RAPID COMMUNICATIONS IN MASS SPECTROMETRY
Rapid Commun. Mass Spectrom. 2003; **17**: 903–908
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/rcm.999

RCM

Primary structural determination of N-terminally blocked peptides from the bark of *Eucommia ulmoides* Oliv by mass spectrometric analysis

Ren-Huai Huang, Ying Zhang and Da-Cheng Wang*

Center for Molecular Biology, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, P.R. China

Received 5 November 2002; Revised 19 February 2003; Accepted 19 February 2003

Sequencing of N-terminally blocked proteins/peptides is a challenge since these molecules inhibit processing by Edman degradation. By using high accuracy Fourier transform ion cyclotron resonance (FTICR) tandem mass spectrometry and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOFMS), the primary structures of two novel N-terminally blocked antifungal peptides (EAFP1 and EAFP2), purified from the bark of *Eucommia ulmoides* Oliv, have been determined. The results show that the high mass accuracy provided by FTICR mass spectrometry is effective to determine the N-terminally blocking group, and can simplify the task of spectral interpretation and improve the precision of sequence determination. The combination of MALDI-TOFMS with carboxyl peptidase Y digestion was used to determine the C-terminal 36- and 27-residue sequences of EAFP1 and EAFP2, respectively, to provide the sequence linkage information for tryptic fragments. Compared with traditional peptide chemistry the advantage of mass spectrometric techniques is their simplicity, speed and sensitivity.

It is known that many proteins and peptides are N-terminally blocked.[1] Over ten kinds of blocking groups have been found,[2] including acetyl, formyl, and pyroglutamyl. The amino acid sequence determinations of these proteins or peptides are still tedious since these modifications inhibit the processing of Edman degradation. Traditionally, a chemical reagent or enzyme is used to cleave the blocking group at the N-terminus before proceeding to Edman degradation.[3–5] Although many kinds of chemical reagents or enzymes have been developed to deblock the N-termini of such sequences, there is still no universal way to cleave all kinds of blocking groups, and each kind of blocked residue can only be deblocked by a corresponding specific chemical reagent or enzyme. A series of chemical reagents or enzymes needs to be tried for sequence analysis of novel N-terminally blocked proteins or peptides, and the efficiencies of some deblocking reactions are so low that a large amount of sample would be consumed. This is truly time-consuming, and sometimes is not feasible in practice if the quantity of sample is insufficient.

With the recent advances in mass spectrometry, such as the ionization techniques matrix-assisted laser desorption/ionization (MALDI)[6] and electrospray,[7] it has become possible to perform primary structural determinations using mass spectrometric analysis.[8,9] Fourier transform ion cyclotron resonance (FTICR) mass spectrometry has been developed to provide high mass accuracy, and can be coupled with electrospray,[10] but to date efforts to analyze novel sequences of N-terminally blocked proteins or peptides by exploiting this feature have been rarely reported.

Recently, two novel antifungal peptides, purified from the bark of *Eucommia ulmoides* Oliv, were found to block Edman degradation.[11] Using high accuracy FTICR-MS we have successfully determined their N-terminally blocked residues as pyroglutamic acid. By the use of MALDI-TOFMS coupled with carboxyl peptidase Y (CPY) digestion, the C-terminal sequences have been obtained to provide sufficient information to assemble the tryptic fragment sequences derived from the traditional Edman degradation. The above results are completely consistent with the amino acid composition. In this paper we report the sequence determination in detail.

EXPERIMENTAL

Materials

The bark of *Eucommia ulmoides* Oliv was collected from the Kunming mountain area of Yun-Nan province in southwest China.

All sequencing reagents including phenylisothiocyanate (PITC), trifluoroacetic acid (TFA) and 1-methylpiperidine were sequence grade from Applied Biosystems. Dithiothreitol (DTT), iodoacetamide (IAA), α-cyano-4-hydroxycinnamic acid (CHCA), TPCK-treated trypsin and carboxypeptidase Y (CPY) were purchased from Sigma. Acetonitrile (ACN) and

**Correspondence to*: D.-C. Wang, Center for Molecular Biology, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, P.R. China.
E-mail: dcwang@sun5.ibp.ac.cn
Contract/grant sponsors: State Basic Research Program of China; Natural Science Foundation of China.

