

A Systematical Analysis of Tryptic Peptide Identification with Reverse Phase Liquid Chromatography and Electrospray Ion Trap Mass Spectrometry

Wei Sun, Shuzhen Wu, Xiaorong Wang, Dexian Zheng, and Youhe Gao*

Proteomics Research Center/National Key Laboratory of Medical Molecular Biology, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences/Peking Union Medical College, Beijing 100005, China.

In this study we systematically analyzed the elution condition of tryptic peptides and the characteristics of identified peptides in reverse phase liquid chromatography and electrospray tandem mass spectrometry (RPLC-MS/MS) analysis. Following protein digestion with trypsin, the peptide mixture was analyzed by on-line RPLC-MS/MS. Bovine serum albumin (BSA) was used to optimize acetonitrile (ACN) elution gradient for tryptic peptides, and Cytochrome C was used to retest the gradient and the sensitivity of LC-MS/MS. The characteristics of identified peptides were also analyzed. In our experiments, the suitable ACN gradient is 5% to 30% for tryptic peptide elution and the sensitivity of LC-MS/MS is 50 fmol. Analysis of the tryptic peptides demonstrated that longer (more than 10 amino acids) and multi-charge state (+2, +3) peptides are likely to be identified, and the hydrophobicity of the peptides might not be related to whether it is more likely to be identified or not. The number of identified peptides for a protein might be used to estimate its loading amount under the same sample background. Moreover, in this study the identified peptides present three types of redundancy, namely identification, charge, and sequence redundancy, which may repress low abundance protein identification.

Key words: tryptic peptide identification, proteomics, reverse phase liquid chromatography, electrospray ion trap mass spectrometry

Introduction

In proteomics research, one of the most commonly used technologies is two-dimensional electrophoretic (2-DE) separations coupled to mass spectrometry (MS) identification of proteins. However, because some specific classes of proteins, such as very acidic or basic proteins, excessively large or small proteins, and membrane proteins, are known to be absent or under-represented in 2-DE patterns, other approaches are necessary to be the complement. Multidimensional protein identification technique (MudPIT; ref. 1), developed by Washburn and his colleagues, is such an approach. The protein mixtures are digested into peptides and separated by strong cation exchange (SCX) and reverse phase (RP) columns. Though the peak capacity of MudPIT is not as high as that of 2-DE, it can identify proteins that cannot be detected in 2-DE

and has been one of the most important approaches in proteomics research.

In the MudPIT method, each elution fraction from SCX column is further separated by RP column. Although it has been developed for a long time, there is no consistent elution method for peptide mixtures and the elution gradient of organic buffer can range from 30% to 60% (1–5). In proteomics research, an SCX-RP run will take about 24 hours (1); most of the time is used for peptides elution in RP column. So a suitable elution gradient for peptides needs to be optimized to save time without compromising the separation and ionization repression. Moreover, a consistent elution method may make different experiment results comparable. Of all the enzymes available, the most commonly used enzyme is trypsin. Trypsin has well defined specificity, yields tryptic peptides of an appropriate size for efficient MS analysis, and locates the basic residues at the terminus of the peptide.

* Corresponding author.

E-mail: gaoyouhe@pumc.edu.cn

So most of the tryptic peptides from different protein mixtures may present certain degree of similarity and can be separated with identical elution gradient, which makes it possible to develop a consistent elution method.

Although there have been studies on tryptic peptides by liquid chromatography and electrospray tandem mass spectrometry (LC-MS/MS), a more comprehensive and systematic analysis of tryptic peptides identified by LC-MS/MS remains helpful for the field. In this study, we systematically studied the elution method of tryptic peptides with bovine serum albumin (BSA) and retested the method with Cytochrome C. In addition, we also analyzed the characteristics of eluted peptides in our experiments.

Results and Discussion

Tryptic peptide elution gradient

In this study two known proteins, BSA and Cytochrome C, were used to develop the elution method for tryptic peptides. The base-peak chromatograms

of eight BSA-digested mixture runs and corresponding identified unique peptides were shown in Figure 1 and Table 1. When the final elution gradient of acetonitrile (ACN) was 5%–10%, 5%–15% and 5%–20% from 10 to 70 min, respectively, many eluted peaks appeared around 80 min, at which the ACN concentration was about 90%. The three above results show that none of the final ACN concentrations 10%, 15% and 20% could elute all peptides. For the other five ACN gradients (5%–25%, 5%–30%, 5%–35%, 5%–40%, and 5%–50%) from 10 to 70 min, the last peak was eluted around 72, 67, 57, 55, and 45 min, respectively, which meant that the last peak of ACN elution concentration was about 28.5% for 25%, 30% and 35% ACN gradients, and 31.5% for 40% and 50% ACN gradients. The identified peptides of eight runs were consistent with the appearance of eluted peaks (data not shown). The final ACN concentration of 30%–35% could elute all tryptic peptides, but higher final ACN concentrations could not help peptide separation and might cause co-elution and ion repression. Therefore, the elution gradient of ACN for digested peptides of BSA was 5%–30% or 5%–35%.

