Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Chromatography B, 803 (2004) 167–172
JOURNAL OF CHROMATOGRAPHY B
www.elsevier.com/locate/chromb

Short communication

Mass spectrometric strategy for primary structure determination of N-terminally blocked peptides

Ren-Huai Huang, Da-Cheng Wang*

Center for Molecular Biology, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, PR China

Abstract

The mass spectrometric strategy including three steps is presented for primary structure determination of the N-terminally blocked peptides. First, the C-terminal sequencing is performed by using matrix-assisted laser desorption/ionization (MALDI) mass spectrometry coupled with carboxypeptidase Y digestion. Then, the peptide is cleaved according to the obtained C-terminal sequence information and the resulting peptides are identified by mass spectrometry and Edman degradation after fractionation by reverse-phase chromatography. Finally, the N-terminal fragment is sequenced by tandem mass spectrometry. The strategy was successfully applied to the sequence determination of two novel N-terminally blocked peptides named EAFP1 and EAFP2.

Keywords: Primary structure determination; Peptides, N-terminally blocked

1. Introduction

The primary structural information is the basis for understanding the three-dimensional structure and function of the protein or peptide. Reliable and fast determination of the amino acid sequence is crucial for the identification and analysis of protein or peptide. Among existing methods for sequence determination, Edman degradation gained almost exclusive application. But many proteins or peptides cannot be directly sequenced by Edman degradation because they have a blocked N-terminal residue [1]. These blockages occur frequently in the eukaryotic proteins [2] and over 10 kinds of blocking groups have been found, such as acetyl, formyl or pyroglutamyl. A series of specific chemical reagents and enzymes has been developed to remove these blocked groups [1,3,4]. When we encounter the N-terminal blockage problem we can try these deblocking reagents and enzymes one by one till the N-terminally blocking group is removed. But some kinds of deblocking reagents have side reactions in some peptide bonds and certain side chains of the amino acids. So, this process will usually spend lots of time and need relative high amount of sample and even analyze with difficulty.

Mass spectrometry provides the capability for sequence determination [5,6] because of the breakthrough of mass spectrometric techniques such as tandem mass spectrometry [7], matrix-assisted laser desorption/ionization (MALDI) [8] and electrospray ionization (ESI) [9]. In this paper, we presented the three-step strategy for primary structure determination of the N-terminally blocked peptides by mass spectrometry. First, the C-terminal sequencing is performed by using MALDI mass spectrometry coupled with carboxypeptidase digestion [10]. And then, the peptides are cleaved according to the obtained C-terminal sequence information, and the resulting peptides are fractionated by reverse-phase HPLC and identified by mass spectrometry and Edman degradation. The N-terminal fragment can be identified by the comparison of the molecular weight between the intact peptide and its hydrolysates. Finally, the N-terminal fragment is sequenced by tandem mass spectrometry. Two novel N-terminally blocked peptides, named EAFP1 and EAFP2 [11–13], were sequenced by using this strategy.

2. Experimental

2.1. Sample and materials

Two antifungal peptides, named EAFP1 and EAFP2, used in the experiment, were purified from the bark of

* Corresponding author. Tel.: +86-10-6488-8547; fax: +86-10-6488-8560.
E-mail address: dcwang@sun5.ibp.ac.cn (D.-C. Wang).

1570-0232/$ – see front matter © 2003 Elsevier B.V. All rights reserved.
doi:10.1016/j.jchromb.2003.07.012

Eucommia ulmoides Oliv (Garryales order, Eucommiaceae family) as described previously [11,12]. The preliminary analysis showed that these two peptides are N-terminally blocked [11–13]. The way of trying the deblocking reagents and enzymes one by one was excluded considering the existence of a number of N-terminal blockages. The three-step strategy was used for their sequencing.

The dithiothreitol (DTT), iodoacetamide (IAA), α-cyano-4-hydroxycinnamic acid (CHCA), TPCK-treated trypsin and carboxypeptidase Y (CPY) were purchased from Sigma Chemical Co. (St. Louis, MO, USA). All sequencing reagents including phenylisothiocyanate (PITC), trifluoroacetic acid (TFA) and 1-methylpiperidine were of sequence grade and from Applied Biosystems (Foster City, USA). The acetonitrile (ACN) was of high-performance liquid chromatography (HPLC) grade. All other reagents were of analytical grade.

