ELSEVIER

International Journal of Mass Spectrometry 209 (2001) 47–55

International Journal of Mass Spectrometry

Sequencing of novel protein from *Bacillus pumilus* PH-01 using a high-resolution hybrid quadrupole-time-of-flight mass spectrometer

Yoon-Seok Chang[a,*], Hyo-Bong Hong[a], Byung-Hoon Kim[a], Steven C. Hall[b], Tina A. Settineri[b]

[a]*School of Environmental Engineering, Pohang University of Science and Technology, San 31, Hyo-Ja dong, Nam-gu, Pohang, 790-784, Korea*
[b]*Applied Biosystems, Foster City, CA, USA*

Received 22 February 2001; accepted 10 May 2001

Abstract

Protein identification can be accomplished with an enzymatic digestion of the protein followed by mass spectrometric analysis of the peptide mass fingerprint followed by database searching. However, if a protein is not in a database, sequence information must be obtained to characterize and identify it. This can be done either by classical Edman sequencing or/and by tandem mass spectrometry. To determine the sequence of an unknown protein from *Bacillus pumilus* PH-01, which adsorbs environmental pollutants such as polychlorinated dibenzo-p-dioxins (PCDDs), dibenzofurans (PCDFs), and biphenyls (PCBs), both sequencing and de novo peptide sequencing by tandem MS/MS of the peptide fragments were performed. Edman sequencing of the reduced and alkylated protein revealed the majority of the sequence; however, all information on disulfide bonding was lost. Therefore, a tryptic digest of the native protein was performed to obtain both complete sequence information and the connectivity of the disulfide bonds. We performed the de novo sequencing using a hybrid quadrupole-time-of-flight mass spectrometer (Q-TOF MS) instrument. The high mass accuracy and sensitivity of the hybrid Q-TOF MS made low-level sequencing of this novel naturally isolated protein possible. (Int J Mass Spectrom 209 (2001) 47–55) © 2001 Elsevier Science B.V.

Keywords: Protein identification; Sequencing; Protein; *Bacillus pumilus*; Hybrid Q-TOF MS; MS/MS

1. Introduction

The isolation of bacterial proteins with important biological function typically results in minute quantities of proteins separated on gels. Traditionally, the gel-separated proteins have been sequenced using repeated cycles of automated Edman degradation [1–2]. However, there are some limitations of Edman sequencing. For example, it is impossible to read the N-terminally blocked proteins, and PTM (posttransnational modified) proteins are difficult to detect and identify conclusively [3].

In the 1970s, the utility of mass spectrometry (MS) for the sequencing of simple peptides was demonstrated [4–5]. MS can provide important information about the sequence of amino acids in a peptide and is used to interrogate the primary structure of proteins at low levels [6–8]. The development of matrix-assisted laser desorption ionization time-of-flight mass spec-

*Corresponding author. E-mail: yschang@postech.ac.kr

PII S1387-3806(01)00463-8

trometers (MALDI TOF MS) has led to the availability of methods for providing valuable information on the structure of proteins [9]. Recent advances in hybrid quadrupole-time of flight mass spectrometer (Q-TOF MS) instrumentation [10], coupled with electrospray ionization (ESI), have well suited biochemical researchers [11].

The unknown peptide described here was actually a small protein, which was isolated from *Bacillus pumilus* PH-01 and which can adsorb environmental pollutants such as polychlorinated dibenzo-p-dioxins (PCDDs), dibenzofurans (PCDFs), and polychlorinated biphenyls (PCBs) [12–13]. The purpose of determining the exact structure of this protein was so that a recombinant form could be made for bioremedation of toxic chemicals [14–16]. In addition, the goal of sequencing this protein was to accurately determine the mechanism of interactions between those toxic compounds and this protein.

De novo sequencing by mass spectrometry is presently limited by the frequency of backbone cleavage for collision-induced dissociation (CID) of large proteins [17–19]. However, this protein is relatively small and can be enzymatically cleaved into peptide fragments for CID experiments.