Table 1 The Identified Peptides from Eight BSA Peptide Mixture Runs

Peptide sequence	Number of amino acids	GRAVY value [#]	Charge state	Identified number
LPQKFPK	7	-1.014	+1	3
			+2	1
LVTDLTK	7	0.429	+1	2
YLYEIAR	7	-0.071	+2	2
AEFVEVTK	8	0.175	+1	23
QTALVELLK*	9	0.644	+2	5
KQTALVELLK*	10	0.190	+2	15
LVVSTQTALA	10	1.390	+2	1
FKDLGEEHFK	10	-1.250	+2	4
			+3	1
LVNELTEFAK	10	0.130	+1	5
			+2	16
HLVDEPQNLIK	11	-0.582	+2	15
YICDNQDTISSK	12	-0.833	+2	11
EYEATLEECCA	12	-0.625	+2	9
LKECCDKP LLEK	12	-0.617	+2	18
			+3	4
SLHTLFGDELCK	12	0.058	+2	10
			+3	4
LGEYGFQNALIVR	13	0.292	+2	17
DAFLGSFLYEYSR	13	-0.085	+2	1
LKPDPNTLCDEFK	13	-0.985	+2	3
			+3	6
ECCHGDLLECADDR	14	-0.621	+3	2

*Sequence redundant peptides. [#] Average predicted hydrophobicity for the peptides (7, 8).

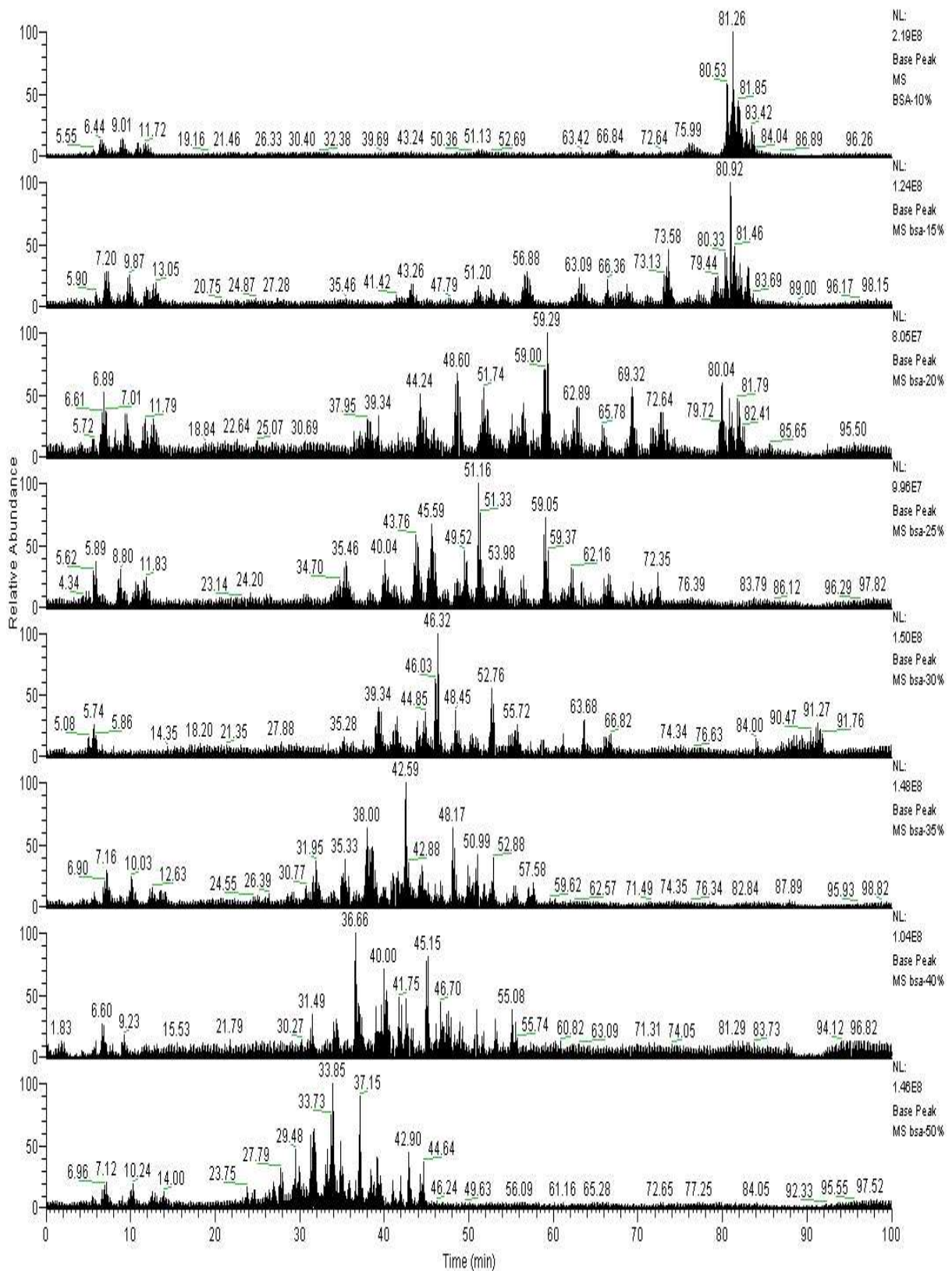


Fig. 1 The base-peak chromatograms of eight BSA-digested mixture experiments with different final gradients of ACN from 10% to 50% at 5 percentage increments.