TFA used in HPLC analysis were HPLC grade. All other reagents were analytical grade.

Isolation and purification of Eucommia antifungal peptides (EAFPs)

Crude EAFPs were isolated from the bark of *E. ulmoides* Oliv by precipitation with ammonium sulfate and fractionation and purification using a CM-cellulose-52 ion-exchange column and a Bio-gel-p-10 column successively, as described previously.[11] Further purification was performed using HPLC with a Waters YWG-Pak reversed-phase column.

MALDI-TOFMS analysis

Matrix-assisted laser desorption/ionization time of flight (MALDI-TOF) mass spectra were recorded using a Bruker Proflex III mass spectrometer (Billerica, MA, USA). The matrix was CHCA, dissolved to saturation in 50% ACN solution containing 0.1% TFA. Peptide solution (2 μL, about 1–5 pmol/μL) dissolved in a 0.1% TFA solution in water was mixed with 20 μL of the CHCA saturated matrix solution. 1 μL of this mixture was deposited on the target sample plate and dried. The spectral data were calibrated with huwentoxin-I (isotope-averaged MW 3750.45 Da, kindly donated by Dr. S. P. Liang[12]).

Reduction and alkylation of cysteine residues

The reduction and carboxyamidomethylation of the peptides were performed as follows: 200 μg of peptide were dissolved in 200 μL of buffer containing 0.1 mol/L Tris-HCl, 6 mol/L guanidinium HCl and 1.7 mg DTT at pH 8.3. This solution was secured under nitrogen gas (N_2), well mixed and allowed to stand at 37°C for 30 min. Then, 4.6 mg IAA were added and the mixture was allowed to stand under N_2 at 37°C; after incubating for 4 h, the reaction mixture was diluted to 1 mL with double-distilled H_2O and purified by reversed-phase HPLC with a linear gradient of 5–45% ACN containing 0.1% TFA over 60 min.

Tryptic digestion and HPLC fractionation of peptides

100 μg of sample dissolved in 200 μL buffer of 0.1 mol/L ammonium bicarbonate with pH 8.0 were mixed with 2 μg of trypsin and incubated at 37°C overnight. The resulting peptides were fractionated by reversed-phase HPLC with a linear gradient of 5–40% ACN containing 0.1% TFA over 60 min. The purified peptides were collected and identified by MALDI-TOFMS and Edman degradation sequencing.

Edman degradation sequence determination

The Edman degradation was performed using a Applied Biosystems 491 gas-phase sequencer with a normal automated program. The phenylthiohydantoin (PTH) amino acids produced by the sequential Edman degradation of the analyzed peptides/proteins were identified on-line using the retention times of a PTH standard mixture run as a reference.

C-terminal ladder sequence analysis

C-terminal sequence analysis was performed by MALDI-TOFMS coupled with carboxypeptidase Y digestion as described by Patterson.[13] 500 pmol of peptide sample and 0.16 units of CPY were suspended in 20 μL and 30 μL of HPLC-grade water, respectively, and then mixed to initiate the reaction at 37°C. The final concentrations were 10 pmol/μL peptide and 0.32×10^{-3} units/μL CPY yielding an enzyme-to-substrate ratio of 3.2×10^{8} units of CPY/mol of peptide sample. Aliquots of 1 μL were taken from the reaction mixture at different reaction times and immediately mixed with 9 μL of CHCA saturated matrix solution in 50% ACN containing 0.1% TFA to quench the reaction. 1 μL of each solution aliquot was placed onto individual wells on the MALDI sample plate, and allowed to evaporate to dryness before insertion into the mass spectrometer.

