ELSEVIER

Journal of Chromatography A, 855 (1999) 695–707

JOURNAL OF CHROMATOGRAPHY A

www.elsevier.com/locate/chroma

Analysis of recombinant and modified proteins by capillary zone electrophoresis coupled with electrospray ionization tandem mass spectrometry

Tao Liu, Xiao-Xia Shao, Rong Zeng, Qi-Chang Xia*

Shanghai Institute of Biochemistry, Chinese Academy of Sciences, Shanghai 200031, China

Received 30 March 1999; received in revised form 2 June 1999; accepted 3 June 1999

Abstract

A method for rapid characterization of recombinant and modified proteins with known sequences is described. The analytical system consists of a capillary zone electrophoresis (CZE) instrument coupled to an electrospray ionization ion trap tandem mass spectrometer via a sheath-flow interface. Following the procedure consists of proteolytic fragmentation, CZE peptide separation, tandem mass spectrometry (MS–MS) analysis of separated peptides, sequence database search and monitoring of the specific peptides, C 125 S mutated interleukin 2 (S-125-IL2) and bovine β-casein were characterized as a model of recombinant protein and naturally modified protein, respectively. A tryptic peptide mixture derived from the synthetic salmon calcitonin (s-CT) was also analyzed to test the performance of the system. Although a conventional sheath-flow interface with much higher flow-rate compared to the microspray interface and nanospray interface was used, the proteins were identified at the low picomole level.

Keywords: Proteins; Interleukins; Caseins; Peptides

1. Introduction

Obtaining the structural information of the biologically important molecules, such as peptides, proteins and DNAs is one of the fundamental and significant steps for achieving a comprehensive understanding of the function and regulation of the complicated biological systems. Large scare sequencing of genomic DNA has undergone significant developments both in knowledge and technology, which has been generating a wealth of information of sequence and regulation for biologists. Knowledge of the large numbers of cDNAs and expressed sequence tags (ESTs) is propelling the investigation of the biological system to the proteome level [1]. The increase in the number and size of protein sequence databases and the extensive developments of mass spectrometry are largely improving the efficiency and velocity of the study in protein sciences. Rapid and conclusive protein identification can be performed by correlating the information contained in the amino acid sequence with the objects in a sequence database.

Mass spectrometry has made great progress in its biological application by the invention of the soft ionization techniques in the last decade. Two of the

*Corresponding author. Tel.: +86-21-6437-4430; fax: +86-21-6433-8357.

E-mail address: xiaqc@sunm.shcnc.ac.cn (Q.-C. Xia)

0021-9673/99/$ – see front matter
PII: S0021-9673(99)00717-7

most commonly used ionization techniques are matrix-assisted laser desorption ionization (MALDI) [2,3] and electrospray ionization (ESI) [3,4]. With the development of the reflectron-type instrument [5–7] and of the post decay technique [8,9], MALDI–MS is becoming outstanding from the other MS in its simplicity, high m/z range and the compatibility with the commonly used buffers in biochemical studies. Unfortunately, it is difficult to couple the ion source on-line with the high performance separation techniques to further expand its applications. In contrast, ESI–MS becomes the most acceptable method for the analysis of peptide mixtures due to the convenience of interfacing the ion source to the high performance separation techniques such as liquid chromatography (LC) and capillary electrophoresis (CE) [10,11]. Triple quadrupole and ion trap mass spectrometers are usually used with this kind of ion source; their advantages and limitations were compared by performing the protein identification [12]. Fragment ion spectra produced by collision-induced dissociation (CID) can be acquired using the tandem mass spectrometer equipped with the ion sources mentioned above, and they were always chosen as the preferred parameter for the sequence database search [13]. The quality of the MS–MS spectra is of significance in improving the quality of the match between the experimental CID spectrum and the predicted CID spectra of any isobaric peptide from the sequence database. From this point of view, the ion trap mass spectrometer is the preferred instrument for the MS–MS experiment.

The technique of coupling capillary electrophoresis to mass spectrometry (CE–MS) has combined the ultrahigh separation efficiency and sensitivity of CE and the ability to obtain the structural information of the analytes of MS in one system [14–16]. Several types of CE–MS interfaces have been developed for the better performance of the coupled analytical system. Liquid-junction and coaxial interfaces were described and compared with regards to ruggedness, ease of use, sensitivity and electrophoretic performance [17]. Some methods for on-line sample preconcentration such as on-line capillary isotachophoresis [18,19] and solid-phase extraction [20,21] were introduced to further improve the sensitivity of the CE–MS system. Sensitivity at low femtomole or subfemtomole level can be achieved using these sample preconcentration techniques in combination with the microspray or nanospray [22] technique. Especially, the sensitivity was further improved to the attomole level when the nanoelectrospray ion source was coupled to the Fourier transform mass spectrometer [23].

Protein modification takes place frequently in many biological events, so methods for identifying those modifications are of great importance. Attempts have been undertaken to develop manual and automatic strategies to use MS–MS for identifying the modifications introduced either artificially or naturally [24–26]. In this study, a method for rapid characterization of recombinant and modified proteins with known sequences using capillary zone electrophoresis-tandem mass spectrometry (CZE–MS–MS) is described. The proteolytic fragments were separated by CZE and were detected and dissociated in the MS–MS analysis. Sequence database search using the resulted CID spectra and monitoring of the specific peptides by extracted ion chromatogram (EIC) could quickly identify the protein. The C 125 S mutated interleukin 2 (S-125-IL2) and bovine β-casein were used as a model of recombinant protein and naturally modified protein, respectively.

2. Experimental

2.1. Chemicals

All chemicals used were of analytical grade if not mentioned otherwise. Acetic acid, ammonium hydroxide, ammonium hydrogencarbonate, formic acid, bovine β-casein were purchased from Sigma (St. Louis, MO, USA). Ammonium acetate was from Shanghai Reagent Manufacture (Shanghai, China). N-Tosyl-L-phenylalanine chloromethyl ketone (TPCK)-inhibited trypsin was made in our laboratory. C 125 S mutated human interleukin 2 was produced genetically by the Institute of Medical Industry (Shanghai, China). Synthetic salmon calcitonin was generously provided by Professor D.F. Cui in our institute. Fused-silica capillary with a cartridge was from Beckman (Fullerton, CA, USA). The water was distilled and deionized (18 MΩ)

using a Milli-Q system from Millipore (Bedford, MA, USA).

