ELSEVIER

Application of LC/MS to proteomics studies: current status and future prospects

Guodong Chen[1] and Birendra N. Pramanik

Schering-Plough Research Institute, 2015 Galloping Hill Road, Kenilworth, NJ 07033, USA

Liquid chromatography/mass spectrometry (LC/MS) has become a powerful technology in proteomics studies in drug discovery, including target protein characterization and discovery of biomarkers. This review article will describe current LC/MS approaches in protein characterization, including a bottom-up method for protein identification and quantitative proteomics. We will discuss the investigation of protein post-translational modifications such as glycosylation (glycoproteomics) and phosphorylation (phosphoproteomics) using LC/MS. Future trends in LC/MS with respect to proteomics studies will also be illustrated.

Introduction

Pharmaceutical and biopharmaceutical companies have invested heavily in the search for novel drugs, including small molecules and therapeutic proteins for the treatment of life-threatening diseases, accounting for approximately $90 billion in 2005 [1]. Over 300 new medicines have been approved by FDA in the past decade alone. The overall drug discovery process is, however, facing enormous challenges. The drug attrition rate in phase III clinical trials is continuing to rise, with about 45% of the compounds failing at this stage [2]. Major causes of drug attrition were lack of efficacy and safety. New approaches have been taken to address these issues [3]. Recent advances in proteomics research have led to better understanding of those biological pathways utilized by target proteins and, hence, a better means of establishing the efficacy of therapeutic interventions. The identification of biomarkers as measurable signals relating to normal and pathological states has been one of the key foci of proteomics research [4–7]. Progress in this area may potentially improve the success rate of translating drug candidates into successful marketed products.

Proteomics research covers the global analysis of the products of gene expression, including the identification, quantification and characterization of proteins [8,9]. The dynamic nature of cell systems dictates a concomitantly changing proteome. Post-translational modifications of proteins also demonstrate temporal variation, further modulating protein and cellular function. Analysis of such sophisticated biological systems requires robust and sensitive techniques. LC/MS has evolved as a major analytical platform for proteomics studies because of its sensitivity, selectivity, accuracy, speed and throughput [10–12].

This review article will provide an overview on current LC/MS methodology for proteomics studies, including bottom-up approaches and quantitative proteomics. Specific applications on post-translational modifications, such as glycoproteomics and phosphoproteomics using LC/MS will also be presented. Future directions in this area, including the use of new technologies, such as ultra-high-performance liquid chromatography (UHPLC), ion mobility and top-down proteomics to enhance further the capability of LC/MS in proteomics research will also be covered.

Current LC/MS methodology

The very first step in protein characterization is molecular weight (MW) determination. This can most easily be accomplished using either electrospray ionization (ESI) [13] or matrix-assisted laser desorption ionization (MALDI) [14]. Protein primary sequence determination is usually performed by a bottom-up method [15].

Bottom-up

This approach involves the enzymatic digestion of purified proteins or mixtures of proteins into smaller peptides. The peptide

Corresponding author: Chen, G. (guodong.chen@bms.com)

[1] Current address: Bioanalytical and Discovery Analytical Sciences, Bristol-Myers Squibb, P.O. Box 4000, Princeton, NJ 08543, USA.

1359-6446/06/$ - see front matter doi:10.1016/j.drudis.2009.02.007

digests are then separated/analyzed by LC/MS or LC/MS/MS. There are two approaches to assist further protein identification. One is to subject the digested peptides (e.g. tryptic peptides) to a direct search against a genome or protein database for protein identification (peptide mass fingerprinting) in the case of purified proteins or very simple mixtures of proteins. The other approach is to carry out tandem mass spectrometric (MS/MS) analysis of the digested peptides using collision-induced dissociation (CID) to obtain fragment ions for a database search (sequence tag) [16]. Amino acid-specific fragment ions from cleavages of amide-bonds are 'b' ions (N-terminus) and 'y' ions (C-terminus) under CID conditions. Several MS systems can be utilized for bottom-up LC/MS and LC/MS/MS studies. These include 3D and linear ion trap [17,18], quadrupole-time-of-flight [19], Orbitrap [20], Fourier-transform ion-cyclotron resonance (FTICR)-MS [21] and other hybrid mass spectrometers (linear ion trap-Orbitrap [22] and linear ion trap-FTICR-MS [23]). The operation principles of these MS systems have been covered extensively in the literature [24,25]. Note that the high-resolution accurate mass capability of Orbitrap and FTICR-MS in both MS and MS/MS modes further enhances the accuracy of protein identification.