High accuracy tandem mass spectrometric analysis

Tandem mass spectra for peptide sequencing were recorded using an APEX™ II FTICR mass spectrometer (Bruker Daltonics Inc., Billerica, MA, USA). The peptide (about 500 pmol) was dissolved in 25 μL of 50% methanol containing 2% acetic acid and infused into the mass spectrometer at a rate of 30 μL/h. The MS/MS fragment ions were generated by introducing a 3500 μs argon pulse to yield 1 mbar pressure, and interpreted according to the nomenclature proposed by Roepstorff and Fohlman.[14,15]

Amino acid analysis

Amino acid analysis was performed using the Pico-Tag protocol provided by Waters Corp. PTC-derivatives were fractionated using a Waters Nova-Pak C18 column (3.9 × 300 mm) connected with a Waters Alliance system.

RESULTS AND DISCUSSION

Purification, alkylation and MALDI-TOF analysis

From the bark of *E. ulmoides* Oliv two antifungal peptides were purified, and named EAFP1 and EAFP2 according to their relative retention times in the reversed-phase chromatogram. These two antifungal peptides can specifically bind with fungal cell wall chitin, and have a potent antifungal activity against a broad spectrum of chitin-containing and chitin-free fungi including eight pathogenic fungi from cotton, wheat, potato, tomato and tobacco.

EAFP1 and EAFP2 were reduced and alkylated with DTT and iodoacetamide as described above. The carboxyamidomethylated EAFPs were purified by reversed-phase HPLC and their molecular masses were determined by MALDI-TOFMS as 4782.4 and 4739.2 Da, respectively (as shown in Figs. 1(C) and (D)), while the molecular masses of the native forms were determined as 4201.4 and 4158.9 Da, respectively (as shown in Figs. 1(A) and (B)). The differences between the masses of the native forms and the carboxyamidomethylated forms can be calculated as ca. 580 Da, indicating that EAFPs contain 10 cysteines which form five pairs of disulfide bridges, since the molecular mass will increase by 58 Da when half of a cystine group is carboxyamidomethylated.

Tryptic digestion and identification of tryptic peptides

PTH signals were not observed when the native and carboxyamidomethylated EAFPs were subjected to Edman degra-

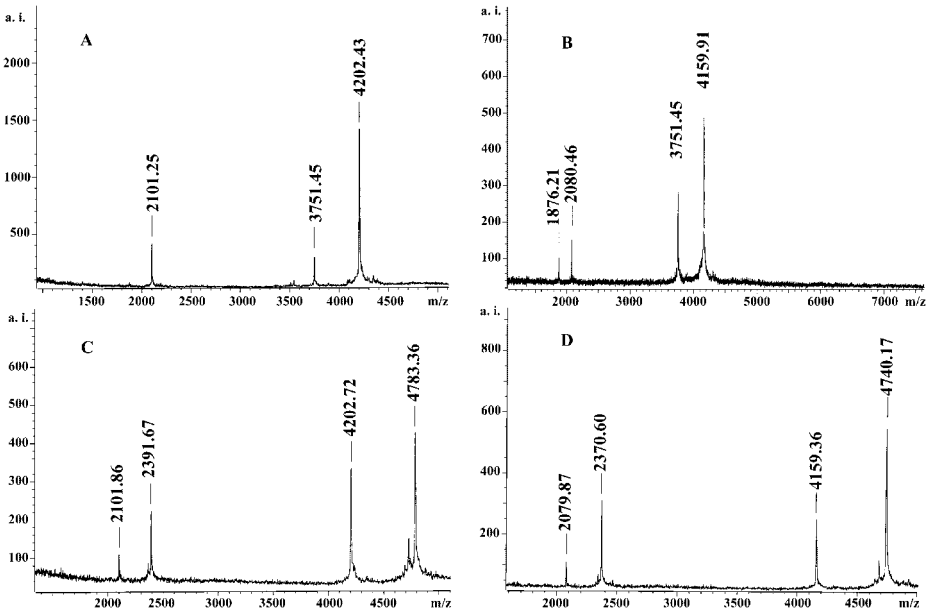

Figure 1. MALDI mass spectra of native and carboxyamidomethyl EAFP1 (A, C) and EAFP2 (B, D). A and B are the native forms, C and D the carboxyamidomethylated forms. The indicated masses correspond to those of $[M+H]^+$. Huwentoxin-I[12] with a molecular mass of 3750.45 Da is used as the internal reference in (A) and (B), while the native forms of EAFP1 and EAFP2 served as the internal references in (C) and (D), respectively.