After determining its exact molecular weight both by MALDI TOF MS and by ESI on the hybrid Q-TOF MS, and getting the majority of the sequence by Edman sequencing, C-terminal sequencing with a commercial C-terminal sequencing kit was performed using MALDI TOF MS. To obtain the sequence of the remaining residues and determine any disulfide bonding of the cysteines, the protein was reduced by TCEP and/or digested with trypsin for MS/MS analysis on the hybrid Q-TOF MS.

In this article, we present a simple and straightforward method for the identification and analysis of a novel bacterial protein using hybrid Q-TOF MS.

2. Experimental

2.1. Cell growth and protein isolation

Bacteria were grown aerobically in 300 mL of nutrient broth medium in a 500-mL Erlenmeyer flask and then cultured overnight in a shacking incubator at 30°C and 160 rpm. Culture medium was transferred to six 50-mL plastic centrifuge tubes and centrifuged at 3000 rpm for 15 min. The supernatant layer was discarded, and then 0.05 M phosphate buffer (pH 7) was added before vortex was followed. We made 0.05 M phosphate buffer by dissolving 4.3 g of $Na_2HPO_4 \cdot 2H_2O$ and 2.3 g of $Na_2HPO_4 \cdot 2H_2O$ in 950 mL of distilled water, and the volume was adjusted to 1000 mL. This washing procedure was repeated two times, and then the final volume was adjusted to 20 mL.

To get protein solution, the concentrated cell solution was boiled at 100°C for 20 min and then centrifuged at 3000 rpm for 20 min. The supernatant layer was filtered by a syringe filter of pore size 0.22 μm. The protein concentration was measured by the Bradford method, and then the solution was stored in a refrigerator [12–13].

2.2. Mass spectrometry analysis

Experiments for molecular determination were carried out on a Voyager delayed extraction (DE) TOF MS (Applied Biosystems, Framingham, MA) equipped with a nitrogen laser operating at 337 nm and an ion source capable of operation in delayed extraction mode.

All samples were also run by nanospray using Protona (Odense, Denmark) tips and source on a QSTAR PULSAR hybrid Q-TOF MS (Applied Biosystems/PE SCIEX, Toronto, Ontario), flowing at 20–50 nL/min. All samples were desalted before analysis by loading them onto an uncoated nanospray tip loaded with 2–3 mm of POROS 20 R2 (Applied Biosystems) and washing with 5 μL, 5% formic acid. Samples ($<$1 pmole/μL) were eluted directly into a coated nanospray tip with 70% MeOH, 5% formic acid. The QSTAR was operated at 8000–10,000 resolution with a mass accuracy of 10–30 ppm using an external calibration maintained for 24 h. MS/MS spectra were acquired using the selection of the parent ion by the quadrupole in low-resolution mode (less than unit resolution) so as to send the entire isotopic cluster of the parent into the collision cell and, thereby, obtain the entire isotopic clusters of all of the

fragments. This allows for charge-state determination of all the ions in the MS/MS spectrum. A typical ion spray voltage for nanospray was 1000 V, and collision energies of 25–40 eV were typical for fragmentation of peptides. Nitrogen gas was used for the collision gas and typical pressures in the collision cell during MS/MS were 4.5×10^{-6} torr.

Sequences were obtained from the MS/MS spectra both manually and by machine, using the BioAnalyst (Applied Biosystems) processing software. These sequences were then compared to the tentative sequences obtained from the Edman sequencing data.

Because of the presence of cysteines in the sequences, disulfide bonds were suspected to be present. Therefore, MS/MS was performed on all of the fragments before and after reduction with TCEP (Tris [2-carboxyethyl]-phosphine-HCL; Pierce, Rockford, IL). Trypsin digestion of the *Bacillus* protein was performed at 0.25 μg/μL in 50 mM ammonium bicarbonate (pH 7.8). Promega trypsin was added at 2% by weight, and the digest was incubated at 37°C for 2.5 h. Reduction was performed by adding 5 μg TCEP to the digest solution and incubating at room temperature for 10 min. Nanospray analysis was then immediately performed. MALDI analysis was done on an Applied Biosystems Voyager DE STR, and Edman sequencing was performed on an Applied Biosystems Procise HT Protein Sequencer.

