Talanta 84 (2011) 614–622

Contents lists available at ScienceDirect

Talanta

journal homepage: www.elsevier.com/locate/talanta

An effective method for de novo peptide sequencing based on phosphorylation strategy and mass spectrometry

Dongmei Zhang[a], Hongxia Liu[a], Shusheng Zhang[c], Xiaolan Chen[c], Shangfu Li[a], Cunlong Zhang[a], Xiangming Hu[a], Kaishun Bi[d], Xiaohui Chen[d], Yuyang Jiang[a,b,*]

[a] The Key Laboratory of Chemical Biology, Guangdong Province, Graduate School at Shenzhen, Tsinghua University, Shenzhen 518055, China
[b] School of Medicine, Tsinghua University, Beijing 100084, China
[c] Department of Chemistry, Zhengzhou University, Zhengzhou 450052, China
[d] Shenyang Pharmaceutical University, Shenyang 110016, China

ARTICLE INFO

Article history:
Received 13 August 2010
Received in revised form
20 December 2010
Accepted 21 December 2010
Available online 8 January 2011

Keywords:
De novo peptide sequencing
Electrospray ionization mass spectrometry
Phosphorylation reaction
The phosphorylated peptides

ABSTRACT

An effective method for peptide sequencing based on phosphorylation strategy and UPLC–MS/MS is proposed in this report. A phosphorylation reaction was carried out by mixing model peptide solution with phosphorylation solution. UPLC–MS/MS was used to analyze and characterize the phosphorylated peptides in the optimized ramp collision energy mode. The results illustrated that this phosphorylation approach significantly strengthened the signal intensity of both a_1 and b series ions of the spectra of the modified peptides. It also can be used to effectively distinguish glutamine (Q) and lysine (K) residues in peptides. The feasibility of this approach was validated by analyzing the trypsin-digested BSA. Data suggested that this proposed method could be a useful tool for the de novo peptide sequencing in proteome research.

1. Introduction

The N-terminal sequence analysis is a key point for identifying a protein [1]. At present, Edman Degradation Method is commonly used for the N-terminal sequencing of peptide and protein and is considered to be an important implement with either manual or automatic instrumentation (protein sequencer) [2–4]. However, the defects are of often causing ambiguous results and low sensitivity. For instance, the residues like tryptophan, cysteine and the first amino acid residue are overlapped by background or cannot be detected [5]. So Edman Degradation Method was difficult to characterize the trace peptides, the blocked N-terminal peptides, or a mixture of peptides. Radioactive reagents can be used in Edman sequencing to improve the sensitivity, but the radioactive waste needs careful disposal. Therefore, the wide application of the Edman degradation method is limited in the growing high-throughput proteome research [6]. With the developments in technology and instrumentation, mass spectrometry (MS) has become a very useful, sensitive and high-throughput tool for the peptide sequencing. The routine peptide mass fingerprinting (PMF) method coupled with both MS and gene/protein database searching tool is of high confidence in the identification of the target proteins [7–11]. Although the N-terminal sequences can be analyzed directly by MS, there still had some significant challenges. For example, the fragmentation patterns of peptides/proteins are very complicated, from which the complete and contiguous series of ions cannot be typically observed [12]. The loss of a_1 ion in the mass spectrum often caused the uncertain N-terminal amino acid segments [13]. In past years, efforts to address these limitations using chemical derivatization method have been conducted to obtain simple and direct peptide fragments to facilitate the interpretation of the MS/MS spectra in proteomics research [14–17]. After introducing a mass shift at either the N-terminus [18,19] or the C-terminus through chemical derivatization [1,20–22], the N-terminal peptide and internal peptide amino acid sequences can be easily analyzed with tandem MS/MS. Kuyama et al. [23] and Watson et al. [24] used p-phenylenediisothiocyanate resin to selectively capture and isolate the N α-amino peptides after reacting with succinimidyloxycarbonylmethyl tris (2,4,6-trimethoxyphenyl) phosphonium bromide (TMPP-Ac-OSu). The derivatization peptides using sulfo-NHS-SS-

Abbreviations: DEP, diethyl phosphonate; DMP, dimethyl phosphonate; DIPP, diisopropyl phosphonate; MeOH, methanol; EtOH, ethanol; CCl_4, carbon tetrachloride; TEA, triethylamine; BSA, bovine serum albumin.

* Corresponding author at: The Key Laboratory of Chemical Biology, Guangdong Province, Graduate School at Shenzhen, Tsinghua University, Shenzhen 518055, China. Tel.: +86 755 26036017; fax: +86 755 26036017.
E-mail addresses: jiangyy@sz.tsinghua.edu.cn, liuhx@sz.tsinghua.edu.cn (Y. Jiang).

0039-9140/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.talanta.2010.12.035

biotin instead of biotinyl cysteic acid (BCA) were successfully interpreted based on tandem MS/MS. And then Chen and his coworkers [25] improved this method for peptide de novo sequencing by MALDI mass spectrometry. N-terminal derivatization of peptides with Edman's reagent PITC can be promoted by gas-phase Edman cleavage that yields abundant complementary b_1 and y_{n-1} ion pairs by tandem MS/MS [26]. Acetylation of free amino groups on N-terminal peptides, sulfonic acid derivatives [27], dimethyl labeling on N-terminal peptides had been commonly used as derivatization strategy for de novo sequencing [14,28,29]. deJong et al. [30] had reported an on-column derivatization method for the derivatization of peptides at sub-pmol quantities as their PA derivatives. The method was selective and quantitative for N-terminal and lysine amino groups in peptides with a partial reaction with the Tyr phenolic hydroxyl group. From previous research in our laboratory, the introduction of a diisopropyloxyphosphoryl (DIPP) group at the N-terminus of proteins can enhance the signal response by 10–100-fold in positive electrospray ionization (+ESI)-MS [31]. Zhao's coworkers have preliminarily studied the N-terminal phosphorylation of peptides with ethoxyphenylphosphinate (EP) for the peptide sequencing [32], but there exist some problems in the actual application of the peptide sequencing, such as the conditions of the phosphorylation reaction, the mode of the collision energy, the characterization of the mass spectrum fragmentation, the application in the mixed peptides and trypsin-digested proteins.

In this study, systematic studies for the phosphorylation reactions of different derivatization agents and their behaviors of MS/MS fragmentation were carried out by ESI-Q-TOF, and the resulting modifications effectively overcome these sophisticated problems. Patterns of the resulting ions, the typical a_1 ion and the entire series of b ions of phosphorylated-peptides can be significantly observed in the mass spectra, allowing successfully de novo peptide sequencing. Meanwhile, this method could identify the internal amino acids such as K/Q/G + A, which have similar molecular mass (K 146.18 Da, Q 146.14, G + A 146.16). Further, the application of this proposed method in the trypsin-digested protein was examined.

2. Materials and methods

2.1. Materials

The used model peptides, LPNLSQP (P1, Mw 767.41 Da), LPNLSGAP (P2, Mw 767.42 Da), LPNLSKP (P3, Mw 767.45 Da), EPVTQAEML (P4, Mw 1016.48 Da), EPVTKAEML (P5, Mw 1016.52 Da), FLSFHLSNL (P6, Mw 1076.57 Da) and LKVGVKQY (P7, Mw 933.56 Da), were purchased from GL Biochem Ltd. (Shanghai, China). The used phosphoryl reagents, DMP, DIPP, diethyl phosphate (DEP) and dipropyl phosphonate (DPP), were prepared according to the reported methods in our laboratory, which were identified by MS and NMR, and their purity were higher than 95%. Pesticide-grade methanol (MeOH) was from Fisher Scientific Company (USA). Ethanol (EtOH), carbon tetrachloride (CCl_4), triethylamine (TEA) and other chemicals were analytical grade and obtained from Tianjin Damao Chemical Reagent Factory. The deionized water was purified using a Milli Q water purification system (Millipore, Bedford, MA, USA).