To illustrate the role of LC/MS in a bottom-up approach, we will focus on the characterization of horse myoglobin. It is a single-chain globular protein of 153 amino acids, containing a heme prosthetic group in its center, around which the remaining apoprotein folds. Its primary protein sequence is shown in Fig. 1. The protein was digested with trypsin at a protease-to-protein ratio of 1:50 overnight at 37°C. Trypsin cleaves at the C-terminal side of arginine (R) and lysine (K). Myoglobin-digested tryptic peptides can then be analyzed by LC/MS on a reverse-phase C18 column using a high-resolution linear ion trap-Orbitrap mass spectrometer. The observed mass values of the tryptic peptides corresponding to those shown in Fig. 1 are listed in Table 1. The mass accuracies for all peptides are less than 3 ppm. Following the interrogation of protein databases using the MOWSE algorithm with high mass accuracy constraints (5 ppm), the top protein match was shown to be horse myoglobin, having a MOWSE score of 5.6E25. The measured sequence coverage is 98.7%. Clearly, this peptide mass fingerprinting method can be very useful to identify proteins, confirm known protein sequences as well as allow the structural identification of post-translational modifications by comparing measured mass values with calculated mass values of predicted tryptic peptides. Alternatively, high-resolution LC/MS/MS experiments can be performed on selected tryptic peptide ions. For example, doubly charged T_3 (LFTGHPETLEK) ions (Table 1 and Fig. 1) can be dissociated under CID conditions and produce several informative fragment ions (Table 2). The mass accuracies for all product ions are less than 3 ppm. The high mass accuracies obtained for both precursor ions and product ions give high confidence in protein identification with greater protein coverage when performing database searches. It is important to note that accurate mass determination coupled with LC/MS retention time can be very useful in identifying peptides in complex mixtures [26–29].

In the case of complex protein mixtures, prefractionation techniques are often used before further protein characterization [30], including ion-exchange, hydrophobic interaction chromatography and affinity chromatography. After the prefractionation of such samples, collected fractions can be further separated and analyzed by LC/MS. Two separation strategies are employed in bottom-up experiments. One is to use one-dimensional (1D) or two-dimensional (2D) gel electrophoresis with the in-gel digestion of specific spots, followed by capillary LC/MS/MS experiments. Gel electrophoresis can provide information on protein molecular weight and pI (in the case of 2D gels) for protein identification. This approach is, however, labor intensive with a correspondingly low throughput. Poor recovery of large or hydrophobic proteins is frequent and loss of proteins during gel separation is another drawback to the technique. Another strategy for protein prefractionation is multidimensional LC (e.g. ion exchange/reverse phase) coupled with MS for analysis. A common approach referred to as 'shotgun' proteomics starts with the enzymatic digestion of proteins in solution to generate a complex mixture of peptides, followed by on-line 1D or 2D LC/MS and data-dependent LC/MS/MS analysis [31]. The MS/MS spectra of eluting peptides derived from these approaches can be used to search protein databases to allow protein identification. To reduce the complexity of digested peptides from mixtures of proteins, on-line multidimensional capillary LC/MS can be performed to fractionate peptides at high resolution to improve sensitivity and sequence coverage. The narrow capillary LC column is usually about 15-cm long with an inner diameter of less than 1 mm. The column is packed with two or more stationary phases with the first phase being a strong cation exchanger, followed by a C18 reverse-phase resin. Under acidic conditions, all digested peptides will be retained on the cation exchange resin. Bound peptides are eluted by a stepwise increase in the ionic strength of the eluant onto C18 resin and further analyzed using a conventional solvent gradient in on-line LC/MS systems. This on-line 2D LC/MS method is termed as multidimensional protein identification technology (MudPIT)

