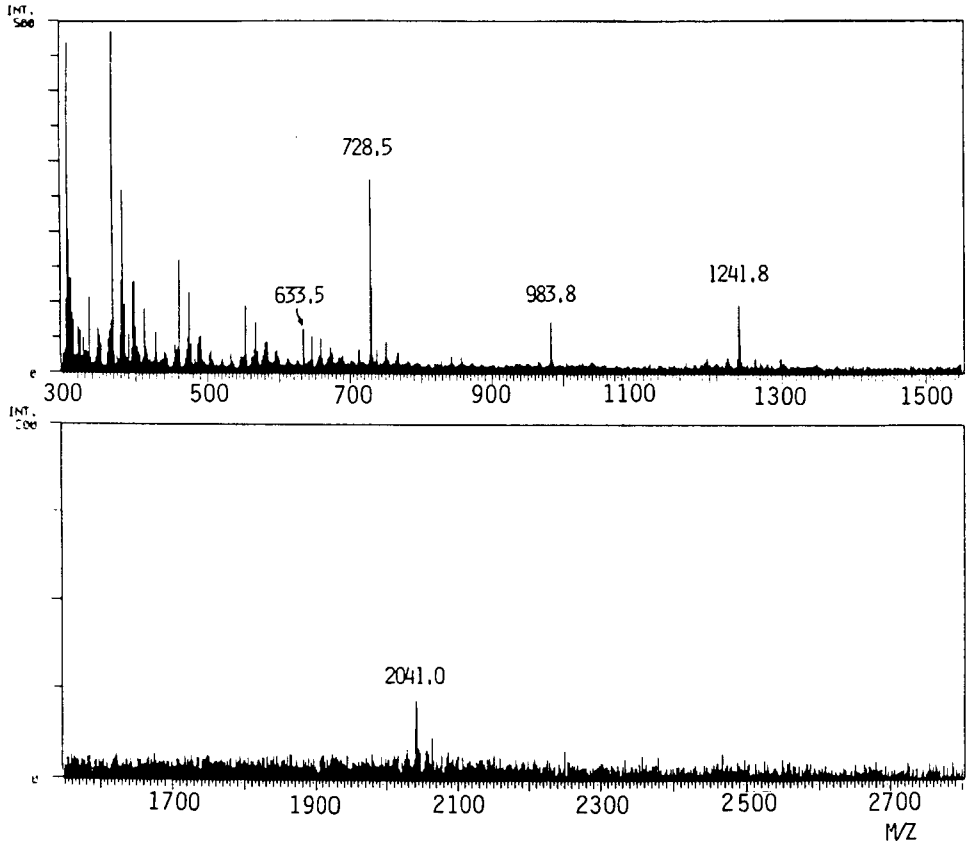

Fig. IV-4. FAB mass spectrum of the protease V8 digest of fraction D (C-terminal BrCN fragment).

In the mass spectrum of the protease V8 digest of fraction D after treatment with carboxypeptidase Y, the signal at m/z=2041.0 had disappeared and a new signal was seen at m/z=1412.5 (data not shown). This new signal probably resulted from removal of the C-terminal sequence (position 128-133, residual mass 628.4) from the original peptide (position 117-133, mass 2041.0) to leave a peptide of mass value 1412.5. The nucleotide-derived mass of the sequence 117-127 is 1414.7. The results therefore indicate that the 2 amu discrepancy is within the sequence 117-127. This

sequence could be further limited to that from position 121 to 127, because peptide 117-120 was found in the mass spectrum of the tryptic digest, as shown in Fig.IV-2.

Recently, human IL-2 isolated from normal peripheral blood and a leukemia T-cell line was examined by the Edman method [15-17]. The partial amino acid sequence determined was identical with the sequence predicted from the nucleotide sequence, although the carboxyl-terminal part of this sequence could not be determined due to its hydrophobic character. Moreover, the amino acid residue at position 3 from the N-terminus was found to be Thr modified with N-acetylgalactosamine [18]. The author isolated four BrCN fragments from IL-2 derived from a cell culture of a human leukemic T-cell line by reversed-phase HPLC (data not shown) and examined them by FAB mass spectrometry. Two of them gave signals at m/z=901.4 and 1841.3, which corresponded to the peptides from position 40 to 46 and from 24 to 39, respectively. The mass value of one of the two remaining fragments (m/z=2737.2) was identical with the theoretical mass value (2737.5) of the sequence from position 1 to 23, in which the Thr residue at position 3 is modified by N-acetylgalactosamine. This observation was further confirmed by mass measurement of the protease V8 digest of this fragment, which gave intense signals at m/z=610.3, 952.3, and 1821.3, corresponding to the peptides from position 16 to 20, 16 to 23, and 1 to 15 (with N-acetylgalactosamine), respectively, as shown in Fig.IV-5. The remaining fragment was digested with protease V8 and examined by FAB mass spectrometry. The spectrum

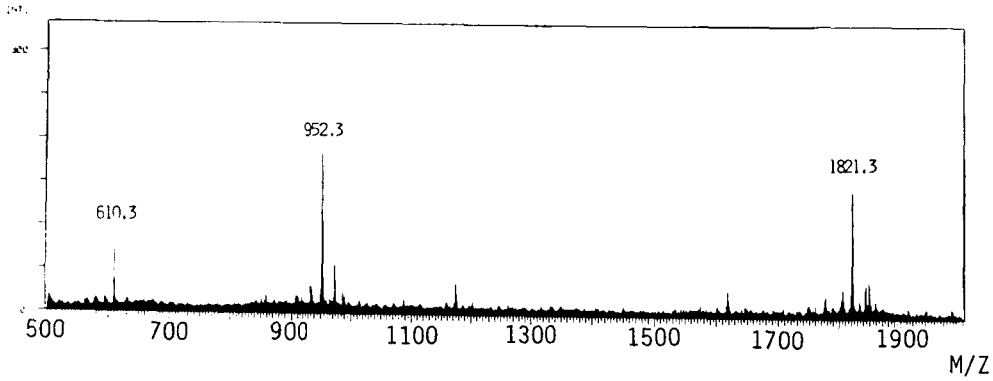

Fig. IV-5. FAB mass spectrum of the protease V8 digest of the N-terminal BrCN fragment of Jurkat IL-2.

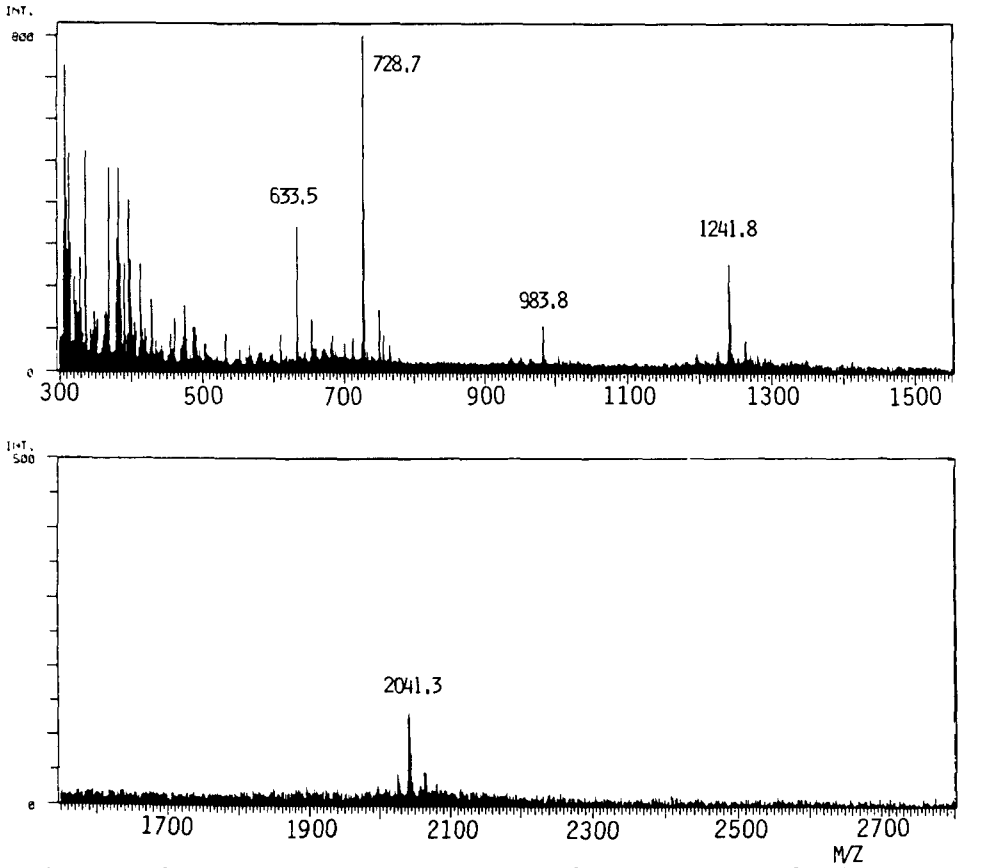

Fig. IV-6. FAB mass spectrum of the protease V8 digest of the C-terminal BrCN fragment of Jurkat IL-2.

(Fig.IV-6) showed intense signals at m/z=633.5, 728.7, 983.8, 1241.8, and 2041.3, like that in Fig.IV-4 of IL-2 expressed by bacteria. The mass value observed at m/z=2041.3 was 2 amu smaller than that calculated from the sequence 117-133.

Next, the author synthesized two peptides with the sequences from positions 121 to 128 and from 117 to 128. These peptides gave signals at m/z=997.4 (Fig.IV-7A) and 1528.2 (data not shown), respectively, which were identical with the mass values calculated from their sequences. However, when these peptides were treated with BrCN under the same conditions as for IL-2 and the products were examined by mass spectrometry, the author found that the original signals at m/z=997.4 and 1528.2 had disappeared with appearance of new signals at m/z=995.2 (Fig.IV-7B) and 1525.7 (data not shown), respectively. These values were 2 amu smaller than those before BrCN treatment. However, when these peptides were dimerized by disulfide formation, the mass values of the dimerized and reduced forms were observed and they were the same before and after BrCN treatment, as shown in Fig.IV-7C. These results indicate that free cysteine residues in peptides were modified by treatment with BrCN without modification of cystines and that after this treatment the mass values of the peptides were 2 amu smaller than the theoretical values. Thus, it was indicated that the Cys residue at position 125 was not involved in disulfide bond formation but present as a free thiol.

To further obtain this evidence, IL-2 was treated with monoiodoacetic acid without prior reduction and treated with BrCN.

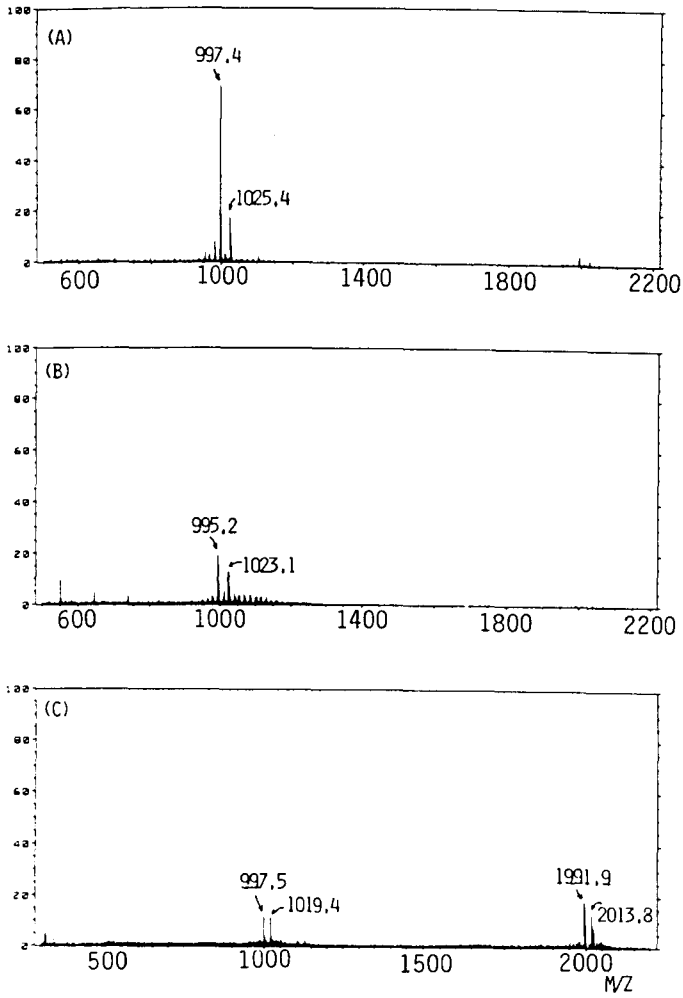

Fig. IV-7. FAB mass spectra of synthetic peptide with the sequence from position 121 to 128 predicted from the nucleotide sequence encoding Jurkat IL-2 (A), synthetic peptide in A after BrCN treatment (B), and a dimer of the synthetic peptide in A (C). The signals at m/z=1025.4 and 1023.1 in A and B, respectively, might correspond to formyl adducts of the parent ions because these peptides were dissolved in formic acid. The signals at m/z=1019.4 and 2013.8 in C corresponded to sodium clusters of the parent ions.