2.2. Phosphorylation reaction

The stock solutions of standard peptides were prepared by dissolving model peptides in water to a concentration of 0.5 mg ml^{-1} (~0.6 mM). And then the same volume of EtOH was added to yield the final concentration of peptide at 0.25 mg ml^{-1} (~0.3 mM). The phosphorylated solution (6.6 mM) was freshly prepared by mixing phosphoryl reagent with MeOH and the final concentration of phosphorylated reagent at 3.3 mM was obtained by adding CCl_4 with the ratio 1:1 (v/v). The phosphorylation reaction was carried out by mixing 100 μl model peptide solution (0.25 mg ml^{-1}) with 10–40 μl phosphorylated reagent and 10–40 μl TEA, and vortexing for 3 min at room temperature, then reacting for 10 min. Afterwards, the reaction mixture was diluted with MeOH to 500 μl and filtered with PTEF membrane for MS analysis without sample cleanup.

2.3. Mass spectrometry

All MS and MS/MS experiments were conducted using Q-TOF Premier mass spectrometer with an ESI ion source (Waters Micromass, Manchester, UK). The samples were directly infused into the mass spectrometer via syringe pump at a flow rate of 10 μl min^{-1}. The positive ionization mode was used to obtain the MS spectra with survey scans. The full MS scan range was from m/z 100 to 1500 and MS/MS from m/z 50 to 1500. The capillary was 3.0 kV. The collision energies were set at 10, 20, 30, 40, and 50 eV, respectively. The source temperature was 120 °C. Desolvation temperature was 300 °C. The MS/MS spectra were calibrated using the LockSpray and 0.1 g L^{-1} of Leu-enkephalin (m/z 556.2771) was introduced into the reference spray with a peristaltic pump at a flow rate of 50 μl/min.

2.4. UPLC–MS analysis

UPLC–MS analysis was performed using an ACQUITY^{TM} UPLC system (Waters, Milford, MA, USA) equipped with a Q-TOF Premier mass spectrometer (Waters Micromass, Manchester, UK). The chromatographic separations were performed on a reversed-phase column (BEH C18, 100 mm × 2.1 mm id, 1.7 μm, Waters Corp., Milford, MA, USA). Mobile-phase A consisted of 0.1% formic acid (FA) in water, mobile-phase B consisted of 95% acetonitrile (ACN) in 0.1% FA. The elution program was a linear gradient of 0–70% mobile-phase B over 10 min at a flow rate of 450 μl min^{-1}. The injection volume was 20 μl.

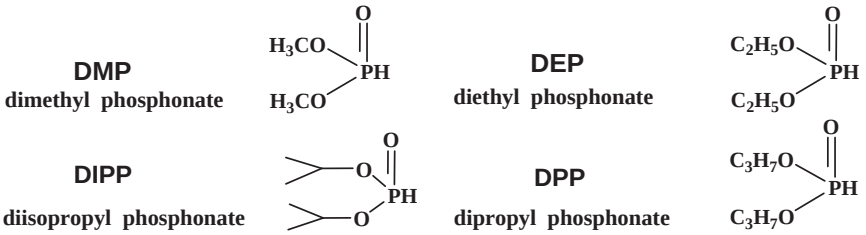
2.5. Data acquisition and analysis

All MS and MS/MS spectra were acquired and analyzed via MassLynx (Waters version 4.1). De novo sequencing was processed using bioLynx either manually or automatically. The multiply charged ions were converted into their singly charged form by using MaxEnt3. All results were further verified using manual interpretation.

3. Results and discussion

3.1. Comparison of different phosphoryl reagents using standard peptides

In our previous work, we have preliminarily investigated the phosphorylation reaction mechanism and characterized some phosphorylation amino acids, peptides and proteins via ion-trap ESI/MS, HPLC/MS and other analytical techniques. The results displayed that the modification sites were on the N-terminal amino acid and the side-chains residues of lysine (Lys) and histidine (His). According to the incremental value of the molecular weight, the numbers of the residues of Lys and His in peptides and proteins could be determined. These findings were very potential to detect the peptide sequence and protein structure characterization (data not published). However, it is necessary to optimize which kind of phosphoryl reagents was the best for the peptide phosphorylation reaction and how about the general adaptability for peptide sequencing.



Scheme 1. . The structures of phosphoryl reagent.

Here, some suitable phosphoryl derivatization reagents for peptide sequencing, including DMP, DEP, DPP and DIPP (refer to the structures shown in Scheme 1), were studied and compared. A synthetic peptide, LPNLSGAP (P2, Mw 767.42 Da), was selected as a model peptide. The phosphorylated processes were easily happened through the modified Atherton-Todd reaction [31]. Different molecular weight shifts of the phosphorylated-peptide products were clearly observed from Fig. 1.

Calculating the molecular weight of the phosphorylated-peptides, the mass increment of peptide was 108 Da for DMP, 136 Da for DEP, 164 Da for DIPP and DPP, respectively. So it was found in Fig. 1 that the ion at *m*/*z* 768.38 was unmodified peptide (P2) precursor ion $[P2+H]^+$ (Fig. 1E), and the phosphorylated-peptide precursor ions at *m*/*z* 876.41, 904.43, 932.46, 932.47 were corresponding to $[P2+DMP+H]^+$, $[P2+DEP+H]^+$, $[P2+DPP+H]^+$ and $[P2+DIPP+H]^+$ (Fig. 1D, C, B, A), respectively. The results clearly showed that all of the phosphoryl reagents used in this work were successfully labeled on the peptides.

Meanwhile the MS/MS spectra of the phosphorylated-peptides were shown in Fig. 1 and the CID fragmental behaviors were investigated. From Fig. 1, it can be seen that the similar fragmental ions, a_1 and b-ions series of phosphorylated-peptides, were more readily identified in Fig. 1D–A than Fig. 1E. For a_1 ion, it was easily observed from the spectrum of phosphorylated-peptides possibly because of the present of phosphoryl group; and for b-ions series, the greatly enhanced signal response might be attributed to the easier breakage site on amide bond and the existence of phosphoryl group in b-ions (as shown Scheme 2). In addition, the stronger signal intensity of b-ions series greatly simplified the fragmentation spectra. It makes b-ions series be found easily from the complex fragmentation spectra than the other ion series. So the entire peptide sequence can easily and rapidly be extrapolated from the a_1 and b-ions series of the MS/MS spectra. Comparing Fig. 1D to Fig. 1C, B and A, the similar mass increments of +28, +56 and +56 Da were detected clearly in a_1 and b-ions series in MS/MS spectra, which result from the mass deviation of the side chain of phosphoryl