2.2. Tryptic digestion

100 μg protein was dissolved in 30 μl ammonium hydrogencarbonate buffer (100 mmol/l, pH 8.5). Salmon calcitonin and S-125-IL2 were firstly reduced with dithiotreitol (5 mmol/l) in ammonium hydrogencarbonate buffer at 50°C for 15 min, followed by alkylation with iodoacetamide (5 mmol/l) in the same buffer at room temperature for 15 min in the dark. After dilution to 100 μl, a 2.5 μl TPCK-inhibited trypsin (1 μg/μl) was added to the mixture and incubated at 37°C for 4 h, then another 2.5 μl enzyme solution was added for overnight digestion at 37°C. β-casein has no disulfide bridges, so 100 μg protein was directly digested overnight with modified trypsin at an enzyme-substrate ratio of 1:20 (w/w). The resulting digestion mixtures were kept at −20°C without further treatment.

2.3. CZE–MS

CZE was performed in an 80 cm×50 μm I.D. fused-silica capillary on the Beckman P/ACE System 5500. The electrospray end of the capillary was inserted into a stainless steel tube mounted towards the orifice of the mass spectrometer via a sheath-flow adapter (Finnigan MAT, San Jose, CA, USA). The polyimine coating was burned off 2–3 mm from this end and washed off with ethanol to ensure the electric contact. The other end of the capillary (injection end) was kept in the buffer reservoir. In order to minimize any influence of gravity flow, the whole capillary electrophoresis apparatus was placed on a height-adjustable platform and care was taken to ensure that the CZE buffer reservoir and the electrospray end of the capillary were at the same height. One percent acetic acid (pH 2.3) was used as the running buffer for the separation of tryptic digestion mixtures. A negative high-voltage (normally 15 kV) was applied to the injection end, and positive high-voltage of 4.25 kV compared to the MS capillary entrance was applied to the electrospray end. Sheath-liquid (ethanol–water–formic acid, 60:39:1) was supplied by a Harvard Apparatus 22 syringe pump (Harvard Apparatus, South Natick, MA, USA) at a flow rate of 2 μl/min, through which the electric contact between the electrospray end of the capillary and the stainless steel tube was established. To ensure the stable electrospray under these conditions, care was taken to adjust the extruded length (0.5–1 mm) of the tip of the electrospray end from the liquid sheath tube by a micrometer head. The silanol groups on the capillary walls were activated by passing a series of solutions through in sequence: hydrochloric acid (1 mol/l) for 15 min, water for 15 min, sodium hydroxide (1 mol/l) for 15 min and water for 15 min. Before each run, the capillary was regenerated with 0.1 mol/l sodium hydroxide for 5 min, then washed with water for 5 min and the running buffer for another 5 min.

Mass spectra were acquired by a model LCQ ion trap mass spectrometer (Finnigan MAT) equipped with a Finnigan MAT electrospray ion source. The LCQ was run with automatic gain control (AGC) for all experiments unless mentioned otherwise. In this mode, the full MS target number of ions was set to $5 \cdot 10^7$, in zoom scan it was set to $1 \cdot 10^5$, and in MS–MS mode to $1 \cdot 10^5$. In full scan mode, the mass analyzer was scanned from 400 to 2000 u.[1] The numbers of microscans in each mode were: 3, 3 and 3 for full MS, zoom scan and MS–MS, respectively. According to these settings, the most intense ion in the spectrum was selected to perform a zoom scan and followed by a CID scan with an isolation width of 3 u after scanning each range in full scan mode. The relative collision energy was set to 45 to optimize the fragmentation of the peptides. Other operating parameters were used as follows: heated capillary temperature, 180°C; capillary voltage, 13 V; tube lens offset, 55 V.

2.4. Sequence database search

The resulting CID spectra were correlated with the sequence database using the SEQUEST database

[1]The masses of all peptide fragments were calculated from the equation: $M = (m/z)n - n$, according to the definition $m/z = [M + nH]^{n+}$. m/z is the mass to charge ratio recorded by the mass spectrometer; n is the charge states which can be judged from the zoom scan spectra of corresponding fragment ions; H is the mass of a hydrogen atom whose value is 1.008; and M is the mass of the peptide fragment.

searching program [21]. The human 072997 sequence database and the OWL nonredundant composite protein sequence database were searched to identify S-125-IL2 and β-casein, respectively.

2.5. Specific peptide monitoring

To further identify the point mutation in S-125-IL2 and the natural phosphorylation in β-casein, specific peptides (e.g. the C-terminal peptide of S-125-IL2 and the phosphoserine containing peptides) were found out by the extracted ion chromatogram (EIC) and their sequences were interpreted from the MS–MS spectra.

3. Results and discussion

Using the coupled separation and detection system, peptide mixtures derived from proteolytic digestion could be analyzed conveniently. All tryptic peptide mixtures were directly applied to the CZE–MS–MS system without further treatment. They were injected into the capillary by applying a positive pressure of 20 p.s.i. (1 p.s.i.=6.89 kPa) at the injection end. Then the peptides were separated in the acetic acid buffer by CZE and ionized at the electrospray end successively according to their respective migration times. After entering the heated capillary, the positively charged ions were detected by the mass spectrometer. Under the AGC state, the zoom scan mode and the MS–MS mode were triggered automatically when the ion intensity in each scan exceeded the predetermined threshold and produced the CID spectra of the most intense ions in that scan. All the resulting uninterpreted CID spectra were used to search the sequence database for protein identification. The migration times of the specific peptides could be revealed by the EICs, and the sequence information could be derived from the corresponding CID spectra. The results in this study demonstrated the convenience and rapidity of this method for analyzing the mutant and modified proteins with their sequences already known.

3.1. Analysis of salmon calcitonin digest

A tryptic digest mixture of s-CT was applied to the CZE–MS–MS system initially to demonstrate its performance for characterizing proteins. Using 1% acetic acid buffer (pH 3.4) as the running buffer, the peptides were separated in a 50 μm I.D. capillary by applying a negative high-voltage of 25 kV to the injection end. Without any further treatment, the tryptic digest was directly injected at 20 p.s.i. for 5 s (total sample amount applied, 5.2 pmol of s-CT). The tryptic cleavage sites on the s-CT sequence are indicated as follows: C*SNLSTC*VLGK$^{\downarrow}$LSQELHK$^{\downarrow}$LQTYPR$^{\downarrow}$TNTGS-GTP-$_{NH_2}$ (C* indicates the cysteine has been alkylated by iodoacetamide). T1, T2, T3, and T4 represented the resulting peptides according to their positions in the sequence from N-terminus to C-terminus.