3. Results and Discussions

3.1. Molecular mass determination

The *Bacillus* protein was analyzed on a MALDI TOF MS. The matrix chosen was sinapinic acid (3, 5-dimethoxy-4-hydroxycinnamic acid), as it generally produced good-quality MALDI spectra with peptides and proteins. The monoisotopic mass of the intact protein, 5314.4 Da, was determined using a high-resolution MALDI TOF mass spectrometer operating in reflector mode (Fig. 1, top). Insulin was used as an internal standard. The mass of the intact protein was subsequently verified by using the high-resolution ESI TOF MS. Fig. 1 (bottom) shows the multiply charged ions, all of which belong to the same protein. Namely, the 4+ ion is 1329.3, the 5+ ion is 1063.5, and the 6+ ion is 886. 5 (*m/z*, respectively).

3.2. Determination of structure of Bacillus pumilus protein

Fig. 2 shows the following information after molecular weight determination of *Bacillus pumilus* protein: After trypsin digestion, MS/MS sequences were obtained for T1 and T4; however, T2 and T5 could not be found at their expected *m/z* values. The four C-terminal amino acid residues of the protein were sequenced by on-plate digestion with carboxypeptidase Y, followed by generation of a ladder mass spectrum using the MALDI TOF MS operating in linear mode. Calculation of the mass differences between adjacent peaks in the ladder spectrum yielded the four C-terminal amino acid residues as FETQ-COOH. T1 was found 2 Da below its expected *m/z* value (667.3, 2^+), and MS/MS showed that it contained a disulfide bond between cysteine residues 6 and 11. Interestingly, sequencing of two *m/z* values that were, respectively, 15.989 and 31. 978 Da higher than the T1 2^+ ion showed that these values were modified versions of T1. The exact mass difference led us to believe that they contained one and two additional sulfur atoms in the disulfide bond. Reduction of the digest with TCEP (shown in Fig. 3) caused these two adduct peaks to disappear, leaving only T1 in reduced form, confirming the presence of the tri- and tetrasulfide bonds. Fig. 4 shows the MS/MS of the reduced T1. Automated sequencing confirmed the sequence predicted by Edman degradation.

After TCEP reduction, the predicted *m/z* values for T2 and T5 also were found in the TOF MS spectrum, indicating that these two values were possibly bound by disulfide (a plausible theory because residue 19 in the sequence was found to be a Cys by Edman sequencing). The MS/MS spectrum of T5 is shown in Fig. 5. Manual and automated sequencing determined the sequence to be NAETDCETFETQ. This confirmed the first four and last four residues found with other sequencing methods and showed that the miss-