dation, indicating that EAFP1 and EAFP2 are blocked at the N-termini. The carboxyamidomethylated EAFP1 and EAFP2 were digested by trypsin, and the resulting peptides were fractionated by reversed-phase HPLC. The purified peptide fragments were identified by MALDI-TOFMS and automated Edman degradation. The results are summarized in Table 1. No PTH peaks were observed for the peptide with molecular mass of 704 Da derived from EAFP2 when subjected to Edman degradation, suggesting that this peptide is N-terminally blocked and thus located at the N-terminus of EAFP2.

Tandem mass spectrum for peptide sequence analysis

The monoisotopic mass of the $[M+H]^+$ ion of the N-terminally blocked fragment derived from EAFP2 was determined as 705.2997 Da by FTICR-MS. The primary structure was determined by high accuracy tandem mass spectrometry (MS/MS). The measurement uncertainties for the masses of all the MS/MS peaks (Fig. 2) were less than 10 ppm due to the use of a reference MS/MS spectrum to recalibrate the MS/MS spectra.[16] The nomenclature used is as described previously[14,15] and on the Matrix Science website.[17] The peaks in the spectra are also assigned in Table 2. The high resolution of the spectra and the high accuracy of the mass data greatly simplify the task of spectral interpretation. The sequence can be derived from the spectrum as pGlu-Thr-Cys-Ala-Ser-Arg, which is N-terminally blocked by the pyroglutamic residue. The molecular mass of the $[M+H]^+$ ion calculated from the sequence is 705.2984 Da, which is consistent with the experimental value determined above to within an accuracy of 1.84 ppm.

Table 1. Sequences and masses of the tryptic fragments

EAFP1				EAFP2			
		Molecular mass				Molecular mass	
Ret. time (min)	Sequence	Exp. $[M+H]^+$	Theor. [M]	Ret. time (min)	Sequence	Exp. $[M+H]^+$	Theor. [M]
9.7	CQCR	608[#]	622.7	9.4	CQCR	ND	622.7
11.0	CGAGNCR	796	794.3	10.3	CGAGNCR	792	794.3
12.9	<QTCASR**	705	704.3	12.3	<QTCASR*	705	704.3
15.1	CPRPC	692	689.3	13.6	CPRPC	689	689.3
24.7	GYCGSGNAY	947	949.4	25.7	GYCGSGAAY	904	905.4
39.8	CPRPCNAGLCCSIY	1727	1727.7	39.4	CPRPCNAGLCCSIY	1728	1727.7
41.1	NAGLCCSIY	1057	1057.4	40.2	NAGLCCSIY	1057	1057.4

<Q denotes pyroglutamic acid; ND, not detected.
[#]The expected mass with loss of 17 mass units.
*The sequence is deduced from the tandem mass spectra, and other sequences are determined by Edman degradation.
**The sequence is believed to be the same as the N-terminal sequence of EAFP2 although not determined experimentally.