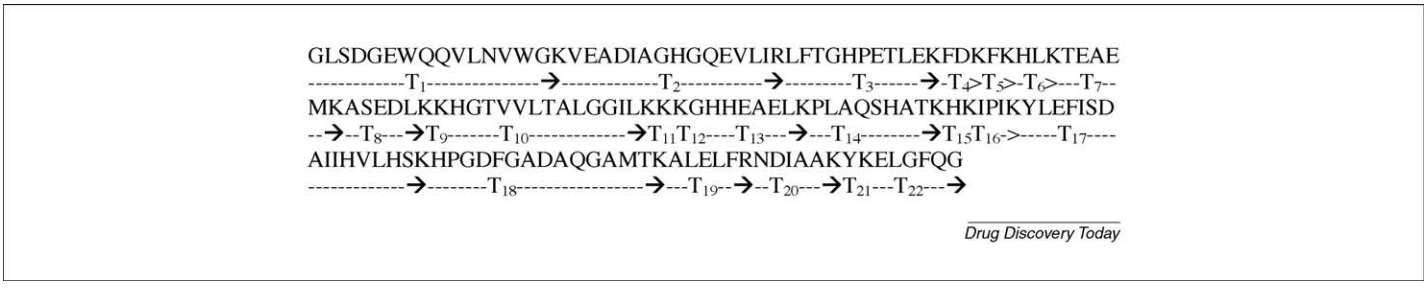

FIGURE 1

Amino acid sequence of horse myoglobin. Tn: expected tryptic peptides.

TABLE 1

Myoglobin tryptic peptide mass accuracies for detected charge states from LC/MS data acquired in linear ion trap-Orbitrap at a resolution of 30,000

Tryptic peptide	Sequence	Charge state	Theoretical *m/z*	Experimental *m/z*	Mass error (ppm)
T_1	GLSDGEWQQVLNVWGK	+1	1815.9024	1815.8986	2.1
		+2	908.4549	908.4541	0.9
T_2	VEADIAGHGQEVLIR	+1	1606.8547	1606.8529	1.1
		+2	803.9310	803.9306	0.5
		+3	536.2898	536.2899	−0.2
T_3	LFTGHPETLEK	+1	1271.6630	1271.6611	1.5
		+2	636.3352	636.3349	0.5
		+3	424.5592	424.5588	0.9
T_3–T_4	LFTGHPETLEKFDK	+2	831.4303	831.4287	1.9
		+3	554.6226	554.6215	2.0
		+4	416.2188	416.2177	2.6
T_4	FDK	+1	409.2082	409.2071	2.7
		+2	205.1077	205.1073	2.0
T_4–T_5	FDKFK	+1	684.3715	684.3699	2.3
		+2	342.6894	342.6885	2.6
		+3	228.7954	228.7948	2.6
T_5	FK	+1	294.1812	294.1805	2.4
		+2	147.5942	147.5941	0.7
T_6	HLK	+1	397.2558	397.2550	2.0
T_7	TEAEMK	+1	708.3233	708.3216	2.4
		+2	354.6653	354.6644	2.5
T_8	ASEDLK	+1	662.3355	662.3343	1.8
		+2	331.6714	331.6705	2.7
T_8–T_9	ASEDLKK	+1	790.4305	790.4290	1.9
		+2	395.7189	395.7181	2.0
		+3	264.1484	264.1478	2.3
T_{10}	HGTVVLTALGGILK	+1	1378.8417	1378.8398	1.4
		+2	689.9245	689.9242	0.4
		+3	460.2854	460.2849	1.1
T_{13}–T_{14}	GHHEAELKPLAQSHATK	+2	927.4845	927.4830	1.6
		+3	618.6587	618.6578	1.5
		+4	464.2459	464.2451	1.7
		+5	371.5982	371.5973	2.4
T_{15}–T_{16}	HKIPIK	+1	735.4876	735.4859	2.3
		+2	368.2474	368.2464	2.7
		+3	245.8340	245.8334	2.4
T_{16}	IPIK	+1	470.3337	470.3326	2.3
		+2	235.6705	235.6699	2.5
T_{17}	YLEFISDAIIHVLHSK	+1	1885.0218	1885.0203	0.8
		+2	943.0145	943.0134	1.2
		+3	629.0121	629.0120	0.2
		+4	472.0109	472.0105	0.8
T_{18}	HPGDFGADAQGAMTK	+1	1502.6693	1502.6670	1.5
		+2	751.8383	751.8375	1.1
		+3	501.5613	501.5608	1.0
T_{19}	ALELFR	+1	748.4352	748.4340	1.6
		+2	374.7212	374.7206	1.6
T_{20}	NDIAAK	+1	631.3410	631.3397	2.1
		+2	316.1741	316.1733	2.5
T_{21}	YK	+1	310.1761	310.1753	2.6
		+2	155.5917	155.5914	1.9
T_{21}–T_{22}	YKELGFQG	+1	941.4727	941.4707	2.1
		+2	471.2400	471.2391	1.9
T_{22}	ELGFQG	+1	650.3144	650.3131	2.0

