Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Chromatography A, 1067 (2005) 73–83
JOURNAL OF CHROMATOGRAPHY A
www.elsevier.com/locate/chroma

Review

Proteomic LC–MS systems using nanoscale liquid chromatography with tandem mass spectrometry

Yasushi Ishihama*

Laboratory of Seeds Finding Technology, Eisai Co. Ltd., 5-1-3 Tokodai, Tsukuba, Ibaraki 300-2635, Japan

Available online 21 November 2004

Abstract

Current nano-scale liquid chromatography–tandem mass spectrometry (nanoLC–MS/MS) approaches in proteome research are reviewed from an analytical perspective. For comprehensive analysis of cellular proteins, analytical methods with higher resolution, sensitivity, and wider dynamic range are required. Miniaturized LC coupled with tandem mass spectrometry is currently one of the most versatile techniques. In this review, the current status of nanoLC–MS/MS systems as well as data management systems is addressed. In addition, the future prospects for complete proteomics are discussed.

Keywords: Proteome; Proteomics; NanoLC–MS/MS

Contents

1. Introduction

Efforts to miniaturize HPLC in the 1980's led to the development of packed microcolumns using fused silica capillaries with a 20–250 μm inner diameter and a flowrate of 0.02–10 μL/min to gain higher sensitivity with higher separation efficiency [1–8]. In such microscale systems, absorbance-based detectors such as UV detectors are not suitable because shorter light paths lead to less sensitive detection. On the other hand, low flowrates from microscale LC were compatible with mass spectrometers with fast atom bombardment (FAB) interfaces as described in 1985 [9] and subsequently applied to peptide mapping [10–12]. At the same time, the most important revolution in atmospheric pressure ionization for MS, electrospray ionization, was

* Tel.: +81 29 847 7192; fax: +81 29 847 7614.
E-mail address: y-ishihama@hhc.eisai.co.jp.

0021-9673/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.chroma.2004.10.107

developed by Yamashita and Fenn [13], which was recognized by the 2002 Nobel Prize in Chemistry. Although the early ESI interfaces allowed a flowrate of 1–10 μL/min, the miniaturization of the electrospray, e.g., microelectrospray [14] and nanoelectrospray [15] produced lower flowrates (<300 nL/min), which were also directly achievable with packed capillary columns of less than 150 μm i.d. Recently, nanoLC separation was coupled to another Nobel Prize winning technique, matrix-assisted laser desorption/ionization (MALDI) [16,17], by continuous spotting of the eluate from the LC onto the MALDI target plates [18–20].

Tandem mass spectrometry has been used as a microscale de novo sequencing tool for peptides because collision-induced dissociation (CID) followed by product ion scanning provides systematic fragment information of amino acid sequences [21]. Further improvement in peptide sequencing sensitivity was accomplished by the development of nanoelectrospray combined with a peptide sequence tag approach for protein identification in databases [22]. Because even one peptide is sufficient to identify a unique protein, this approach is more powerful for protein identification in proteome-scale experiments than the peptide fingerprinting approach where several peptide masses from one protein are used for identification.

The recent completion of genome sequencing and annotation of various organisms has increased the importance of analysis for possible gene products. One of the primary goals of proteomics is to address all functions in the cell at the protein level, including protein expression, modification, localization and protein–protein networks [23]. Therefore, one of the biggest challenges for analytical scientists is to analyze these proteomic samples, in which more than 10^6 human tryptic peptides may be present in concentrations varying by more than 10^6 in a human cell, with as much throughput and sensitivity as possible. To fulfill this analytical demand, two-dimensional gel electrophoresis has been used to separate proteins followed by in-gel digestion of excised spots and MALDI MS or nanoelectrospray MS/MS to identify these proteins. However, this approach has several disadvantages in that only a limited range of proteins are analyzable, it suffers from low dynamic range and lower throughput or difficulty of automation [24]. Nevertheless, a differing opinion championing its merits as a proteomics platform has also been published [25]. An alternate and perhaps more powerful approach is nanoLC combined with tandem mass spectrometry. Most of the current publications about large-scale protein identification are performed not by two-dimensional gel electrophoresis, but by nanoLC–MS/MS combined with different prefractionation approaches [26,27]. In addition, post nanoLC–MS/MS technologies such as database searching algorithms have been combined to increase the analytical performance of the method as a whole.

Here, current analytical technologies, including nanoLC–MS/MS, as well as the data management for proteomics are reviewed. Furthermore, the future demands and direction of this field is discussed.

2. LC–MS/MS as protein sequencer

One key technology in large-scale genome sequencing is a high-throughput DNA sequencer based on multiplex capillary electrophoresis to separate the fluorophore-tagged oligonucleotide ladders produced by the dideoxy method. Similarly, tandem MS allows for separating oligopeptide ladders, which are generated inside the mass spectrometer from parent peptides by CID, to determine the partial amino acid sequences of these peptides (Fig. 1). In the DNA sequencer, a nucleotide sequence can be determined by sequentially identifying the corresponding fluorophores in the electropherogram. Multiplexing capillaries with a 96 or 384 capillary format allows for high-throughput parallel analysis. In addition, low sample amounts can be overcome easily by PCR amplification. On the other hand, proteins are normally difficult to apply directly to a mass spectrometer to obtain sequence information because of the relatively lower efficiency of ionization and fragmentation, compared to smaller molecules such as peptides. Peptides with some length, produced by sequence-specific cleavage reactions such as trypsin digestion, are unique enough to identify their source proteins from fragmented ion spectra produced by CID. In addition, although CID does not always produce perfect peptide ladders, partial reaction products recorded in MSMS spectra are specific enough to identify one unique peptide out of the candidates obtained from a protein sequence database [28]. Therefore, digested peptides instead of proteins are generally analyzed with tandem mass spectrometry to identify proteins with the help of protein databases and the various search engines [29–31]. Because, however, digestion of protein mixtures provides a larger number of solutes with a wider dynamic range than that of DNA, more efficient approaches for introduction of the sample to the mass spectrometer are required to reduce sample complexity and to widen the dynamic range of analysis. Currently, direct coupling of nanoLC with tandem MS is the most powerful approach. ESI as well as MALDI interfaces are currently used between LC and MS. The main difference between ESI and MALDI for protein identification is the ease of fragmentation of peptides. ESI normally produces multiply charged parent ions, whereas singly charged peptides are dominant in the MALDI process. Therefore, in general, MSMS spectra by ESI have more fragmented ions than those by MALDI. In addition, the influence of ionization suppression should be considered when the samples are highly complex. In general, this suppression effect is more severe in MALDI than ESI. On the other hand, MALDI has advantages over ESI in terms of the flexibility of the front-end separation tool. Because the coupling between MALDI and LC is not perfectly on-line, parallel separation is easily accomplished. In addition, empty fractions caused by gradient delay and column equilibration/washing can be omitted from analysis. Furthermore, re-analysis of pertinent fractions and optimization of analysis conditions to improve data quality from previously analyzed spots is possible.