Fig. 1 shows the base peak (representing the most abundant ion in a peak) of the separation of the salmon calcitonin tryptic digest consisting of 4 peptides ranging in length from 6 to 11 amino acids by CZE. The results indicated that the 4 peptides were separated well without peak broadening and peak tailing. Under these conditions, singly charged ions of the 4 peptides predominated in the mass spectra were selected automatically for CID. The sequence information derived from the CID spectra indicated that the peptide ions were well fragmented by setting the relative collision energy at 45. So the transfer of the conditions used in this experiment to the protein characterization experiments was straightforward. Another running buffer of 15 mmol/l ammonium acetate contained 15% (v/v) methanol (pH 2.5) was also tried to optimize the separation. The results showed significant peak broadening compared to the results obtained in the 1% acetic acid (pH 3.4) buffer (data not shown). In addition, the peptides were eluted from the capillary in a sequence of T2, T3, T1 and T4 using this buffer system, with T1 and T4 exchanged their positions compared to the result obtained by using the 1% acetic acid buffer system. This might due to the buffer pH shift, which led to the change of the charging state of the analytes.

3.2. Analysis of S-125-IL2 digest

Conditions used in the CZE–MS–MS analysis of

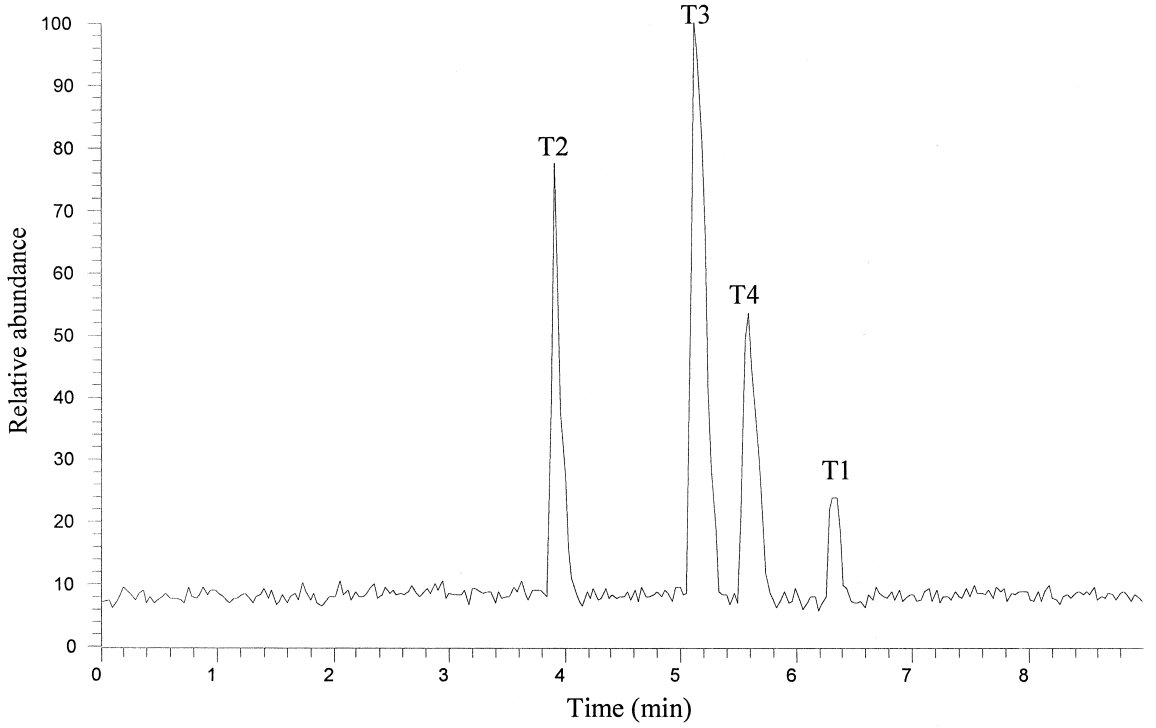

Fig. 1. CZE–MS analysis of the 5.2 pmol tryptic digest of synthetic salmon calcitonin on an 80 cm×50 μm I.D. fused-capillary with −30 kV at the injection end and +4.25 kV at the electrospray end. CZE was carried out in 1% acetic acid (pH 3.4) buffer.

the S-125-IL2 tryptic digest were much the same as those used before. After applying the sample at 20 p.s.i. for 20 s, a total amount of 4.6 pmol of S-125-IL2 tryptic digest was injected into the capillary. Optimized separation was carried out in the 1% acetic acid buffer (pH 2.3) by applying a negative high-voltage of 15 kV to the injection end of the capillary. Higher pH value or separation voltage resulted in partial separation of the tryptic peptide mixture. All peptides in the mixture eluted in 40 min (Fig. 2).

Table 1 summarizes the results of the CZE–MS–MS analysis. From the separation pattern shown in Fig. 2, it can be seen that all theoretical tryptic peptides were recovered except one small peptide (LTR, 388.5 u, 36–38). It is because the trap was set to scan from 400 to 2000 u in order to decrease the influence of some strong noises in the low m/z region. Although T3 (MLTFK) and T4 (FYMPK) have the similar mass/charge ratio and therefore coeluted in the same peak, rapid scanning of the ion trap mass spectrometer in both the MS and MS–MS mode still allowed the detection and acquisition of CID spectra for them, respectively.

A total of 24 different CID spectra were generated during the experiment. All uninterpreted CID spectra were used to identify the protein by screening the human 072997 sequence database using the SEQUEST program. Tables 2 and 3 summarize the database search results. Four of the CID spectra resulted in matches with the IL2 sequence (Table 2). Their correlation coefficients had relative high values (>1.5), which indicated the high quality of the match between the experimental CID spectrum and the predicted CID spectra of any isobaric peptide from the sequence database. The delta correlation factors higher than 0.1 indicate the good distinction between the top and the next best match [21]. Closer inspection of the results discovered that the N-terminal tryptic peptides T1 (MAPTSSSTK) and MAPTSSSTKK did not resulted in matches with the IL2 sequence, but resulted in matches with the