The above 5%–30% ACN gradient was tested by Cytochrome C. The elution time was from 10 to 90 min. Figure 2 shows the base-peak chromatograms of Cytochrome C peptide mixture with sequentially titration from 50 to 1,000 fmol. The final peak was eluted around 62 min (22% ACN concentration) and the identified unique peptides were also consistent with the appearance of eluted peaks (data not shown). The results from Cytochrome C also proved that the elution gradient of 5%–30% could ensure complete peptide elution and should be a proper elution gradient.

Based on the above results in our study, we can conclude that the proper elution gradient of ACN for tryptic peptides is 5% to 30%. However, because our results only came from the digested mixtures of two proteins (BSA and Cytochrome C), whether it is suitable for complex protein mixtures or not needs to be further proved. Moreover, the gradient only resulted from tryptic peptides, while whether peptides from other enzymes could be best eluted by such gradient is also unknown. In general, according to our results and other reports (1–5), we recommend the ACN elution gradient be ranged up to 30%.

Apparatus reproducibility and sensitivity

We performed three times of LC-MS/MS analyses for Cytochrome C of 200 fmol (Figure 2). One of the peaks was eluted at about 52 min in three experiments and the corresponding standard deviation was 0.42 min. The elution time and standard deviation of other two main peaks were at 45 ± 0.58 min and 58 ± 0.21 min, respectively, which showed good reproducibility among different runs.

The number of identified peptides and the coverage with different loading amount of Cytochrome C are shown in Figure 3. When the loading amount was below 50 fmol, the coverage of identified peptides was $30.13\pm 2.7\%$ and the number was 4 ± 1 , so the detected loading amount could reach 50 fmol by using 0.18×100 mm capillary HPLC column and LCQ Deca XP^{plus} ion trap mass spectrometer. As an example of peptide identification, an LC-MS/MS analysis is shown in Figure 4. Figure 4B shows a single MS survey scan at the point of 62.29 min during the analysis. Figure 4C shows an MS/MS scan of the precursor ion, m/z 748.8, triggered from the survey scan, which demonstrates a good signal-to-noise ratio. Therefore, LC-MS/MS should be a highly sensitive method for protein detection.

Protein coverage and identified peptide number

It is known that the peptide hits of protein are related to protein's abundance, but there are only a few studies on it (6). In our study we addressed the question with known proteins. Figure 3A shows that the protein coverage with different loading amount was about 30% for 50 fmol, 50% for 100 and 200 fmol, and 60% for 500 and 1,000 fmol, respectively, which indicated that there was no linear increase in protein coverage with the increase of protein loading. The number of identified peptides with different loading amount is shown in Figure 3B. The numbers of identified peptides from 50 to 1,000 fmol were 4 ± 1 , 9.3 ± 1.15 , 15 ± 2.65 , 24 ± 2.65 , and 53.6 ± 23.46 , respectively. Although the number of identified peptides was not in proportion to the loading amount ($R^2=0.91$), it could be used as a rough estimate for preliminary protein quantitation because the number of peptides did increase with the loading amount.

However, we think the followings should be noticed when it is used as a quantitative method. First, different proteins might be cleaved to different numbers of tryptic peptides, so the number of identified peptides may also be different even with the same loading amount. Second, because peptide co-elution and following redundancy identification are often detected in ESI-MS/MS, and the peptides from other proteins may interfere with the identification of targeted peptide, the number of targeted peptides might be different in different peptide mixtures with the same loading amount. So we recommend that the number of peptides should only be used as a rough estimate of the quantity comparison method for one protein under the same sample background.

Characteristics of tryptic peptides

Although many proteomics researches were based on tryptic peptide identification, there is no comprehensive analysis of tryptic peptides up to now to our knowledge. We analyzed the characteristics of identified peptides from BSA and Cytochrome C on peptide length, charge state, and hydrophobicity (Figure 5). Figure 5A and 5B show the identified number of total and unique peptides with the amino acid number. For unique peptides, the percentage of peptides with 10–20 amino acids was about 65%. For total peptides, the percentage reached to 80%. The above results