BrCN-treated IL-2 was digested with protease V8. The digest was separated by reversed-phase HPLC, and each peptide was subjected to FAB mass spectrometry. The observed mass value 2100.9 was consistent with the theoretical one (2101.1) of the C-terminal peptide from position 117 to 133 containing carboxymethylcysteine at position 125. Furthermore, the observation of a signal at m/z=1967.4, corresponding to a fragment with the sequences from position 53 to 62 and 105 to 110, provided evidence that the Cys residues at positions 58 and 105 are linked by a disulfide bond in IL-2. These results indicated that the Cys residue at position 125 is free, while the Cys residues at positions 58 and 105 are linked by a disulfide linkage.

The present results indicated that human IL-2 expressed in E. coli carrying a recombinant DNA has the same amino acid sequence as that predicted from the nucleotide sequence of cDNA for Jurkat IL-2, except that the Thr residue at position 3 is not modified. On the other hand, this residue was confirmed to be modified with N-acetylgalactosamine in IL-2 isolated from a leukemic T-cell line, as reported by other investigators [18]. An interesting finding was that the method could determine the positions of the disulfide bond and free cysteine residue in IL-2, i.e. the Cys residues at positions 58 and 105 are linked by a disulfide bond, while the Cys residue at position 125 is free. This result is consistent with the suggestion of disulfide bond formation between the Cys residues at positions 58 and 105 deduced in an investigation of the structure-activity relationships of IL-2 by a

recombinant method [19] and determined by recent sequence analysis [20].

The direct measurement of protein digests by FAB mass spectrometry as described in this chapter is a simple and useful procedure for identifying the primary structure of proteins produced by a recombinant DNA technique in bacteria.

References

1) M. G. Grutter, W. Marki, and H. P. Walliser, J. Biol. Chem., 260, 11436 (1985).

2) K. Axelsson, Acta Chem. Scand., B39, 69 (1985).

3) T. Takao, M. Kobayashi, O. Nishimura, and Y. Shimonishi, J. Biol. Chem., to be submitted.

4) D. A. Morgan, F. W. Ruscetti, and R. Gallo, Science, 193, 1007 (1976).

5) S. Gillis, M. M. Ferm, W. Ou, and K. A. Smith, J. Immunol., 120, 2027 (1979).

6) S. Gillis and K. A. Smith, Nature, 268, 154 (1977).

7) J. Watson, J. Exp. Med., 150, 1510 (1979).

8) S. Gillis, K. A. Smith, and J. Watson, J. Immunol., 124, 1954 (1980).

9) J. Watson, S. Gillis, J. Marbrook, D. Mochizuki, and K. A. Smith, J. Exp. Med., 150, 849 (1979).

10) T. Taniguchi, H. Matsui, T. Fujita, C. Takaoka, N. Kashima, R. Yoshimoto, and J. Hamuro, Nature, 302, 305 (1983).

11) R. Devos, G. Plaetinck, H. Cheroutre, G. Simons, W. Degrave, J. Tavernier, E. Remaut, and W. Fiers, Nucleic Acids Res., 11, 4307 (1983).

12) S. Maeda, N. Nishino, K. Obaru, S. Mita, H. Nomiyama, K. Shimada, K. Fujimoto, T. Teranishi, T. Hirano, and K. Onoue, Biochem. Biophys. Res. Commun., 115, 1040 (1983).

13) T. Tsuji, A. Yabuki, H. Morimoto, K. Fukuhara, and K. Toi,

Abstracts of the 104th Annual Meeting of the Pharmaceutical Society of Japan, Sendai, 468 (1984).

14) R. B. Merrifield, J. Am. Chem. Soc., 85, 2149 (1963).

15) R. J. Robb, R. M. Kutny, and V. Chowdhry, Proc. Natl. Acad. Sci., 80, 5990 (1983).

16) A. S. Stern, Y. -C. E. Pan, D. L. Urdal, D. Y. Mochizuki, S. DeChiara, R. Blacher, J. Wideman, and S. Gilles, Proc. Natl. Acad. Sci., 81, 871 (1984).

17) S. C. Clark, S. K. Arya, F. Wong-Staal, M. Matsumoto, M. Kobayashi, R. M. Kay, R. J. Kaufman, E. L. Brown, C. Shoemaker, T. Copeland, S. Oroszlan, K. Smith, M. G. Sarngadharan, S. G. Linder, and R. C. Gallo, Proc. Natl. Acad. Sci., 81, 2543 (1984).

18) R. J. Robb, R. M. Kutny, M. Panico, H. R. Morris, W. F. DeGrado, and V. Chowdhry, Biochem. Biophys. Res. Commun., 116, 1049 (1983).

19) A. Wang, S. -D. Lu, and D. F. Mark, Science, 224, 1431 (1984).

20) R. J. Robb, R. M. Kutny, M. Panico, H. R. Morris, and V. Chowdhry, Proc. Natl. Acad. Sci., 81, 6486 (1984).

CHAPTER V

Sequence Determination of Heat-Stable Enterotoxins Produced by Enteric Bacteria

V-1 Introduction

Enterotoxigenic *Escherichia coli* produces two types of enterotoxins that are responsible for diarrhea in human children and in domestic animals. One is a high-molecular-weight heat-labile enterotoxin (LT) which is similar to cholera toxin in their biological and physicochemical properties. The other is a low-molecular-weight heat-stable enterotoxin (ST) [1]. Burgess and his co-workers [2] reported two types of ST: the first, STI, is methanol-soluble and active in neonatal piglets and infant mice, but inactive in weaned pigs; the second, STII, is methanol-insoluble and active in weaned pigs, but inactive in infant mice. STI stimulates the guanylate cyclase-cyclic GMP system in enterotoxin-sensitive cells and raises the level of cyclic GMP [3-6], while the biological activities of STII are still unknown. Moreover, there exists two genetically distinct forms in STI [7], which are named STh and STp (derived from strains SK-1 and 18D of *E. coli*, respectively) in this study.

STI-like enterotoxins have been also known to be produced by *Yersinia enterocolitica* [8], *Klebsiella pneumoniae* [9], and *Enterobacter cloacae* [10]. The properties of *Yersinia*-ST resemble

those of STs of enterotoxigenic E. coli. For instance, its enterotoxigenic activity is neutralized by both antisera against E. coli-ST and antisera against Yersinia-ST, just as the toxicity of E. coli-ST is also neutralized by antisera against Yersinia-ST [11]. These phenomena suggested that the structures of at least the active sites of the two STs are very similar.

Furthermore, recently, an ST-like enterotoxin was isolated from Vibrio cholerae non-01 (a strain no. A-5) [12], which has similar, but not identical, physicochemical and immunological properties to those of E. coli-STs. These findings indicate that the toxin has a similar structure to E. coli-STs and Yersinia-ST.

Thus, it seemed interesting to determine the primary structures of these enterotoxins and clarify the structure-activity relationships among them. In this chapter, the author describes the purification and sequence determination of heat-stable enterotoxins produced by E. coli (strains SK-1 and 18D), Y. enterocolitica, and V. cholerae non-01. In particular, FAB mass spectrometry was used for determination of exact molecular weights and sequence analysis from C-termini of these enterotoxins.

V-2 Materials and Methods

Abbreviations. ST: heat-stable enterotoxin; STh, STp: heat-stable enterotoxins produced by enterotoxigenic *E. coli* strains SK-1 and 18D, respectively; *Yersinia*-ST, NAG-ST: heat-stable enterotoxins produced by *Y. enterocolitica* and *V. cholerae* non-01, respectively.

V-2-1 Materials

SP-Sephadex C-50 and DEAE-Sephadex A-25 were purchased from Pharmacia Japan. Aminopeptidase M and carboxypeptidase Y were purchased from the Protein Research Foundation (Minoh, Japan). Microsorb 5C18 column (5 μm, 4.6 X 250 mm) was obtained from Rainin Instrument Co. Inc. (MA, USA). Reversed-phase columns were packed with Lichrosorb RP-8 (5 μm, Merck) or YMC-ODS S-5 in the author's laboratory. All other reagents were the same as described in the preceding chapters.

V-2-2 Preparation of toxins

(STh and STp of *E. coli*) The cells (strains SK-1 and 18D) were grown in CAYE medium [13] containing kanamycin (50 μg/ml) at pH 8.6 without glucose at 37 °C for 24 h with vigorous shaking. CAYE medium was composed of 2% casamino acids (Difco), 0.6% yeast extract (Difco), 0.25% NaCl, 0.871% K_2HPO_4, and 0.1% (v/v) of a trace salt solution containing 5% $MgSO_4$, 0.5% $MnCl_2$ and 0.5% $FeCl_3$.

The culture supernatant was adjusted to 90% saturation of ammonium sulfate with stirring and the resulting suspension was kept overnight in a refrigerator at 2 °C. Then it was centrifuged at 16300 X g for 30 min at 4 °C, and the precipitate was dissolved in a small volume of water, and dialyzed against two changes of distilled water for 2-h periods. Then, the dialyzed solution was batch-treated with hydroxyapatite. The supernatant was decanted and concentrated to a small volume under reduced pressure. Ethanol was added to a final concentration of 90% and the mixture was kept overnight in a freezer at -20 °C. The resulting precipitate was removed by centrifugation (16300 X g) for 10 min at 4 °C, and the supernatant was concentrated to dryness (crude toxin) under reduced pressure.

(Yersinia-ST) The crude toxin which was isolated from the culture supernatant of a strain 23 of Y. enterocolitica in Ref. 14 was used.

(NAG-ST) The toxin which was isolated and purified from the culture supernatant of a strain (no. A-5) of V. cholarae non-01 in Ref. 12 was used.

V-2-3 Purification by ion-exchange chromatography

The crude toxin (derived from strains SK-1 and 18D of E. coli) was dissolved in a small amount of distilled water and applied to a column (0.9 X 40 or 3 X 46 cm) of SP-Sephadex C-50 (H^+ form) equilibrated with distilled water. The column was washed with distilled water and then the adsorbed toxin was eluted with

0.05 or 0.1 M ammonium acetate (pH 6.6 or 6.7) and fractions showing toxic activity were collected.

These fractions were diluted with distilled water and applied to a column (2 X 49 or 1.2 X 53 cm) of DEAE-Sephadex A-25 (acetate form) equilibrated with distilled water. Material was eluted with a gradient of 0-1 M acetic acid and then with 1 M formic acid. The fractions containing toxin were collected, freez-dried and stored in a refrigerator. All the fractions eluted from columns were examined by biological assay and/or HPLC.

V-2-4 Purification by high-performance liquid chromatography

(STh and STp of E. coli) HPLC was performed in the same apparatus as described in chapter I. The material obtained by column chromatography on DEAE-Sephadex A-25 was further purified on a preparative Lichrosorb RP-8 column (5 μm, 8 X 300 mm). The column was first equilibrated with 10% acetonitrile containing 0.01 M ammonium acetate (pH 5.7) and then materials dissolved in distilled water were injected onto the column. The column was developed with a linear gradient of 10-35% acetonitrile containing 0.01 M ammonium acetate (pH 5.7) at 30°C for 40 min at a flow rate of 2 ml/min. The main fractions containing toxic activity were collected and lyophilized. The resulting powder was redissolved in a small amount of water and rechromatographed on a Lichrosorb RP-8 column (5 μm, 4 X 250 mm) at 30°C at a flow rate of 1 ml/min in the same way as described above. Fractions of eluate were monitored for absorbance at 220 nm. The main fractions containing

toxic activity were collected and lyophilized.

(Yersinia-ST) HPLC was performed in the same apparatus as described in Chapter IV. The crude toxin was purified by two steps of reversed-phase HPLC. In the first step, the HPLC column (YMC-ODS, S-5, 4 X 250 or 8 X 300 mm) was equilibrated with 0.1% trifluoroacetic acid (pH 2.0) and developed with a linear gradient of 10-60% or 15-60% acetonitrile with increase in acetonitrile concentration of 0.5 %/min at a flow rate of 1 or 2 ml/min. The toxic fractions were separately pooled and purified further by the second step of HPLC, in which the column was developed with a linear gradient of 10-50% acetonitrile in 0.01 M ammonium acetate (pH 5.7). Fractions of eluate were monitored for absorbance at 220 and 280 nm. All procedures were done at room temperature. The fractions containing toxic activity were collected and lyophilized.

(NAG-ST) HPLC was performed in the same apparatus as described in Cnapter IV. Native and carboxymethylayed toxins were purified on a column (4 X 250 mm) of YMC ODS S-5 packed in the author's laboratory. The column was equilibrated with 10% acetonitrile in 0.05% trifluoroacetic acid (pH 2.35) and developed with a linear gradient of 10-40% acetonitrile at a flow rate of 1 ml/min. Fractions or eluate were monitored for absorbance at 220 and 280 nm. All procedures were done at room temperature.

V-2-5 Biological assay

ST activity was assayed in suckling mice of 2-4 days old, as

described in Ref. 15. The fluid accumulation ratio of each animal was calculated as the ratio of the weight of the entire intestine to that of the rest of the body. The minimal amount of ST giving a fluid accumulation ratio of over 0.09 was designated as 1 mouse unit (MU), as described in Ref. 15.

V-2-6 Amino acid analysis

The amino acid compositions of peptides were determined in a Hitachi type-835 analyzer using samples hydrolyzed in 4 M methanesulfonic acid at 110 °C for 24 h in evacuated sealed tubes.

V-2-7 Fast atom bombardment mass spectrometry

FAB mass spectra were recorded with a JEOL double-focusing mass spectrometer DX-300 or JMS-HX100 equipped with an FAB ion source. The ion source was a 1-3 or 5 kV accelerating potential. Other analytical conditions were the same as described in chapter I.