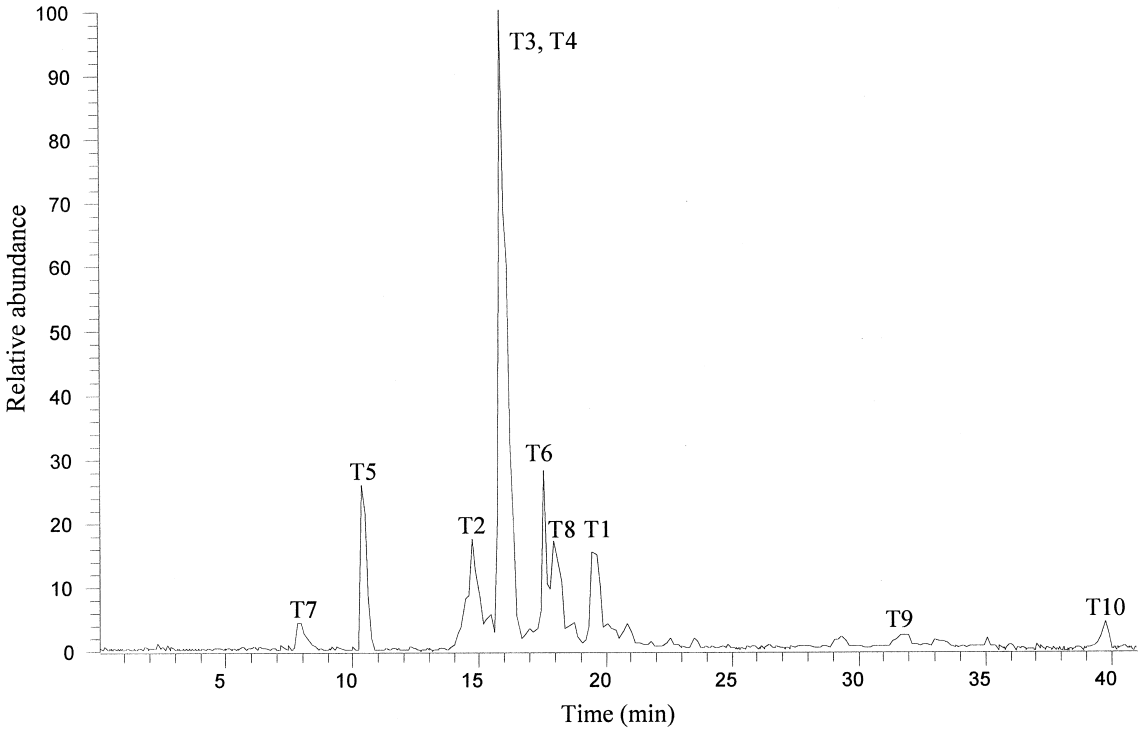

Fig. 2. Separation of an S-125-IL2 tryptic digest on a 80 cm×50 μm I.D. fused-capillary with −15 kV at the injection end and +4.25 kV at the electrospray end. CZE was performed in 1% HAc (pH 2.3) buffer. A total amount of 4.6 pmol sample was injected into the capillary by applying 20 p.s.i. at the injection end for 20 s. Each peak that correlates as resulting from S-125-IL2 tryptic digest (Table 1) was labeled with Tn.

Table 1
Tryptic peptides of S-125-IL2 analyzed by CZE–MS

Tryptic peptide	Position	Sequence	Migration time (min)	Calculated mass	Measured mass
T1	−1–8	MAPTSSSTK	19.48	909.0	908.5
T2	9–35	KTQLQLEHLLLDLQ–MILNGINNYKNPK	14.71	3192.7	3192.4
	36–38	LTR		388.5	
T3	39–43	MLTFK	15.88	638.8	638.4
T4	44–48	FYMPK	16.37	684.8	684.5
T5	49–54	KATELK	10.37	688.8	688.5
T6	55–76	HLQC*[a]LEEELKPLEE–VLNLAQSK	17.55	2621.0	2621.4
T7	77–83	NFHLRPR	7.97	939.1	939.2
T8	84–97	DLISNINVIVLELK	17.97	1582.9	1581.8
T9	98–120	GSETTFMC*[a]EYADET–ATIVEFLNR	31.97	2684.9	2684.2
T10	121–133	WITFS*[b]QSIISTLT	39.73	1496.7	1496.8

[a] C* indicates that the cysteine has been treated with iodoacetamide.
[b] S* indicates that the 125 cysteine has been replaced by serine.

Table 2
Results of the search of the human 072997 sequence database with CID spectra of tryptic peptides derived from S-125-IL2

Mass (MH^{1+})[a]	Calculated mass	Correlation coefficient	Δ correlation factor	Sequence	Position
639.5	638.8	1.4927	0.394	MLTFK	39–43
685.5	684.8	1.6293	0.129	FYMPK	44–48
689.5	688.8	1.6207	0.246	KATELK	49–54
940.2	939.1	2.2338	0.224	NFHLRPR	77–83

[a] All the masses are presented in their MH^{1+}. The sequence position of the peptides within the S-125-IL2 sequence (134 amino acids) is indicated.

Table 3
Cumulative score and ranking of IL2

Protein	Total score	No. 1	No. 2	No. 3	No. 4	No. 5
IL2_HUMAN	40	4	0	0	0	0
TOP1_HUMAN	20	1	0	1	1	0
A21785	20	2	0	0	0	0
LMB1_HUMAN	18	1	1	0	0	0
ARDH_HUMAN	18	1	1	0	0	0

A21785 sequence instead, because the S-125-IL2 sample has an additional methionine at its N-terminus. Table 3 showed that the top-ranked IL2 had a cumulative score of 40 and correlated with 4 top-ranked individual CID spectra, while 2 other top-ranked individual CID spectra were correlated with A21785 which has partial sequence of IL2 (residues −1–61). So actually there were 6 individual CID spectra correlated with IL2 and were ranked in the