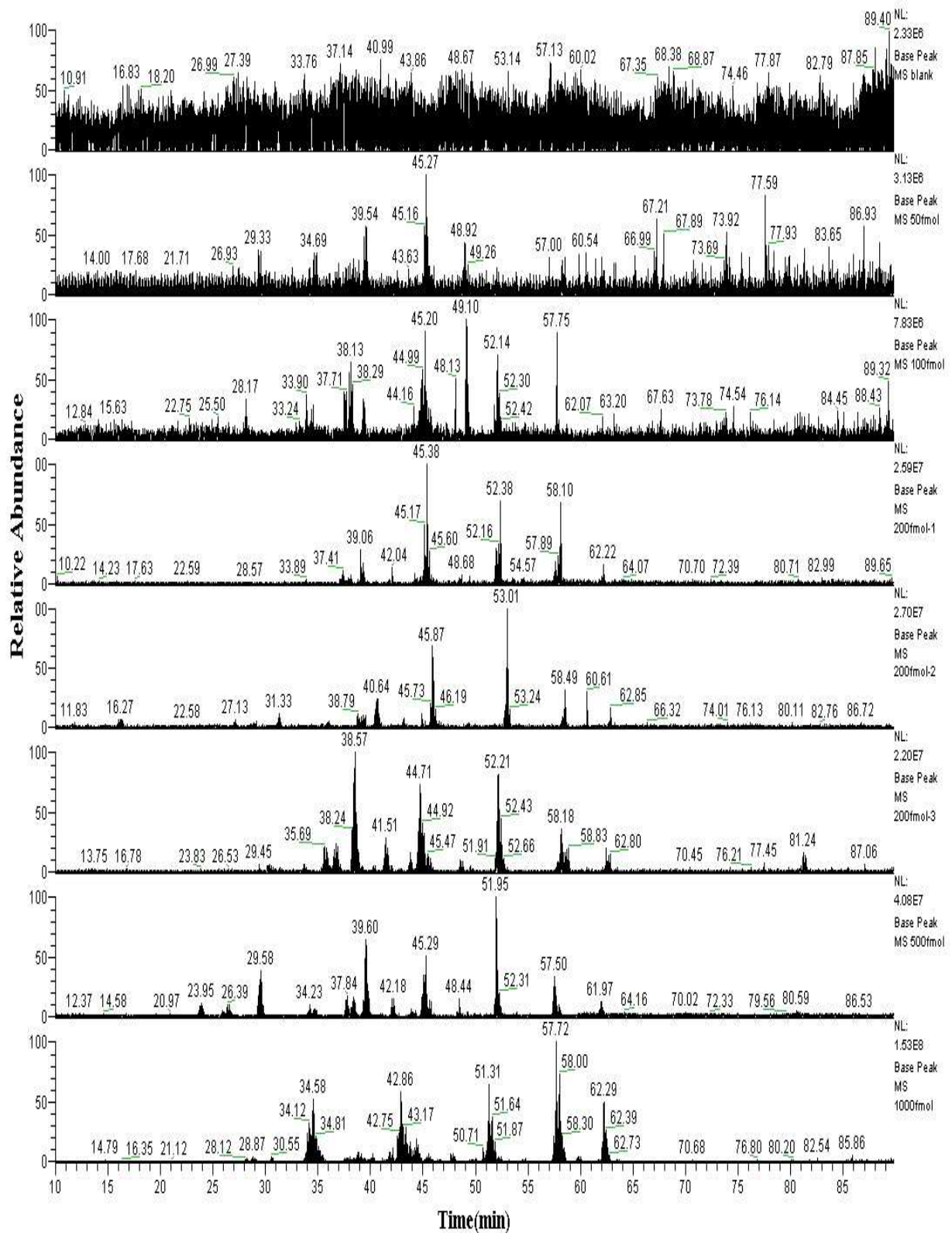


Fig. 2 The base-peak chromatograms of Cytochrome C peptide mixture with sequentially titration from 50 to 1,000 fmol. The amounts of loaded peptides are indicated.

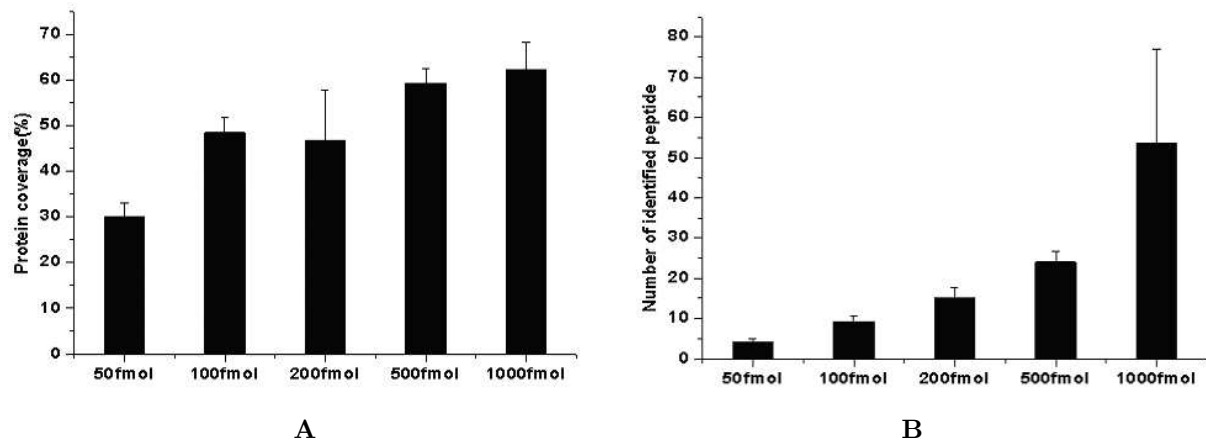


Fig. 3 The coverage (A) and the number (B) of different amount of peptides loaded from 50 to 1,000 fmol in Cytochrome C analysis.

demonstrate that longer peptides (longer than 10 amino acids) were easier to be detected by LC-MS/MS. The possible reason is that trypsin cleaves proteins by every 10–11 amino acids (lysine and arginine distribution probability in proteins), so the number of shorter peptides is relatively less than that of the longer ones.