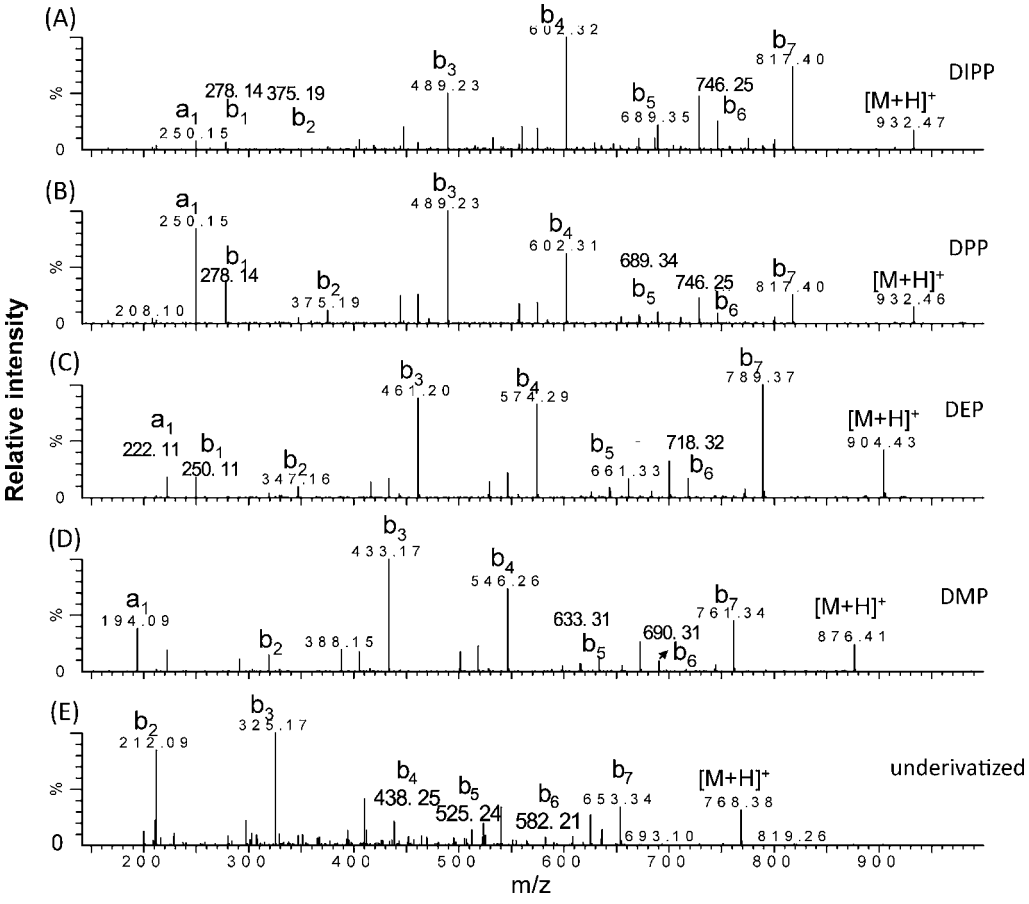
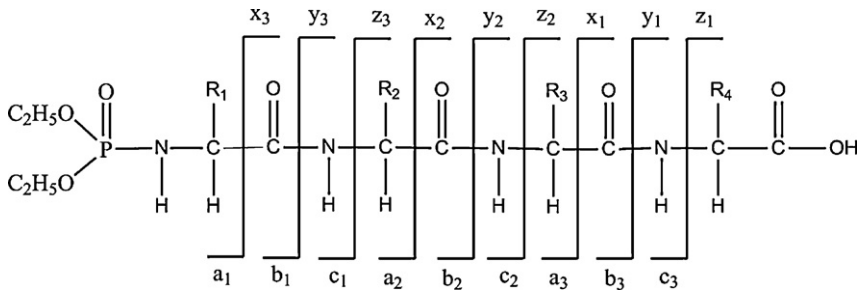


Fig. 1. MS/MS spectra of the phosphorylated-peptide, *LPNLSGAP. (A) DIPP labeled peptide, (B) DPP labeled peptide, (C) DEP labeled peptide, (D) DMP labeled peptide and (E) underivatized peptide.



Scheme 2. The mechanism of sequence peptide.

reagents (DEP-DMP, DPP-DMP, DIPP-DMP). Therefore, the derivatization of peptide with phosphoryl reagents mentioned here was the potential strategy for correctly identifying the peptide sequence.

It is also noticed that DIPP could easily form propylene because of the existence of isopropyl, which makes MS/MS spectrum become complicated. DMP was the lower stability during storage; and DEP and DPP exhibited the similar superiority in the peptide sequencing. Thus in the following work we use DEP as an exampled derivatization reagent.

3.2. Optimization of phosphoryl reagent labeling conditions

To obtain the optimal labeling conditions, a systematic investigation of reaction experiments was carried out using a model peptide LPNLSKP (P3). As indicated in Section 2.2, the final DEP phosphorylation reagent (3.3 mM) was prepared by mixing the DEP phosphoryl solution (6.6 mM) with CCl_4 at the ratio 1:1 (v:v). Because the excess DEP phosphoryl reagent could produce more complex by-products, and the insufficient DEP would lead to the incomplete phosphorylation reaction of peptide P3, the appropriate concentration of DEP related to the molar value of the lysine and the N-terminal residues in the peptide P3 was optimized. For the reaction between P3 and DEP, the different reagent volume ratio of P3 (0.33 mM) and DEP (3.3 mM) was set between 10:1 and 10:4, and the optimized solution volume ratio was obtained at 10:1.9–2.3. In this case, the molar ratio of P3 and DEP was about 1:1.9–2.3. It is noteworthy that the phosphorylation products would be hydrolyzed in the acidic environment and the addition of TEA was necessary to neutralize the acids generated during the reaction. The molar ratio of the DEP reagent and TEA at 1:1 could ensure the phosphorylation products stable. Therefore, to obtain the high efficiency of phosphorylation reaction, the appropriate molar ratio of P3, DEP and TEA was about 1:1.9–2.3:1.9–2.3.

The other model peptides were used to optimize the labeling conditions and the results shown that the molar ratios of peptide and DEP were related to the free $-NH_2$ numbers in the peptide chains. So the appropriate range ratio of peptide, phosphorylated reagent and TEA was very important for the any peptide to complete the phosphorylation reaction.

3.3. Kinetics of phosphorylation reaction

With LPNLSKP (P3) as a model peptide, the kinetics of phosphorylation reaction of P3 was studied by calculating the relative amount (peak area ratio) of reaction products at the different reaction times (0, 5, 10, 15, 20, 25 and 30 min). The amount–time curve was shown in Fig. 2. From Fig. 2, at the beginning of the reaction, only the signal of peptide P3 could be observed. At 5 min, the single-phosphorylated-P3 and double-phosphorylated-P3 were clearly observed, and the relative amount of single-phosphorylated-P3 was higher than that of P3. With increase of the reaction time, P3 were gradually changed to single-phosphorylated-P3 [(DEP)LPNLSKP] and to double-phosphorylated-P3 [(DEP)LPNLS(DEP)KP]. At 10 min, the relative amount of (DEP)LPNLS(DEP)KP was higher than that of both (DEP)LPNLSKP and P3. After 20 min, the amounts of the reaction products remained unchangeable, suggesting that the reaction was nearly completed within about 20 min. The single and double-phosphorylated-P3 were accurately identified according to the incremental mass of 136 and 272 Da, respectively. By comparing with the relative amount of products, the yields of (DEP)LPNLS(DEP)KP were about 80%.