RCM

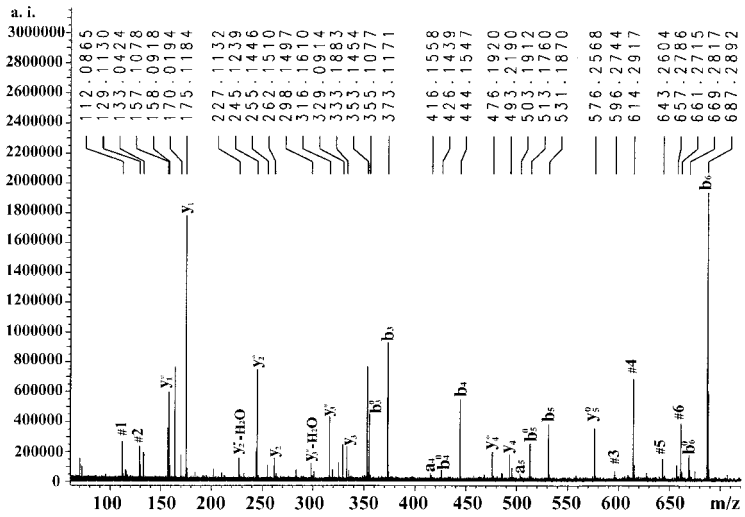

Figure 2. Tandem mass spectra of the 704-Da peptide derived from EAFP2. Peaks labeled 661.2715 and 614.2917 correspond to the peaks for the precursor ions with side-chain cleavages of threonine and cysteine, respectively; 643.2904 and 596.2744 correspond to the peaks for the dehydrated product ions of the above two fragments; 112.0865 and 129.1130 represent the masses of the arginine-related ions.

A peptide with molecular mass of 705 Da ($[M+H]^+$) is also found in the peptide map of EAFP1, and this peptide has the same retention time and the same mass to within experimental uncertainty as the 704 Da peptide of EAFP2 (Table 1). Its sequence is believed to be same as that of the N-terminal sequence of EAFP2, although it was not determined experimentally.

C-terminal ladder sequence analysis

Due to lack of a complementary second set of cleavage sites in the sequences, the tryptic fragments (Table 1) of the EAFPs cannot be linked into a unique sequence without using a different second set of cleavage fragments. MALDI-TOFMS coupled with CPY digestion has been shown to be an effective method for obtaining C-terminal sequence information.[13]

Table 2. Interpretation of peaks in the tandem mass spectra of the 704-Da peptide from EAFP2

Ion type	Description	Exp. mass	Theor. mass	Accuracy (ppm)
b_6		687.2892	687.2878	2.04
b_6^0	b_6-H_2O	669.2817	669.2772	6.67
#6	MH^+-CH_2CHOH	661.2715	661.27218	1.04
#5	#6-H_2O	643.2604	643.26166	1.96
#4	MH^+-NH_2COCH_2SH	614.2917	614.28921	6.05
#3	#4-H_2O	596.2744	596.27869	7.19
y_5^0	y_5-H_2O	576.2568	576.2558	1.74
b_5		531.1870	531.1867	0.56
b_5^0	b_5-H_2O	513.1760	513.17618	0.35
a_5		503.1912	503.1918	1.19
y_4		493.2190	493.2187	0.61
y_4^*	y_4-NH_3	476.1920	476.19214	0.29
b_4		444.1547	444.1547	0.00
b_4^0	b_4-H_2O	426.1439	426.14418	0.66
a_4		416.1558	416.1598	9.62
b_3		373.1171	373.1176	1.34
b_3^0	b_3-H_2O	355.1077	355.10708	1.75
y_3		333.1883	333.1881	0.60
y_3^*	y_3-NH_3	316.1610	316.16154	1.71
	y_3^*-H_2O	298.1497	298.15102	4.43
y_2		262.1510	262.15098	0.08
y_2^*	y_2-H_2O	245.1239	245.1244	2.12
	y_2^*-H_2O	227.1132	227.1139	3.08
y_1		175.1184	175.11895	3.14
y_1^*	y_1-NH_3	158.0918	158.0924	3.73
y_1^0	y_1-H_2O	157.1078	157.1084	4.01
#2	y_1^0-CO	129.1130	129.11353	4.11
#1	y_1^*-H_2O-CO	112.0865	112.08697	4.20

#4 and #6 are the precursor ions with the side-chain cleavage of carboxyamidomethyl cysteine and threonine.