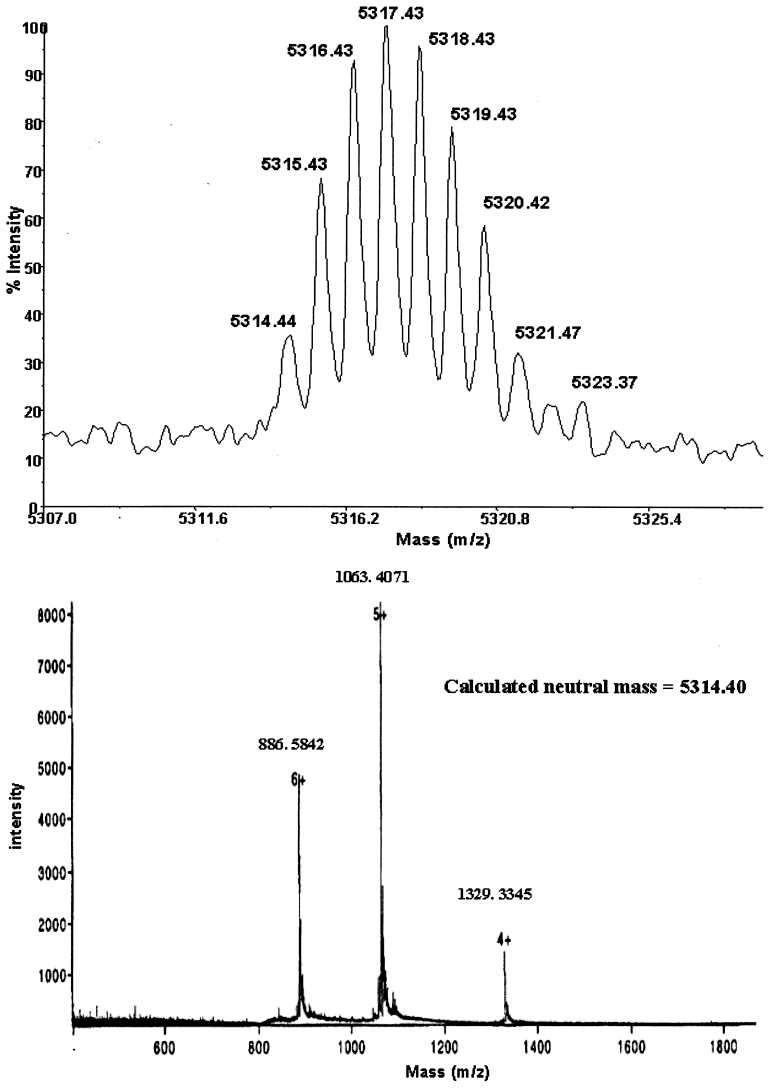

Fig. 1. Mass spectrum of protein from *Bacillus pumilus* PH-01 utilizing sinapinic acid matrix (top; HR matrix-assisted laser desorption ionization time-of-flight [TOF] mass spectrometers [MS] [a], bottom; high resolution electrospray ionization TOF MS [b]).

ing four residues in the middle of T5 were –DCET-. The presence of the cysteine in this peptide confirmed that there was a disulfide bond linking T2 and T5. Indeed, on looking at the original TOF MS data for the unreduced material, a triply charged ion at 943.388 (MW = 2827.164) was found, which corresponds to the correct mass for the disulfide bound (T2 + T5) peptide (MW = 2827.154). The data for before and after reduction are shown in Fig. 6. Fig. 7 shows the summation of all of the data after sequencing by MS/MS on the Q-TOF MS before and after reduction of the di-, tri-, and tetrasulfide bonds.

Finally, to obtain more information on the new protein sequence, the BLAST (http://www.ncbl.nlm.nih.gov/BLAST/) resource protein sequence database was used to perform a homology database search.

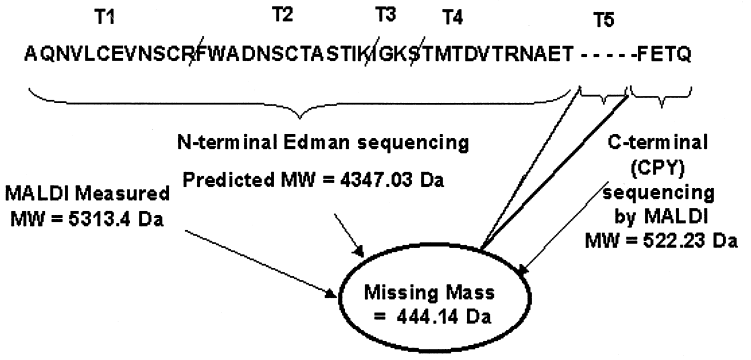

Fig. 2. Partial sequence data from Edman degradation and C-terminal sequencing by matrix-assisted laser desorption ionization.

This database provides significant information on homologues with nonidentical sequences of protein. The result of this database search (data not shown here) led us to believe that the newly discovered protein is a novel protein from *Bacillus pumilus*. It has never been reported previously that such a bacteria-secreted protein adsorbs dioxins and dioxin-like compounds.

Clearly, the sequence data obtained from a hybrid Q-TOF instrument are extremely powerful and are