2.2. Reduction and alkylation of the peptides

The routine reduction and carboxyamidomethylation of the peptides was performed as per the following steps. One hundred micrograms of peptide was dissolved in 100 μl of buffer containing 0.1 mol/l Tris–HCl, 6 mol/l guanidinium HCl with pH 8.3. Dithiothreitol was added according to the molar ratio of 1:100 and allowed to stand under nitrogen gas at 37 °C for 30 min. Then, iodoacetamide was added according to the molar ratio of 1:500 and allowed to stand under nitrogen gas at 37 °C. After incubating for 4 h, the reaction mixture was diluted to 1 ml with double-distilled water and purified by reverse-phase column (Waters YWG-PAK C18, 3.9 mm × 250 mm) on the Waters (Milford, USA) Alliance HPLC system with a linear gradient of 5–45% acetonitrile containing 0.1% TFA over 60 min.

2.3. C-terminal ladder sequence analysis

C-terminal sequence analysis was performed by MALDI-TOF mass spectrometry coupled with carboxypeptidase Y digestion as described by Patterson et al. [10]. Five hundred picomoles of peptide sample and 0.16 units of CPY were suspended with 20 and 30 μl double-distilled water, respectively, and 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 of 20, 60 s, 2, 5, 10, 30 min, 2 and 16 h and immediately mixed with 9 μl of α-cyano-4-hydroxycinnamic acid saturated matrix solution in 50% ACN containing 0.1% TFA to quench the reaction. One microliter of each aliquot solution was placed onto individual wells of the MALDI sample plate and allowed to evaporate till dryness before insertion into the mass spectrometer.

2.4. Enzymatic cleavage and HPLC fractionation of the peptides

The peptide was cleaved by trypsin. One hundred micrograms of sample dissolved in 200 μl buffer of 0.1 mol/l ammonium bicarbonate with pH 8.0 was mixed with 2 μg of trypsin and incubated at 37 °C overnight. The resulting peptides were fractionated by Waters Nova-Pak C18 column (3.9 mm × 250 mm) with a linear gradient of 5–40% acetonitrile containing 0.1% TFA over 60 min on Waters Alliance HPLC system. The purified peptides were identified by MALDI-TOF mass spectrometry on Bruker (Billerica, USA) Proflex III mass spectrometer and automated Edman degradation sequencing on Applied Biosystems (Foster City, USA) 491 gas-phase sequenator with a standard program.

2.5. High-accuracy tandem mass spectrometric analysis

Tandem mass spectra for peptide sequencing were recorded on an APEX^{TM} 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 injected into the mass spectrometer at the rate of 30 μl/h. The MS/MS fragment ions were generated by introducing 3500 μs argon pulse with 1 mbar pressure and interpreted according to the nomenclatures proposed by Roepstorff and Fohlman [14], and Johnson et al. [15].

3. Results and discussion

3.1. C-terminal ladder sequencing

In our previous work, we purified two kinds of antifungal peptides named EAFP1 and EAFP2, which blocked the Edman degradation [11,12]. We gave up the way of trying the deblocking reagents and enzymes one by one considering the existence of over 10 kinds of N-terminal blockages. The C-terminal sequence analysis was selected as the first step for sequence determination of these two peptides, EAFP1 and EAFP2. The methods for C-terminal sequencing mainly include chemical methods and carboxypeptidase methods. The sequencing length of chemical methods is still restricted after many years of development [16]. Sequencing by carboxypeptidase digestion is based on the fact that the amino acid will be released one by one from the C-terminus of the peptide with CPY enzymatic reaction processing after the peptide was mixed with CPY. Traditionally, the sequence information is obtained by direct chromatographic analysis of the released amino acid, which is complicated by the contaminants and the difference of the releasing rate between residues. With the advancement of mass spectrometric techniques such as MALDI [8], Patterson et al. [10] presented the C-terminal ladder sequencing via MALDI