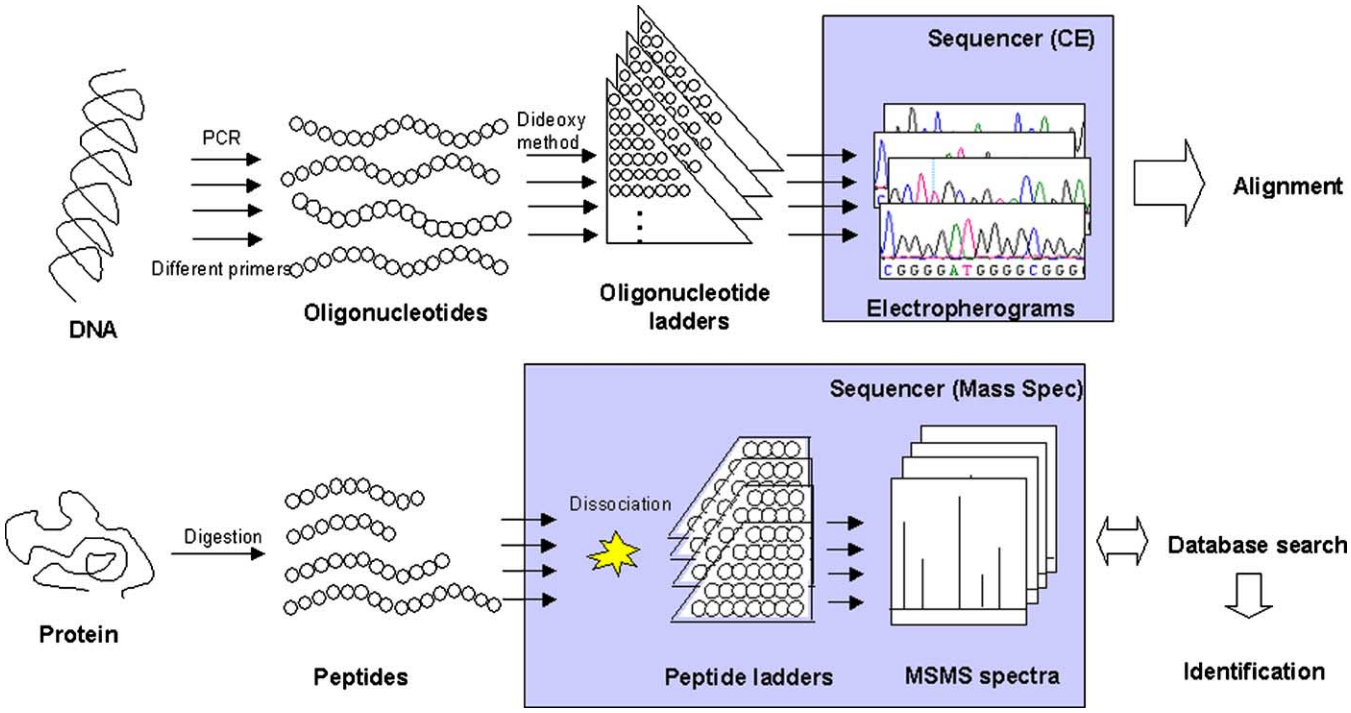

Fig. 1. Comparison between DNA and protein sequencing. Top: DNA sequencing by capillary array electrophoresis. Bottom: Protein sequencing by tandem mass spectrometry.

3. NanoLC–MS/MS systems

3.1. Column and interface to MS

Typical microcolumns for nanoLC are prepared using reversed phase materials with a 3–10 μm diameter packed into fused silica capillaries with a 12–100 μm diameter, in which sintered silica particles or silicate-polymerized ceramics have been used as frits [7,14,32–34]. The post-column connections affect peak broadening, and usually zero dead-volume unions are used, in which tubing ends are closely adjoined to one another. In this case, how the tubing is cut is critical to avoid peak broadening [35] and customized unions are used in some cases [36,37].

In ESI-MS, a spray needle with a tapered outlet is used as a restrictor for packed particles to prepare a fritless column. This approach is quite attractive for LC–MS because the post-column dead volume is minimized. So far, several groups have reported that the opening size of the column must be smaller than or equal to the average diameter of the packing materials to retain them [38–40]. The column, however, is easily blocked during the packing process [41]. This is because the particle size was almost equal to that of the outlet, and a single particle often completely blocked the outlet. To overcome this problem, we developed a “stone-arch” column, where the opening size is two- to five-fold larger than the average particle size, and particles at the end of the column arch over the opening and these self-assembled particles function as a frit. A similar approach is to make a silica frit at the outlet of the ESI needle, in which the assembled silica particles are fused by a pulsed laser beam to make permanent frits [42]. Monolithic columns are another type of fritless column. Both silica-based and organic polymer-based materials were reported [43,44]. It should be considered that the lower loading capacity of monoliths is a potential problem in some cases.

Generally, smaller columns at a lower flowrate combined with real nanoelectrospray conditions give higher sensitivity [36,45]. It is quite difficult, however, to routinely prepare packed columns with a small diameter (<30 μm), because particles stop in the middle of the column during packing. Removal of larger particles improves the efficiency of the packing [32,36]. Some groups reported that the use of loosely packed transfer tubing helps to pack smaller columns more tightly [42,46].

In ESI interfaces, liquid junctions are mostly used in both packed needles and columns with empty needles to apply the spray voltage. With the former, an inlet connector is used for the liquid junction using a metal union or a tee with a platinum electrode inserted. Because of the large difference in electrical resistance between liquid and gas, the drop in electrical field along the packed needle is negligible and sufficient voltage for spraying is maintained along the needles. In latter cases, conductive unions with distally coated emitters are most often used. Conductive tips, such as stainless steel needles [47] and glass tips coated with conductive materials [48–51], are also used.

There are two approaches of the liquid phase-MALDI interface. One is based on the on-line introduction of analytes

into MS using a continuous flow. Another is based on the off-line fractionation of analytes from LC eluent. In both modes, the MALDI matrix is mixed with analytes before introduction into the MS. In on-line aerosol liquid MALDI, an eluent mixed with the MALDI matrix is sprayed into a vacuum chamber of the MS to generate an aerosol, and the analyte is ionized from the aerosol by an irradiating laser [52]. The continuous flow MALDI with a porous frit at the capillary end inside the mass spectrometer, similar to continuous flow FAB, is also used for an on-line mode. The MALDI matrix is mixed with column eluent before crystallization occurs on the frit, which is used as a target for laser irradiation in MALDI [53,54].

In off-line interfaces, analytes are directly deposited onto the target. For a target, a moving belt of cellulose membrane containing the matrix was reported first [18]. Another approach is to use a rotation wheels in the vacuum region of the MS and to continuously deposit the mixture of analyte with the matrix [20]. Non-continuous deposition approaches using piezoelectric microdispensers [55] and electrosprayers [56] are also used. The simplest non-continuous deposition approach is to spot the mixture solution of analyte and the matrix on the target by controlling the distance between the flow output and the target or by applying an intermittent negative potential to the plates [19]. This approach is routinely used in many laboratories with commercial instruments. In these off-line approaches, the efficiency of the front LC separation is to some extent compromised. Because LC and MS become fully independent, however, more flexible MS analysis is possible, such as skipping or repeating analyses as described earlier.

3.2. Mobile phase conditions

In LCMS for peptides, acidic conditions with ion-pair reagents are usually used in combination with C18 stationary phases to suppress peak broadening. Trifluoroacetic acid (TFA) is one of the most popular reagents because of higher peak capacity with smaller peak width. However, signal suppression by TFA was often reported in LCMS analysis [57] although some groups use a 0.05–0.1% level [58] or lower [41] because signal suppression highly depends on the MS instrument. Wolters et al. reported great improvements by the addition of 0.02% heptafluorobutyric acid (HFBA) in LC/LC–MS/MS analysis [59] and later decreased down to 0.012% [60] or even without HFBA [61], and Peng et al. used 0.005% HFBA [27]. Other ion-pairs with longer fluorocarbon chains, including perfluorooctanoic acid, have also been used [57,62]. In our laboratory, 0.5% acetic acid is mainly used, and 0.005% HFBA as well as 0.005% TFA are used to increase the number of identified proteins by changing the retention behavior [63]. Modern C18 stationary phases have generally good features to suppress peak tailing with formic acid or acetic acid. In our case, “stone-arch” needles packed with ReproSil C18 give Gaussian symmetric peaks with approximately 5 s half-height peak widths (Fig. 2). Acetonitrile has been used as an organic solvent in many applications, but methanol gives better sensitivity in off-line infusion nanoES experiments; although in LC analysis the peak width is generally broader in methanolic buffers. Giorgianni et al. reported that Michrom Magic C18 with methanol gives higher sensitivity in peptide LCMS analysis [64]. The influence of the gradient time on sensitivity and the resolution were evaluated for a malaria proteome study [65]. Under the conditions employed, a 90 min gradient from 5 to 20% acetonitrile gave the best results in terms of the number of identified proteins. The optimum gradient is highly dependent on sample complexity as well as the amounts loaded because a shallower gradient gives better resolution, while a steeper gradient give better sensitivity. For neuropeptides with less complexity, a steeper gradient gives better results, as described by Haskins et al. [66].