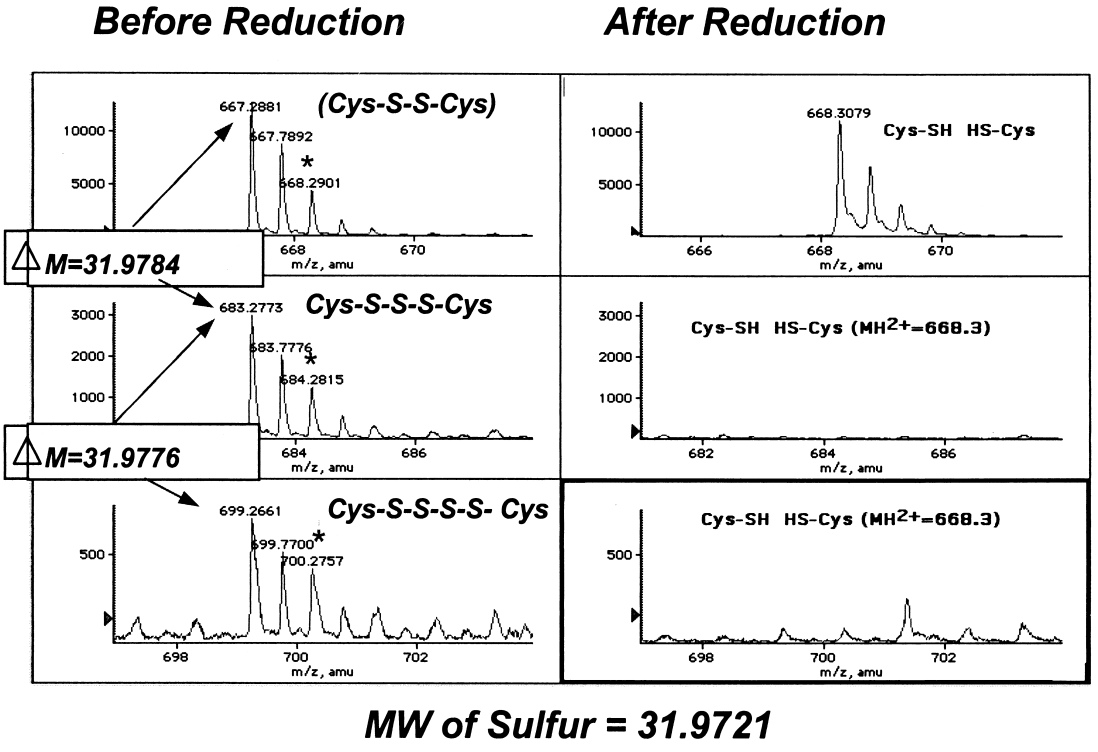

Fig. 3. Evidence for disulfide as well as tri- and tetrasulfide bonds formed between Cys-6 and Cys-11 of T1. The increase in intensity of second isotope is caused by additional sulfur.