V-2-8 Aminoethylation, carboxymethylation and pyridylethylation of purified toxins

(STh and STp of *E. coli*) STs, 0.7 mg (STh) and 1.2 mg (STp), were dissolved in 0.4 M Tris/HCl (pH 8.6) containing 0.2% EDTA and 6 M guanidine/HCl and stirred for 45 min under a stream of nitrogen. Then, the solution was heated with dithiothreitol for 1 min in a boiling water and kept at about 50°C for 6 h. The solution was divided in half. One half was treated with ethylene

imine three times at 10-min intervals at room temperature. Then it was adjusted to pH 2 with 1 M formic acid, and passed through a column ot Sephadex G-10 using 1 M formic acid as an effluent. The fractions of eluate containing aminoethylated STs were collected and lyophilized. The recoveries were 210 μg (STh) and 540 μg (STp). The other half was treated with monoiodoacetic acid in 1 M NaOH for 15 min at room temperature in the dark. Then it was adjusted to pH 3 with 1 M formic acid, and passed through Sephadex G-10 using 1 M formic acid as an effuluent; fractions of eluate containing carboxymethylated STs were collected and lyophilized. The recoveries were 240 μg (STh) and 310 μg (STp).

(Yersinia-ST) A sample of native toxin (ca. 150 μg) was reduced and carboxymethylated, as described above. Carboxymethylated toxin was isolated by reversed-phase HPLC. A sample of native toxin (ca. 200 nmol) was reduced, as described above, and the resulting sample solution was mixed with 4-vinylpyridine (4.2 μl), stood for 1.5 h at room temperature and then adjusted to pH 3 with 1 M formic acid. The solution was applied to a column of Sephadex G-25 fine (0.9 X 33 cm) and pyridylethylated toxin was eluted with 1 M formic acid, collected and lyophilized.

(NAG-ST) A sample of native toxin (ca. 100 μg) was reduced and carboxymethylated, as described above. Carboxymethylated toxin was purified by reversed-phase HPLC.

V-2-9 Enzymatic digestion

(*S. aureus* protease V8) A peptide was dissolved in 0.5% ammonium bicarbonate and mixed with protease V8 with a substrate : enzyme ratio of approximately 30 : 1 (w/w). The mixture was deaerated with a stream of nitrogen and kept at 37 °C for 18 h.

(Carboxypeptidase B) A peptide or a peptide mixture was dissolved in 0.01 M pyridine acetate (pH 7.0) and treated with carboxypeptidase B with a substrate : enzyme ratio of 10 : 1 - 50 : 1 (w/w) at 37 °C. Aliquots were withdrawn from the digest at appropriate times for mass measurement.

(Carboxypeptidase Y) A peptide (2 - 5 μg) was digested with carboxypeptidase Y with a substrate : enzyme ratio of 20 : 1 or 5 : 1 (w/w) at 37 °C for 90 min or 15 h in 0.01 M pyridine acetate (pH 7.0).

(Aminopeptidase M) A peptide (4 - 5 μg) was dissolved in 0.01 M pyridine acetate (pH 7.0) and treated with aminopeptidase M with a substrate : enzyme ratio of 20 : 1 (w/w) at 37 °C for 5 h.

V-2-10 Edman degradation

(STh and STp of *E. coli*) Carboxymethylated STs, 95 μg (STh) and 155 μg (STp), were dissolved in a mixture of pyridine and water (1 : 1, v/v) adjusted to pH 9.6 with N-methylmorpholine. Edman degradations were performed manually. The resulting phenylthiohydantoin (PTH) derivatives of amino acids were analyzed by HPLC on a Zorbax ODS column (4.6 X 250 mm) with a mixture of acetonitrile and 0.01 M sodium acetate (pH 4.5) (42 : 58, v/v) as solvent. HPLC was performed at 62 °C and a flow rate of 1.0 ml/min.

(Yersinia-ST) The sequential degradation of peptides (1 nmol of native and 2.4 nmol of carboxymethylated toxins), purified from the culture supernatant treated with protamine sulfate, was performed with an Applied Biosystems gas-phase sequenator model 470A (California, USA), as described in Ref. 16. The PTH derivatives were analyzed by HPLC (Spectra-Physics SP8100 Liquid Chromatograph, California, USA) on a SEQ-4 column (4.6 X 300 mm, Senshu Scientific Co. Ltd., Tokyo, Japan) with a step-wise gradient of acetonitrile in 40 mM sodium acetate.

The native toxins (10 - 15 nmol), purified from the culture supernatant treated with protamine sulfate and DEAE-Sephacel, were dissolved in a mixture of pyridine and water (1/1, v/v) adjusted to pH 9.6 with N-methylmorpholine and degraded manually by the Edman method. Pyridylethylated peptide (ca. 7 nmol) was degraded in a JEOL JAS-570K sequence analyzer. After three precycling operations with Polybrene of the sequence analyzer, an aqueous solution of the sample was introduced into the cup of the analyzer. The apparatus was operated as recommended by the manufacturer. For identification of the resulting PTH derivatives, the column (Microsorb 5C18 column, 4.6 X 250 mm) was developed isocratically with a mixture of acetonitrile and 0.0125 M sodium acetate (pH 4.5) (40 : 60, v/v) as solvent. HPLC was performed at 40°C and a flow rate of 1.4 ml/min.

(NAG-ST) The native toxin (ca. 12 nmol) was manually degraded by the Edman method. The carboxymethylated toxin (2.45 nmol) was degraded sequentially with an Applied Biosystems

gas-phase sequenator model 470A, as described above. The resulting PTH derivatives were analyzed by HPLC (Vista 5500, Varian) on a Varian micropack SP-ODS column (4.6 X 150 mm).

V-3 Results

V-3-1 Heat-Stable Enterotoxin Produced by Enterotoxigenic *E. coli* Strain SK-1

(Purification) Crude toxin extracted from the culture supernatant of *E. coli* SK-1 was first chromatographed on SP-Sephadex C-50 (Fig.V-1). The main toxic activity was recovered in a peak eluted with 0.05 M ammonium acetate. Since the peak fraction still contained a large amount of pigmented material, the fraction shown by a horizontal bar were then subjected to ion-exchange chromatography on DEAE-Sephadex A-25. The toxic activity emerged in the fractions shown by a horizontal bar in Fig.V-2. The material was further purified on a preparative reversed-phase column by HPLC (Fig.V-3). Most of the toxic activity was obtained in the well-separated peak eluted with about 28% acetonitrile. Since the fraction was still slightly coloured after lyophilization and its purity had to be examined, it was subjected to repeated HPLC under the same conditions as described above. The toxin was eluted as a single peak (not shown). The purification of STh is summarized in Table V-1. The amino acid compositions of purified native ST, reduced and carboxymethylated

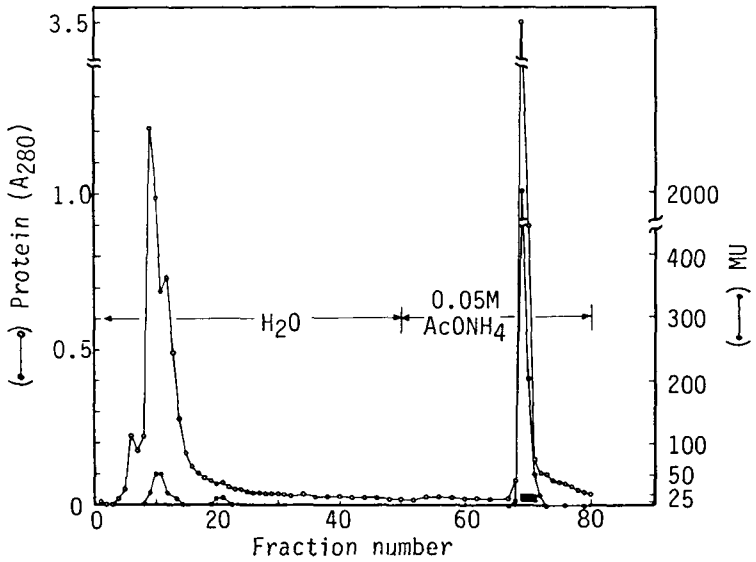

Fig. V-1. Chromatography on SP-Sephadex C-50 of crude toxin. The column was developed at a flow rate of 20 ml/h and 2.38-ml fractions were collected. The fractions containing toxin, shown by a horizontal bar, were pooled.

ST, and performic-acid-oxidized ST are summarized in Table V-2. From the results, the total number of amino acid residues of the purified STh was deduced to be 19. The smallest effective amount of purified STh for fluid accumulation in suckling mice was found to be about 2.5 ng, which was almost the same value as described previously [15].

(Molecular weight determination) Fig.V-4 shows the FAB mass spectrum of native STh, giving an intense signal at m/z=2041. The result indicates that the molecular weight of STh is 2040. This value was consistent with the theoretical value calculated from the amino acid composition. Since the amino acid composition and observed molecular weight of STh suggested that STh molecule

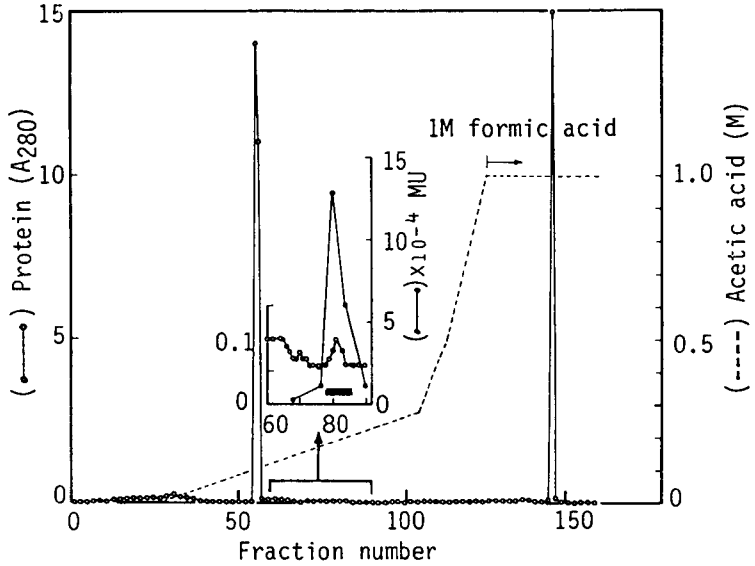

Fig. V-2. Chromatography on DEAE-Sephadex A-25 of the toxin fraction obtained from SP-Sephadex C-50. The column was developed at a flow rate of 19.5 ml/h and 12-ml fractions were collected.

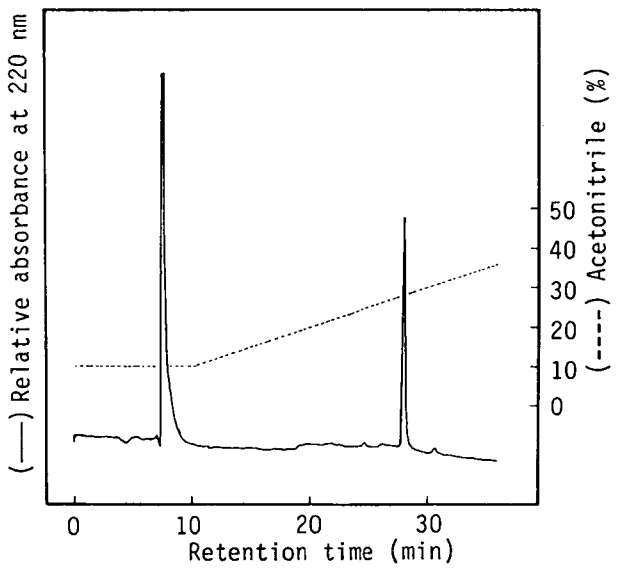

Fig. V-3. Preparative HPLC of the toxin from DEAE-Sephadex A-25 on a Lichrosorb RP-8 column. Conditions were as described in Materials and Methods (V-2-4).

Table V-1. Summary of purification of STh. One mouse unit (MU) is defined at the minimum amount that gives a fluid accumulation ratio of 0.09 or more.

	Specific activity	Relative activity	Recovery (%)
Culture supernatant (4 L)	1000 ng/mu	1	100
Crude toxin	16	62.5	62.5
SP-Sephadex C-50	17	58.5	51.5
DEAE-Sephadex A-25	3.4	294.1	9.7
HPLC	2.5	400	6.8
Final product			1.4 mg

Table V-2. Amino acid composition of purified STh.

	Native ST_h	Carboxymethylated ST_h	Performic acid oxidized ST_h	Nearest integer
$CySO_3H$			5.09	
Cys(CM)		6.00		
Asp Asn	2.96	3.09	2.91	3
Thr	0.96	0.98	0.89	1
Ser	1.91	1.93	1.78	2
Glu Gln	1.16	1.11	1.09	1
Gly	1.21	1.11	1.09	1
Ala	1.00	1.00	1.00	1
$\frac{1}{2}$Cys	4.76	0.11		6
Leu	1.10	1.09	0.94	1
Tyr	1.92	2.02		2
Pro	1.26	1.16	0.94	1
His				
Phe				
Total				19

a) Values are shown as calculated mol per 1 mol of alanine.