3.3. Injection system

Because of the low flowrate, injectors with a smaller dead volume such as injection valves with 100 μm bore and 20–25 μm i.d. transfer lines, should be used for nanoLC. In typical cases, the proteomic sample size ranges from a few microliters to 100 μL. Trap columns are useful to reduce the injection time. Because the diameter of the analytical column is quite small, the size of the trap should be carefully selected to provide sufficient loading capacity whilst maintaining separation efficiency [67]. In general, it is difficult to prepare a reliable short trap less than 5 mm length (0.3–1 mm i.d.) reproducibly. It was also reported that the use of longer trap columns with smaller i.d. affects the elution order (trap with 100 μm i.d. and 25 mm length combined with analytical column with 75 μm i.d.) [68]. Licklider et al. reported an interesting system called a v-column, in which the trap columns and analytical columns are directly connected via a tee with an open/close switch [69]. A similar system was also reported using a custom-made butt tee connector between the trap and column [37]. We developed triple columns, where strong cation exchange chromatography (SCX) is inserted between the C18 trap and the C18 analytical column, and the waste line is between the SCX and the analytical C18 columns [70]. Then hydrophilic as well as hydrophobic contaminants, such as Coomassie dye, can be removed. These systems are not practical, however, because of the difficulties in assembly.

Direct injection was accomplished by concentrating sample volumes using pipette tip-based microcolumns. For instance, stop and go extraction tips (StageTips) are used because of the higher capacity, higher recovery, and smaller elution volumes required [71]. These tip-columns allow samples to be processed in parallel, and consequently, reduce the total analysis time. In addition, the robustness of the LC system is improved by filtering the sample solutions. A splitter between injector and column was effective in avoiding the influence of dead volume inside the injector if the pump can generate the direct flow for loading as shown in Fig. 3.

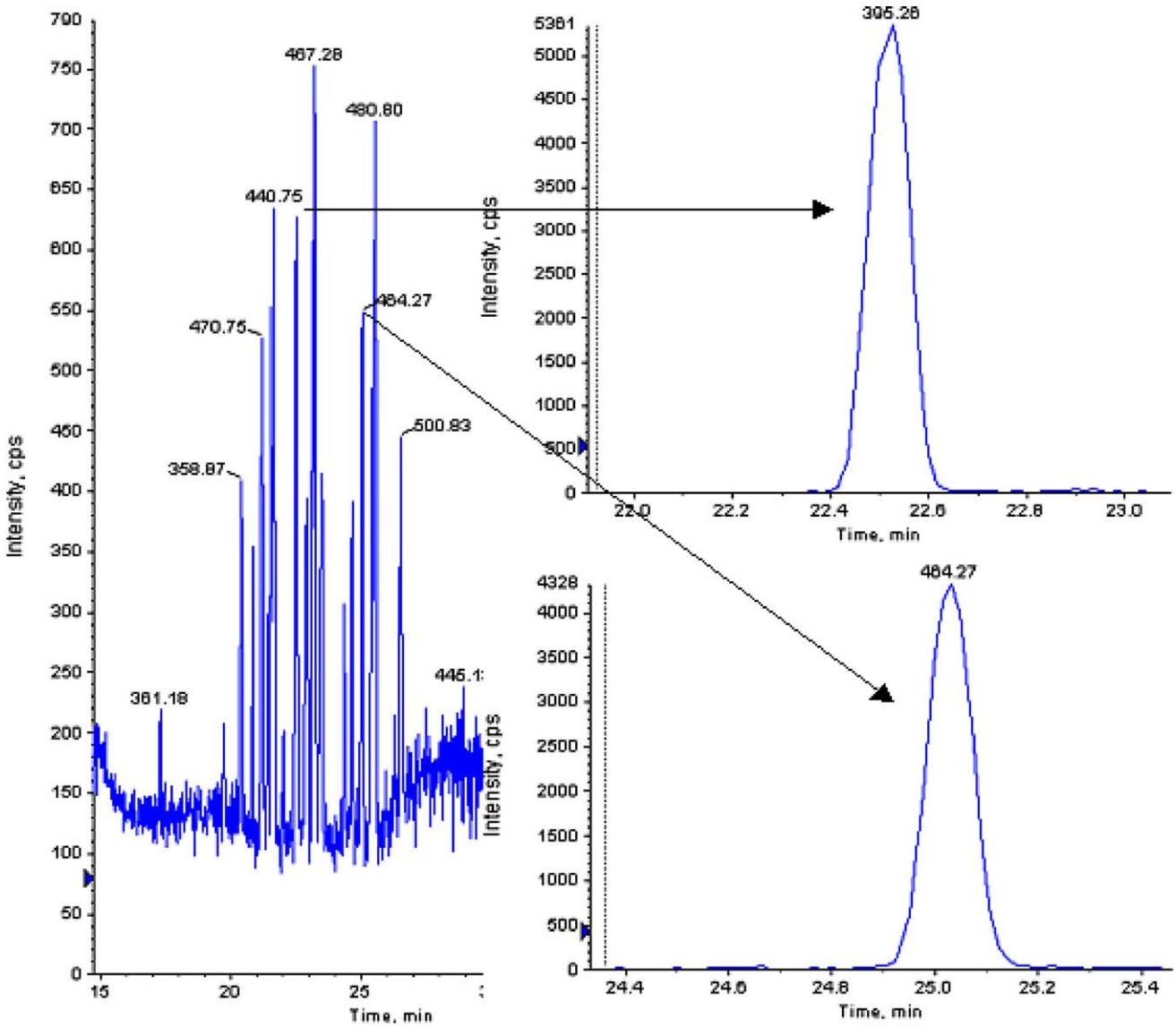

Fig. 2. NanoLC–MS/MS analysis for 50 fmol tryptic digest of human serum albumin. Conditions: column, arch-stone C18 (100 μm i.d., 150 mmL, 6 μm opening, ReproSil C18-3 μm); flowrate, 300 nL/min; mobile phase, 0.5% acetic acid with acetonitrile; gradient, acetonitrile 4–24% in 10 min; MS, AB-Sciex QSTAR pulsar i.

Direct loading is also possible using an air-pressured cell where the column is directly immersed into the sample solution [38]. Because this is a "true" direct loading from sample solution to the column without transfer tubing, carry-over caused by the injector is avoided. In addition, the loading time is negligible when loading is performed during another analysis using another column. However, it has the disadvantage of being a fully manual process.

3.4. Nano-flow gradient system

Few commercial pumps can generate low flowrates of less than 1 (L/min in a gradient mode. Two types of nanoflow pumps are currently available. One is a split type, where a splitter divides the higher flowrate generated by the pump into nanoflow. Because the split device has a flow monitoring function, the final flowrate becomes constant during gradient elution, even if the backpressure of the column changes. A simple homemade splitter consisting of a tee and a restriction line has also been used in many applications. Without a feedback system, however, the split ratio changes during gradient elution. Therefore, one serious problem that can occur during automated analysis is to lose all samples if the column becomes blocked, because the total backpressure does not change when the column is blocked. Therefore, filtration of samples helps to make the system robust. StageTip has been used not only to desalt but also filtrate samples prior to nanoLC–MS/MS analysis [65,72–74].