The charge state distribution analysis in identified total peptides showed that more than 80% peptides were of +2 or +3 charge (Figure 5C). The reason is that a tryptic peptide often forms multi-charge ions with N-termini ammonia and C-termini or side chain basic residues (His, Arg, Lys) in acidic buffer (0.1% formic or acetic acid). And a fully tryptic peptide usually form +2 charge ion at N-termini ammonia and C-termini, so the number of +2 peptides is much more than that of +3 ones.

The hydrophathy analysis of unique peptides revealed that 75% unique peptides for Cytochrome C and 56% ones for BSA had GRAVY (Grand average of hydrophaticity; ref. 7, 8) value of -2 to 0 (Figure 5D). But there was a peptide (MIFAGIK) identified for 14 times with GRAVY value of +1.6. Therefore, whether a peptide is easier to be identified or not might not be related with its hydrophaticity.

Identified peptide redundancy

The complexity of identified peptides would be increased not only with the increase of sample complexity, but also with the increased internal redundancy. Although there was a research on this topic (9), a systematical analysis is still not reachable. We addressed this problem with two known protein identification results. According to the analysis result shown in Tables 1 and 2, the internal redundancy of identified

peptides can be classified into three categories. First, identification redundancy, which means that one peptide with one certain charge state can be repeatedly fragmented at different times. Many of the identified peptides during one run were identified for several times. For example, a peptide from Cytochrome C, KTGQAPGFTYTDANK, were identified with +2 charges for 20 times and with +3 charges for 21 times. Second, charge redundancy, which means that one peptide may be identified with different charge states (+1, +2, +3, *etc.*). In our experiments, 11 out of 34 unique peptides were fragmented with two different charge states, which proved it a common phenomenon in peptide identification by tandem mass spectrometry. Third, sequence redundancy, which means that more than one peptide may be identified from one root sequence due to different enzyme cleavage sites or ionization. The different retention time observed for TEREDLIAYLKK (70.13 min) and TEREDLIAYLK (79.88 min) suggest these two forms coexisted in solution, while the nearly identical retention time of the peptides TGPNLHGLFGRK (70.16 min) and TGPNLHGLFGR (70.28 min) suggest they might be formed by ionization.

The identification redundancy in BSA and Cytochrome C experiments was shown in Figure 6. In eight BSA experiments, the number of redundant peptides was close to that of unique peptides (Figure 6A). However, in three Cytochrome C experiments, with the increase of peptide concentration, the number of identified redundant peptides also increased dramatically (Figure 6B). When the loading amount increased from 50 to 1,000 fmol, the number of redundant peptides increased from 3 to 124 while the number of unique peptides increased from 5 to 16. So the redun-