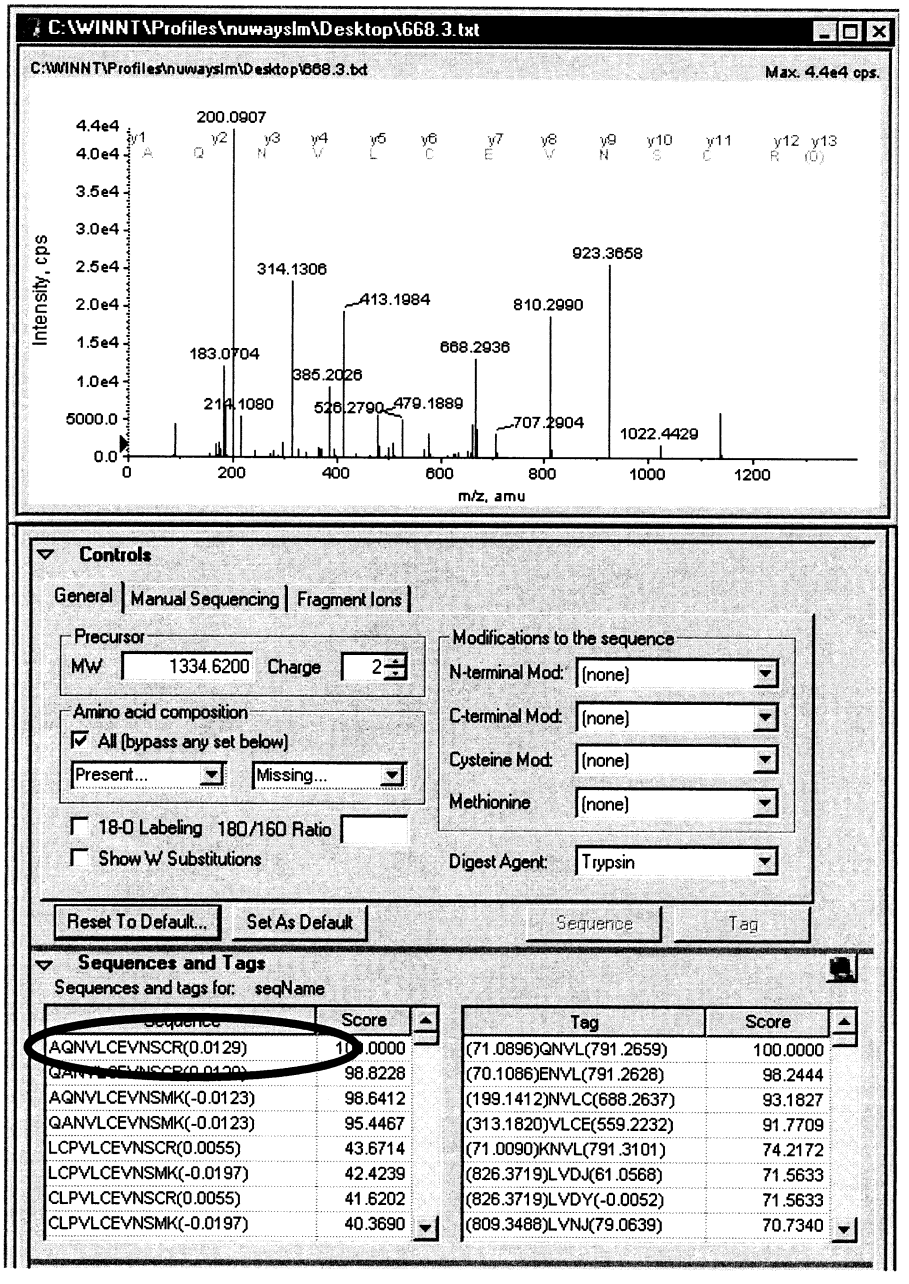

Fig. 4. MS/MS of T1 (*MW* = 1334.62) after reduction with TCEP (Tris [2-carboxyethyl]-phosphine- HCL). The sequence, AQNVLCEVNSCR (in bold circle), is determined by the BioAnalyst.

excellent complements to Edman sequencing. The Edman sequencing data could not yield any information on disulfide bonding. In addition, ambiguity often exists in the calling of residues by Edman sequencing, especially when the sample is contaminated by other trace peptides, so having confirmatory information from MS/MS data is very useful. In addition, the sensitivity of this mass spectrometric technology is