Another type of pump utilizes direct flow without flow splitting. Generation of a low flowrate less than 1 μL/min in gradient mode is accomplished using a large mixing chamber in which the initial solvent is exponentially replaced with

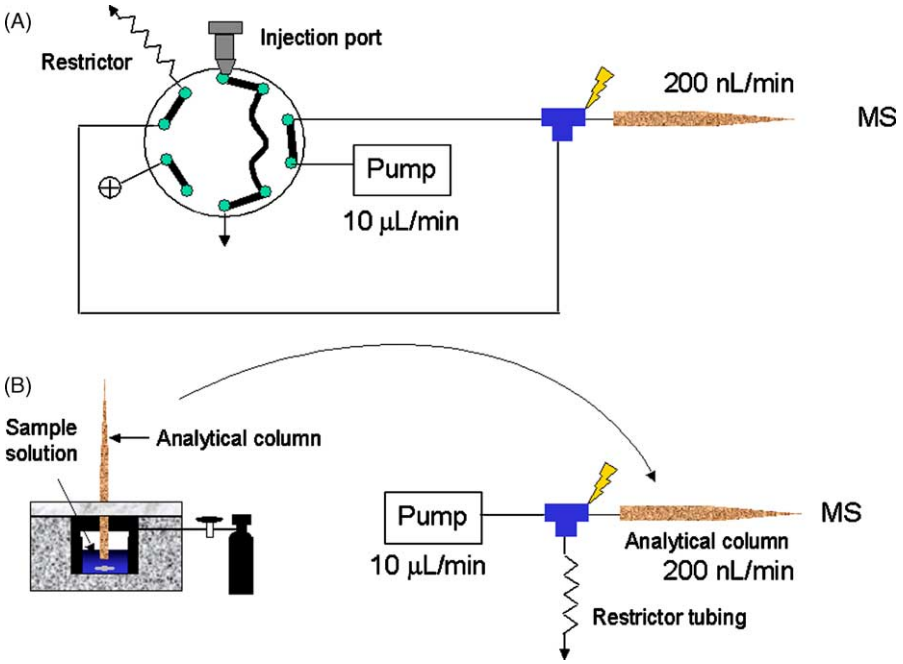

Fig. 3. Direct injection systems without trap columns. (A) Using a 10-port injection valve. (B) Using an air-pressure cell.

the final solvent [75]. Both the flowrate and the volume of the mixing chamber control the gradient curve profile. This system was miniaturized for nanoLC [76,77]. Shen et al. reported the use of ultrahigh pressure-tolerant pump systems [78]. Although the retention time reproducibility was not satisfactory due to the constant pressure conditions in their systems, the run-to-run difference in retention times was minimized using a genetic algorithm [58]. We accomplished the linear gradient using split tubing array (STAR) systems [79] (Fig. 4). Natsume et al. also developed a multistep gradient with a relatively large mixing cell to obtain a linear gradient [45]. Stepwise gradient elution is performed using two filled loops with two different solvents in microLC [80]. Using the same principle, linear gradient elution in nanoLC was accomplished using a loop filled with solvents from a conventional gradient pump [81,82]. This was also applied in a ultrahigh pressure nanoLC system [83].

3.5. *Multidimensional separation*

Although current nanoLC–MS/MS has a throughput of approximately 2000–4000 peptides per run within 1–2 h, it is not sufficient to analyze complex peptide mixtures such as that obtained from a whole cell lysate. Therefore, additional steps prior to nanoLC–MS/MS are necessary to reduce the complexity as well as the dynamic range of the peptide abundance. Subcellular fractionation using ultracentrifugation and sucrose gradient separation effectively increases the number of identified proteins and purifies the cellular component of interest. Protein separation or selective enrichment using immunoprecipitation, size exclusion chromatography, ultrafiltration, SDS-PAGE, ion-exchange chromatography, chromatofocusing, and isoelectric focusing (IEF) have been reported. Peptide separation after digestion is also performed using orthogonal modes of reversed phase separation that is used for final steps before MS/MS. So far, SCX is mostly often used [26,84,85]. Theoretically, separation at the protein

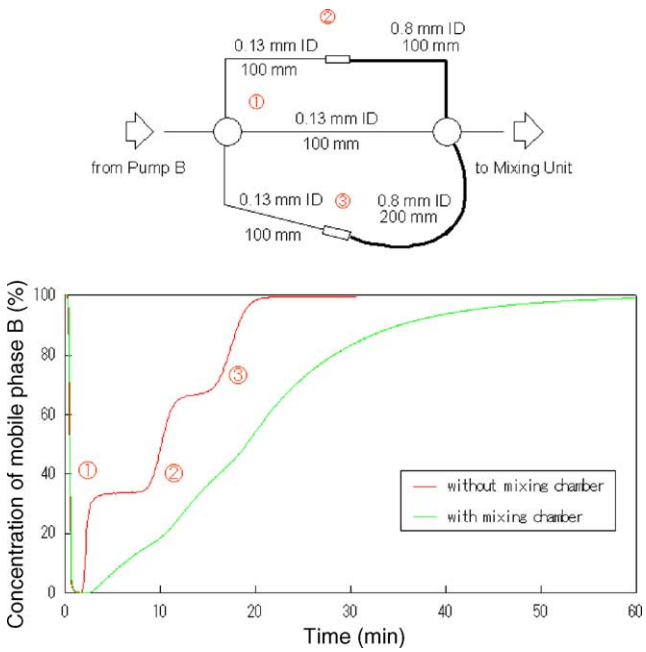

Fig. 4. Split tubing array (STAR) gradient systems. Top: STAR gradient generator using three split tubes with different restrictions and volumes. Bottom: Obtained gradient profiles with and without a mixing chamber.

level is more effective than that at peptide level in reducing the dynamic range problem as simulated in Fig. 5. Actually, a GeLCMS approach, in which SDS-PAGE followed by slicing the gels, digeston, and LC–MS analysis, was performed, which solved the dynamic range problem in the malaria proteome by isolating huge amounts of hemoglobins from red blood cell samples [65]. On the other hand, the complexity problem can be solved by fractionation at both the protein and peptide levels. This leads to an increase in the total analysis time and the number of identified proteins becomes saturated as fraction number increases [27,86]. Therefore, selection of the fractionation method is important to maximize the efficiency of the identification process, and two factors, peak resolution and orthogonality to C18 separation, should be considered for peptide fractionation. For example, in SCX, to increase the resolution, linear gradient salt elution is preferable to step gradients, and the addition of organic solvents is effective to suppress hydrophobic interactions, i.e., to increase the orthogonality to C18 separation. Therefore, working in an off-line mode would more easily achieve the optimum conditions. In addition, in an off-line mode, larger bore columns can be used to handle larger sample volumes to increase the dynamic range, whereas the on-line mode can be fully automated and provides potentially more reproducible results. Recently, IEF separation for peptide prefractionation was reported as an alternative to SCX [87].

For different situations, such as immnoprecipitation experiments where only a few micrograms of moderately complex samples are available, a more flexible off-line system may more easily be adapted to the best fractionation conditions. We employed a StageTip with C18/SCX/C18 stacked disks and the resultant four fractions increased the number of identified peptides by up to 240% for *Escheriachia coli* soluble lysate [63].

Selective enrichment at the peptide level reduces the complexity for ICAT peptides, which are biotin-modified cysteines, and an avidin column was used to fish out the ICAT peptides in combination with SCX and C18 separation [27]. Phosphopeptides were also enriched using pipette-tip-based immobilized metal-ion-affinity chromatography (IMAC) column [88]. Methylation of carboxyl moieties effectively reduced the non-phosphopeptide and IMAC interaction [89]. We also performed phosphopeptide fishing using a C18/titania/C18 StageTip and phosphopeptides in whole cell lysate were successfully enriched (Fig. 6) [90].

3.6. Mass analyzers

Mass analyzers of various design and performance are currently used for proteome research [23]. Factors for comparison are sensitivity for resolution of peptides, mass accuracy, and the ability to generate information-rich peptide fragment ion mass spectra. In general, ion-trap (IT) instruments are relatively robust, sensitive, and inexpensive. In addition, they generate more fragment ions and even MS^n ($n > 2$) is possible, although resolution and mass accuracy is relatively lower. The linear IT, where ions are stored in a cylindrical volume that is considerably larger than that of the traditional three-dimensional ITs, was recently introduced with increased sensitivity, resolution and mass accuracy. This instrument was modified to perform electron transfer dissociation (ETD) to generate c, z-series fragment ions and successfully applied to phosphorylated peptide analysis without loss of a phosphomoiety during fragmentation [91].