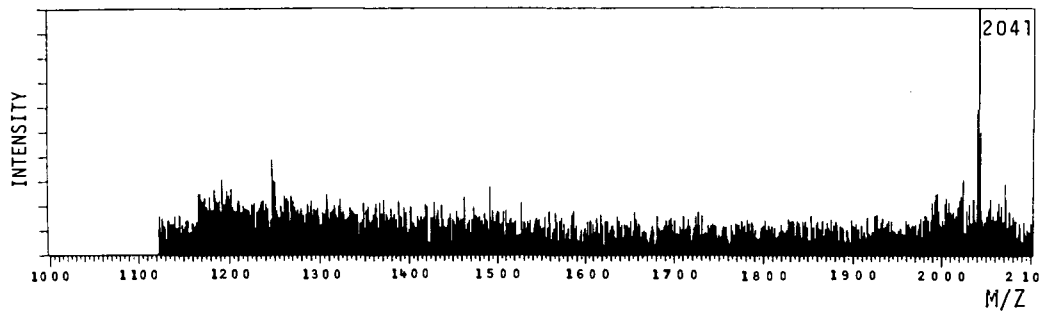

Fig. V-4. FAB mass spectrum of purified native STh.

contains at least one acidic amino acid residue, reduced and aminoethylated STh was digested with protease V8 and subjected to mass measurement. Fig.V-5A shows the FAB mass spectrum of the protease V8 digest of reduced and aminoethylated STh, giving prominent signals at m/z=1005 and 1319. These mass values were consistent with that of native STh.

(Amino acid sequence) Reduced and carboxymethylated ST was degraded manually in 14 steps by the Edman method. The results revealed 13 amino acid residues except for the 11th position from the N-terminal end (Fig.V-6). The theoretical mass value corresponding to the sequence from Asn-1 to Glu-8 was completely identical to that (1005) of a peptide fragment observed in Fig.V-5A. The sequence starting from Leu-9 seemed to correpond to a peptide fragment with a mass value of 1319 observed in Fig.V-5A.

A combination of FAB mass spectrometry and carboxypeptidase digestion is a simple and accurate method for determining C-terminal amino acid sequences of peptides, as described in

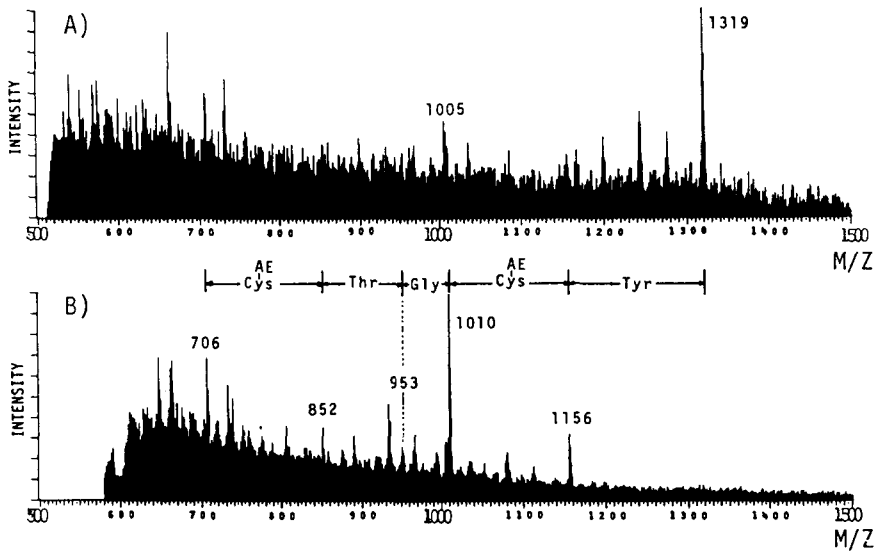

Fig. V-5. FAB mass spectra of (A) aminoethylated STh digested with protease V8 and (B) a mixture of (A) treated with carboxypeptidase B for 5, 10, and 15 min.

chapter II. The author used this method to determine the C-terminal amino acid sequence of STh. The protease V8 digest of reduced and aminoethylated ST was treated with carboxypeptidase B without separation of the components and subjected to FAB mass measurement. Since a commercial carboxypeptidase B used was found to be contaminated by other carboxypeptidases, the author used it for sequence determination of ST, although the toxin does not contain any basic amino acid residues. Fig.V-5B shows the FAB mass spectrum of a mixture withdrawn from the digest after 5, 10 and 15 min. From the observed mass values, all possible mass shifts of peptides were sought, taking into consideration the fact that the author had observed fragmentation except for that of the

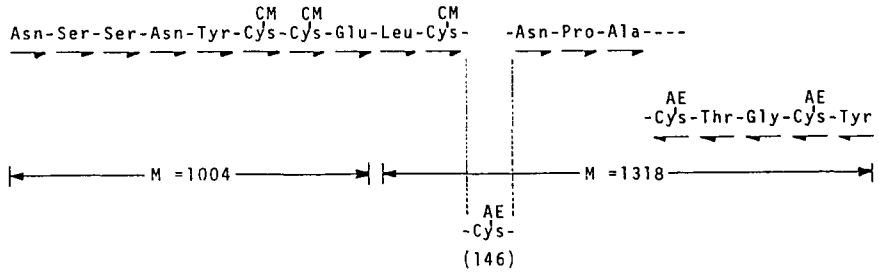

Fig. V-6. Summary of sequencing of STh. ⟶ indicates residues determined by manual Edman degradation and ⟵ indicates residues determined by a combination of carboxypeptidase B digestion and FAB mass spectrometry.

side-chain of S-aminoethyl-cysteine residues with very weak intensities in FAB mass spectra using a xenon atom beam; therefore, prominent mass shifts were due to amino acid residues cleaved by digestion. The results indicated that the C-terminal sequence of the peptide with the mass value 1319 was -Cys(Ae)(146)-Thr(101)-Gly(57)-Cys(Ae)(146)-Tyr(163) (Ae means an aminoethyl group and numbers in parentheses denote residue weights of amino acids) (Fig. V-5B). Mass shifts of the peptide with the mass value 1005 could not be identified. From the results of Edman degradation and a combination of FAB mass spectrometry and carboxypeptidase digestion, the sequence of the peptide with the mass value 1319 was determined to be as shown in Fig.V-6. The missing residue must be a Cys, because subtraction of the residue weights of the determined amino acid residues from the molecular weight (1318) resulted in the residue weight of Cys. Sequencing of STh is summarized in Fig.V-6. In addition, it was demonstrated by measurement of the exact molecular weight (2040) of STh that six half-cystine residues in the ST molecule are joined by disulfide

linkages, although the positions of disulfide linkages are still unknown.

V-3-2 Heat-Stable Enterotoxin Produced by Enterotoxigenic E. coli Strain 18D

(Purification) Crude toxin (STp) extracted from the culture supernatant of E. coli 18D was purified by the same procedure as that used for STh. The crude toxin was first chromatographed on SP-Sephadex C-50 (Fig.V-7). The peak fractions eluted with 0.1 M ammonium acetate was further purified by ion-exchange chromatography on DEAE-Sephadex A-25. The toxic activity emerged in the area shown by a horizontal bar in Fig.V-8. Final

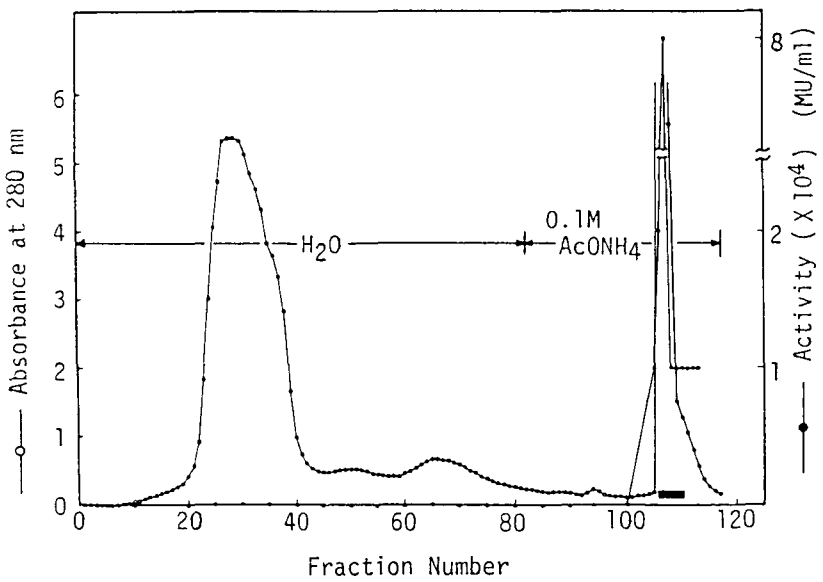

Fig. V-7. Chromatography on SP-Sephadex C-50 of crude toxin. The column was developed at a flow rate of 150 ml/h and 13.5 ml-fractions were collected. The fractions containing toxin, shown by a horizontal bar, were pooled.

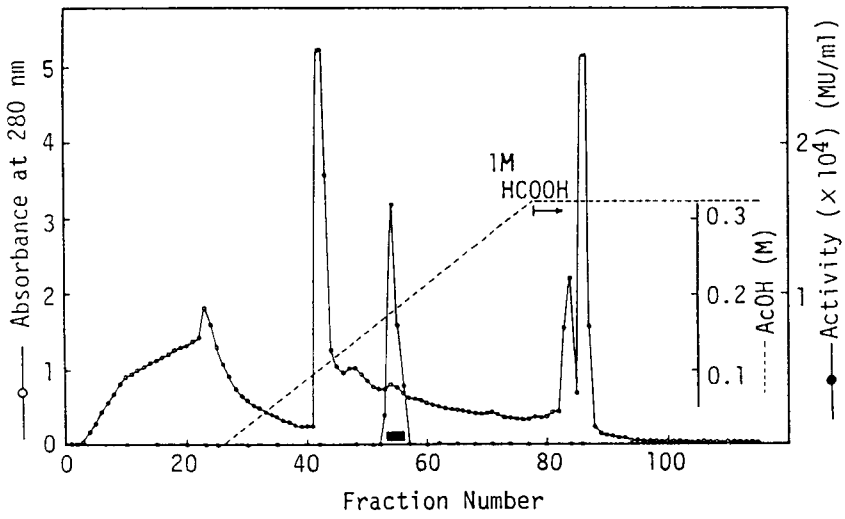

Fig. V-8. Chromatography on DEAE-Sephadex A-25 of the fraction shown by a horizontal bar in Fig. V-7. The column was developed at a flow rate of 75 ml/h and 13.5 ml-fractions were collected. The fractions containing toxin, shown by a horizontal bar, were pooled.

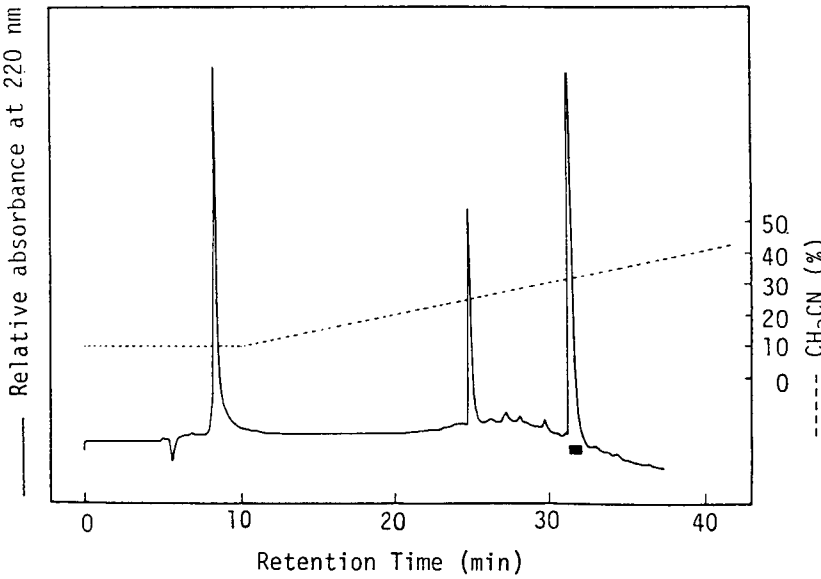

Fig. V-9. Reversed-phase HPLC of the toxin fraction obtained in Fig. V-8. The toxic activity was found in the peak fraction marked by a horizontal bar. Chromatographic conditions were the same as those for STh.

Table V-3. Summary of purification of STp.

	Total unit	Specific activity[a)]	Recovery (%)
Culture supernatant (20L)	3.2×10^6	17,800 ng/mu	100
Crude toxin	2.48×10^6	224	77.5
SP-Sephadex C-50	1.89×10^6	172	59.1
DEAE-Sephadex A-25	4.05×10^5		12.7
HPLC		2.5	(1.49 mg)

a) One mouse unit (mu) is defined at the minimum amount that gives an FA ratio of 0.09 or more.

Table V-4. Amino acid composition of purified STp.

	Native ST_p	Nearest integer
Asp	2.01	2
Thr	0.99	1
Glu	1.16	1
Gly	1.16	1
Ala	2.00	2
1/2Cys	4.92	6
Leu	1.13	1
Tyr	1.92	2
Phe	0.99	1
Pro	1.27	1
Total		18

Values were calculated as mol/mol Ala

purification was performed by HPLC on a reversed-phase column (Fig.V-9). The main toxic fraction shown by a horizontal bar was confirmed to be pure by repeated HPLC (not shown). The purification procedure and recovery of STp are summarized in Table V-3. The yield of STp was less than that of STh. The amino acid composition of purified native STp is summarized in Table V-4.