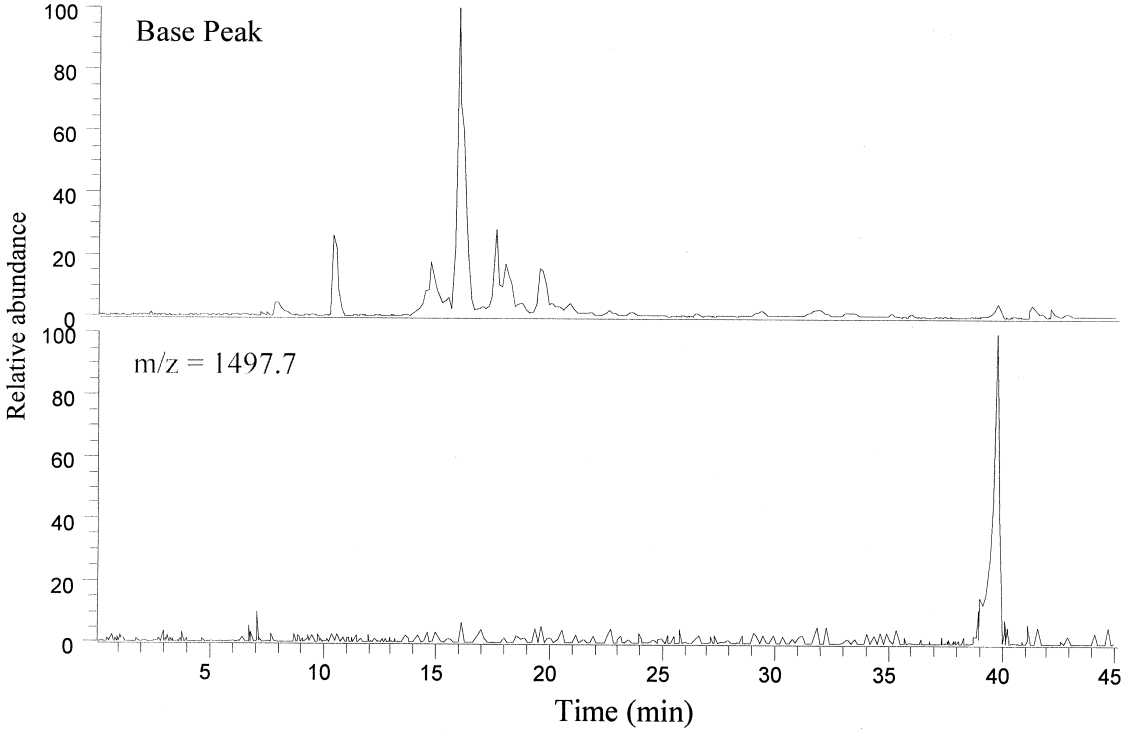

Fig. 3. CZE–MS analysis of a peptide mixture derived from tryptic digestion of S-125-IL2. Conditions as in Fig. 2. Top panel shows the base peak versus time, bottom panel is the EIC of the ion with $m/z = 1497.7$.

first position, which indicated a highly significant identification of IL2. The same data were used to search the OWL database (data not shown). IL2 still ranked in the first position and correlated with 4 top matches even in the much bigger database. T6 and T7 peptides were not correlated to IL2 using SEQUEST program because their cysteine residues had been alkylated by iodoacetamide and their mass were shifted by 57 u. Other tryptic peptides were identified by their mass and the corresponding CID spectra.

In order to confirm the point mutation at position 125 of the IL2 amino acid sequence, the C-terminal peptide T10 that contains the mutation site was monitored by the EIC (Fig. 3). T10 (WITFS*QSIISTLT) has a theoretical mass of 1496.7, so the ion chromatogram of its singly charged precursor ion with $m/z=1497.7$ was reconstructed to find out its migration time. The result displayed in Fig. 3 clearly revealed that T10 eluted at 39.7 min. The amino acid sequence information of this peptide was obtained by interpreting the corresponding CID spectrum (Fig. 4). The MS–MS spectrum of T10 was characterized by the successive dehydration of the b ions (the N-terminal ions from a peptide bond breakage) due to its high serine and threonine content, which was confirmed by several repetitive experiments. Sometimes the peptide bond broke at the N-terminus of the 125-serine residue and produced the peptide S*QSIISTLT, whose CID spectrum was also characterized by the successive dehydration of the b ions.

3.3. Analysis of β-casein digest

Natural modification presented widely in some proteins, either in the post-translational processing or in the metabolic progress, and influenced their biological functions to a large extent. Phosphorylation and glycosylation were two of the most common types of protein modification. A tryptic digest of β-casein from bovine milk was used as a model peptide mixture to characterize the modified proteins with known sequence.

A total amount of 1.5 pmol β-casein tryptic digest was injected by applying a pressure of 20 p.s.i. for

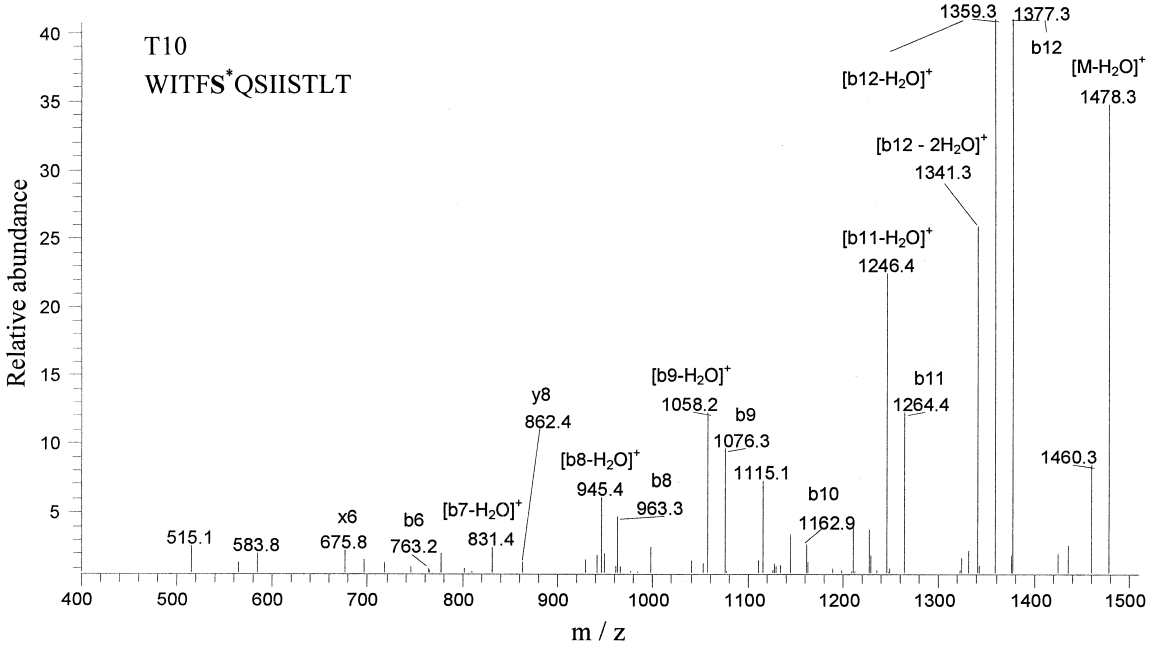

Fig. 4. CID spectrum recorded on the $[M+H]^+$ ion at m/z 1497.7 of the mutation site-containing C-terminal peptide derived from trypsin digestion of the mutated protein S-125-IL2. The b ions are the N-terminal ions from a peptide bond breakage, the y ions are the C-terminal ions from a peptide bond breakage.