3.4. Selectivity of phosphorylation sites

For exploring the competitiveness of two kinds of phosphorylation site, the mixture of phosphorylation products at 5 min were directly analyzed by UPLC/MS/MS without the further purification. Fig. 3 shows the typical total ion current chromatogram (TIC) and extraction ion current chromatograms (EIC) of P3 peptide labeled with DEP tag. Data showed that four peaks were detected with a retention time at 2.57, 3.96, 4.27 and 5.34 min, respectively. The peak at 2.57 min corresponded to original P3 peptide (m/z 768.38), the peak at 5.34 min represented to double-phosphorylated P3 [(DEP)LPNLS(DEP)KP] (m/z 1040.43) with two DEP (+272.05 Da), and the peaks at 3.96 and 4.27 min were characterized as single-phosphorylated P3 (m/z 904.41), (DEP) LPNLSKP or LPNLS(DEP)KP with one DEP (+136.03 Da). From the MS/MS spectrum (not shown—you may need to show if compared to another) of the peak at 3.96 min, the fragmentation feature of b-ions was same as (DEP)LPNLSKP (Fig. 4). However, although the MS/MS spectrum for the peak at 4.27 min was not optimal, the fragmentation feature of b-ions was similar to that of LPNLS(DEP)KP in Fig. 4. Comparing the peak heights of (DEP)LPNLSKP at 3.96 and LPNLS(DEP)KP at 4.27 min at different reaction times (5, 10 and 15 min), the peak of LPNLS(DEP)KP at 4.27 min was always small while (DEP)LPNLSKP

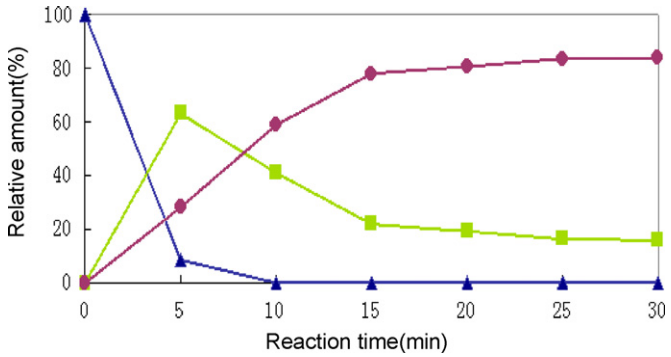


Fig. 2. The time curve of phosphorylated reaction between LPNLSKP peptide and DEP. ▲, LPNLSKP peptide; ■, single-phosphorylated-peptide; ●, double-phosphorylated-peptide. Chromatographic conditions: separation column, BEH C18, 100 mm × 2.1 mm, mobile-phase A: 0.1% formic acid in water; mobile-phase B: 95% acetonitrile in 0.1% formic acid. The elution program was a linear gradient.

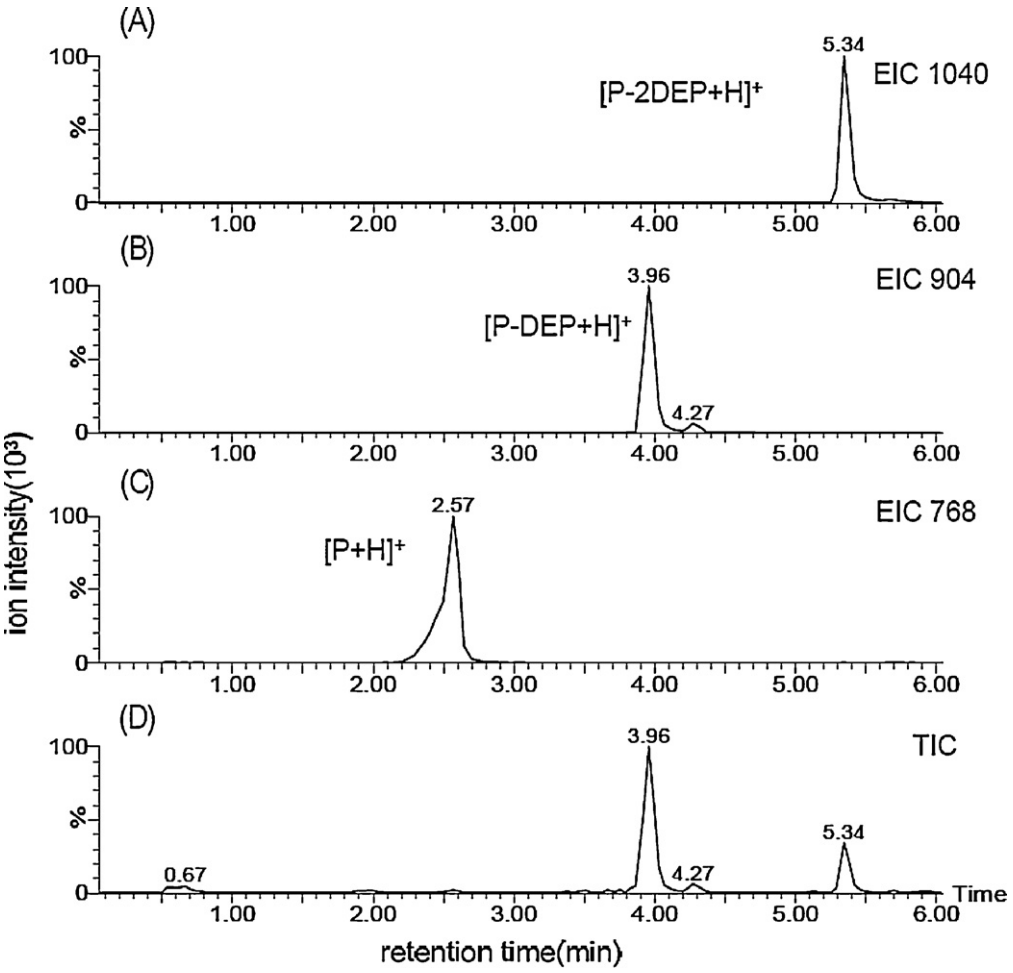


Fig. 3. The chromatograms of the mixture samples with UPLC/MS. (A) The extraction ion chromatogram of double-phosphorylated-peptides; (B) the extraction ion chromatogram of single-phosphorylated-peptides; (C) the extraction ion chromatogram of LPNLSKP peptide; (D) total ion current chromatograms (TIC) of the mixture samples; chromatographic conditions: separation column, BEH C18, 100 mm × 2.1 mm, mobile-phase A: 0.1% formic acid in water; mobile-phase B: 95% acetonitrile in 0.1% formic acid. The elution program was a linear gradient.

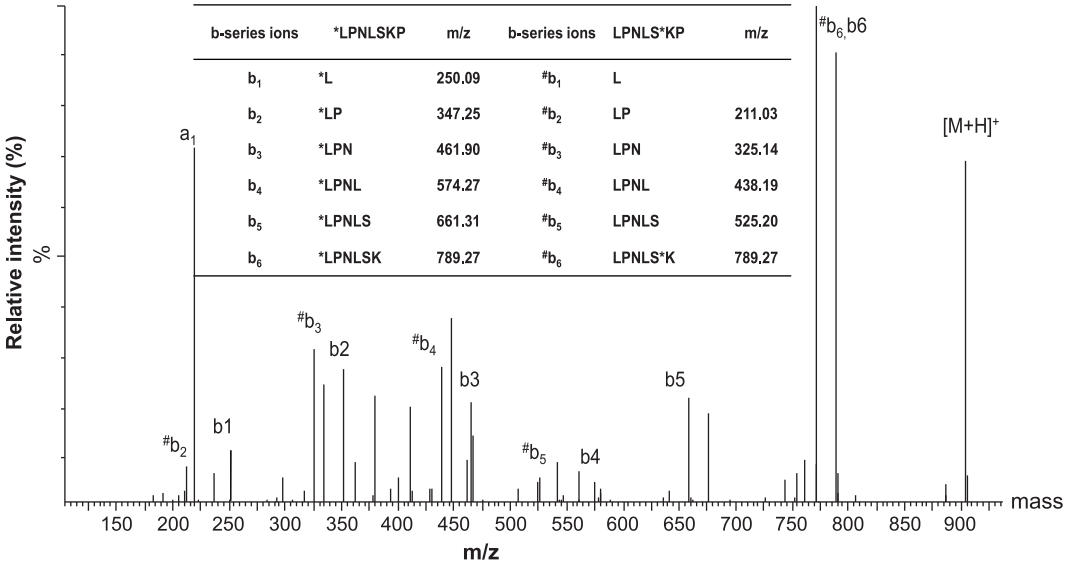


Fig. 4. The fragmentation spectrum of single-phosphorylated-peptide in the mixture reaction. * *the site of phosphorylation* [(DEP) LPNLSKP or LPNLS(DEP)KP). #b series ion from peptide LPNLS(DEP)KP.