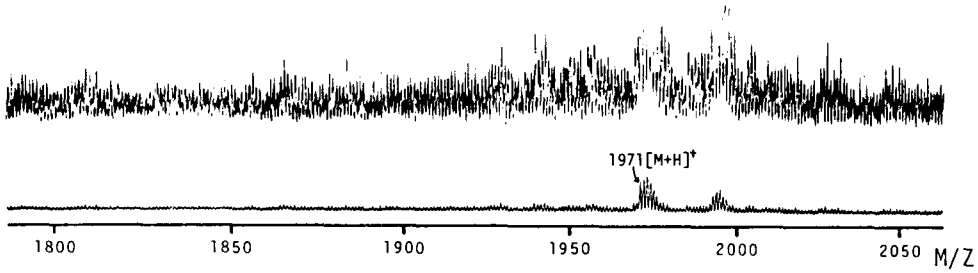

Fig. V-10. FAB mass spectrum of purified STp.

From the result, the total number of amino acid residues was deduced to be 18. The minimum effective dose of purified STp in suckling mice assay was 2.5 ng, which is almost the same value as that found for STh.

(Molecular weight determination) Fig.V-10 shows the FAB mass spectrum of purified STp, giving an intense signal at m/z=1971. The result indicates that the molecular weight of STp is 1970. This value was consistent with the theoretical value calculated from the amino acid composition.

(Amino acid sequence) Fig.V-11A shows the FAB mass spectrum of reduced and aminoethylated STp digested with protease V8, giving intense signals at m/z=965 and 1289. To determine the C-terminal amino acid sequences of the resulting two peptide fragments, the protease V8 digest of aminoethylayed STp was directly treated with a commercial preparation of carboxypeptidase B and subjected to mass measurement. Fig.V-11B shows the FAB mass spectrum of a mixture of aliquots withdrawn from the digest after 30 sec, 1 and 5 min. From the newly observed signals at m/z=980

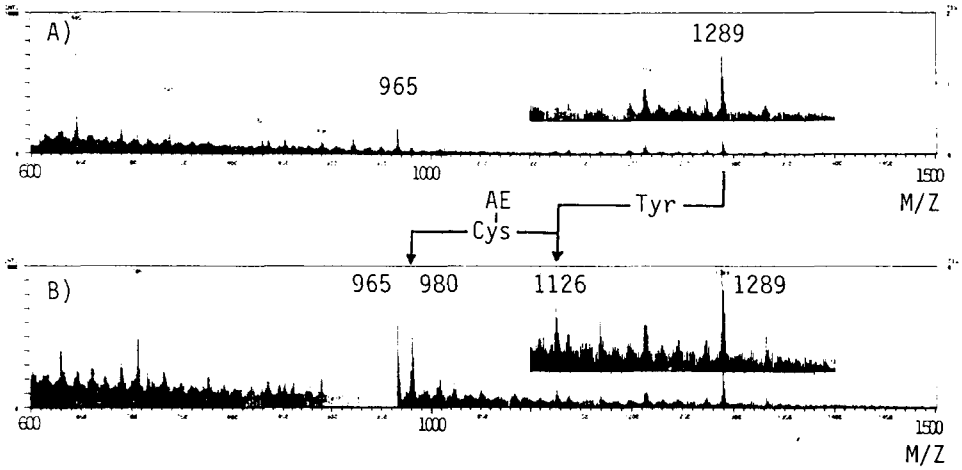

Fig. V-11. FAB mass spectra of: (A) aminoethylated STp digested with protease V8; (B) a mixture of (A) treated with carboxypeptidase B for 30 s, 1, and 5 min.

and 1126, the C-terminal sequence of the peptide with the mass value 1289 was -Cys(AE)(146)-Tyr(163). The FD mass spectrum of the carboxypeptidase B digest gave intense signals at m/z=165 and 182 (not shown), indicating that S-aminoethyl-cysteine (165) and tyrosine (182) were released from the peptide with the mass value 1289.

Reduced and carboxymethylated STp was degraded manually by the Edman method. The results revealed the sequence of 16 amino acid residues from the N-terminus, although the C-terminal two residues were uncertain. The sequence data were compatible with the amino acid composition and the data obtained from the mass spectra shown in Figs.V-10, 11A and 11B. Sequencing of STp is summarized in Fig.V-12. In addition, it was indicated by measurement or the exact molecular weight (1970) of STp that six

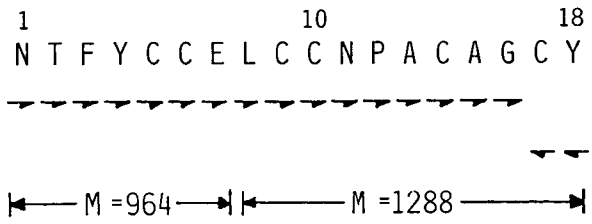

Fig. V-12. Summary of sequencing of STp. → indicates residues determined by manual Edman method and ← indicates residues determined by a combination of carboxypeptidase B digestion and FAB mass spectrometry.

half-cystine residues in the ST molecule are linked by disulfide bonds, although the positions of disulfide bonds are still unknown.

V-3-3 Heat-Stable Enterotoxin Produced by Y. enterocolitica

(Purification) Previously, Yersinia-ST was purified by ion-exchange chromatography and gel-filtration [17]. Since this purification procedure is time-consuming and the yield is low, in this study, the author attempted to purify Yersinia-ST by reversed-phase HPLC from the culture supernatant of this bacteria treated with protamine sulfate or with protamine sulfate and DEAE-Sephacel.

First, the culture supernatant of cells treated with protamine sulfate was purified by reversed-phase HPLC. As shown in Fig.V-13, the enterotoxigenic activity was eluted in several fractions with about 28% acetonitrile. Since these fractions were not separated completely from each other, peak fraction (a) and

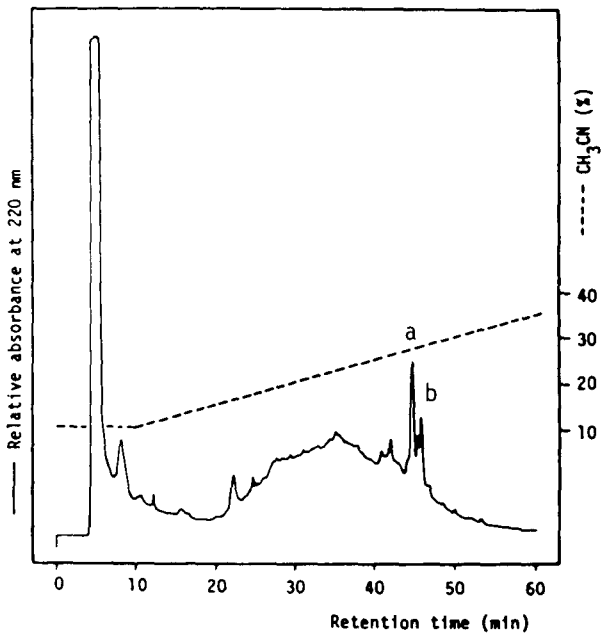

Fig. V-13. Reversed-phase HPLC of the culture supernatant of Y. enterocolitica treated with protamine sulfate. Chromatographic conditions were as described in Materials and Methods (V-2-4).

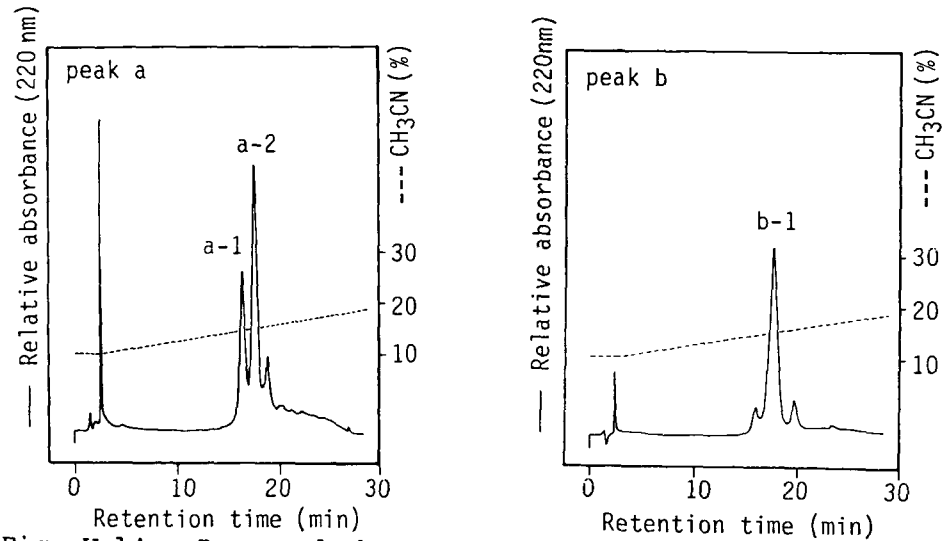

Fig. V-14. Reversed-phase HPLC of peak fractions (a) and (b) isolated in Fig. V-13. Chromatographic conditions were as described in Materials and Methods (V-2-4).

Table V-5. Amino acid compositions of fractions a-1, a-2, and b-1 in Fig. V-14.

	a-1	a-2	b-1	Carboxymethylated a-2	Nearest integer
CMC-Cys	-	-	-	6.95	
Asp	5.37	5.25	5.04	5.37	5
Thr	0.12	0.08	0.04	0.06	
Ser	2.97	2.95	2.92	3.16	3
Glu	3.10	2.27	2.11	2.07	2
Gly	1.88	1.06	1.00	1.09	1
Ala	4	4	4	4	4
$\frac{1}{2}$Cys	5.94	4.91	5.11	-	7
Val	1.89	1.88	1.82	2.27	2
Met	-	0.01	-	-	
X	-	1.04	0.65	-	(1)
Ile	0.14	0.10	0.06	0.05	
Leu	0.18	0.13	0.08	0.08	
Tyr	0.07	0.04	0.03	0.05	
Phe	0.10	0.05	0.03	0.06	
Lys	0.18	0.16	0.06	0.07	
His	0.04	0.02	-	0.07	
Trp	0.88	0.91	0.89	0.94	1
Arg	-	-	-	-	
Pro	5.19	5.09	5.11	5.18	5
Total					30

Values were calculated as mol/mol of Ala. The values for X (unknown component) were calculated assuming that its color value was the same as that of Ala.

(b) were pooled and purified further by reversed-phase HPLC under different conditions (Fig.V-14). Fractions a-1, a-2 and b-1 shown in Fig.V-14 were collected. The amino acid compositions of the purified fractions were determined to be as shown in Table V-5. Peptides a-2 and b-1 gave an unknown peak at an elution position just after that of methionine on amino acid analysis, besides usual amino acids, but peptide a-1 did not. Peptide a-2 was reduced and carboxymethylated, and the amino acid composition was determined to be as shown in Table V-5. The carboxymethylated ST

did not contain the unknown component (X) found on amino acid analysis of the intact ST. The nearest integer values of the amino acid residues in peptide a-2 were calculated from the observed amino acid compositions of intact and carboxymethylated peptides. Peptide a-2 was deduced to consist of 30 amino acid residues besides the unknown component (X).

Then, the culture supernatant of Y. enterocolitica treated with protamine sulfate and DEAE-Sephacel was examined by reversed-phase HPLC (Fig.V-15) under similar conditions to those described in Fig.V-13. Enterotoxigenic activity emerged in five peak fractions , which were isolated and purified to homogeneity

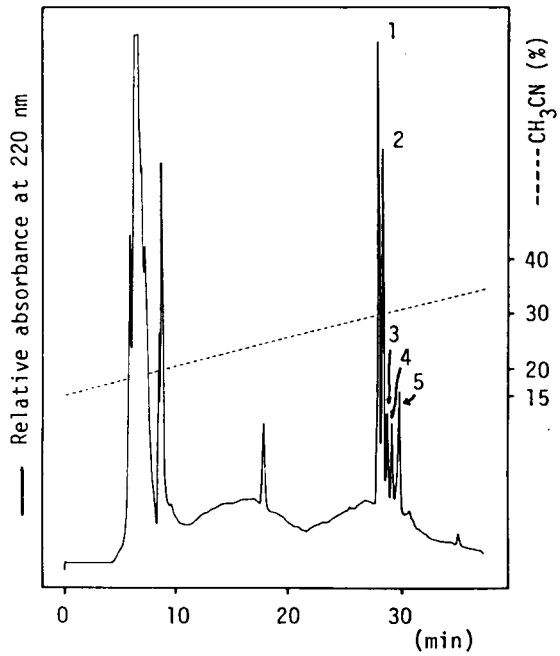

Fig. V-15. Reversed-phase HPLC of the culture supernatant of Y. enterocolitica treated with protamine sulfate and DEAE-Sephacel. Chromatographic conditions were similar to those described in Fig. V-13.

by repeated HPLC (not shown). The amino acid compositions of the separated fractions (1, 2, 3, 4 and 5) were determined to be as shown in Table V-6. These peptides did not contain the unknown component (X) found in the peptides isolated from the culture supernatant treated with protamine sulfate. Furthermore, these peptides contained fewer amino acid residues than those obtained in Fig.V-13.

(Molecular weight determination) The molecular weights of purified toxins obtained in Figs.V-14 and 15 were examined by FAB

Table V-6. Amino acid compositions of fractions 1-5 in Fig. V-15.