10 s to the injection end. CZE separation was performed using 1% acetic acid buffer (pH 2.3) as the running buffer, the negative high-voltage applied to the injection end was 15 kV. Fig. 5 displayed the base peak of the separation of the tryptic peptide mixture. All components were resolved and eluted within 30 min. T5, T6, T7 and T8 had the similar mass/charge ratio, they eluted one by one with small separation distance apart. T10, T11, T12 and T13 peptides eluted in the same way due to the homogeneity in their sequences. The results summarized in Table 4 indicate that 16 of the 22 peptides derived from the tryptic digest were detected by CZE–MS and their respective spectra were acquired by CZE–MS–MS. Also, because the mass spectrometer was set to scan from 400 to 2000 u, several small peptides with a mass below 400 u did not be recorded by MS. Under these conditions, the peptides localized at 49–97 and 114–169 in the protein sequence with masses of 5319.3 and 6362.4 were also not detected by MS. Similar to the results obtained by acquiring a parent ion scan of the tryptic digest of β-casein, the predominant peak of phosphorylated peptide corresponded to the doubly charged peptide T2 (FQS^{P}EEQQQTEDELQDK, S^{P} corresponds to the phosphorylated serine residue) with a m/z of 1031.5, whereas the other known phosphorylated peptide (R^{1}ELEELNVPGEIVES^{P}LS^{P}S^{P}S^{P}EESITR^{25}) is not easily discerned from the background [27]. In our experiment, another longer peptide T3 (F^{33}QS^{P}EEQQQTEDELQDKIHPFAQTQSLVYPFPGPIPN^{68}) which contained the whole T2 peptide sequence was also detected by CZE–MS.

The doubly charged phosphorylated peptide T2 can be found conveniently in the mass spectra using the EIC of the ion with $m/z=1031.5$ (Fig. 6). It can be seen easily that the phosphorylated peptide eluted at 29.1 min. Under certain MS–MS conditions, the CID spectrum of the phosphorylated peptide was

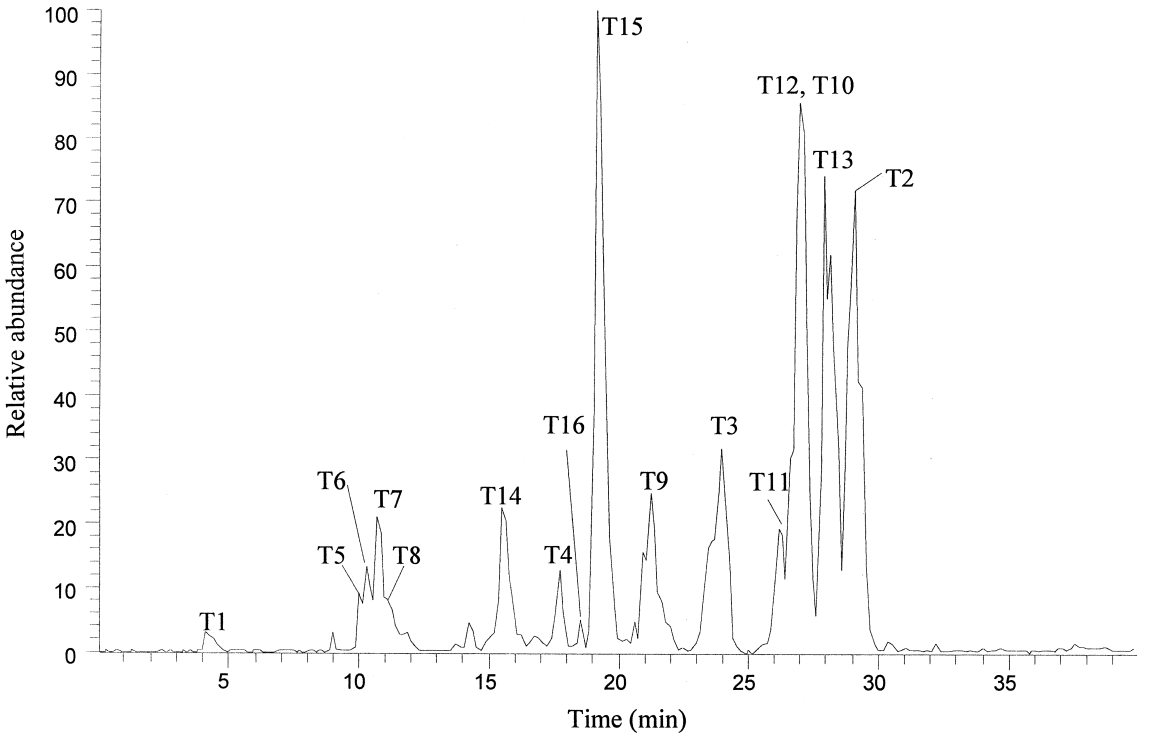

Fig. 5. Separation of a β-casein tryptic digest on an 80 cm×50 μm I.D. fused-capillary with −15 kV at the injection end and +4.25 kV at the electrospray end. CZE was performed in 1% HAc (pH 2.3) buffer. A total amount of 1.5 pmol was injected into the capillary. Each peak that correlates as resulting from β-casein tryptic digest (Table 4) was labeled with Tn.

Table 4
Tryptic peptides of β-casein analyzed by CZE–MS

Tryptic peptide	Position	Sequence	Migration time (min)	Calculated mass	Measured mass
	1–25	RELEELNVPGEIVES^{P}LS^{P}–S^{P}S^{P}EESITR[a]		3119.1	
	26–28	INK		373.4	
T1	29–32	KIEK	4.26	516.6	516.4
T2	33–48	FQS^{P}EEQQQTEDELQDK	29.10	2061.0	2061.5
T3	33–68	FQS^{P}EEQQQTEDELQDKIH–PFAQTQSLVYPFPGPIPN	23.99	4266.6	4267.1
T4	49–68	IHPFAQTQSLVYPFPGPIPN	17.73	2223.6	2223.4
	49–97	IHPFAQTQSLVYPFPGPIP–NSLPQNIPPLTQTPVVVPP–FLQPEVMGVSK		5319.3	
	98–99	VK		245.3	
T5	100–105	EAMAPK	10.27	645.8	645.5
	106–107	HK		283.3	
T6	108–113	EMPFPK	10.70	747.9	747.5
	114–169	YPVEPFTESQSLTLTDVEN–LHLPLPLLQSWMHQPHQP–LPPTVMFPPQSVLSLSQSK		6362.4	
T7	170–176	VLPVPQK	10.97	780.0	779.6
T8	177–183	AVPYPQR	11.26	830.0	829.7
T9	184–202	DMPIQAFLLYQEPVLGPVR	21.23	2186.6	2185.8
T10	184–190	DMPIQAF	26.98	821.0	820.3
T11	184–191	DMPIQAFL	26.19	934.1	933.3
T12	184–192	DMPIQAFLL	26.60	1047.3	1046.5
T13	184–193	DMPIQAFLLY	27.92	1210.5	1209.5
T14	191–202	LLYQEPVLGPVR	15.52	1383.7	1382.8
T15	203–209	GPFPIIV	19.18	741.9	741.5
T16	205–209	FPIIV	18.56	587.8	587.5

[a] S^{P}: phosphoserine.

predominated by the signal of the ion with m/z= 982.7 u, which represented the $[M-H_2O-PO_3]^{2+}$ ion of the peptide of mass 2061.5 (Fig. 7). In addition, an intact series of y ions with the similar intensity can clearly be discerned from the CID spectrum.