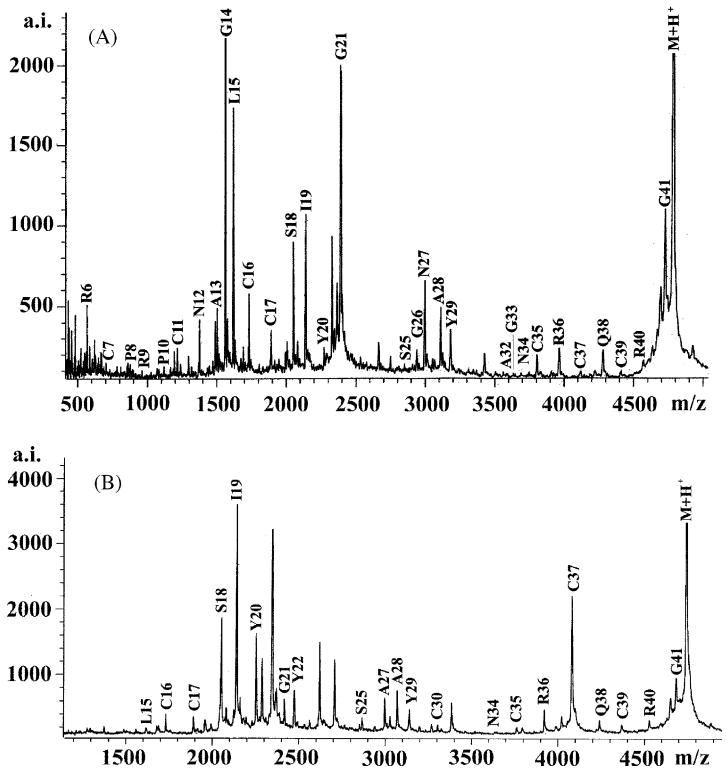

Fig. 1. MALDI-TOF mass spectrometric analysis of the CPY digesting mixtures of EAFP1 (A) and EAFP2 (B) with reaction overnight. The nomenclature of the peak labels denotes the peptide populations resulting from the loss of the indicated amino acids.

mass spectrometry coupled with carboxypeptidase Y digestion, which is shown to successfully obtain the sequence of 19 C-terminal residues. The masses of the peptides in CPY digestion mixture can be directly determined by MALDI-TOF mass spectrometry. The sequence can be read from the mass differences of adjacent peaks of the released peptides. Using this procedure, the C-terminal 36- and 30-residue sequences of alkylated EAFP1 and EAFP2 have been obtained, respectively. Fig. 1 showed the mass spectra of the reaction mixture with quenching time of about 16 h and the peaks representing the residue loss of C-terminus would disappear gradually by further prolonging the CPY digesting time. There were only few peaks in the mass spectra at shorter CPY digesting time, which showed the low CPY digesting rate of EAFPs. This low CPY digesting speed should be due to the existence in the sequences of relatively high amount of the glycyl residues which are hydrolyzed slowly. From Fig. 1, the C-terminal sequences of EAFP1 and EAFP2 can be read, which are R CPRPC NAGLC CSIYG □□gSG NAYcg AGNCR CQCRG and nagLC CSIYG Y□□Sg AAYC□ □□NCR CQCRG, respectively. Leucine and isoleucine in the sequences would be distinguished in the next Edman degradation analysis of the tryptic fragments since they have the same mass. The lower cases in the sequences represent the residues identified in the spectra of different quenching time. The sign "□" indicates the gaps unidentified in the mass spectra. This is mainly because the hydrolysis rates are varied according to the residual kinds at the C-terminus and penultimate position [10]. Glycine is released at such a low rate that the signal of the peptide population representing the loss of certain glycyl residue is relatively low or even disappears. The sequence analysis can frequently be processed until the detection of the truncated peptide peaks was impaired by the presence of CHCA matrix ions (<600 Da). This C-terminally sequencing method often provides longer C-terminal sequence with smaller consumption of sample than the traditional chemical methods [16]. The shortcoming is that CPY digestion still relies on the amino acid composition as well as their order along the sequence. Certainly, the correctness of the sequence should be still confirmed by further experiments.

3.2. The tryptic cleavage and fractionation of peptides

The possible hydrolysis sites in the peptide sequences can be known from the C-terminal sequencing results, which will help us to determine which kind of chemical reagents or enzymes should be selected to cleave the peptides. From the above C-terminal sequence results, there exist only the trypsin cleaving sites in the sequences of EAFP1 and EAFP2. So, trypsin had to be used to cleave the peptides. The resulting peptides were fractionated by reverse-phase HPLC. EAFP1 and EAFP2 have similar peptide map as shown in Fig. 2. The fractionated peptides were identified by MALDI-TOF mass spectrometry and their sequences