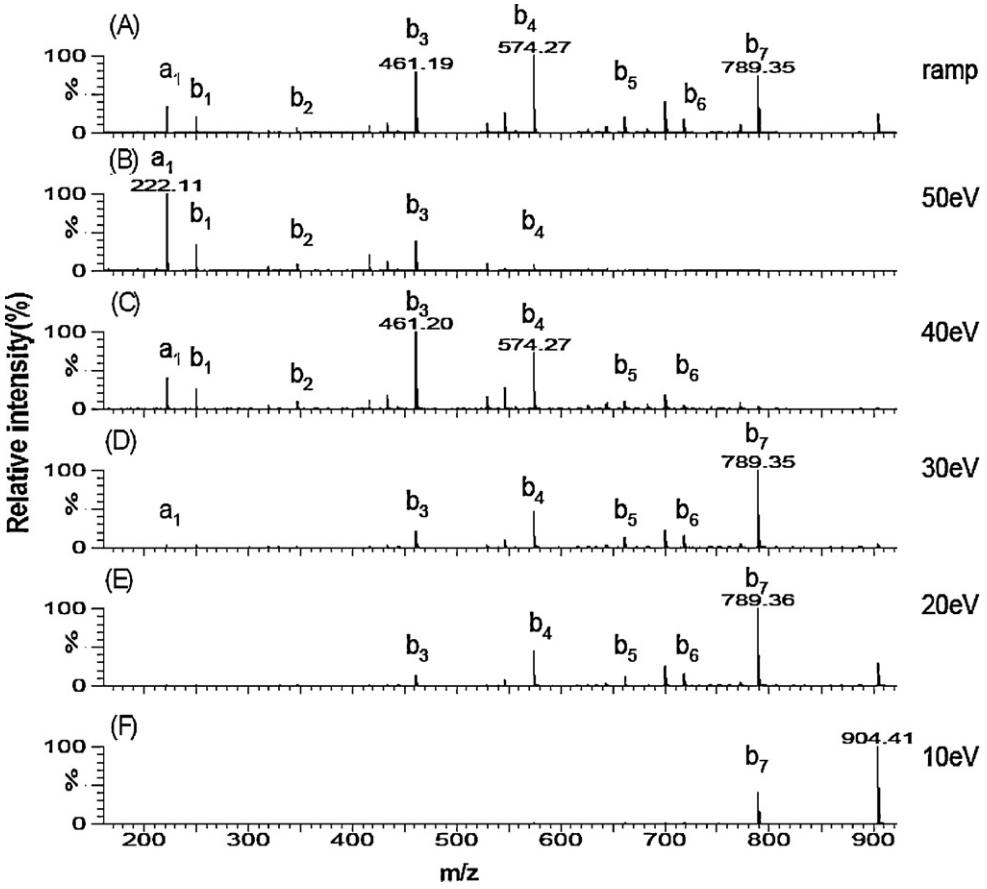


Fig. 5. The CID spectrum of DEP-LPNLSGAP at different collision energy modes. (A) The collision energy ramp mode from initial energy 20 to final energy 40 eV, (B) 50 eV, (C) 40 eV, (D) 30 eV, (E) 20 eV, (F) 10 eV.

at 3.96 min were very high. It implied that (DEP)LPNLSKP was easily produced.

In addition, the model peptide FLSFHLSNL (P6), containing both an N-terminal α-amino group and an imidazolyl group of histidine side chain, was also studied as same as model peptide P3 as above, and results were obtained. These results indicate that the labeling was occurred firstly at N-terminal amino-group, which was possessed the higher selectivity in the phosphorylation reaction process.

3.5. Optimized collision energy

In this section the effect of collision energy on the fragmentation of single-phosphorylated P2 (DEP-LPNLSGAP) was investigated in detail. Fig. 5 shows the CID spectra of single-phosphorylated P2 (DEP-LPNLSGAP) at different collision energies. From Fig. 5, the fixed either lower (10, 20 or 30 eV) or higher collision energy (40 or 50 eV) could not induce the sufficiently meaningful fragmentation. However, with the equipped ramp mode changing collision energy from 20 to 40 eV, the perfect fragmentation was obtained as shown in Fig. 5A. The similar fragmentation variations for single- and double-phosphorylated P1–P7 were observed with the fixed and ramp energy modes (data not shown). These results illustrate that the collision energy was a key factor for obtaining the correct sequencing peptide and the ramp collision energy mode was optimized.

3.6. Distinguishing glutamine (Q) and lysine (K) residues

In general, the K and Q were very hard to be distinguished in peptide sequencing by mass spectrometry due to their identical molecular weights. We found that the behaviors of phosphorylation reaction on both lysine and glutamic acid side-chain in the peptides were significantly different. When Q and K existed in the middle position of peptide sequence, only the K side-chain can be labeled by phosphoryl reagent. To verify the credibility of this finding, seven model peptides, LPNLS**Q**P, LPNLS**GA**P, LPNLS**K**P, EPVT**Q**AEML, EPVT**K**AEML, FLSF**H**LSNL, L**K**VGV**K**QY, were phosphorylated with DEP. Table 1 shows the mass shift information before or after the phosphorylation reactions. Data show that LPNLS**Q**P, LPNLS**GA**P and LPNLS**K**P peptides have the similar *m*/*z* at 768.39, 768.38, 768.41 when not phosphorylated. However, after phosphorylation, these peptides had clearly different *m*/*z* at 904.39, 904.41 and 1040.48, respectively. They were labeled with one DEP, one DEP or two DEP, respectively. We noticed that the main difference for three peptides lied on the **Q** in the LPNLS**Q**P, **GA** in the LPNLS**GA**P and **K** in the LPNLS**K**P. From their fragmentation features of b-ions, LPNLS**K**P was labeled with two DEP on the **L** (N-terminal α-amino group) and **K** (ε-amino group of side chain) sites. LPNLS**Q**P and LPNLS**GA**P were labeled one DEP on the L (N-terminal α-amino group). The **Q** and GA sites were not labeled with DEP. With the same way, EPVT**K**AEML and EPVT**Q**AEML with the similar *m*/*z* at 1017.53 and 1017.49 were labeled with two DEP (on E and K sites) and one DEP (on E site), respectively. For FLSF**H**LSNL with *m*/*z* at 1077.57, double-phosphorylated peptide were seen and labeled on **F** and **H** sites. As expected. L**K**VGV**K**QY, including two **K** residues, was observed to be labeled with three DEP groups on one L and two K sites (data not shown). Therefore, it is suggested that this strategy could be employed to exactly distinguish the **K** or **H** residue no matter whether existed in the peptide, and in sequencing of the peptide by MS technique, this strategy could be a promising way to distinguish the **K** or **Q**/**GA** in peptides. Probably, this strategy could conventionally iden-

Table 1
The numbers of phosphoryl labeled with DEP on peptide models.