TABLE 2

Theoretical and experimental mass-to-charge values for product ions derived from CID experiments of myoglobin doubly charged T_3 (*m/z* 636, LFTGHPETLEK) at a resolution of 30,000 in a linear ion trap-Orbitrap mass spectrometer

Product ion	Theoretical *m/z*	Experimental *m/z*	Mass error (ppm)
b_2	261.1598	261.1593	1.9
y_2	276.1554	276.1549	1.8
b_3	362.2074	362.2067	1.9
y_3	389.2395	389.2388	1.8
y_7^{2+}	427.2243	427.2236	1.6
y_8^{2+}	455.7351	455.7343	1.8
y_4	490.2871	490.2864	1.4
y_9^{2+}	506.2589	506.2581	1.6
b_5	556.2878	556.2868	1.8
b_{10}^{2+}	563.2824	563.2816	1.4
y_{10}^{2+}	579.7931	579.7919	2.1
y_5	619.3297	619.3300	−0.5
y_6	716.3825	716.3810	2.1
b_7	782.3832	782.3818	1.8
y_7	853.4414	853.4397	2.0
b_8	883.4308	883.4310	−0.2
y_8	910.4629	910.4610	2.1
b_9	996.5149	996.5130	1.9
y_9	1011.5106	1011.5084	2.2
b_{10}	1125.5575	1125.5552	2.0
y_{10}	1158.5790	1158.5759	2.7

and has found wide applications in global expression profiling, large-scale tissue profiling and membrane profiling owing to its robustness and high-throughput format [32].

The main drawback of a bottom-up approach is the lack of measurement of the intact molecular protein species. Post-translational modifications are not likely to be maintained during CID at the peptide level. Dynamic range might be an issue when working with extremely complex protein mixtures. In spite of these limitations, the bottom-up method is the current method of choice in protein identification in proteomics studies because of the availability of advanced LC/MS instrumentation and excellent software development.

Quantitative proteomics

Another important area in proteomics research is quantitative measurement of changes in protein expression between different states [33]. It is not uncommon to have several fold changes in protein expression between normal and disease states. There are several strategies employed in the field of quantitative proteomics.

Traditional gel-based methods employ a 2D gel-matching procedure (gel imaging) to compare two sets of protein mixtures run under standard operation conditions from different cell states [34]. Hundreds and thousands of proteins can be separated and visualized with conventional protein stains or by immunological approaches. Proteins in gel spots with different intensities are cut out and digested, followed by MS identification using the bottom-up method. It is worth noting that proteins with post-translational modifications that modify protein charge can also be readily detected by 2D gel electrophoresis. Problems associated with gel-to-gel reproducibility can be minimized by using differential gel electrophoresis (DIGE). A good dynamic range over four orders of magnitude can be achieved by gel-based methods. Limitations of this approach include low-throughput, labor-intensive operation with high dependence on the expertise of the operator. In addition, 2D gel electrophoresis does not work well for certain classes of protein, such as very large and very small proteins, hydrophobic proteins and very acidic or basic species.