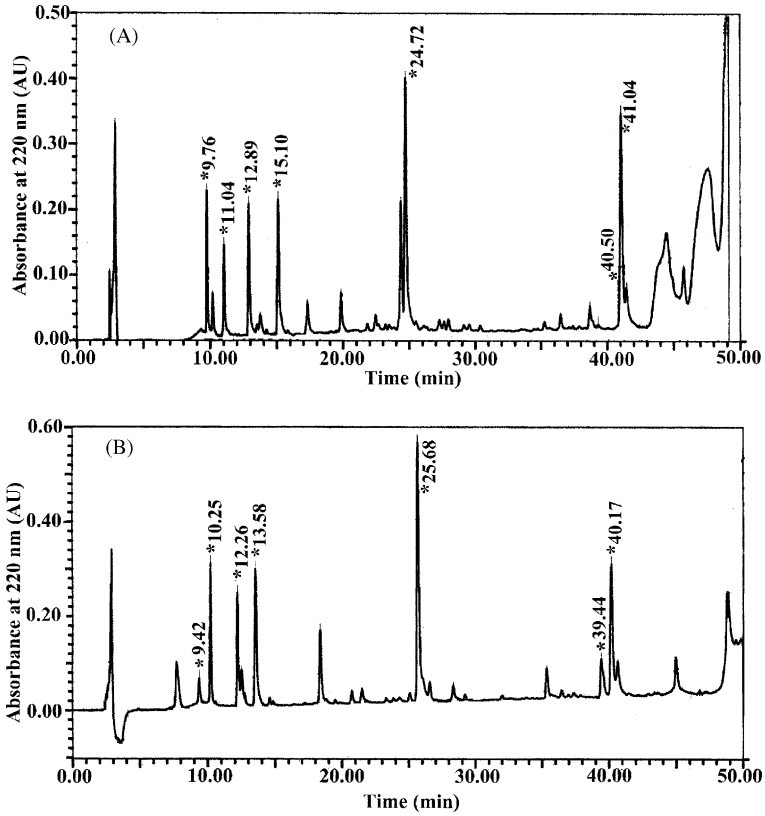

Fig. 2. The chromatograms of the tryptic digesting mixtures of carboxyamidomethyl EAFP1 (A) and EAFP2 (B). (∗) Denotes the analyzed peaks shown in Table 1 and the corresponding retention times.

were determined by automated Edman degradation. The results were summarized in Table 1. The sequences in Table 1 could be aligned with the C-terminal sequences determined above. The C-terminal 35-residue sequences of EAFP1 and EAFP2 were derived as R CPRPC NAGLC CSIYG YCGSG NAYCG AGNCR CQCRG and R CPRPC NAGLC CSIYG YCGSG AAYCG AGNCR CQCRG, respectively. There were no PTH signals observed when the peaks with retention time of about 12 min in the peptide maps of EAFP1 and EAFP2 (Fig. 2) were applied to Edman degradation. The molecular weights of the peptides contained in these two peaks are ca. 704 Da (Table 1), which is the difference of molecular weights between the above sequences of C-terminal 35-residue part and their corresponding intact

Table 1
The sequences and masses of the tryptic fragments of alkylated EAFP1 and EAFP2

EAFP1				EAFP2			
Retention time (min)	Sequence	Molecular weight		Retention Time (min)	Sequence	Molecular weight	
		Experimental $[M + H]^+$	Theoretical [M]			Experimental $[M + H]^+$	Theoretical [M]
9.76	CQCR	608[a]	622.7	9.42	CQCR	nd	622.7
11.04	CGAGNCR	796	794.3	10.25	CGAGNCR	792	794.3
12.89	nd[b]	705	704.3	12.26	<QTCASR[c]	705	704.3
15.10	CPRPC	692	689.3	13.58	CPRPC	689	689.3
24.72	GYCGSGNAY	947	949.4	25.68	GYCGSGAAY	904	905.4
40.50	CPRPCNAGLCCSIY	1727	1727.7	39.44	CPRPCNAGLCCSIY	1728	1727.7
41.04	NAGLCCSIY	1057	1057.4	40.17	NAGLCCSIY	1057	1057.4

<Q: the pyroglutamic residue; nd: not detected.
[a] The mass with loss of 17 mass units.
[b] The sequence is believed to be same as the N-terminal sequence of EAFP2 though not determined experimentally, since they have same mass and same retention time in the reverse-phase chromatogram.
[c] The sequence is determined by the tandem mass spectrometry (Fig. 2).