Peptide	[Peptide+H]$^+$ (m/z)	DEP labeled peptide	[DEP labeled peptide+H]$^+$ (m/z)	Mass drift (Da)	Labeled DEP number
LPNLSQP (P1)	768.39	*LPNISQP	904.39	+136.00	1
LPNLSGAP (P2)	768.38	*LPNISGAP	904.41	+136.03	1
LPNLSKP (P3)	768.41	*LPNIS*KP	1040.48	+272.07	2
EPVTQAEML (P4)	1017.49	*EPVTQAEML	1153.49	+136.00	1
EPVTKAEML (P5)	1017.53	*EPVT*KAEML	1289.42	+271.89	2
FLSFHLSNL (P6)	1077.57	*FLSF*HISNL	1348.59	+272.02	2
LKVGVKQY (P7)	934.33	*L*KVGV*KQY	1342.62	+408.29	3

* represents the site of phosphorylation.

tify and speculate the number of **K** or **H** residue in the target peptide.

3.7. De novo sequencing of phosphorylated-peptides

Fig. 6 shows the results of De novo sequencing of phosphorylated-peptides (P1–P6). In Fig. 6A, B and C, they have the similar a_1 ions at m/z 222.10 and b_1–b_5 series ions at 250.09, 347.13, 461.15, 574.27 and 661.23, respectively, but their b_6 ions were significantly different. In Fig. 6A, the b_6 ion at m/z 789.25 was from the fragmentation of (DEP) LPNLS**Q**. In Fig. 6B, the b_6 ion at m/z 718.32 was from the fragmentation of (DEP) LPNLSG, and b_7 ion at m/z 789.35 (similar to the b_6 ion in Fig. 6A) was from the fragmentation of LPNLSGA. In Fig. 6C, the b_6 ion at m/z 925.32 was

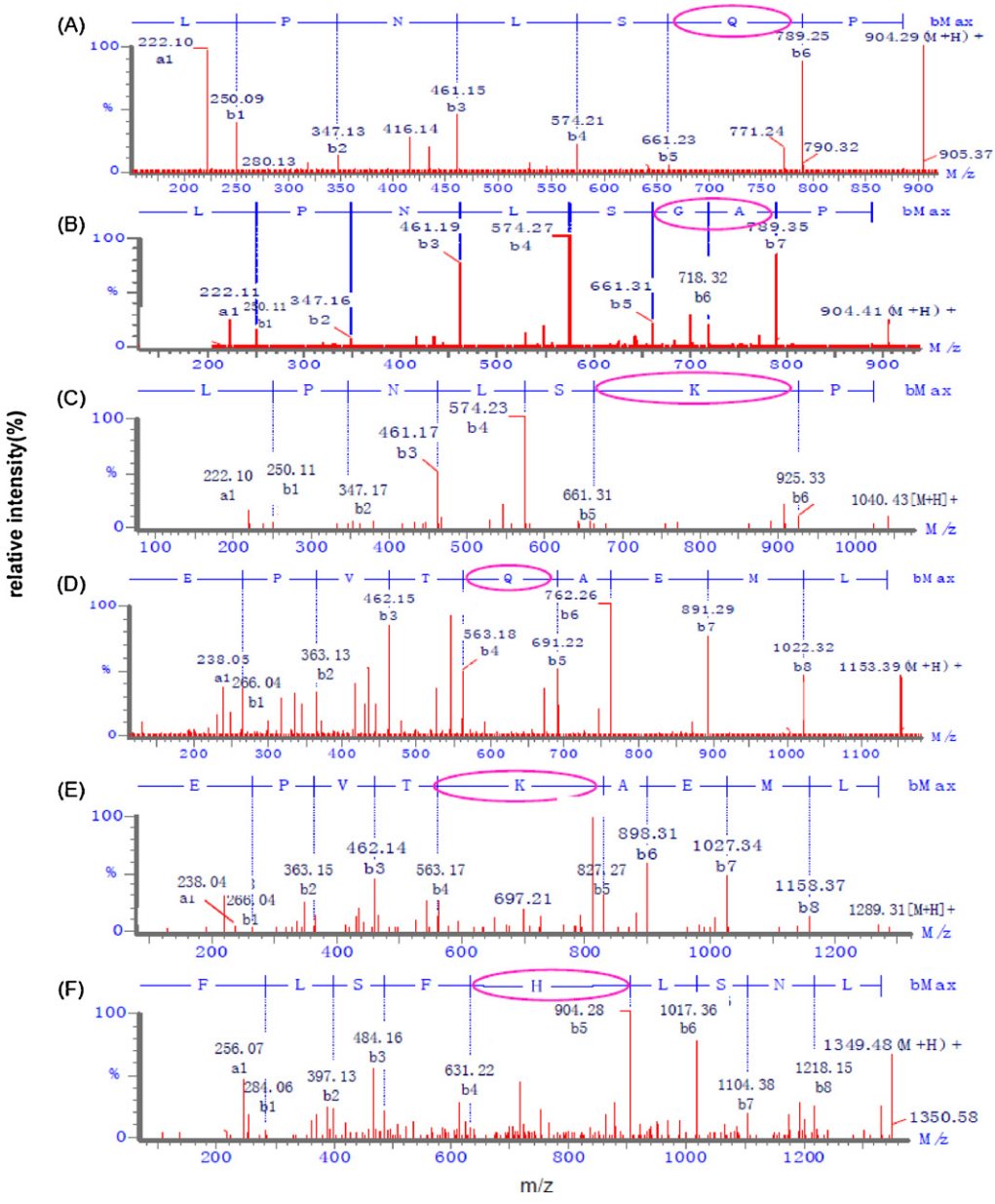


Fig. 6. The MS/MS fragmentation spectra for de novo sequencing of the phosphorylated-peptides.

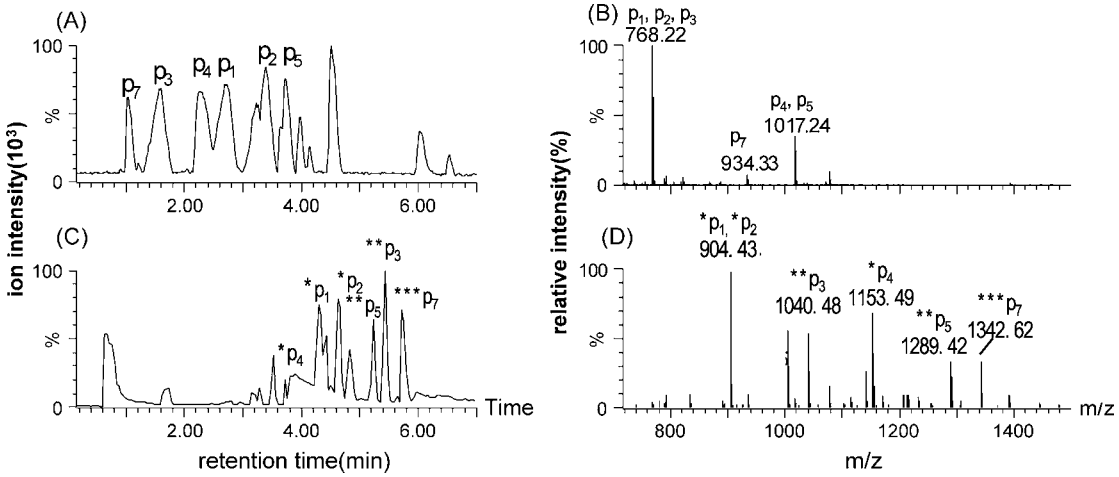


Fig. 7. The analysis of peptides and phosphorylated-peptides mixtures. (A) Total ion current (TIC) of the peptides mixtures; (B) the full scan MS from A1 range 1–4 min; (C) total ion current (TIC) of the phosphorylated-peptides mixtures; (D) the full scan MS from B1 range 3–6 min; *, labeled one DEP; **, labeled two DEP; ***, labeled three DEP. Chromatographic conditions: separation column, BEH C18, 100 mm × 2.1 mm, mobile-phase A: 0.1% formic acid in water; mobile-phase B: 95% acetonitrile in 0.1% formic acid. The elution program was a linear gradient.

from the fragmentation of LPNLS**K**. By calculating the deviation of their b_6 or b_7 ions, it turned out that the **K** site of fragment LPNLS**K** in Fig. 6C was phosphorylated with the second DEP group with 136 Da mass shift. As the same way, phosphorylated EPVT**Q**AEML and EPVT**K**AEML peptides had the similar fragmentation behaviors as shown in Fig. 6D and E, and the residues of **Q** and **K** were easily distinguished from the mass drift. In addition, FLSF**H**LSNL, the model peptide including more **H** residues, was tested. It was found that its peptide sequence was very easily obtained as shown in Fig. 6F. The results significantly illustrated that the N-terminal phosphorylation of the peptides was a critical method to enhance a_1 and b ions series and simplify the MS/MS fragmentation pattern, which ensures peptide sequences to be correctly analyzed. These results were critical for correctly de novo sequencing analysis of peptides.