Fourier Transform Ion Cyclotron Resonance MS (FTMS) captures the ions under very high vacuum in a high magnetic field. The advantages are ultrahigh mass accuracy, resolution, sensitivity, and dynamic range. However, the instrument is complex and requires constant maintenance to keep the per-

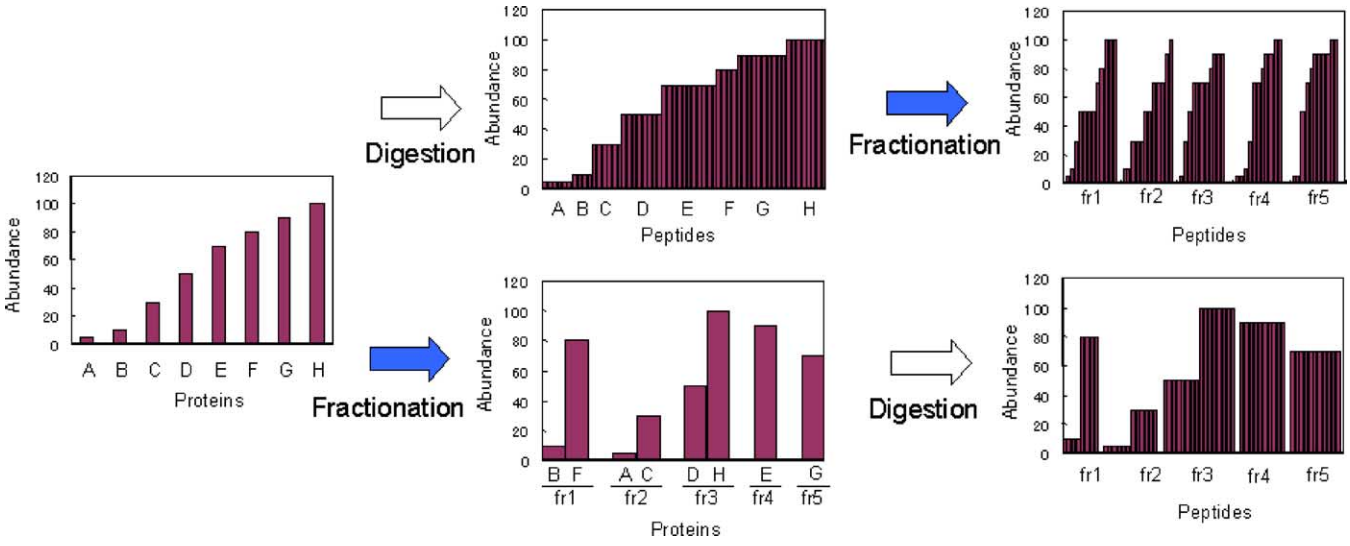

Fig. 5. Simulated prefractionation of eight proteins with different amounts into five fractions. Top: Fractionation at the peptide level after digestion. Bottom: Fractionation at the protein level before digestion. Simulated fractionation was performed using a random sampling function in Microsoft EXCEL.

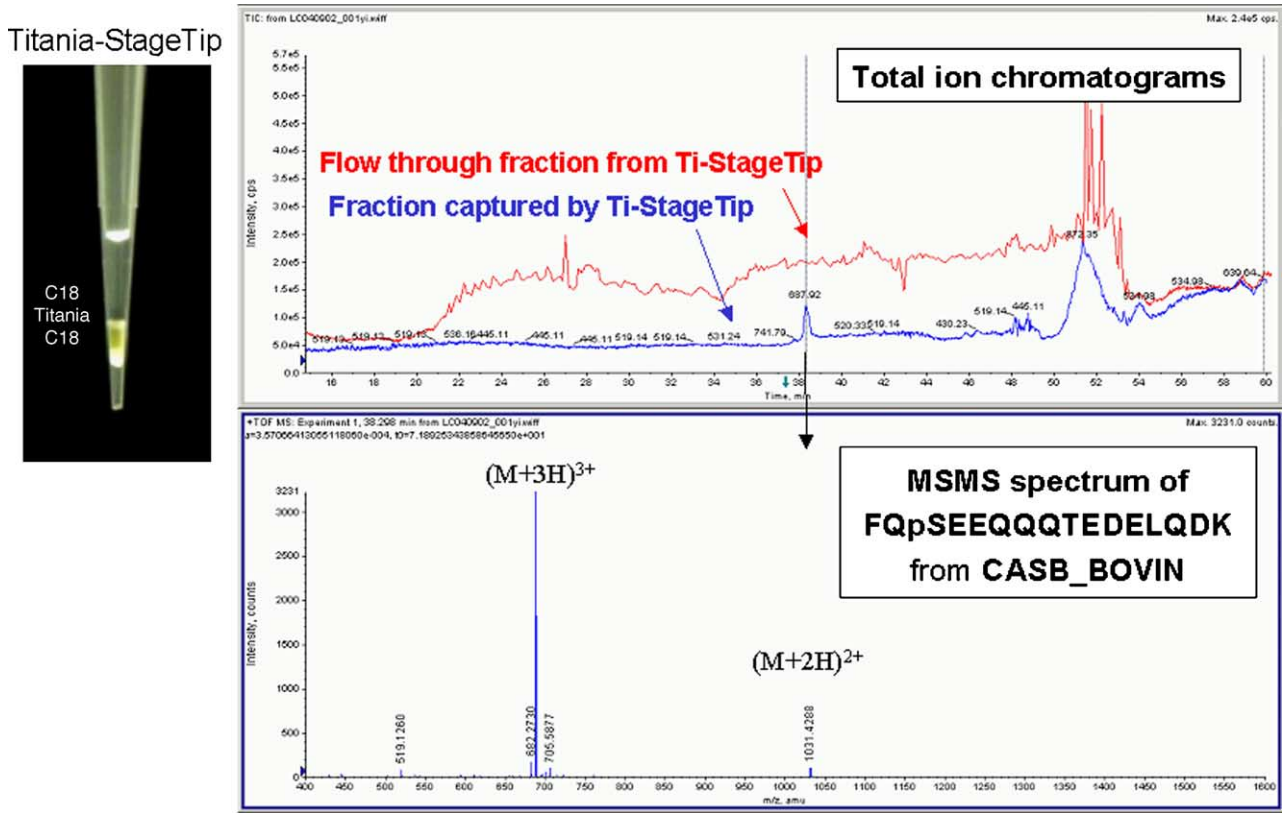

Fig. 6. Phosphopeptide enrichment using C18/titania/C18 StageTip. Left: C18/titania/C18 StageTip. Right: LC–MSMS analysis for tryptic peptides from Hela cell lysate spiked with β-casein. Right-top: Total ion chromatograms of the flow through and the enriched fractions. Right-bottom: MSMS spectrum of an enriched peak. Mascot protein identification engine identified phosphoserine-containing peptides from β-Casein.

formance. In addition, fragmentation efficiency is generally not very high. So far, FTMS has not been used routinely in proteomics. Very recently, a hybrid linear IT–FTMS was introduced, in which MS/MS is performed in an IT and FT is used for high accuracy simultaneous measurement of parent ions [92,93].

Time-of-flight (TOF) instruments are used with MALDI as well as ESI. TOF also has high mass accuracy, resolution and sensitivity. For MS/MS measurement, it requires another analyzer and CID source in front of TOF. Hybrid quadrupole-TOF, IT–TOF and TOF–TOF have been developed with ESI and MALDI interfaces.

While TOF, IT, and hybrid TOF instruments are currently widely used, the development of new instruments with higher performance is quite rapid because of the increasing demands of proteomics. Along with LC having a higher resolution, faster scanning capabilities will be more important in the near future.