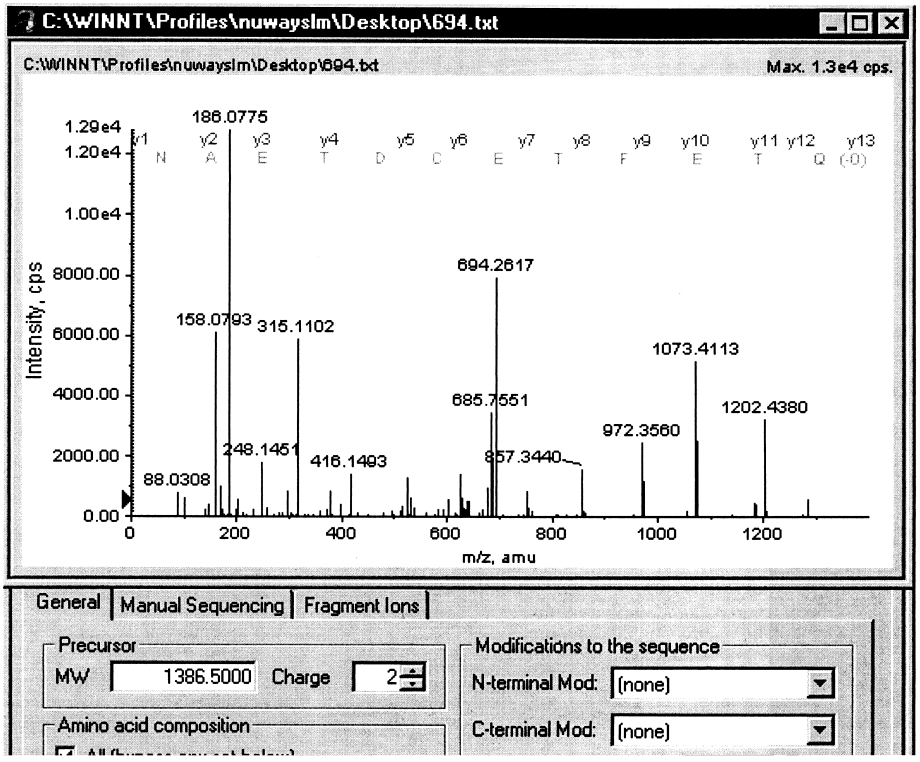

Fig. 5. MS/MS of T5 (C-terminal peptide) after TCEP (Tris [2-carboxyethyl]-phosphine-HCL) reduction. The sequence, NAETDCETFETQ (in bold circle), is determined by the BioAnalyst.

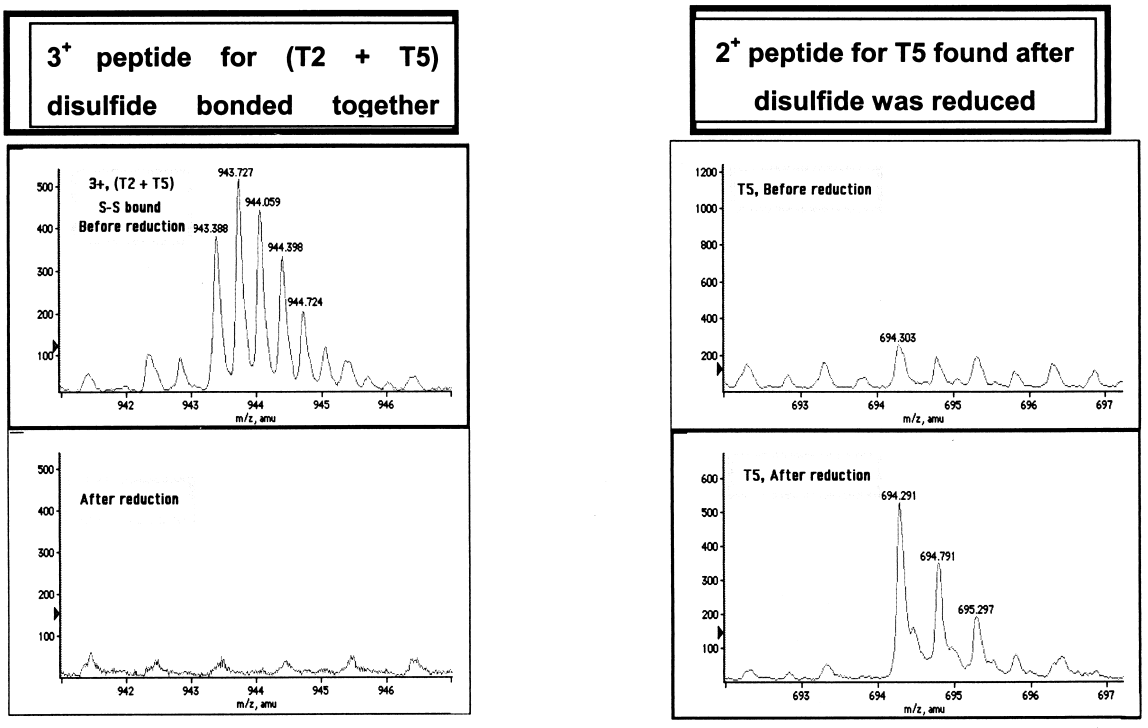

Fig. 6. High-resolution time-of-flight mass spectrometry spectra of T5 (C-terminal peptide) before reduction.

one or two orders of magnitude more sensitive than Edman sequencing. This is very important information for those that typically rely only on Edman type data.