3.8. Analysis of the mixed standard peptides

To validate the feasibility of this strategy, the mixture of six standard peptides, LPNLS**Q**P (P1), LPNLS**GA**P (P2), LPNLSKP (P3), EPVT**Q**AEML (P4), EPVT**K**AEML (P5), and L**K**VGV**K**QY (P7), was phosphorylated and analyzed by UPLC–MS/MS (Fig. 7). Fig. 7A and B were the chromatogram and ESI/MS of the untagged mixed peptides, and Fig. 7C and D were the chromatogram and ESI/MS of the mixture of phosphorylated peptides. Each chromatographic peak was identified with the molecular weight of peptides. As shown in Fig. 7A, six peptides could be separated in baseline from retention time 1 to 4 min. In Fig. 7B, the precursor ion at *m*/*z* 768.22 corresponded to [P1, P2 or P3+H]$^+$, the ion at *m*/*z* 1017.24 corresponded to [P4 or P5+H]$^+$, and ion at *m*/*z* 934.33 corresponded to [P7+H]$^+$, respectively. It was noticed that P1, P2, and P3 (or P 4 and P5) were not distinguished in full scan mass spectrometry. As expected, when the mixed peptides were phosphorylated with DEP, the different phosphorylated-products were obtained and their resolution could be greatly improved. As shown in Fig. 7C, six phosphorylated peptides were eluted from 3.5 to 6 min and had good resolutions. In Fig. 7D, each phosphorylated peptide was identified according to the molecular weight of phosphorylated peptides. The ion at *m*/*z* 904.43 corresponded to [phosphorylated-P1 or -P2+H]$^+$, the ion at *m*/*z* 1040.48 denoted to [phosphorylated-P3+H]$^+$, the ion at *m*/*z* 1153.49 indicated to [phosphorylated-P4+H]$^+$, the ion at *m*/*z* 1289.42 represented to [phosphorylated-P5+H]$^+$, the ion at *m*/*z* 1342.62 corresponded to [phosphorylated-P7+H]$^+$, respectively. These peaks and fragment ions were very easily identified because of labeling different numbers of phosphoryl reagents and having the longer and different retention times of phosphorylated-products. Therefore, this strategy was a significant advantage in analysis of the mixed peptides.

3.9. Phosphorylation of trypsin-digested BSA

To verify this strategy, BSA was digested with TPCK-trypsin and the mixed-peptides were phosphorylated with DEP. Two kinds of digested-BSA solutions, un-phosphorylated and phosphorylated, were analyzed and characterized by MS/MS, respectively. Fig. 8 showed the full-scan MS of the trypsin-digested BSA (A) and its phosphorylated-products (B). The sequences for these peptides could be obtained according to the de novo sequencing of phosphorylated-peptides as previous described. As exhibited in Fig. 8A, the trypsin-digested BSA included many peptides, and the MS pattern of the peptides was complicated. Some of peptides had the lower intensity of MS signal, even was not noticed out, and thus it is very difficult to obtain these peptides' sequences. After these peptides were phosphorylated with DEP, however, the MS pattern was simplified and the MS signal intensity was enhanced, which provided much more useful information for obtaining their sequences. For example, the fragmental peptides at *m*/*z* 147.05(K), 276.16(EK), 294.18(FK) and 347.23(VTK) with low intensity (Fig. 8A) were not characterized well. When they were phosphorylated with one or two DEP and transferred as the phosphorylated-products, they were correctly identified at *m*/*z* 419.15(*K*), 548.19(*EK*), 566.20(*FK*) and 619.28(*VTK*) with high intensity (Fig. 8B). Again for the peptide at *m*/*z* 649.34 in Fig. 8A, its corresponding sequence was the TETMR and CASIQK fragmentation, here two kinds of fragmentation could not distinguished easily by MS spectra. After they were phosphorylated with DEP, different phosphorylated-products of (DEP)TETMR at *m*/*z* 785.32 and (DEP)CASIQ(DEP)K at *m*/*z* 921.38 were clearly seen in Fig. 8B. In addition, the fragmental peptides LAK and ALK had the same molecular weight at *m*/*z* 331.23 for un-phosphorylated-peptides in Fig. 8A and *m*/*z* 603.25 for their phosphorylated-products Fig. 8B. It seems impossible to distinguish LAK from ALK with or without phosphorylation in MS spectra, but in the MS/MS mode, the peptide sequence of LAK and ALK were easily

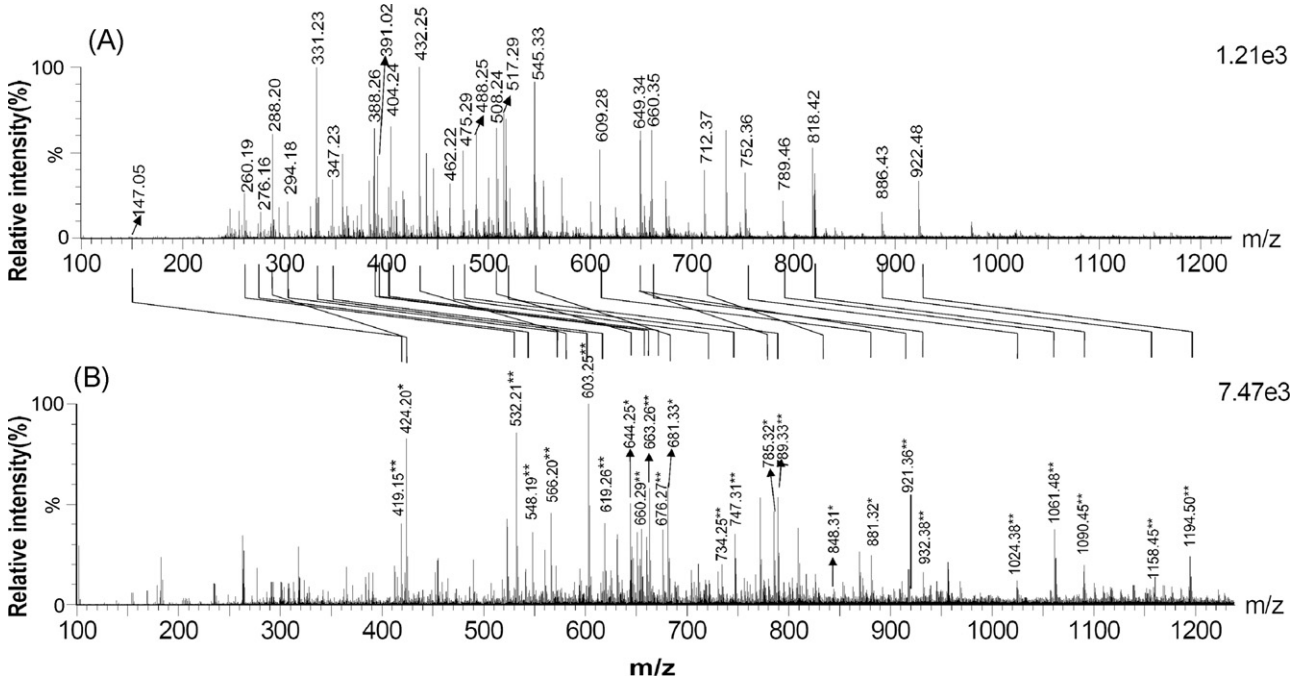


Fig. 8. The full-scan MS of the trypsin-digested BSA (A) and its phosphorylated-products (B). *, labeled one DEP; **, labeled two DEP; ***; labeled three DEP.

distinguished from the a_1 ion at m/z 222.10 and 180.05, respectively.