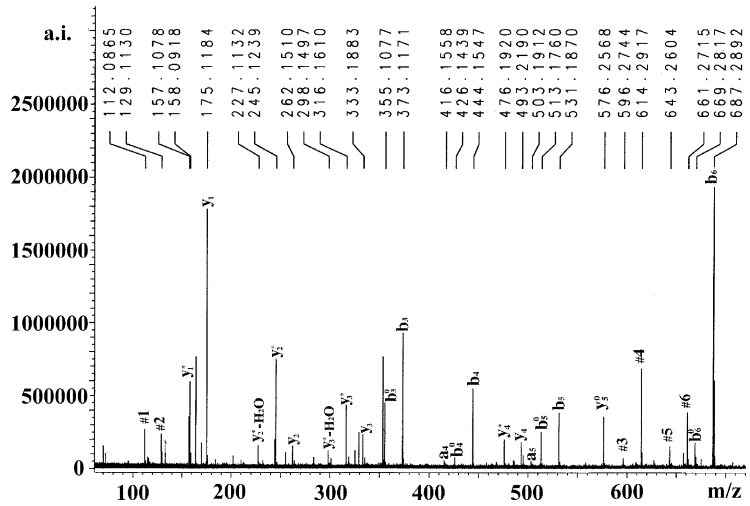

Fig. 3. Tandem mass spectrum of the 704 Da peptide derived from the EAFP2. The peaks of the fragment ions are labeled according to the nomenclature proposed by Roepstorff and Fohlman [14], and Johnson et al. [15]. (#) 661.2715 and 614.2917 correspond to the peaks of the parent ions with side-chain cleavage of threonine and cysteine, respectively; 643.2904 and 596.2744 correspond to the peaks of the dehydrate products of the above two fragments; 129.1130 represent the mass of the arginine-related ions, which formed by a combination of *a*-type and *y*-type cleavage and 112.0865 is the corresponding ion with loss of ammonia produced from the above ion with mass of 129.1130.

carboxyamidomethyl EAFPs. These results suggested that these 704 Da peptides should have the identical sequence and should be the fragments located in the N-termini of EAFPs. It is worth noting that there were more peaks in the reverse-phase chromatogram (Fig. 2) than those expected according to the number of the cleaving site of trypsin in the sequences, which is likely due to the impurity of trypsin.

3.3. The tandem mass analysis

The Fourier transform ion cyclotron resonance (FTICR) mass spectrometer was employed to determine the sequence of 704 Da peptide mainly because the instrument has high accuracy of less than 1–10 ppm and can perform the tandem mass analysis conveniently [17]. The high accuracy of the mass spectrometric data facilitates the identification of fragmentized ions and greatly simplifies the process of the spectrum interpretation because it is easy to find those ions which belong to the same series in the spectra according to the principle that the value of decimal part of the ion mass will decrease by less than 0.11 Da one by one in the same series of ions. This principle can be seen from the monoisotopic masses of the amino acid residues. The monoisotopic mass ($[M + H]^+$) of the parent ion was determined as 705.2997 Da. The *y*-serial ions including y_1, y_2, y_3, y_4, y_5^0 can be easily identified in the tandem mass spectrum (Fig. 3). y_5^0 is product formed by dehydration of y_5 ion. The interpretations of the ions referred to the nomenclature described by Roepstorff and Fohlman [14], and Johnson et al. [15]. The sequence were derived as PyroGlu–Thr–Cys–Ala–Ser–Arg, which is N-terminally blocked by pyroglutamic group.

From above results, the sequences of EAFP1 and EAFP2 can be completely deduced as shown in Fig. 4, which are in good agreement with the results of amino acid analysis as shown in Table 2. The pyroglutamic residue is derived to glutamic acid when the peptide is hydrolyzed with hydrochloride acid. Their average molecular masses calculated from the sequence are 4211.80 and 4168.77 Da for EAFP1 and EAFP2, respectively, which are higher (about 10 Da) than the experimental value (data not shown) and can be explained by five pairs of disulfide bonds formed by 10 cysteines of EAFPs.