	1	2	3	4	5
Asp	4.05 (4)	3.98 (4)	3.99 (4)	3.83 (4)	4.23 (4)
Thr	0.02	0.02	0.04	0.08	0.07
Ser	1.92 (2)	1.09 (1)	1.84 (2)	1.76 (2)	0.23
Glu	0.10	0.10	0.16	1.12 (1)	0.30
Gly	0.98 (1)	1.02 (1)	1.04 (1)	1.02 (1)	1.06 (1)
Ala	2	2	2	2	2
$\frac{1}{2}$Cys	4.75 (6)	4.83 (6)	4.83 (6)	4.36 (6)	4.76 (6)
Val	0.74 (1)	0.76 (1)	1.67 (2)	1.58 (2)	0.83 (1)
Met	0.03	0.02	0.03	0.04	0.05
Ile	0.03	0.03	0.04	0.09	0.09
Leu	0.02	0.03	0.04	0.10	0.09
Tyr	0.02	0.01	0.02	0.04	0.03
Phe	0.02	-	0.03	0.06	0.07
Lys	0.03	0.06	0.04	0.09	0.07
His	-	0.01	-	0.03	0.02
Trp	0.95 (1)	0.90 (1)	0.90 (1)	0.81 (1)	0.96 (1)
Arg	0.02	0.03	0.02	0.05	0.04
Pro	1.05 (1)	1.04 (1)	1.04 (1)	1.12 (1)	1.21 (1)
Total	18	17	19	20	16
$[M+H]^+$	1845.4	1758.6	1944.7	2073.6	1671.4

Values were calculated as mol/mol of Ala; numbers in parentheses indicate nearest integer values.

mass spectrometry. Fig.V-16A shows the FAB mass spectrum of peptide a-2, giving signals at m/z=3037.7 and 3170.8. The mass value of the ST excluding the unknown component (X) was calculated to be from 3032.1 to 3039.0 from its amino acid composition (On amino acid analysis of the acid hydrolysate the ratios of Glu to Gln and of Asp to Asn cannot be determined). The mass value observed at m/z=3037.7 is within the range of theoretical mass

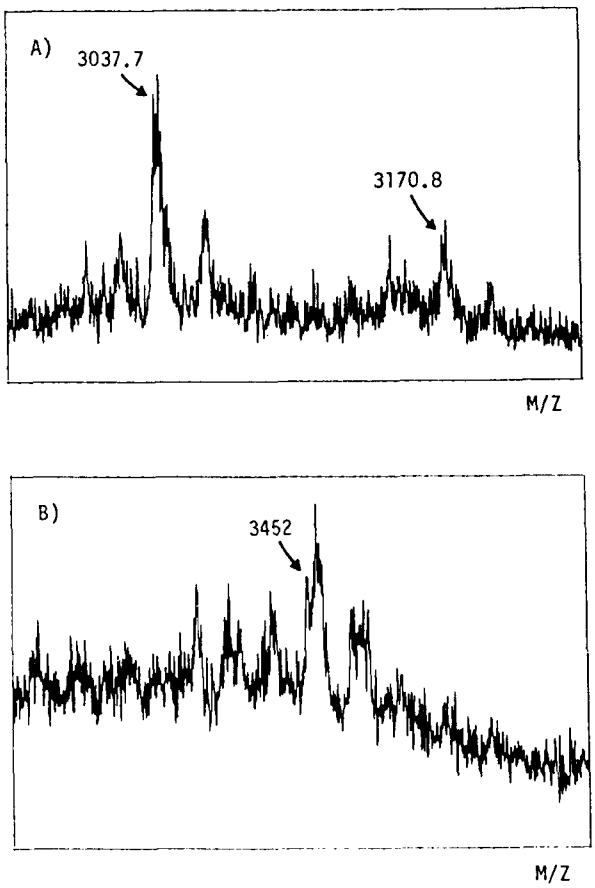

Fig. V-16. FAB mass spectra of A) purified Yersinia-ST (peptide a-2 in Fig. V-14) and B) its carboxymethylated derivative.

values. As described in chapter I, when a cystine-containing peptide is examined by FAB mass spectrometry, it gives the signal of the mother peptide together with that of the fragments formed by reductive cleavage of the disulfide bond. This fact suggests that the mass values observed at m/z=3170.8 and 3037.7 in Fig.V-16A corresponded to the signals of the mother peptide and the fragment, respectively. Namely, the molecular weights of peptide a-2 with and without the X component were 3169.8 and 3036.7, respectively. In addition, the residue weight of the X component was 134. Fig.V-16B shows the mass spectrum of carboxymethylated ST, which gave a signal at m/z=3452, indicating that seven carboxymethyl groups were introduced into the ST molecule; that is, seven half-cystine residues are present in the ST molecule.

The mass spectra of peptides 1-5 gave signals at m/z=1845.4, 1758.6, 1944.7, 2073.6, and 1671.4, respectively, which coincided well with the theoretical mass values calculated from the amino acid compositions described in Table V-6. The result indicates that the molecular weights of peptides 1-5 are 1844.4, 1757.6, 1943.7, 2072.6, and 1670.4, respectively.

(Amino acid sequence) The amino acid sequence of peptide a-2 was first analyzed with a gas-phase sequenator. Seventeen amino acid residues could be determined from the N-terminus of the intact ST except for the residue at position 3 (Fig.V-20). Fig.V-17 shows the recoveries of PTH-amino acids obtained by

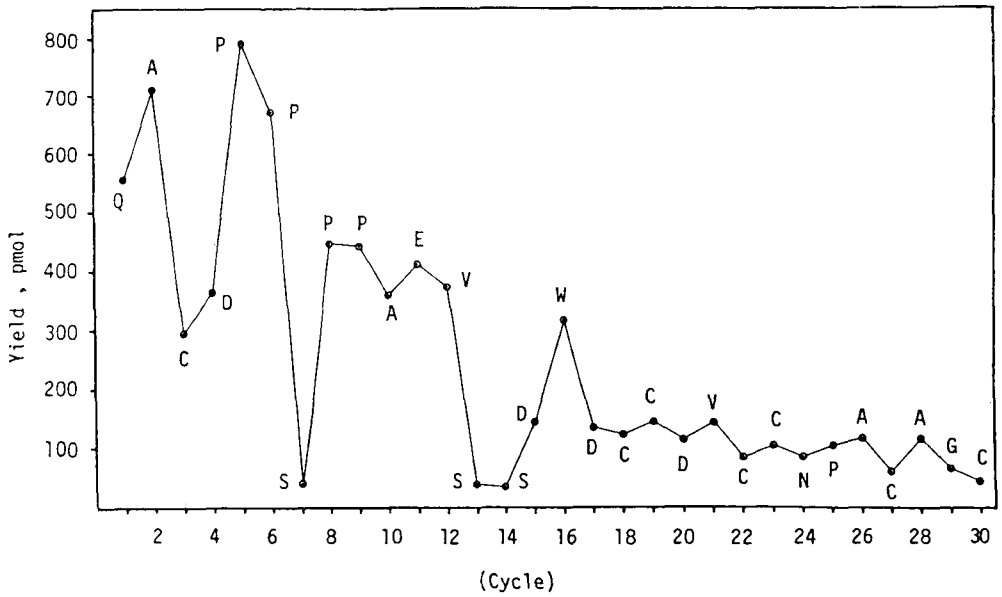

Fig. V-17. Recoveries of PTH-amino acids released from carboxymethylated peptide a-2 (2.4 nmol) in Fig. V-14 with a gas-phase sequenator.

sequential Edman degradation of carboxymethylated ST. Thirty amino acid residues could be assigned to the sequence of carboxymethylated ST from its N-terminus (Fig.V-20). Except at position 3, the sequence of 17 N-terminal amino acid residues of the intact ST was completely identical with that of carboxymethylated ST. The sequence data were compatible with the amino acid composition and the mass data.

For determination of the positions of the half-cystine residues bound to the unknown component (X) and linked intramolecularly by disulfide bonds in peptide a-2, it was treated with proline-specific endopeptidase [18], which specifically cleaves the carboxyl-sites of proline residues in a peptide. Fig.

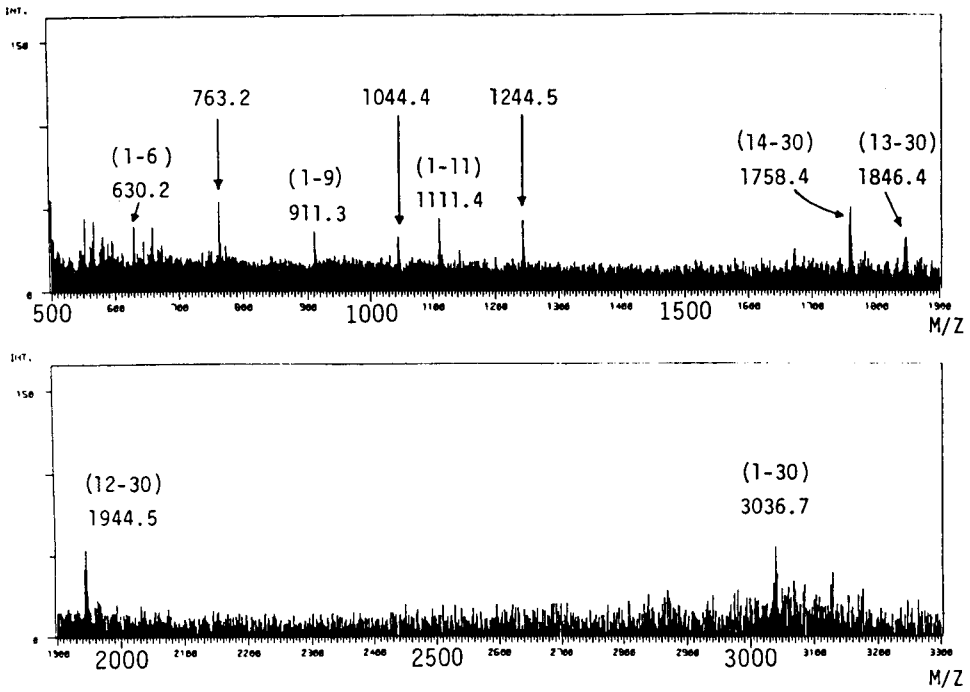

Fig. V-18. FAB mass spectrum of peptide a-2 in Fig. V-14 treated with proline-specific endopeptidase for 1.5 h. Numbers with parentheses indicate the position of the sequence in Fig. V-20.

V-18 shows the FAB mass spectrum of the hydrolysate, giving signals at m/z=763.2, 1044.4, 1244.5, and 1944.5 with the signal at m/z=3036.7 (theoretical value is 3037.0) of the undigested peptide. These mass values corresponded to the sequences from 1 to 6, 1 to 9, 1 to 11, and 12 to 30, respectively, indicating that the X component binds to the half-cystine residue at position 3 and that the six half-cystine residues in the C-terminal portion form three disulfide linkages.

Then, the amino acid sequence of the ST, purified from the culture supernatant treated with protamine sulfate and

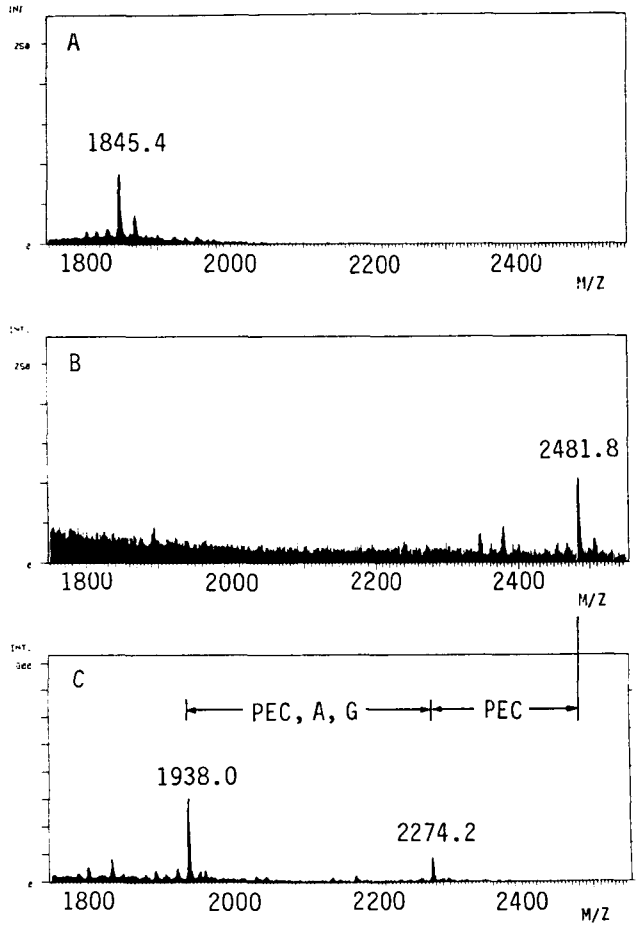

Fig. V-19. FAB mass spectra of A) peptide 1 in Fig. V-15, B) its pyridylethylated derivative, and C) its digest with carboxypeptidase B at 37°C for 100 min.

DEAE-Sephacel, was examined. Since peptide 1 was obtained in the largest amount in the toxic fractions shown in Fig.V-15, its amino acid sequence was analyzed. Figs.V-19A and 19B show the mass spectra of the intact and pyridylethylated STs, respectively, indicating that six half-cystine residues were pyridylethylated. Fig.V-19C shows the mass spectrum of a carboxypeptidase B digest of the pyridylethylated ST. The original signal at m/z=2481.8 disappeared and new signals were observed at m/z=2274.2 and

1938.0. The result indicates that pyridylethyl-cysteine was first released from the C-terminus of the ST and that glycine, alanine, and pyridylethyl-cysteine were released, although their order could not be determined. Then, the pyridylethylated ST was degraded with a liquid-phase sequenator. Fourteen amino acid residues from the N-terminus could be determined. The amino acid sequence thus determined is summarized in Fig.V-20. The sequence from residues 2 to 4 of the C-terminus could not be determined, but was deduced by comparison with that of peptide a-2. The sequence is identical with that from position 13 to position 26 and C-terminus of peptide a-2 (Fig.V-20).