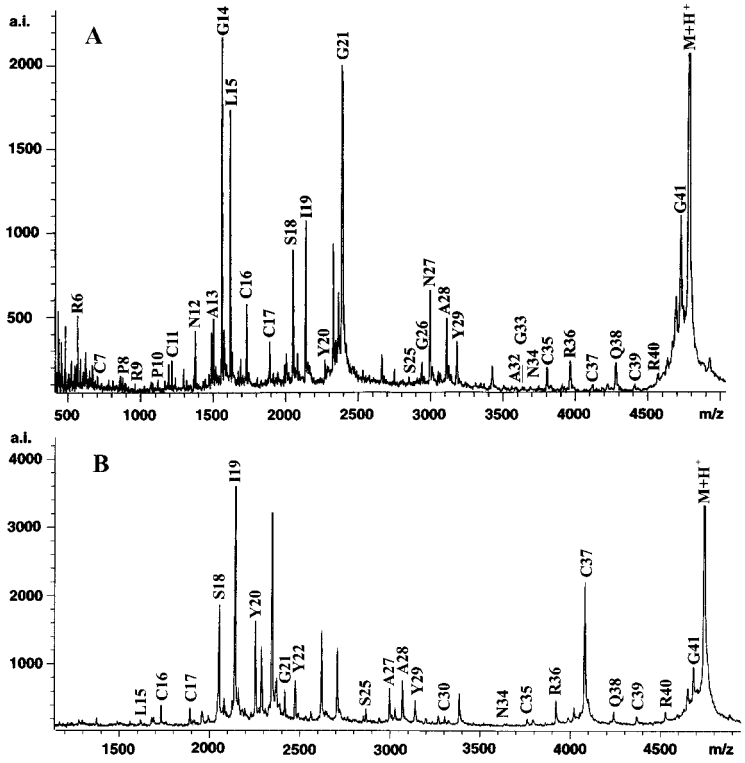

Figure 3. MALDI mass spectra of the CPY digest mixtures of EAFP1 (A) and EAFP2 (B). C-Terminal sequence of EAFP1: R CPRPC NAGLC CSIYG ycgSG NAYcg AGNCR CQCRG derived from (A); C-terminal sequence of EAFP2: LC CSIYG YcgSg AAYCg agNCR CQCRG derived from (B). The lower-case letters in the sequences indicate the unidentified residues in the spectra. The nomenclature used to label the peaks denotes the peptide populations resulting from the loss of the indicated amino acid residues.

The amino acid residues can be released one by one from the C-terminus of the peptide by CPY enzymatic processing, and the masses of the resulting peptides in the mixture can be determined by MALDI-TOFMS. The sequence can be read from the mass differences between adjacent peptide peaks.[13] Using this method, the C-terminal 36- and 27-residue sequences of EAFP1 and EAFP2, respectively, were obtained. Figure 3 shows the mass spectra of the reaction mixture quenched after reaction overnight. When the reaction mixture was quenched after a shorter time only a few peaks could be observed, mainly due to the presence of a C-terminal glycine which is only slowly digested by CPY. From Fig. 3 the C-terminal sequences of EAFP1 and EAFP2 can be read off respectively as:

R CPRPC NAGLC CSIYG ycgSG NAYcg AGNCR CQCRG
and LC CSIYG YcgSg AAYCg agNCR CQCRG.

The lower cases in these sequences indicate the gaps unidentified in these mass spectra. These arise mainly because the hydrolysis rates vary with the nature of the amino acids at the C-terminus and at the penultimate position.[13] Glycine is hydrolyzed at a very low rate, which results in the signal of the peptide population representing the loss of some glycine residue being small or even disappearing. However, the peaks representing the losses of residues in the gap regions could be generated by adjusting the quantity of enzyme and the quenching time appropriately (data not shown). The sequence analysis can frequently be processed until the detection of the truncated peptide peaks is impaired by the presence of CHCA matrix ions (<600 Da). These C-terminal sequence results not only provided the information required for full sequence assembly, but also confirmed the results summarized in Table 1.

From the results reported above the sequences of EAFP1 and EAFP2 can be completely deduced as shown in Fig. 4. These are entirely consistent with the results of amino acid analysis (data not shown). The isotope-averaged molecular masses calculated from the sequences are 4211.80 and 4168.77 Da for EAFP1 and EAFP2, respectively, which are both higher by about 10 Da than the experimental values. This result is consistent with the conclusion that the 10 cysteines of each EAFP lost a hydrogen atom, respectively, to be included in five pairs of disulfide bonds. To our knowledge this is the first finding of a plant antifungal peptide with a five-disulfide motif.