3.7. Data analysis

Data analysis is a key step in "-omics" research because huge amounts of information-rich data are easily generated. The first post-MS step is to produce peak lists from MS raw data consisting of MS scans with three-dimensional axes (time, *m*/*z*, and ion counts) and MSMS scans with parent ion masses, acquired time, and fragment ions with ion counts. Peak extraction is generally performed using scripts attached to the MS operation software or identification software. However, the quality of peak extraction software varies with each MS instrument. In our laboratory, therefore, we developed our own software that is used to generate the same quality of MSMS peak lists from various MS instruments from different vendors including quadrupole-TOF, IT, and TOF–TOF. The peak lists are then automatically submitted to database search engines for protein identification.

There are two types of search engines developed so far to identify proteins via tandem mass spectra. The "peptide sequence tag" approach uses partial sequence information as well as the parent ion mass and the sequence specificity of the cleavage reaction are used as "tags" to constrain searches of the sequence database [28]. It requires a pre-interpretation step to obtain these tags before database searching, although this can be automated [94,95]. On the other hand, another approach does not require any pre-steps before starting the database search because these algorithms are based on comparisons between observed and theoretical spectra. The program based on cross-correlation between observed and theoretical spectra is known as Sequest [29], while other programs based on probability of random match-

ing between the measured and the theoretical peaks have been commercialized as Mascot [30] and Sonar [31], for instance. These fully automated search engines are now widely used for large-scale protein identification via LC–MS/MS data and for searching sequence databases. The output, however, should be manually verified to avoid false positive proteins in the list. A 1% false positive rate at peptide level leads to more than a 10% false positive rate in protein levels if manual verification is not performed in the case of large-scale yeast proteome [27], because the present automated algorithms are not versatile enough to accurately distinguish false positive identifications from true positive ones especially when the quality of MS/MS spectra is poor. Other concerns are to remove some constraints or to add possible variations such as cleavage specificity or additional modifications, because it would also cause a drastic increase in the false positive rate. Recently, a hybrid linear IT–FTMS was employed to evaluate the specificity of trypsin cleavage using IT fragmentation with ultrahigh accuracy of parent ions measured by FTMS [93]. As a result, under the conditions employed, they concluded that trypsin cleaves C-terminal to both Arg and Lys, exclusively.

The next step is to validate the results, remove the redundancy, and quantify proteins if necessary. The use of multiple algorithms based on different principles would be helpful to validate protein identification. Other parameters such as mass accuracy of parent ions, peptide retention times, isoelectric points if IEF is used, and protein molecular weight if protein separation based on the size is performed are also helpful to remove the false positives. In quadrupole-TOF instruments, re-calibration using top-ranked peptides improves the mass accuracy an average of up to 10–20 ppm [65]. Software such as MSQuant (http://msquant.sourceforge.net) can perform recalibration automatically. Approaches based on estimation of peptide retention times in reversed phase separation were developed in the 1980s, based on the amino acid composition and other parameters by Meek [96] and Sakamoto et al. [97]. Using current proteomic LCMS, it is much easier to obtain datasets of more than 1000 peptides in a single run. Recently, Palmblad et al. used the estimated retention times for protein identification [98]. Also Petritis et al. estimated the retention times using a neural network based on Meek's equation [58]. The coefficients for the equation depend on the LC system including the mobile and stationary phases. In our laboratory, a peptide mixture from digested whole cell lysate was prepared and analyzed by LCMS with 90 min linear gradient elution. Usually more than 3000 sequencing events were performed and approximately 1500 peptides were identified with 95% confidence. Using this dataset, multiple linear regression analysis was performed to calculate Meek's coefficients for each of the 20 amino acids. The obtained results were used to eliminate the false positive proteins (Fig. 7).

Grouping of redundant proteins or protein families is based on shared peptides. This should be done carefully for quantitation because two isomers with different expressions often have shared regions. Then, biological information such as cellular localization, the biological process, and the molecular function are added to the identified proteins. These data validation and mining steps should be performed automatically to avoid human error. The tools, however, are not standardized at present, with individual laboratories developing their own tools and some commercial products are now being introduced.

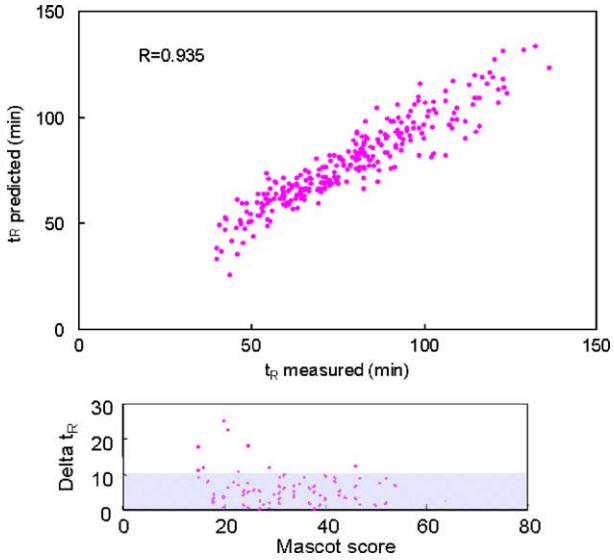

Fig. 7. False positive removal using predicted retention times. Upper: Relationship between predicted and measured retention times of peptides from *E. coli* soluble lysate. The prediction was based on Meek's equation [96]. Lower: Relationship between the differences between predicted and measured retention times and Mascot probability scores of peptides. The outside of the shadow zone indicates false-positive identified peptides.

4. Conclusion

Current nanoLC–MS/MS technology has been successfully applied to proteomics research and has provided dramatic improvement in protein identification, although complete coverage of the proteins for any organism has not yet been accomplished. Analytical challenges remain in trying to resolve the dynamic range problem as well as the complexity as mentioned. Currently, the Human Proteome Organization (HUPO) is focusing on human plasma proteome (http://www.hupo.org/hpp/hppp.htm), where these problems are emphasized [99]. In addition, large numbers of false positive proteins have been reported as identified proteins from different laboratories because of different criteria for protein identification using different search engines. A non-redundant protein list obtained from four different approaches is helpful in reducing the number of false positive identifications [100]. The Proteomics Standards Initiative has been organized in HUPO to standardize various aspects including a data format [101]. Quantitation is another hot issue in this area although it is not reviewed in this article. Comprehensive gene expression analysis is now easily per-

formed using DNA microarray as well as RT–PCR. Unlike the transcriptome, proteomic quantitation is not sufficient enough in terms of the coverage, although stable isotope labeling approaches are widely employed [102–104]. Another application of LCMS in proteomics would be the study of protein–protein interactions [72,105]. While proteomics has so far provided significant insights in cell biology, extensive improvement in analytical science is still required for comprehensive understanding of cellular function.

Acknowledgments

I thank Professor Matthias Mann and all the members in the Center for Experimental BioInformatics (CEBI, University of Southern Denmark-Odense) for part of my proteome research, Shao-En Ong (CEBI) for critical reading of the manuscript, and Hiroyuki Katayama, Yoshiya Oda, and other members in the LSFT (Eisai) for fruitful discussions. I also thank Eisai for giving me a chance to take a sabbatical at CEBI.