4. Conclusion

The sensitivity and high mass accuracy of the hybrid Q-TOF mass spectrometer has tremendous advantages for de novo sequencing of low levels of novel proteins and peptides. The ability to generate MS/MS data that contains sequence information with very accurate masses (20 ppm or less) gives one the ability to characterize structures of peptides with disulfide bonds and modifications, which is virtually impossible to perform by Edman sequencing. However, the connectivity of the peptides needed to make up the sequence would be much harder to determine

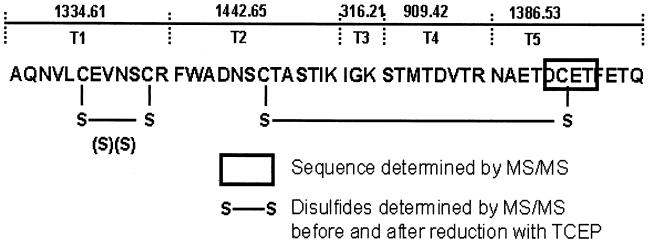

Fig. 7. Final structure of polypeptide determined by hybrid quadrupole-time-of-flight mass spectrometer in conjunction with Edman sequencing.

without Edman data. Therefore, this highly accurate MS/MS data is a necessary complement to Edman data when de novo sequencing is required, especially at femtomole peptide/protein levels.

In addition, the use of de novo sequencing in conjunction with searching a homology-based protein sequence database identified a significant novel protein from *Bacillus pumilus* PH-01. This sequence is significantly important for elucidating the mechanism of this biomolecule's binding to environmental pollutants, such as PCDD/Fs and PCBs. At present, we are conducting three-dimensional structure analysis by using cellular modeling, NMR, and crystallography.

Acknowledgements

This work was financially supported in part by the Korea Science and Engineering Foundation (KOSEF) through the Advanced Environmental Monitoring Research Center (ADERC) at Kwangju Institute of Science and Technology through the and G-7 Project from the National Institute of Environmental Research (2000).

References

[1] W.J. Henzel, J. Tropea, D. Dupont, Anal. Biochem. 267 (1999) 148.
[2] L. Stryer, Biochemistry, W.H. Freeman, New York 1995.
[3] D.B. Kassel, J.T. Stults, P. Thibault, Mass Spectrometric Characterization of Peptide and Proteins. The Short Course 46th ASMS Conference on Mass Spectrometry and Allied Topics, Orlando, FL, 1994.
[4] P. Roepstorff, J. Fohlman, Biomed. Mass Spectrom. 11 (1984) 601.
[5] K. Biemann, Biomed. Environ. Mass Spectrom. 16 (1988) 99.
[6] R.S. Johnson, S.A, Martin, K. Biemann, J.T. Stults, J.T. Watson, Anal. Chem. 59 (1987) 2621.
[7] B. Christian, Biomed. Environ. Mass Spectrom. 19 (1990) 363.
[8] G. Siuzdak, Mass Spectrometry for Biotechnology, Academic Press, San Diego 1996.
[9] G. Gritchley, B. Worster, VG ORGANIC Applications Note Number AN43, VG ORGANIC, Manchester, M23 9LZ, UK.
[10] P. Chaurand, F. Luetzenkirchen, J. Am. Soc. Mass Spectrom. 10 (1999) 91.
[11] S. Andrej, C. Igor, E. Werner, G.S. Kenneth, T. Bruce, W. Matthias, M. Matthias, Rapid Commun. Mass Spectrom. 11 (1997) 1015.
[12] H-B. Hong, S-D. Choi, Y-S. Chang, Organohalogen Compounds. 41 (1999) 369.
[13] H-B. Hong, Y-S. Chang, S-D. Choi, Y-H. Park, Water Research. 34 (1999) 2404.
[14] B.E. Gillesby, T.R. Zacharewski, Environ. Toxicol. Chem. 17 (1998) 3.
[15] S.P. Bradbury, O.G. Mekenyan, G.T. Ankley, Environ. Toxicol. Chem. 17 (1998) 15.
[16] A.J. Murk, J. Legier, M.S. Denison, J.P. Giesy, C. Guchte, A. Brouwer, Fund. Appl. Toxicol. 33 (1998) 149.
[17] J.A. Taylor, R.S. Johnson, Rapid Commun. Mass Spectrom. 11 (1997) 1067.
[18] C.D. Elenor, P.B, Ida, C.C. Rosanne, T.L. Victor, G. Richard, O. Kenneth, L.G. Beth, Anal. Chem. 68 (1996) 4044.
[19] D.C. Reiber, R.S. Brown, S. Weingerger. J. Kenny, J. Bailey, Anal. Chem. 70 (1998) 1214.