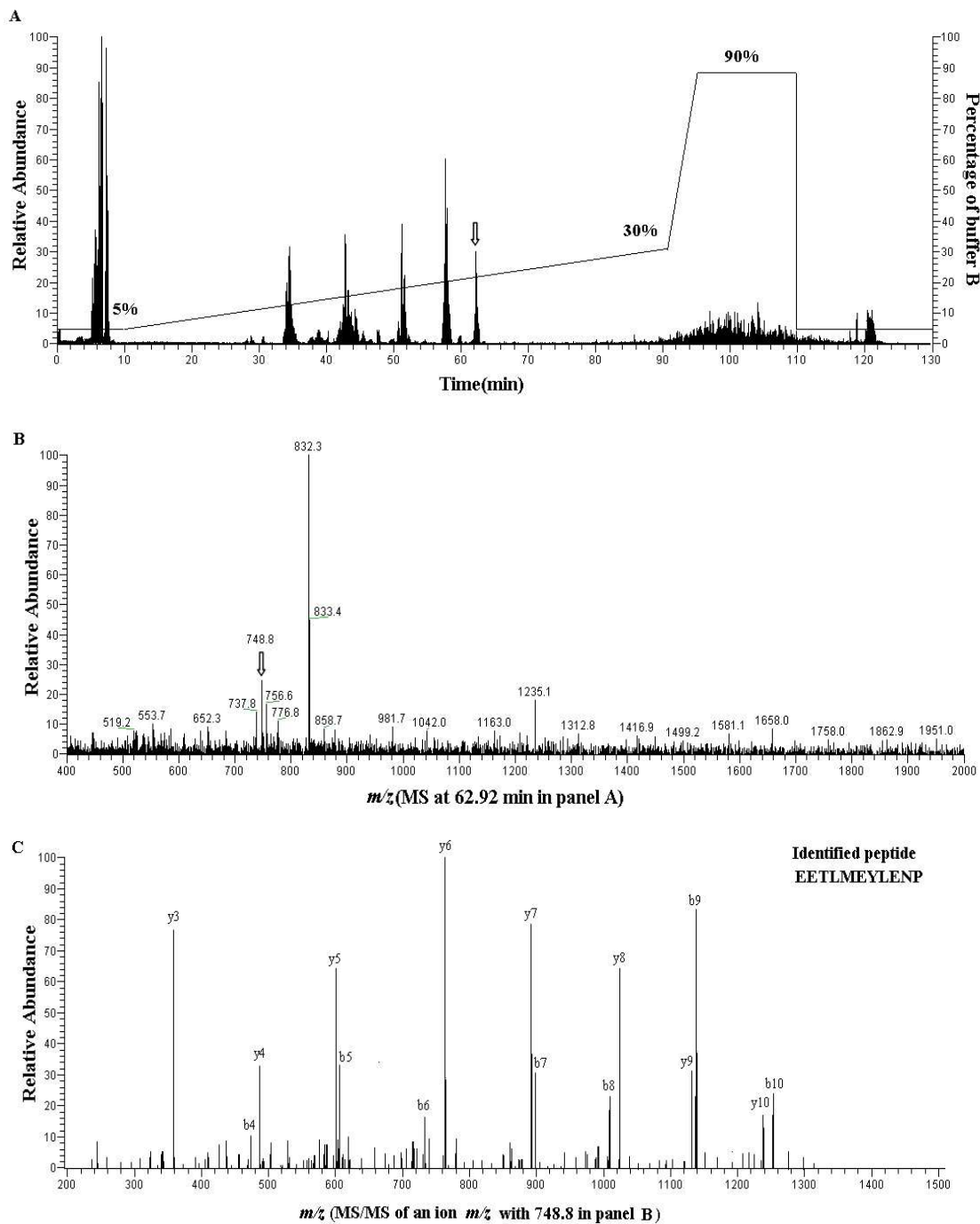


Fig. 4 An analysis of Cytochrome C by LC-MS/MS. **A.** Base-peak chromatogram of 1,000 fmol Cytochrome C during RP LC-MS/MS analysis. Peptides were loaded and desalted in the first 10 min and eluted in a 5%-30% gradient of buffer B over 80 min. Selected peptides were subjected to MS/MS analysis. **B.** An MS survey scan during LC-MS/MS analysis at time 62.29 min. The mass spectrometer sequentially selected three peptide ions for further sequence analysis of collision-induced dissociation. **C.** MS/MS scan of precursor ion m/z 748.8. This was one of the three peptide ions chosen for analysis that matched to the peptide amino acid sequence shown in the up right corner. Some b- and y- ions derived from the peptide ion are also indicated.