References

[1] T. Takeuchi, D. Ishii, J. Chromatogr. 190 (1980) 150.
[2] T. Takeuchi, D. Ishii, J. Chromatogr. 218 (1981) 199.
[3] T. Takeuchi, D. Ishii, J. Chromatogr. 213 (1981) 25.
[4] T. Takeuchi, D. Ishii, J. Chromatogr. 238 (1982) 409.
[5] F.J. Yang, J. Chromatogr. 236 (1982) 265.
[6] D.C. Shelly, J.C. Gluckman, M. Novotny, Anal. Chem. 56 (1984) 2990.
[7] R.T. Kennedy, J.W. Jorgenson, Anal. Chem. 61 (1989) 1128.
[8] K.-E. Karlsson, M. Novotny, Anal. Chem. 60 (1988) 1662.
[9] Y. Ito, T. Takeuchi, D. Ishii, M. Goto, J. Chromatogr. 346 (1985) 161.
[10] R.M. Caprioli, B. DaGue, T. Fan, W.T. Moore, Biochem. Biophys. Res. Commun. 146 (1987) 291.
[11] L.J. Deterding, M.A. Moseley, K.B. Tomer, J.W. Jorgenson, Anal. Chem. 61 (1989).
[12] W.J. Henzel, J.H. Bourell, J.T. Stults, Anal. Biochem. 187 (1990).
[13] M. Yamashita, J.B. Fenn, J. Phys. Chem. 88 (1984) 4451.
[14] M.R. Emmett, R.M. Caprioli, J. Am. Soc. Mass Spectrom. 5 (1994) 605.
[15] M. Wilm, M. Mann, Int. J. Mass Spectrom. Ion Processes 136 (1994) 167.
[16] K. Tanaka, H. Waki, Y. Ido, S. Akita, Y. Yoshida, T. Yoshida, Rapid Commun. Mass Spectrom. 2 (1988) 151.
[17] M. Karas, F. Hillenkamp, Anal. Chem. 60 (1988) 2299.
[18] H. Zhang, R.M. Caprioli, J. Mass Spectrom. 31 (1996) 1039.
[19] H. Lee, T.J. Griffin, S.P. Gygi, B. Rist, R. Aebersold, Anal. Chem. 74 (2002) 4353.
[20] J. Preisler, F. Foret, B.L. Karger, Anal. Chem. 70 (1998) 5278.
[21] D.F. Hunt, R.A. Henderson, J. Shabanowitz, K. Sakaguchi, H. Michel, N. Sevilir, A.L. Cox, E. Appella, V.H. Engelhard, Science 255 (1992) 1261.
[22] M. Wilm, A. Shevchenko, T. Houthaeve, S. Breit, L. Schweigerer, T. Fotsis, M. Mann, Nature 379 (1996) 466.
[23] R. Aebersold, M. Mann, Nature 422 (2003) 198.
[24] J. Rappsilber, U. Ryder, A.I. Lamond, M. Mann, Genome Res. 12 (2002) 1231.
[25] W.F. Patton, B. Schulenberg, T.H. Steinberg, Curr. Opin. Biotechnol. 13 (2002) 321.
[26] M.P. Washburn, D. Wolters, J.R. Yates 3rd, Nat. Biotechnol. 19 (2001) 242.
[27] J. Peng, J.E. Elias, C.C. Thoreen, L.J. Licklider, S.P. Gygi, J. Proteome Res. 2 (2003) 43.
[28] M. Mann, M. Wilm, Anal. Chem. 66 (1994) 4390.
[29] J.K. Eng, A.L. McCormack, I. Yates, R. John, J. Am. Soc. Mass Spectrom. 5 (1994) 976.
[30] D.N. Perkins, D.J. Pappin, D.M. Creasy, J.S. Cottrell, Electrophoresis 20 (1999) 3551.
[31] H.I. Field, D. Fenyo, R.C. Beavis, Proteomics 2 (2002) 36.
[32] S. Hsieh, J.W. Jorgenson, Anal. Chem. 68 (1996) 1212.
[33] N.W. Smith, M.B. Evans, Chromatographia 38 (1994) 649.
[34] H.J. Cortes, C.D. Pfeiffer, B.E. Richter, T.S. Stevens, J. High Resolut. Chromatogr. Chromatogr. Commun. 10 (1987) 446.
[35] C.C. Benevides, R. Collamati, J. Granger, D. DellaRovere, R. Plumb, K. Fadgen, H. Liu, E.S.P. Bouvier, Proceedings of the 51st ASMS Conference on Mass Spectrometry and Allied Topics, Montreal, Que., Canada, 2003, p. TPP294.
[36] Y. Shen, R. Zhao, S.J. Berger, G.A. Anderson, N. Rodriguez, R.D. Smith, Anal. Chem. 74 (2002) 4235.
[37] H.D. Meiring, E. van der Heeft, G.J. ten Hove, A.P.J.M. de Jong, J. Sep. Sci. 25 (2002) 557.
[38] C.L. Gatlin, G.R. Kleemann, L.G. Hays, A.J. Link, J.R. Yates 3rd, Anal. Biochem. 263 (1998) 93.
[39] S.E. Martin, J. Shabanowitz, D.F. Hunt, J.A. Marto, Anal. Chem. 72 (2000) 4266.
[40] S.P. Gygi, B. Rist, T.J. Griffin, J. Eng, R. Aebersold, J. Proteome Res. 1 (2002) 47.
[41] M.T. Davis, T.D. Lee, J. Am. Soc. Mass Spectrom. 9 (1998) 194.
[42] G.A. Valaskovic, J.P. Murphy III, Proceedings of the 51st ASMS Conference on Mass Spectrometry and Allied Topics, Montreal, Que., Canada, 2003, p. MPX472.
[43] R.E. Moore, L. Licklider, D. Schumann, T.D. Lee, Anal. Chem. 70 (1998) 4879.
[44] N. Ishizuka, H. Minakuchi, K. Nakanishi, N. Soga, H. Nagayama, K. Hosoya, N. Tanaka, Anal. Chem. 72 (2000) 1275.
[45] T. Natsume, Y. Yamauchi, H. Nakayama, T. Shinkawa, M. Yanagida, N. Takahashi, T. Isobe, Anal. Chem. 74 (2002) 4725.
[46] M.W.H. Pinkse, P.M. Uitto, M.J. Hilhorst, B. Ooms, A.J.R. Heck, Anal. Chem. 76 (2004) 3935.
[47] Y. Ishihama, H. Katayama, N. Asakawa, Y. Oda, Rapid Commun. Mass Spectrom. 16 (2002) 913.
[48] G.A. Valaskovic, N.L. Kelleher, F.W. McLafferty, Science 273 (1996) 1199.
[49] D.R. Barnidge, S. Nilsson, K.E. Markides, Anal. Chem. 71 (1999) 4115.
[50] E.P. Maziarz 3rd, S.A. Lorenz, T.P. White, T.D. Wood, J. Am. Soc. Mass Spectrom. 11 (2000) 659.
[51] S. Nilsson, M. Wetterhall, J. Bergquist, L. Nyholm, K.E. Markides, Rapid Commun. Mass Spectrom. 15 (2001) 1997.
[52] K.K. Murray, D.H. Russell, Anal. Chem. 65 (1993) 2534.
[53] L. Li, A. Wang, L.D. Coulson, Anal. Chem. 65 (1993) 493.
[54] Q. Zhan, A. Gusev, D.M. Hercules, Rapid Commun. Mass Spectrom. 13 (1999) 2278.
[55] T. Miliotis, S. Kjellstrom, J. Nilsson, T. Laurell, L.E. Edholm, G. Marko-Varga, J. Mass Spectrom. 35 (2000) 369.
[56] R.R. Hensel, R.C. King, K.G. Owens, Rapid Commun. Mass Spectrom. 11 (1997) 1785.
[57] Y. Ishihama, H. Katayama, N. Asakawa, Anal. Biochem. 287 (2000) 45.
[58] K. Petritis, L.J. Kangas, P.L. Ferguson, G.A. Anderson, L. Pasa-Tolic, M.S. Lipton, K.J. Auberry, E.F. Strittmatter, Y. Shen, R. Zhao, R.D. Smith, Anal. Chem. 75 (2003) 1039.
[59] D.A. Wolters, M.P. Washburn, J.R. Yates 3rd, Anal. Chem. 73 (2001) 5683.