The N-terminal amino acid sequences of peptides 2-5 were also analyzed by the Edman method, as shown in Fig.V-20. Furthermore, peptides 1-4 were digested with aminopeptidase M and the resulting digests were compared by reversed-phase HPLC. The degradation products were all found to be the same as peptide 5 by chemical and physical criteria such as HPLC, amino acid analyses and FAB mass spectrometry (data not shown). These results indicate that peptide 1-5 have the same amino acid sequence except in their N-terminal portions (Fig.V-20), and that they might be derived from peptide a-2 enzymatically either in the cytoplasma or after their secretion from the cells.

The results of sequence analysis of Yersinia-ST are summarized in Fig.V-20. In addition, by comparison of the theoretical mass values calculated from the amino acid sequences with the observed mass values of these peptides, the six

Y. enterocolitica ST

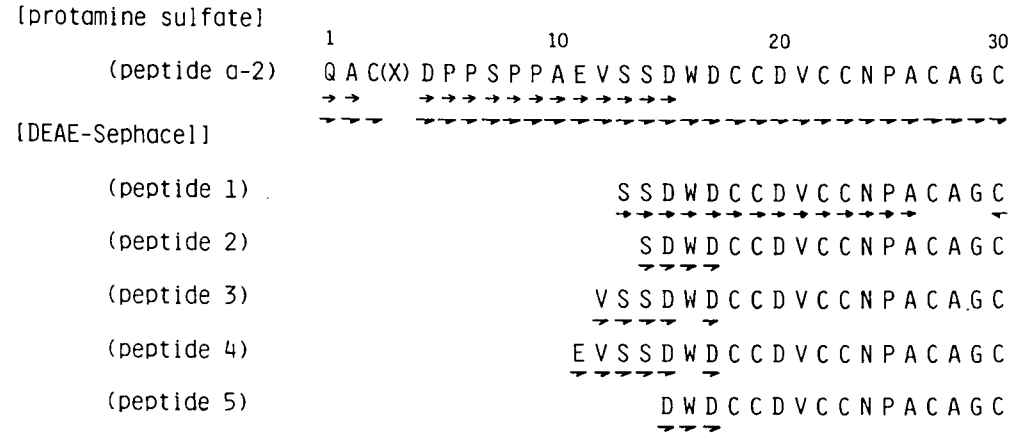

Fig. V-20. Summary of sequencing of Yersinia-ST. ⟶ and ⟶: residues determined by Edman degradation of native and carboxymethylated ST (peptide a-2 in Fig. V-14) with a gas-phase sequenator, respectively; ⟶: residues determined by Edman degradation of pyridylethylated ST (peptide 1 in Fig. V-15) with a liquid-phase sequenator; ⟵: residues determined by a combination of carboxypeptidase B digestion and FAB mass spectrometry; ⟶: residues determined by manual Edman degradation of native ST (peptide 2-5 in Fig. V-15).

half-cystine residues in the C-terminal portion were concluded to be intramolecularly linked by three disulfide bonds, although the positions of disulfide bonds are still unknown.

V-3-4 Heat-Stable Enterotoxin Produced by V. cholerae non-01

NAG-ST purified by Honda and his co-workers [12] was used for sequence determination. A part of purified toxin was reduced and carboxymethylated. Carboxymethylated ST was further purified by reversed-phase HPLC (Fig.V-21). The fraction shown by a

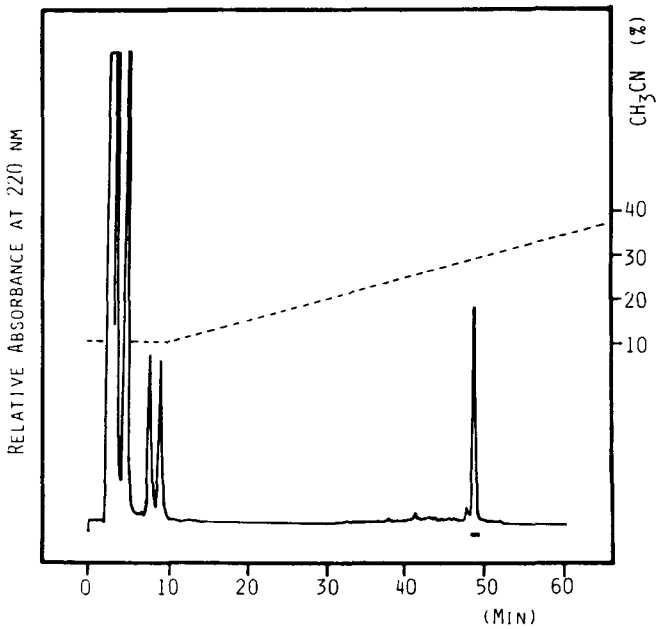

Fig. V-21. Reversed-phase HPLC of reductively carboxymethylated NAG-ST. Chromatographic conditions were as described in Materials and Methods (V-2-4).

Table V-7. Amino acid composition of purified NAG-ST.

	Native	Carboxy-methylated	Nearest integer
CM-Cys	–	6.28	
Asp	3.20	3.10	3
Ser	0.12	0.05	
Glu	1.20	1.08	1
Gly	1.11	1.14	1
Ala	1.00	1.00	1
½Cys	5.10	0.07	6
Val	0.04	–	
Ile	1.82	1.84	2
Leu	1.11	1.02	1
Tyr	0.05	–	
Phe	1.17	1.03	1
Arg	–	0.10	
Pro	0.83	1.05	1
Total			17

Values were calculated as mol/mol Ala

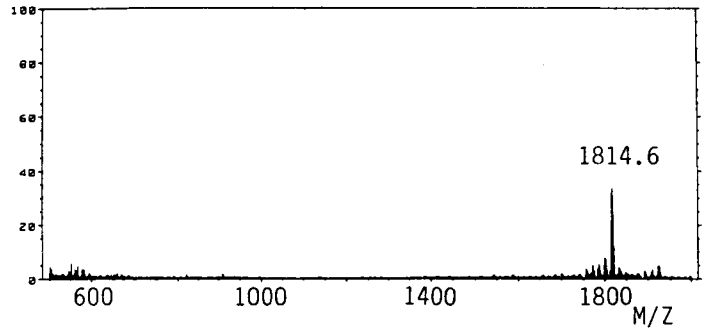

Fig. V-22. FAB mass spectrum of native NAG-ST.

horizontal bar was collected. The amino acid compositions of native and carboxymethylated STs are shown in Table V-7.

(Molecular weight determination) The molecular weight of NAG-ST was determined to be 1813.6 by measurements of native and carboxymethylated STs by FAB mass spectrometry, which gave intense signals at m/z=1814.6 (Fig.V-22) and 2168.6 (not shown), respectively. The results clearly indicated that NAG-ST contains six half-cystine residues which are intramolecularly linked by three disulfide bonds and that two of five carboxyl groups of three Asp, one Glu and the C-terminus are present as amides.

(Amino acid sequence) The native ST was degraded manually by the Edman method, and Ile and Asp were found in positions 1 and 2, respectively, from the N-terminus (Fig.V-24). The C-terminal amino acid sequence of NAG-ST was examined by measurement of the FAB mass spectra of the carboxypeptidase Y and B digests of the native ST (Fig.V-23). The differences between the mass values of native ST and the digests suggested that the C-terminal sequence

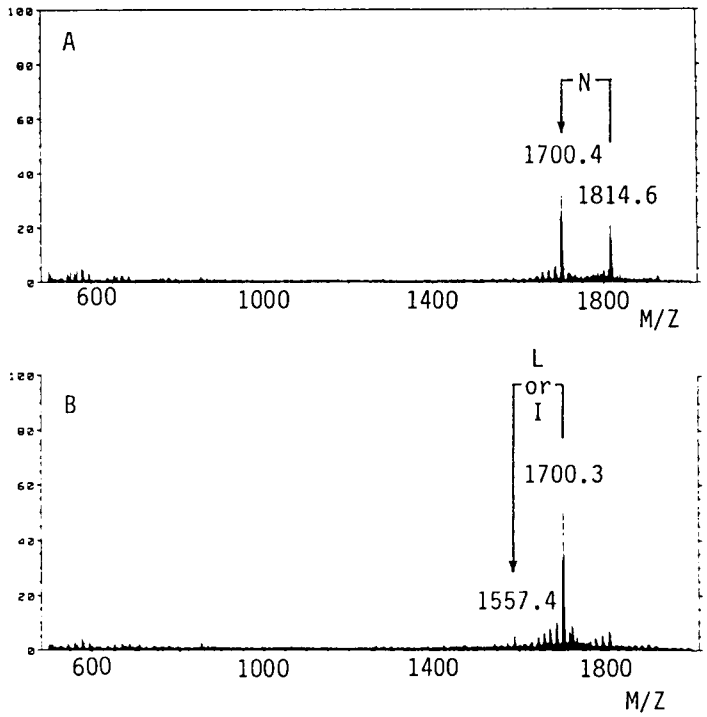

Fig. V-23. FAB mass spectra of (A) a digest of native NAG-ST with carboxypeptidase Y at 37°C for 90 min and (B) a digest of (A) with carboxypeptidase B at 37°C for 20 h.

is -Leu(or Ile)-Asn (Leu and Ile cannot be distinguished, because they have the same residue weights). Moreover, native ST (2.24 nmol) released Asn (2.14 nmol) on amino acid analysis after digestion with carboxypeptidase Y. Thus, the C-terminal amino acid residue was concluded to be Asn.

Thereafter, carboxymethylated ST was analyzed in a gas-phase sequenator. As shown in Fig.V-24, the carboxymethylated ST had a sequence of 17 amino acid residues from the N-terminus. The sequence data were compatible with the amino acid composition and the mass data. The results of sequence analysis of NAG-ST are

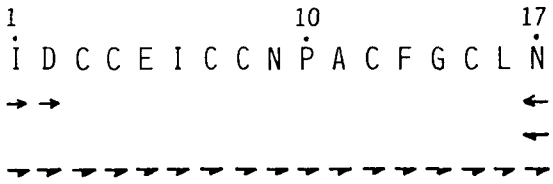

Fig. V-24. Summary of sequencing of NAG-ST. ⟶: residues determined by manual Edman degradation of native NAG-ST; ⟵ and ⟵: residues determined by FAB mass measurement and amino acid analysis of a carboxypeptidase digest of native NAG-ST, respectively; ⟶: residues determined by Edman degradation of carboxymethylated NAG-ST with a gas-phase sequenator.

summarized in Fig.V-24.

V-4 Discussion

The sequences of the heat-stable enterotoxins determined above were summarized and compared with one another, as shown in Fig.V-25. The sequences (boxed region) are identical in these heat-stable enterotoxins and six half-cystine residues are located at the same relative positions. The six half-cystine residues are intramolecularly joined by disulfide bonds, as demonstrated by measurement of the exact molecular weight of each enterotoxin, although the positions of disulfide bonds are still unknown. These findings strongly indicate that these enterotoxins have the common secondary structure and that the tertiary structure formed by the three intramolecular disulfide bonds may be responsible for heat stability and the common biological and immunological properties

of these enterotoxins. Furthermore, the present results suggest that the sequence (boxed region) is general and conserved in the heat-stable enterotoxins produced by enteric bacteria.

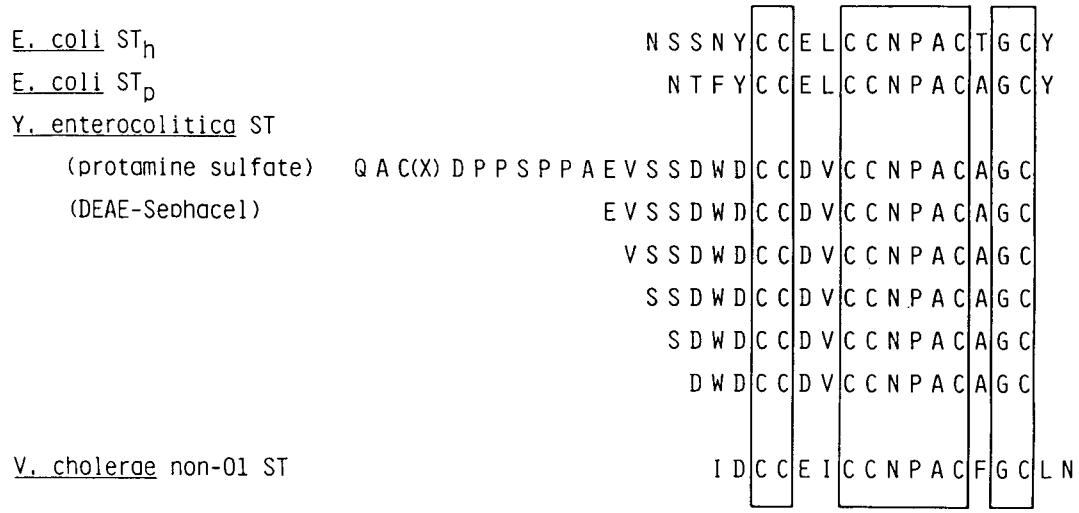

Fig. V-25. Amino acid sequences of four kinds of heat-stable enterotoxins produced by E. coli, Y. enterocolitica, and V. cholerae non-01. Boxed region indicates an identical sequence in these heat-stable enterotoxins.