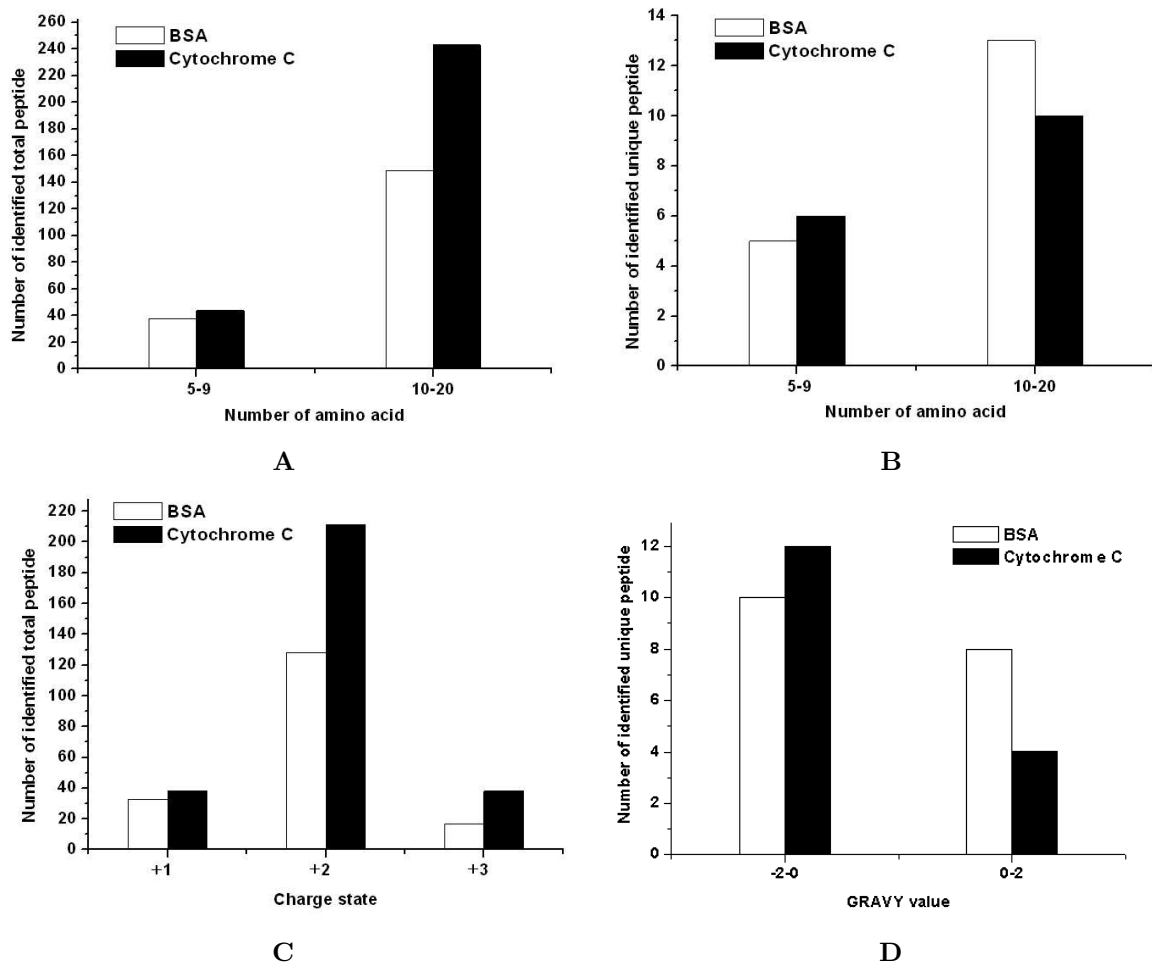


Fig. 5 The characteristics of identified peptides in BSA and Cytochrome C experiments. **A** and **B**. The length of identified total and unique peptides. **C**. The charge state of identified total peptides. **D**. The hydrophobicity of identified unique peptides.

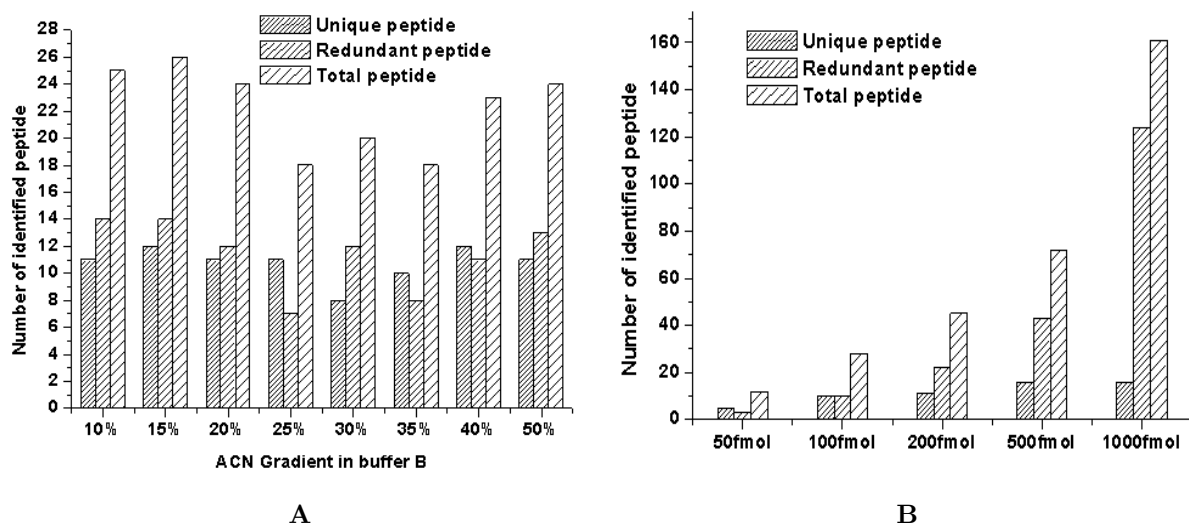


Fig. 6 The number of unique, redundant and total peptides identified in BSA (A) and Cytochrome C (B) experiments.

Table 2 The Identified Peptides from Fifteen Cytochrome C Peptide Mixture Runs

Peptide sequence	Number of amino acids	GRAVY value [#]	Charge state	Identified number
KIFVQK	6	0.033	+1	3
KYIPGTK	7	-1.043	+1	1
			+2	1
MIFAGIK* ¹	7	1.600	+1	14
MIFAGIKK* ¹	8	0.912	+2	5
EDLIAYLK* ²	8	0.212	+1	9
			+2	10
EDLIAYLKK* ²	9	-0.244	+2	11
TEREDLIAYLK* ³	11	-0.636	+3	5
			+2	23
TEREDLIAYLKK* ³	12	-0.908	+3	6
KTEREDLIAYLK* ³	12	-0.908	+2	10
TGPNLHGLFGR* ⁴	11	-0.391	+1	11
			+2	40
TGPNLHGLFGRK* ⁴	12	-0.683	+3	4
EETLMEYLENPK* ⁵	12	-1.292	+2	36
EETLMEYLENPKK* ⁵	13	-1.492	+2	24
GITWKEETLMEYLENPKK* ⁵	18	-1.156	+3	3
TGQAPGFTYTDANK* ⁶	14	-1.986	+2	30
KTGQAPGFTYTDANK* ⁶	15	-1.180	+2	21
			+3	20

*¹⁻⁶Sequence redundant peptides. # Average predicted hydrophobicity for the peptides (7, 8).

dancy of the peptides was highly related with the loading amount.

The presence of redundant peptides demonstrates the issue of increasing internal complexity without any new information being added, and it will also be detrimental to the analysis of tissue or body fluid proteome by preventing the selection of peptides from less abundant proteins. Therefore, eliminating any of these redundant peptides is essential for the selection, isolation, fragmentation, and identification of a less abundant ion, which is an important issue to be addressed for expression proteomics research.