[60] M.P. Washburn, R. Ulaszek, C. Deciu, D.M. Schieltz, J.R. Yates 3rd, Anal. Chem. 74 (2002) 1650.
[61] M.J. MacCoss, W.H. McDonald, A. Saraf, R. Sadygov, J.M. Clark, J.J. Tasto, K.L. Gould, D. Wolters, M. Washburn, A. Weiss, J.I. Clark, J.R. Yates 3rd, Proc. Natl. Acad. Sci. U.S.A. 99 (2002) 7900.
[62] K. Petritis, S. Brussaux, S. Guenu, C. Elfakire, M. Dreux, J. Chromatogr. A 957 (2002) 173.
[63] Y. Ishihama, J. Rappsilber, M. Mann, Proceedings of the 51st ASMS Conference on Mass Spectrometry and Allied Topics, Montreal, Que., Canada, 2002, p. MPX489.
[64] F. Giorgianni, A. Cappiello, P. Palam, S. Beranova-Giorgianni, D. Desiderio, Proceedings of the 52nd ASMS Conference on Mass Spectrometry, Nashville, TN, USA, 2004, p. MPV429.
[65] E. Lasonder, Y. Ishihama, J.S. Andersen, A.M. Vermunt, A. Pain, R.W. Sauerwein, W.M. Eling, N. Hall, A.P. Waters, H.G. Stunnenberg, M. Mann, Nature 419 (2002) 537.
[66] W.E. Haskins, Z. Wang, C.J. Watson, R.R. Rostand, S.R. Witowski, D.H. Powell, R.T. Kennedy, Anal. Chem. 73 (2001) 5005.
[67] S.P. Gygi, B. Rist, T.J. Griffin, J. Eng, R. Aebersold, J. Proteome Res. 1 (2002) 47.
[68] J.P. Murphy III, G.A. Valaskovic, Proceeding of the 51st ASMS Conference, Montral, Canada, 2003, p. TPP293.
[69] L.J. Licklider, C.C. Thoreen, J. Peng, S.P. Gygi, Anal. Chem. 74 (2002) 3076.
[70] Y. Ishihama, J. Rappsilber, J.S. Andersen, M. Mann, Proceeding of the 50th ASMS Conference, Orlando, FL, USA, 2002, p. ThPA009.
[71] J. Rappsilber, Y. Ishihama, M. Mann, Anal. Chem. 75 (2003) 663.
[72] B. Blagoev, I. Kratchmarova, S.E. Ong, M. Nielsen, L.J. Foster, M. Mann, Nat. Biotechnol. 21 (2003) 315.
[73] L.J. Foster, C.L. De Hoog, M. Mann, Proc. Natl. Acad. Sci. U.S.A. 100 (2003) 5813.
[74] W.X. Schulze, M. Mann, J. Biol. Chem. 279 (2004) 10756.
[75] T. Takeuchi, D. Ishii, J. Chromatogr. 253 (1982) 41.
[76] A. Ducret, N. Bartone, P.A. Haynes, A. Blanchard, R. Aebersold, Anal. Biochem. 265 (1998) 129.
[77] T. Le Bihan, D. Pinto, D. Figeys, Anal. Chem. 73 (2001) 1307.
[78] Y. Shen, R. Zhao, M.E. Belov, T.P. Conrads, G.A. Anderson, K. Tang, L. Pasa-Tolic, T.D. Veenstra, M.S. Lipton, H.R. Udseth, R.D. Smith, Anal. Chem. 73 (2001) 1766.
[79] K. Murata, Y. Ishihama, N. Mano, N. Asakawa, in PCT/JP00/03057, 1999.
[80] T. Takeuchi, T. Niwa, D. Ishii, J. Chromatogr. 405 (1987) 117.
[81] M.T. Davis, D.C. Stahl, T.D. Lee, J. Am. Soc. Mass Spectrom. 6 (1995) 571.
[82] K. Deguchi, S. Ito, S. Yoshioka, I. Ogata, A. Takeda, Anal. Chem. 76 (2004) 1524.
[83] J.W. Finch, H. Liu, S. Vazquez, S.A. Cohen, T.A. Dourdeville, D. DellaRovere, S. Koziol, S. Ciavarini, C.C. Benevides, Proceedings of the 52nd ASMS Conference on Mass Spectrometry, Nashville, TN, USA, 2004, p. WOAam1055.
[84] N. Takahashi, N. Ishioka, Y. Takahashi, F.W. Putnam, J. Chromatogr. 326 (1985) 407.
[85] A.J. Link, J. Eng, D.M. Schieltz, E. Carmack, G.J. Mize, D.R. Morris, B.M. Garvik, J.R. Yates 3rd, Nat. Biotechnol. 17 (1999) 676.
[86] H. Liu, R.G. Sadygov, J.R. Yates, Anal. Chem. (2004).
[87] B.J. Cargile, D.L. Talley, J.L. Stephenson Jr., Electrophoresis 25 (2004) 936.
[88] A. Stensballe, S. Andersen, O.N. Jensen, Proteomics 1 (2001) 207.
[89] S.B. Ficarro, M.L. McCleland, P.T. Stukenberg, D.J. Burke, M.M. Ross, J. Shabanowitz, D.F. Hunt, F.M. White, Nat. Biotechnol. 20 (2002) 301.
[90] Y. Ishihama, M. Mann, Chromatography (in Japanese) 24 (Suppl. 1) (2003) 12.
[91] J.E. Syka, J.J. Coon, M.J. Schroeder, J. Shabanowitz, D.F. Hunt, Proc. Natl. Acad. Sci. U.S.A. (2004).
[92] J.E.P. Syka, J.A. Marto, D.L. Bai, S. Horning, M.W. Senko, J.C. Schwartz, B. Ueberheide, B. Garcia, S. Busby, T. Muratore, J. Shabanowitz, D.F. Hunt, J. Proteome Res. 3 (2004) 621.
[93] J.V. Olsen, S.E. Ong, M. Mann, Mol. Cell Proteom. 3 (2004) 608.
[94] V. Dancik, T.A. Addona, K.R. Clauser, J.E. Vath, P.A. Pevzner, J. Comput. Biol. 6 (1999) 327.
[95] M. Mann, S.E. Ong, J. Rappsilber, Y. Ishihama, J.S. Anderson, Proceeding of the 50th ASMS Conference, Orlando, FL, USA, 2002, p. ThOBpm320.
[96] J.L. Meek, Proc. Natl. Acad. Sci. U.S.A. 77 (1980) 1632.
[97] Y. Sakamoto, N. Kawakami, T. Sasagawa, J. Chromatogr. 442 (1988) 69.
[98] M. Palmblad, M. Ramstrom, K.E. Markides, P. Hakansson, J. Bergquist, Anal. Chem. 74 (2002) 5826.
[99] N.L. Anderson, N.G. Anderson, Mol. Cell Proteom. 1 (2002) 845.
[100] N.L. Anderson, M. Polanski, R. Pieper, T. Gatlin, R.S. Tirumalai, T.P. Conrads, T.D. Veenstra, J.N. Adkins, J.G. Pounds, R. Fagan, A. Lobley, Mol. Cell Proteom. 3 (2004) 311.
[101] C.F. Taylor, N.W. Paton, K.L. Garwood, P.D. Kirby, D.A. Stead, Z. Yin, E.W. Deutsch, L. Selway, J. Walker, I. Riba-Garcia, S. Mohammed, M.J. Deery, J.A. Howard, T. Dunkley, R. Aebersold, D.B. Kell, K.S. Lilley, P. Roepstorff, J.R. Yates III, A. Brass, A.J. Brown, P. Cash, S.J. Gaskell, S.J. Hubbard, S.G. Oliver, Nat. Biotechnol. 21 (2003) 247.
[102] Y. Oda, K. Huang, F.R. Cross, D. Cowburn, B.T. Chait, Proc. Natl. Acad. Sci. U.S.A. 96 (1999) 6591.
[103] S.P. Gygi, B. Rist, S.A. Gerber, F. Turecek, M.H. Gelb, R. Aebersold, Nat. Biotechnol. 17 (1999) 994.
[104] S.E. Ong, B. Blagoev, I. Kratchmarova, D.B. Kristensen, H. Steen, A. Pandey, M. Mann, Mol Cell Proteom. 1 (2002) 376.
[105] J.A. Ranish, E.C. Yi, D.M. Leslie, S.O. Purvine, D.R. Goodlett, J. Eng, R. Aebersold, Nat. Genet. 33 (2003) 349.