All the 29 different uninterpreted CID spectra resulted during the experiment were correlated with the OWL protein sequence database. Tables 5 and 6 summarize the results of the database search by SEQUEST program. Five peptide ions with masses of m/z=645.5, 747.5, 779.6, 741.5 and 2186.8 u, respectively, resulted in matches with the β-casein sequence with high correlation coefficients and delta correlation factors even in the nonredundant composite sequence database (Table 5). The match of the peptide ion with m/z=748.5 u had a Δ correlation factor lower than 0.1, which suggested that peptides different from the one matched to bovine β-casein were matched with a similar correlation coefficient. Especially, the CID spectrum obtained for the $[M+2H]^{2+}$ ion of the peptide of mass 2186.8 has the highest correlation coefficient of 5.6. It indicated that this peptide was well fragmented under the MS–MS conditions, which resulted in the high quality of the match between the experimental CID spectrum and the predicted CID spectrum of this peptide from the sequence database. The cumulative score for the identification of β-casein is shown in Table 6. β-casein scored in the first rank with five CID spectra matched most strongly to it whereas no other protein

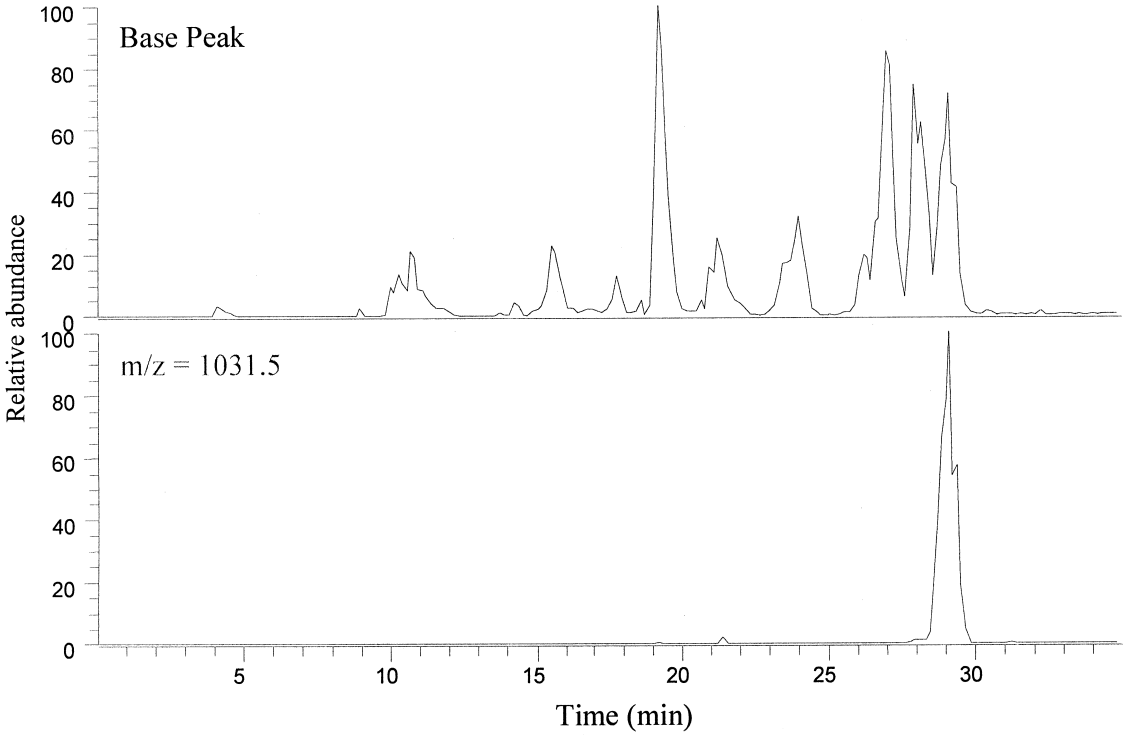

Fig. 6. CZE–MS analysis of a peptide mixture derived from tryptic digestion of β-casein. Conditions as in Fig. 5. Top panel shows the base peak versus time, bottom panel shows the EIC of the ion with $m/z = 1031.5$.

in the sequence database correlated with more than one top-ranked individual CID spectrum. Both the quality of the CID spectra and the precision of the database search results were confirmed by several repetitive experiments (data not shown).

4. Conclusions

Post-translational modification and point mutation are very common in protein engineering. MS offers high speed and great flexibility for identifying these changes in protein structure. Although database search routines can provide so much help and some modified sequence databases can recognize the modified peptides, but they are not so popular and available at present, so manual interpretation is still being used as a complementary method for protein identification, especially when the complete characterization of a proteolytic digestion mixture is desired. In this study, peptide mixtures derived from the digestion of S-125-IL2 and β-casein were separated by CZE and were analyzed by MS–MS using the CZE–MS–MS system described in Sections 2 and 3. The sequence database search using the resulted CID spectra can reveal the protein source rapidly. The mutation and modification can be discovered by monitoring the precursor ions of the specific peptides using EIC and be further identified by interpreting their respective CID spectra. These indicate that our method can be used to characterize the mutated and modified proteins with know sequence both conveniently and rapidly. It is very useful in quality control of the recombinant proteins. What make this method distinguished from the other analytical methods of protein identification are its high speed, easy performance, low sample amount requirement and high precise.