CONCLUSIONS

The amino acid sequence analysis of N-terminally blocked proteins and peptides is still a challenge since they resist processing by Edman degradation. The traditional method uses

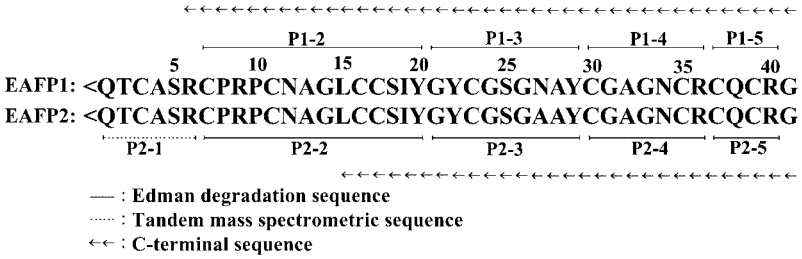

Figure 4. Proof of the amino acid sequence assembly of EAFP1 and EAFP2. The pyroglutamic residue is denoted by <Q.

specific chemicals or enzymes to cleave the blocking group before Edman degradation, which can be time-consuming and sometimes not feasible when there are insufficient quantities of sample. The results reported in this paper suggest that mass spectrometry can be used as a universal method with simplicity, fast speed and high sensitivity to solve these problems. Tandem mass spectrometry is shown to be an effective method to analyze the N-terminally blocking group, while the high mass accuracy of the FTICR mass spectrometer can simplify the task of interpretation of fragment ions and improve the reliability of sequence determination. MALDI-TOFMS coupled with carboxypeptidase Y digestion is shown to be a alternative highly efficient way to perform C-terminal sequence analysis and provide sufficient information for assembling the tryptic fragments into the sequence. In addition the molecular mass determination by high-accuracy mass spectrometry coupled with the thiol modification can easily determine the number of disulfide bridges.

Acknowledgements

This work was supported by the State Basic Research Program of China and the Natural Science Foundation of China. We are grateful to Mr. Xin B of the Institute of Chemistry of the Chinese Academy of Sciences for help in measurements of FTICR mass spectra.

REFERENCES

1. Brown JL, Roberts WK. *J. Biol. Chem*. 1976; **251**: 1009.
2. Tsunasawa S, Hirano H. In *Methods in Protein Sequence Analysis*, Imahori K, Sakiyama F (eds). Plenum: New York, 1993; 43.
3. Wellner D, Panneerselvam C, Horecker BL. *Proc. Natl. Acad. Sci. USA* 1990; **87**: 1947.
4. Miyatake N, Kamo M, Satake K, Uchiyama Y, Tsugita A. *Eur. J. Biochem*. 1993; **212**: 785.
5. Bergman T, Gheorghe MT, Hjelmqvist L, Jornvall H. *FEBS Lett*. 1996; **390**: 199.
6. Stults JT. *Curr. Opin. Struct. Biol*. 1995; **5**: 691.
7. Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM. *Science* 1989; **246**: 64.
8. Hunt DF, Yates JR, Shabanowitz J, Winston S, Hauer CR. *Proc. Natl. Acad. Sci. USA* 1986; **83**: 6233.
9. Wilm M, Shevchenko A, Houthaeve T, Breit S, Schweigerer L, Fotsis T. *Nature* 1996; **379**: 466.
10. Hendrickson CL, Emmett MR. *Annu. Rev. Phys. Chem*. 1999; **50**: 517.
11. Liu XZ, Hu Z, Li Y, Yang JB, Li BJ. *Acta Botanica Yunnanica* 1994; **16**: 385.
12. Liang SP, Zhang DY, Pan X, Chen Q, Zhou PA. *Toxicon* 1993; **31**: 969.
13. Patterson DH, Tarr GE, Regnier FE, Martin SA. *Anal. Chem*. 1995; **67**: 3971.
14. Roepstorff P, Fohlman. *J. Biomed. Mass Spectrom*. 1984; **11**: 601.
15. Johnson RS, Martin SA, Biemann K, Stults JT, Watson JT. *Anal. Chem*. 1987; **59**: 2621.
16. Hau J, Stadler R, Jenny TA, Fay LB. *Rapid Commun. Mass Spectrom*. 2001; **15**: 1840.
17. Available: www.matrixscience.com.