Materials and Methods

Apparatus

An LCQ Deca XP^{plus} ion trap mass spectrometer and a 0.18×100 mm RP C18 (5 μm, 300Å) capillary column were purchased from Thermo Finnigan (San Jose, USA).

Reagents

Deionized water from a MilliQ RG ultrapure water system (Millipore, Bedford, USA) was used in all our experiments. HPLC grade acetonitrile and formic acid, ammonium bicarbonate, iodoacetamide and dithiothreitol were purchased from Merck (San Diego, USA). Sequencing grade modified trypsin, bovine serum albumin (BSA) and Cytochrome C (horse) were purchased from Sigma (St. Louis, USA) and used without further purification.

Sample preparation

BSA and Cytochrome C of 1 mg were dissolved in a 25 mmol/L ammonium bicarbonate buffer (pH 8.5) to a concentration of 1 mg/mL, respectively. The reduction of disulfide bonds was performed at 57°C for 60 min with 10 mmol/L DTT. After cooling the sample solution to room temperature, cysteine residues were alkylated with 50 mM iodoacetamide in the dark for 30 min. A 20-μg trypsin sample (1:50) was added to digest the proteins at 37°C overnight.

RP liquid chromatography

The tryptic peptides were acidified with formic acid to 0.1% and then loaded onto a 0.18×100 mm RP C18 (5 μm , 300Å) capillary column. For gradient elution, a 120 $\mu\text{L}/\text{min}$ flow from the pump was split to 2 $\mu\text{L}/\text{min}$ flow through the capillary column. Eight iterative BSAs (500 fmol) were loaded to capillary column and peptides were eluted from the column with a linear gradient of following methods from 10 to 70 min. Eight iterative elutions were run according to different gradients of buffer B: 5%–10%, 5%–15%, 5%–20%, 5%–25%, 5%–30%, 5%–35%, 5%–40%, and 5%–50% (A, 0.1% formic acid in water; B, 0.1% formic acid in ACN).

After the above gradient elution, the proper gradient was further tested by different concentrations (5, 10, 20, 50, and 100 fmol/ μL) of Cytochrome C with the same loading volume (10 μL). Each concentration was run for three times to test the reproducibility.

Tandem mass spectrometry and data processing

Eluting peptides were analyzed by an LCQ Deca XP^{plus} ion trap mass spectrometer equipped with an electrospray ion source. The mass spectrometer was operated in a data-dependent MS/MS mode, in which the precursor ion was selected from the previous full-scan mass spectrum. Peptide ions were detected in a survey scan from 400 to 1,500 amu (3 μ scan) followed by three data-dependent MS/MS scans (3 μ scans each, isolation width 3 amu, 35% normalized collision energy, dynamic exclusion for 3 min).

Tandem mass spectra were correlated using TurboSequest software with a database containing 4,400 sequences of horse and bovine proteins downloaded from National Center for Biotechnology Information (NCBI) web page. Output files from the correlation analysis were summarized using Xcorr and DeltaCn (1) to produce a list of identified peptides.

Conclusion

Data-dependent LC-MS/MS analyses were performed on standard proteins BSA or Cytochrome C to test the gradient of organic buffer in RP column. A gradient from 5% to 30% for ACN is suitable for tryptic peptide separation. Our analysis of identified peptides

shows that the number of identified peptides increased with the increase of loading amount and may be used as rough quantitation method under the same sample background. Moreover, longer (more than 10 amino acids) and multi-charge (+2, +3) peptides tend to be identified, and peptide hydrophobicity might not be related to identification. An important problem that needs to be solved for low abundance protein identification is the existence of three kinds of internal redundancy, namely identification, charge, and sequence redundancy.

References

1. Washburn, M.P., *et al.* 2001. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotech.* 19: 242-247.
2. Gatlin, C.L., *et al.* 2000. Automated identification of amino acid sequence variations in proteins by HPLC/microspray tandem mass spectrometry. *Anal. Chem.* 72: 757-763.
3. Pang, J.X., *et al.* 2002. Biomarker discovery in urine by proteomics. *J. Proteome Res.* 1: 161-169.
4. McDonald, W.H., *et al.* 2002. Comparison of three directly coupled HPLC MS/MS strategies for identification of proteins from complex mixtures: single-dimension LC-MS/MS, 2-phase MudPIT, and 3-phase MudPIT. *Int. J. Mass Spectrom.* 219: 245-251.
5. Peng, J.M., *et al.* 2003. Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the yeast proteome. *J. Proteome Res.* 2: 43-50.
6. Yates, J.R., *et al.* 2004. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal. Chem.* 76: 4193-4201.
7. Gasteiger, E., *et al.* 2003. ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* 31: 3784-3788.
8. Kyte, J. and Doolittle, R.F. 1982. A simple method for displaying the hydrophobic character of a protein. *J. Mol. Biol.* 157: 105-132.
9. Spahr, C.S., *et al.* 2001. Towards defining the urinary proteome using liquid chromatography tandem mass spectrometry. *Proteomics* 1: 93-107.

This work was partially supported by grants from the National Basic Research Program (973) of China (2004CB520804), and National Natural Science Foundation of China (No. 30270657, 30230150, 3037030).