References

1) H. W. Smith and C. L. Gyles, J. Med. Microbiol., 3, 387 (1970).

2) M. N. Burgess, R. J. Bywater, C. M. Cowley, N. A. Mullan, and P. M. Newsome, Infect. Immun., 21, 526 (1978).

3) M. Field, L. H. Graf, W. J. Laird, and P. L. Smith, Proceedings of the National Academy of Sciences of the United States of America, 75, 2800 (1978).

4) J. M. Hugnes, F. Murad, B. Chang, and R. L. Guerrant, Nature, 271, 755 (1978).

5) P. M. Newsome, M. N. Burgess, and N. A. Mullan, Infection and Immunity, 22, 290 (1978).

6) R. A. Giannella and K. W. Drake, Infection and Immunity, 24, 19 (1979).

7) M. So and B. J. McCarthy, Proc. Natl. Acad. Sci., 77, 4011 (1980).

8) C. H. Pai and V. Mors, Infect. Immun., 19, 908 (1978).

9) F. A. Klipstein and R. F. Engert, Infect. Immun., 13, 373 (1976).

10) F. A. Klipstein and R. F. Engert, Infect. Immun., 13, 1307 (1976).

11) K. Okamoto, A. Miyama, T. Takeda, Y. Takeda, and T. Miwatani, FEMS Microbiol. Lett., 16, 85 (1983).

12) T. Honda, M. Arita, T. Takeda, M. Yoh, and T. Miwatani, Lancet, 2, 163 (1985).

13) D. H. Mundell, C. R. Anselmo, and R. M. Wishnow, Infect. Immun., 14, 383 (1976).

14) K. Okamoto, T. Inoue, H. Ichihara, Y. Kawamoto, and A. Miyama, Infect. Immun., 31, 554 (1981).

15) Y. Takeda, T. Takeda, T. Yano, K. Yamamoto, and T. Miwatani, Infect. Immun., 25, 978 (1979).

16) M. W. Hunkapiller, R. Hewick, W. J. Dreyer, and L. E. Hood, Methods Enzymol., 91, 399 (1983).

17) K. Okamoto, T. Inoue, K. Shimizu, S. Hara, and A. Miyama, Infect. Immun., 35, 958 (1982).

18) T. Yoshimoto, R. Walter, and D. Tsuru, J. Biol. Chem., 255, 4786 (1980).

SUMMARY

In this thesis (chapter I - IV), the strategy for protein sequence analysis by FAB mass spectrometry is described. The strategy, which consists of direct measurement of protein digests by FAB mass spectrometry, is concluded to be very useful for 1) peptide mapping of proteins, 2) confirmation of the primary structure of a protein that has already been determined by a conventional method, 3) verification of a protein sequence deduced from the nucleotide sequence, 4) obtaining information on post-translational modifications of a protein, 5) determination of a primary structure that is difficult to determine by a conventional method.

In addition, a combination of FAB mass spectrometry and carboxypeptidase digestion or Edman degradation provides an easy and relible method for determination of a C- or N-terminal amino acid sequences of peptides or peptide mixtures, similarly to the method by FD mass spectrometry.

However, it should be noted that not all peptides in a protein digest can be identified under the conditions used in the present study. There are three reasons for this: 1) With the present apparatus, the quasi-molecular ions of peptides can be detected with accuracy and reliance in a mass range of only up to about 4000 amu. 2) The peptides in a protein digest do not ionize equally. This seems to be due to differences in the properties of the peptides, as discussed in chapter I. 3) The ions of

low-molecular-weight peptides of less than about 300 amu are difficult to detect under the present experimental conditions, because they are buried under intense noise signals derived from the matrix (glycerol, α-thioglycerol, etc.) in the low mass region. These problems may be overcome by measuring two or more kinds of protein digests obtained by cleavage with different enzymes, or by use of a larger amount of sample, or a different technique for mass spectrometry.

Chapter V describes the determinations of the amino acid sequences of four kinds of heat-stable enterotoxins produced by enteric bacteria. The results indicate that these enterotoxins are highly homologous, especially in their 13 C-terminal amino acid residues.

LIST OF PUBLICATIONS

1) A New Method for Caboxyl-Terminal Sequence Analysis of a Peptide Using Carboxypeptidases and Field-desorption Mass Spectrometry.
Yasutsugu SHIMONISHI, Yeong-Man HONG, Toshifumi TAKAO, Hisashi MATSUDA, and Yoshiharu IZUMI
Proc. Japan Acad., 57B, 304-308 (1981).

2) Amino-Acid Sequence of a Heat-Stable Enterotoxin Produced by Human Enterotoxigenic Escherichia coli
Saburo AIMOTO, Toshifumi TAKAO, Yasutsugu SHIMONISHI, Saburo HARA, Tae TAKEDA, Yoshifumi TAKEDA, and Toshio MIWATANI
Eur. J. Biochem., 129, 257-263 (1982).

3) Application of FD/FAB Mass Spectrometry for Determination of Amino Acid Sequences of Heat-Stable Enterotoxins Produced by Human and Porcine Strains of Enterotoxigenic Escherichia coli
Saburo AIMOTO, Toshifumi TAKAO, Motomu HANE, Haruo IKEMURA, Yasutsugu SHIMONISHI, Tae TAKEDA, Yoshifumi TAKEDA, and Toshio MIWATANI
Peptide Chemistry 1982, 219-224 (1983).

4) Simultaneous Analysis of C-Terminal Amino Acid Residues of Peptide Mixtures by Combination of Field Desorption Mass Spectrometry and Carboxypeptidase Digestion
Yeong-Man HONG, Toshifumi TAKAO, Saburo AIMOTO, and

Yasutsugu SHIMONISHI
Biomed. Mass Spectrom., 10, 450-457 (1983).

5) Amino Acid Sequence of a Heat-Stable Enterotoxin Isolated from Enterotoxigenic Escherichia coli Strain 18D
Toshifumi TAKAO, Tetsuo HITOUJI, Saburo AIMOTO, Yasutsugu SHIMONISHI, Saburo HARA, Tae TAKEDA, Yoshifumi TAKEDA, and Toshio MIWATANI
FEBS Lett. 152, 1-5 (1983).

6) Peptide Mapping of a Protein by Soft-Ionization Mass Spectrometry
Toshifumi TAKAO, Tetsuo HITOUJI, Masakuni YOSHIDA, Saburo AIMOTO, and Yasutsugu SHIMONISHI
Peptide Chemistry, 1983, 139-144 (1984).

7) Fast Atom Bombardment (FAB) Mass Spectra of Protein Digests: Hen and Duck Egg-White Lysozymes
Toshifumi TAKAO, Masakuni YOSHIDA, Yeong-Man HONG, Saburo AIMOTO, and Yasutsugu SHIMONISHI
Biomed. Mass Spectrom., 11, 549-556 (1984).

8) Chemical Synthesis of Fully Active and Heat-Stable Fragments of Heat-Stable Enterotoxin of Enterotoxigenic Escherichia coli Strain 18D
Shoko YOSHIMURA, Toshifumi TAKAO, Haruo IKEMURA, Saburo AIMOTO, Yasutsugu SHIMONISHI, Saburo HARA, Tae TAKEDA, Yoshifumi TAKEDA, and Toshio MIWATANI
Bull. Chem. Soc. Jpn., 57, 1381-1387 (1984).

9) Verification of Protein Sequence by Fast Atom Bombardment Mass

Spectrometry (Amino Acid Sequence of Protein S, A Development-Specific Protein of *MYxococcus Xanthus*)
Toshifumi TAKAO, Tetsuo HITOUJI, Yasutsugu SHIMONISHI, Tadashi TANABE, Sumiko INOUYE, and Masayori INOUYE
J. Biol. Chem., 259, 6105-6109 (1984).

10) Primary Structure of Heat-Stable Enterotoxin Produced by *Yersinia Enterocolitica*
Toshifumi TAKAO, Noriko TOMINAGA, and Yasutsugu SHIMONISHI
Biochem. Biophys. Res. Commun., 125, 845-851 (1984).

11) Mass Measurement of High-Molecular-Weight Peptides by Fast Atom Bombardment Mass Spectrometry --- Various Insulins ---
Toshifumi TAKAO, Yasutsugu SHIMONISHI, and Ken INOUYE
Peptide Chemistry, 1984, 187-192 (1985).

12) Facile Identification of Protein Sequence by Mass Spectrometry --- B subunit of *Vibrio cholerae* Classical Biotype Inaba 569B Toxin ---
Toshifumi TAKAO, Hirofumi WATANABE, and Yasutsugu SHIMONISHI
Eur. J. Biochem., 146, 503-508 (1985).

13) Complete Amino Acid Sequence of the Ubiquinone Binding Protein (QP-C), a Protein Similar to the 14,000-Dalton Subunit of the Yeast Ubiquinol-Cytochrome c Reductase Complex
Sadao WAKABAYASHI, Toshifumi TAKAO, Yasutsugu SHIMONISHI, Seiki KURAMITSU, Hiroshi MATSUBARA, Tsing-ying WANG,

Zhen-ping ZHANG, and Tsoo E. KING
J. Biol. Chem., 260, 337-343 (1985).

14) Verification by Mass Spectrometry of the Primary Structure of Human Interleukin-2
Ken-ichi FUKUHARA, Takashi TSUJI, Koji TOI, Toshifumi TAKAO, and Yasutsugu SHIMONISHI
J. Biol. Chem., 260, 10487-10494 (1985).

15) Subtilosin A, a New Antibiotic Peptide Produced by Bacillus subtilis 168: Isolation, Structural Analysis, and Biogenesis
Katsuhiko BABASAKI, Toshifumi TAKAO, Yasutsugu SHIMONISHI, and Kiyoshi KURAHASHI
J. Biochem., 98, 585-603 (1985).

16) Isolation, Primary Structure and Synthesis of Heat-Stable Enterotoxin Produced by Yersinia Enterocolitica
Toshifumi TAKAO, Noriko TOMINAGA, Shoko YOSHIMURA, Yasutsugu SHIMONISHI, Saburo HARA, Takashi INOUYE, and Akio MIYAMA
Eur. J. Biochem., 152, 199-206 (1985).

17) Amino Acid Sequence of Heat-Stable Enterotoxin Produced by Vibrio cholerae non-01
Toshifumi TAKAO, Yasutsugu SHIMONISHI, Makoto KOBAYASHI, Osamu NISHIMURA, Michiko ARITA, Tae TAKEDA, Takashi HONDA, and Toshio MIWATANI
FEBS Lett., 193, 250-254 (1985).

18) Verification by Fast Atom Bombardment (FAB) Mass Spectrometry

of the Primary Structures of Proteins Produced by a Recombinant DNA Technique
Toshifumi TAKAO, Hironobu HOJO, and Yasutsugu SHIMONISHI
Peptide Chemistry, 1985, 409-412 (1986).

19) Application of S-1-Adamantylcysteine to the Synthesis of a 37-Residue Peptide Amide Corresponding to Human Calcitonin Gene-related Peptide (hCGRP)
Nobutaka FUJII, Haruaki YAJIMA, Akira OTAKA, Susumu FUNAKOSHI, Motoyoshi NOMIZU, Kenichi AKAJI, Itsuo YAMAMOTO, Kanji TORIZUKA, Kouki KITAGAWA, Tadashi AKITA, Kenshi ANDO, Tatsuhiko KAWAMOTO, Yasutsugu SHIMONISHI, and Toshifumi TAKAO
J. Chem. Soc., Chem. Commun., 602-603 (1985).

20) Amino Acid Sequences of Allophycocyanin α- and β-subunits Isolated from Anabaena Cylindrica
Y. MINAMI, F. YAMADA, T. HASE, H. MATSUBARA, A. MURAKAMI, Y. FUJITA, T. TAKAO, and Y. SHIMONISHI
FEBS Lett., 191, 216-220 (1985).

21) Studies on Peptides. CXXXVI. Solution-Phase Synthesis of a 37-Residue Peptide Amide Corresponding to the Entire Amino Acid Sequence of Human Calcitonin Gene-Related Peptide (hCGRP)
Nobutaka FUJII, Akira OTAKA, Susumu FUNAKOSHI, Motoyoshi NOMIZU, Kenichi AKAJI, Haruaki YAJIMA, Itsuo YAMAMOTO, Kanji TORIZUKA, Kouki KITAGAWA, Tadashi AKITA, Kenshi ANDO, Tatsuhiko KAWAMOTO, Yasutsugu SHIMONISHI, and

Toshifumi TAKAO

Chem. Pharm. Bull., 34, 613-620 (1986).

22) A Heat-Stable Enterotoxin of Vibrio cholerae non-01: Chemical Synthesis and Biological and Physicochemical Properties

Shoko YOSHIMURA, Toshifumi TAKAO, Yasutsugu SHIMONISHI, Saburo HARA, Micniko ARITA, Tae TAKEDA, Hidenori IMAISHI, Takashi HONDA, and Toshio MIWATANI

Biopolymers, in press (1986).

23) Primary Structure of Limulus Anticoagulant Anti-lipopolysaccharide Factor

Jun AKETAGAWA, Toshiyuki MIYATA, Sadami OHTSUBO, Takanori NAKAMURA, Takashi MORITA, Hidenori HAYASHIDA, Takashi MIYATA, Sadaaki IWANAGA, Toshifumi TAKAO, and Yasutsugu SHIMONISHI

J. Biol. Chem., 261, 7357-7365 (1986).