4. Conclusions

The present work showed an effective approach for de novo peptide sequencing using DEP labeling and UPLC–MS/MS. The phosphorylation sites may occur at α-amino groups of N-termini, ε-amino groups of lysines and side chain of histidines in peptides. The MS signal of a_1 and b series ions were significantly enhanced with DEP labeling, allowing de novo sequencing of target peptides with high confidence. This method was validated with seven model peptides and the trypsin-digested BSA. This work demonstrated that the DEP labeling facilitates the interpretation of peptide sequences from tandem mass spectra, which resolves the main problems in peptide sequencing of proteomic research, such as the complexes of fragmentation patterns of peptides, and do not typically yield the complete and contiguous series of ions. The higher intensity of a_1 ion can improved the credibility of the N-terminal amino acid residue in peptide. Lys and Gln in the peptide can be accurately distinguished by calculating the mass shift before or after the phosphorylation.

Acknowledgements

We are grateful to the Ministry of Science and Technology of China (2007AA02Z160, 2009ZX09501-004) and the Chinese National Natural Science Foundation (20672068, 20872077, 90813013) for financial support. The authors also would like to thank Dr. Deliang Cao (Medicine School of Southern Illinois University) for his assistance with editing the manuscript.

References

[1] M.R. Wilkins, E. Gasteiger, L. Tonella, K. Ou, M. Tyler, J.C. Sanchez, A.A. Gooley, B.J. Walsh, A. Bairoch, R.D. Appel, K.L. Williams, D.F. Hochstrasser, J. Mol. Biol. 278 (1998) 599–608.
[2] C.G. Fields, V.L. Vandrisse, G.B. Fields, Pept. Res. 6 (1993) 39–47.
[3] F.R. Masiarz, B.A. Malcolm, Methods Enzymol. 241 (1994) 302–310.
[4] J.W. Metzger, Angew. Chem. Int. Ed. Engl. 33 (1994) 723–725.
[5] P. Rydberg, B. Luning, C.A. Wachtmeister, L. Eriksson, M. Tornqvist, Chem. Res. Toxicol. 15 (2002) 570–581.
[6] R. Kellner, G. Talbo, T. Houthaeve, M. Mann, Tech. Protein Chem. 6 (1995) 47–54.
[7] D.J.C. Pappin, P. Hojrup, A.J. Bleasby, Curr. Biol. 3 (1993) 327–332.
[8] H.W. Lahm, H. Langen, Electrophoresis 21 (2000) 2105–2114.
[9] O.N. Jensen, A.V. Podtelejnikov, M. Mann, Anal. Chem. 69 (1997) 4741–4750.
[10] M.J. Chalmers, S.J. Gaskell, Curr. Opin. Biotechnol. 11 (2000) 384–390.
[11] J.R. Yates, S. Speicher, P.R. Griffin, T. Hunkapiller, Anal. Biochem. 214 (1993) 397–408.
[12] R.L. Beardsley, L.A. Sharon, J.P. Reilly, Anal. Chem. 77 (2005) 6300–6309.
[13] M. Yamaguchi, T. Nakazawa, H. Kuyama, T. Obama, E. Ando, T. Okamura, N. Ueyama, S. Norioka, Anal. Chem. 77 (2005) 645–651.
[14] J.L. Hsu, S.Y. Huang, J.T. Shiea, W.Y. Huang, S.H. Chen, J. Proteome Res. 4 (2005) 101–108.
[15] M. Yamaguchi, T. Obama, H. Kuyama, D. Nakayama, E. Ando, T.A. Okamura, N. Ueyama, T. Nakazawa, S. Norloka, O. Nishimura, S. Tsunasawa, Rapid Commun. Mass Spectrom. 21 (2007) 3329–3336.
[16] C.X. Zhou, Y.J. Zhang, P.B. Qin, X. Liu, L.Y. Zhao, S.H. Yang, Y. Cai, X.H. Qian, Rapid Commun. Mass Spectrom. 20 (2006) 2878–2884.
[17] M.B. Strader, N.C. Verberkmoes, D.L. Tabb, H.M. Connelly, J.W. Barton, B.D. Bruce, D.A. Pelletier, B.H. Davison, R.L. Hettich, F.W. Larimer, G.B. Hurst, J. Proteome Res. 3 (2004) 965–978.
[18] T.Y. Samgina, S.V. Kovalev, V.A. Gorshkov, K.A. Artemenko, N.B. Poljakov, A.T. Lebedev, J. Am. Soc. Mass Spectrom. 21 (2010) 104–111.
[19] A. Semmler, R. Weber, M. Przybylski, V. Wittmann, J. Am. Soc. Mass Spectrom. 21 (2010) 215–219.
[20] S. Ovenden, S.A. Fredriksson, C.K. Bagas, T. Bergstrom, S.A. Thomson, C. Nilsson, D.J. Bourne, Anal. Chem. 81 (2009) 3986–3996.
[21] T. Bergman, E. Cederlund, H. Jornvall, Anal. Biochem. 290 (2001) 74–82.
[22] T. Nakazawa, M. Yamaguchi, T. Okamura, E. Ando, O. Nishimura, S. Tsunasawa, Proteomics 8 (2008) 673–685.
[23] H. Kuyama, K. Sonomura, O. Nishimura, S. Tsunasawa, Anal. Biochem. 380 (2008) 291–296.
[24] Z.H. Huang, J. Wu, K.D.W. Roth, Y. Yang, D.A. Gage, J.T. Watson, Anal. Chem. 69 (1997) 137–144.
[25] W.B. Chen, P.J. Lee, H. Shion, N. Ellor, J.C. Gebler, Anal. Chem. 79 (2007) 1583–1590.
[26] D.X. Wang, S.A. Fang, R.M. Wohlhueter, Anal. Chem. 81 (2009) 1893–1900.
[27] M. Salek, S. Costagliola, W.D. Lehmann, Anal. Chem. 76 (2004) 5136–5142.
[28] Q. Fu, L.J. Li, Anal. Chem. 77 (2005) 7783–7795.
[29] P.T. Shen, J.L. Hsu, S.H. Chen, Anal. Chem. 79 (2007) 9520–9530.
[30] M.S. Cardenas, E. vanderHeeft, A.P.J.M. deJong, Rapid Commun. Mass Spectrom. 11 (1997) 1271–1278.
[31] Y. Chen, J.C. Zhang, J. Chen, X.Y. Cao, J. Wang, Y.F. Zhao, Rapid Commun. Mass Spectrom. 18 (2004) 469–473.
[32] J.Y. Bao, H. Fu, Y.Y. Jiang, Y.F. Zhao, Int. J. Mass Spectrom. 251 (2006) 82–84.