The results demonstrated that this strategy is fast and efficient method for the sequence determination of the N-terminally blocked peptide with simplicity and low consumption of sample. C-terminal ladder sequencing by

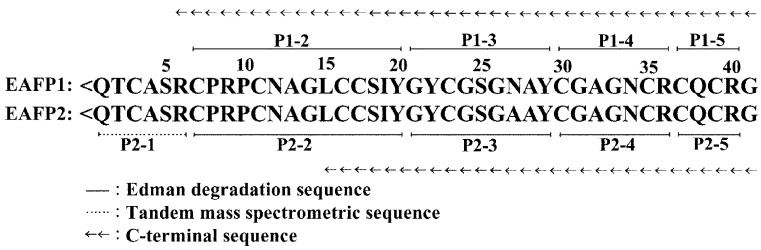

Fig. 4. The proof of the amino acid sequences assembling of EAFP1 and EAFP2. <Q: pyroglutamic residue.

Table 2
The amino acid composition of native EAFP1 and EAFP2

Amino acid	EAFP1		EAFP2	
	Experimental	Theoretical	Experimental	Theoretical
D/N	3.32	3	2.49	2
E/Q	1.98	2	2.11	2
S	2.67	3	2.63	3
G	7.15	7	6.96	7
H	0	0	0	0
R	3.77	4	3.81	4
T	0.98	1	1.00	1
A	3.03	4	4.39	5
P	2.05	2	2.11	2
Y	3.45	3	3.59	3
V	0	0	0	0
M	0	0	0	0
C	nd	10	nd	10
I	1.31	1	1.37	1
L	1.55	1	1.67	1
F	0.60	0	0.51	0
K	0	0	0	0
Total		41		41

nd: no data.

coupling the MALDI mass spectrometry with carboxypeptidase Y digestion can provide enough information to direct the enzymatic cleavage of the peptide and the assembling of the resulting fragmental sequences. The N-terminal blocked fragment can be easily identified by comparing the mass difference between the assembled sequence of C-terminal part and intact sequence and can be sequenced directly by tandem mass spectrometry. The strategy presented in this report is independent of the use of specific chemical reagent or enzyme and should have broad application in the amino acid sequence determination of the N-terminally blocked peptides.

Acknowledgements

This work was supported by State Basic Research Program of China (G19990756) and the Natural Science Foundation of China (Grant 30070162).

References

[1] D. Wellner, C. Panneerselvam, B.L. Horecker, Proc. Natl. Acad. Sci. U.S.A. 87 (1990) 1947.
[2] J.L. Brown, W.K. Roberts, J. Biol. Chem. 251 (1976) 1009.
[3] N. Miyatake, M. Kamo, K. Satake, Y. Uchiyama, A. Tsugita, Eur. J. Biochem. 212 (1993) 785.
[4] T. Bergman, M.T. Gheorghe, L. Hjelmqvist, H. Jornvall, FEBS Lett. 3909 (1996) 199.
[5] D.F. Hunt, J.R. Yates, J. Shabanowitz, S. Winston, C.R. Hauer, Proc. Natl. Acad. Sci. U.S.A. 83 (1986) 6233.
[6] M. Wilm, A. Shevchenko, T. Houthaeve, S. Breit, L. Schweigerer, T. Fotsis, Nature 3799 (1996) 466.
[7] K. Biemann, H.A. Scoble, Science 237 (1987) 992.
[8] J.T. Stults, Curr. Opin. Struct. Biol. 5 (1995) 691.
[9] J.B. Fenn, M. Mann, C.K. Meng, S.F. Wong, C.M. Whitehouse, Science 246 (1989) 64.
[10] D.H. Patterson, G.E. Tarr, F.E. Regnier, S.A. Martin, Anal. Chem. 67 (1995) 3971.
[11] X.Z. Liu, Z. Hu, Y. Li, J.B. Yang, B.J. Li, Acta Bot. Yunnanica 16 (1994) 385.
[12] R.H. Huang, Y. Xiang, X.Z. Liu, Y. Zhang, Z. Hu, D.C. Wang, FEBS Lett. 521 (2002) 87.
[13] Y. Xiang, R.H. Huang, W. Liu, G.P. Li, X.Z. Liu, D.C. Wang, Acta Crystallogr., Sect. D: Biol. Crystallogr. 58 (2002) 1838.
[14] P. Roepstorff, J. Fohlman, Biomed. Mass Spectrom. 11 (1984) 601.
[15] R.S. Johnson, S.A. Martin, K. Biemann, J.T. Stults, J.T. Watson, Anal. Chem. 59 (1987) 2621.
[16] J. Li, S. Liang, Anal. Biochem. 302 (2002) 108.
[17] C.L. Hendrickson, M.R. Emmett, Annu. Rev. Phys. Chem. 50 (1999) 517.