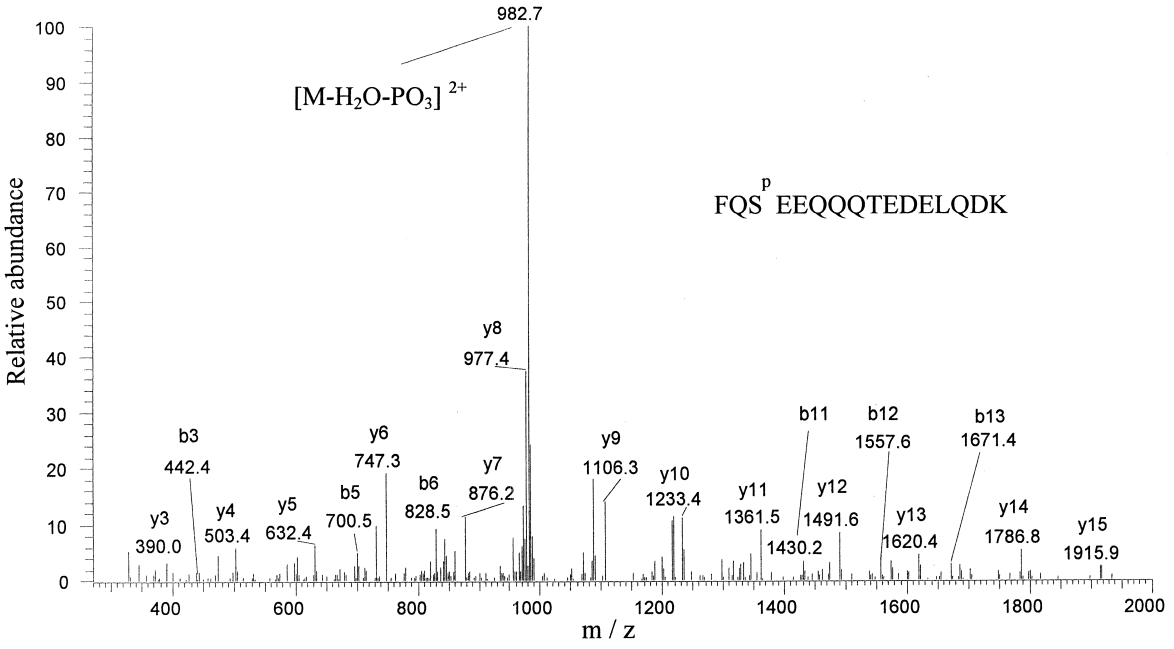

Fig. 7. CID spectrum recorded on the $[M+2H]^{2+}$ ion at $m/z=1031.5$ of the phosphorylated peptide derived from trypsin digestion of β-casein.

Table 5
Results of the search of the OWL composite sequence database with CID spectra of tryptic peptides derived from bovine β-casein

Mass (MH^{1+})[a]	Calculated mass	Correlation coefficient	Δ Correlation factor	Sequence	Position
646.5	645.8	2.1162	0.246	EAMAPK	100–105
748.5	747.9	1.7142	0.088	EMPFPK	108–113
780.6	780.0	1.9041	0.253	VLPVPQK	170–176
1093.9[b]	2186.6	5.5982	0.401	DMPIQAFLLYQEPVLGPVR	184–202
742.5	741.9	1.6002	0.641	GPFPIIV	203–209

[a] All the masses are presented in their MH^{1+}. The sequence position of the peptides within the β-casein sequence (209 amino acids) is indicated.

[b] This mass is presented in its $[M+2H]^{2+}$.

Table 6
Cumulative score and ranking of β-casein

Protein	Total score	No. 1	No. 2	No. 3	No. 4	No. 5
CASB_BOVIN	50	5	0	0	0	0
LYC_FELCA	10	1	0	0	0	0
C64011	10	1	0	0	0	0
S31952	10	1	0	0	0	0
GRP_DAUCA	10	1	0	0	0	0

Acknowledgements

The authors thank Professor Da-fu Cui for her kind gift of synthetic salmon calcitonin and Mr. Ke-yu Zhou for the expert technical assistances. A grant from ‘‘863’’ Biotechnology Project of China, No. 102-08-06-02, and a grant for key project from the National Natural Science Foundation of China are gratefully acknowledged for support of the work.

References

[1] P. Roepstorff, Curr. Opin. Biotechnol. 8 (1997) 6.
[2] R.C. Beavis, B.T. Chait, Methods Enzymol. 270 (1996) 520.
[3] J.S. Anderson, B. Svensson, P. Roepstorff, Nature Biotechnol. 14 (1996) 450.
[4] J.B. Fenn, M. Mann, C.K. Meng, S.F. Wong, C.M. Whitehouse, Science 246 (1989) 67.
[5] M. Mann, G. Talbo, Curr. Opin. Biotechnol. 7 (1996) 11.
[6] K. Gevaert, J.L. Verschelde, M. Puype, J.V. Damme, M. Goethals, S.D. Boeck, J. Vandekerckhove, Electrophoresis 17 (1996) 918.
[7] H. Lee, D.M. Lubman, Anal. Chem. 67 (1995) 1400.
[8] S. Hsieh, K. Dreisewerd, R.C. van der Schors, C.R. Jimenez, J. Stahl-Zeng, F. Hillenkamp, J.W. Jorgenson, W.P.M. Geraerts, K.W. Li, Anal. Chem. 70 (1998) 1848.
[9] M.D. Jones, S.D. Patterson, H.S. Lu, Anal. Chem. 70 (1998) 138.
[10] J.A. Loo, C.G. Edmonds, R.D. Smith, Anal. Chem. 63 (1991) 2489.
[11] S.J. Locke, P. Thibaault, Anal. Chem. 66 (1994) 3438.
[12] D. Figeys, R. Aebersold, Electrophoresis 18 (1997) 361.
[13] S.D. Patterson, R. Aebersold, Electrophoresis 16 (1995) 1792.
[14] D. Figeys, R. Aebersold, Electrophoresis 19 (1998) 888.
[15] J. Cai, J. Henion, J. Chromatogr. A 703 (1995) 667.
[16] R.D. Smith, H.R. Udseth, J.H. Wahl, D.R. Goodlett, S.A. Hofstadler, Methods Enzymol. 271 (1996) 448.
[17] S. Pleasance, P. Thibault, J. Kelly, J. Chromatogr. 591 (1992) 325.
[18] F. Foret, E. Szoko, B.L. Karger, J Chromatogr. 608 (1992) 3.
[19] R.L. Chien, D.S. Burgi, Anal. Chem. 64 (1992) 489A.
[20] D. Figeys, A. Ducret, J.R. Yates III, R. Aebersold, Nature Biotechnol. 14 (1996) 1579.
[21] D. Figeys, I.V. Oostveen, A. Ducret, R. Aebersold, Anal. Chem. 68 (1996) 1822.
[22] M. Wilm, M. Mann, Anal. Chem. 68 (1996) 1.
[23] G.A. Valaskovic, N.L. Kelleher, F.W. McLafferty, Science 273 (1996) 1199.
[24] K.M. Swiderek, M.T. Davis, T.D. Lee, Electrophoresis 19 (1998) 989.
[25] P.T. Jedrzejewski, W.D. Lehmann, Anal. Chem. 69 (1997) 294.
[26] J.R. Yates III, J.K. Eng, A.L. McCormack, D. Schieltz, Anal. Chem. 67 (1995) 1426.
[27] M. Wilm, G. Neubauer, M. Mann, Anal. Chem. 68 (1996) 527.