Other MS-based approaches include metabolic labeling and chemical derivatization. For metabolic labeling, isotopically defined and distinct medium containing unique isotopically labeled amino acids are exposed to cells. It is then possible to separate digested peptides with identical primary sequence from mixed cell lysates purely on the basis of the mass difference conferred by the incorporation of isotopically labeled amino acids [35]. One of the commonly used approaches is stable-isotope labeling by amino acids in cell culture (SILAC) [36]. In SILAC, only a specific amino acid is isotopically labeled. This allows light and heavily labeled peptides to be readily analyzed by MS. Metabolic labeling techniques require much less sample preparation for quantitative studies, although they cannot be used on tissue samples or body fluids. Chemical derivatization strategies rely upon the addition of an isotopically unique functional group to a peptide to distinguish separate samples by their unique mass. This allows measurement of changes in expression levels between two proteomes in a single experiment. Several methods have been developed for the relative quantitation of peptides from proteolytic digestion of complex proteins, including isotope-coded affinity tags (ICAT) [37], using $^{18}O/^{16}O$-labeled water [38] and isobaric tags for relative and absolute quantification (iTRAQ) [39]. Note that absolute quantification using iTRAQ is possible through differentially tagging and spiking known amounts of standard peptides, although it is not commonly used. One of the main issues in chemical derivatization approaches is that the dynamic range of the method may not be sufficient to cover expression levels spreading over a few orders of magnitude.

A third approach in quantitative proteomics is label-free quantitation [40]. This method correlates peptide mass spectral peak data with the abundance of the protein in the sample. It uses either mass spectral peak intensities of peptide ions to represent protein amount or number of MS/MS spectra assigned to a protein as a measure of protein abundance. In the former case, at least one peptide shared by a pair of samples is utilized for peak area calculation from extracted ion chromatograms, typically three or more peptides in common per protein needed for reproducible quantitation. For spectral counting based quantitation, at least one spectrum in either sample pair is used from MS/MS data for any peptide in a given protein. Four or more spectra per protein are often required for accurate quantitation. In label-free LC/MS studies, the analysis of raw LC/MS data is an important step in finding and determining differences in complex mixtures. A fully automated and computationally efficient method in differential MS was developed for detecting differences using complete LC/MS datasets (intensity at each time and *m/z* ratio), followed by the targeted MS/MS analysis of identified differences [41,42]. This

method selectively finds statistically significant differences in the intensity of both high- and low-abundance ions, accounting for the variability of measured intensities. It reduces the number of false positives in identifying differences and provides an effective strategy for large-scale biomarker studies. The label-free method does not require metabolic labeling or labor-intense stable isotope labeling. It works with various sample types (tissues and body fluids) for direct comparison experiments on multiple samples with various conditions.

Post-translational modification

Glycoproteomics

Glycosylation is an important post-translational modification of proteins, playing crucial roles in biological events such as cell recognition and cell–cell interaction. Glycosylation can impact the properties of proteins (charge, conformation and stability) and introduce heterogeneity into the protein as a result of the generation of various glycoforms. It represents the most common modification for recombinant proteins.

MS-based glycoproteomics methods involve the analysis of enzymatic-digested glycopeptides, followed by LC/MS and LC/MS/MS analysis [43]. Some common diagnostic fragment ions for glycopeptide ions under CID conditions are low MW oxonium ions, such as *m/z* 163 (Hex), *m/z* 204 (HexNAc), *m/z* 292 (NANA), *m/z* 366 (Hex-HexNAc) and *m/z* 657 (HexHexNAcNANA). Such MS/MS experiments are normally performed in precursor-ion mode using a triple quadrupole mass spectrometer. Selected ion monitoring experiments can also be carried out to identify glycopeptides with ion trap MS and quadrupole-time-of-flight MS. There are some potential issues related to the MS approaches, such as gas-phase deglycosylation for the location of sugar site attachment and poor ionization efficiency of glycopeptides with respect to their unmodified forms. Thus, several methods have been employed to overcome these difficulties: (1) The release of glycans by β-elimination for O-linked glycans or digestion using peptide-N-glycosidase F or A for N-linked glycans [44]. The conversion of serine to alanine and threonine to aminobutytic acid in O-linked glycan elimination results in a loss of 16 mass units. This can be used to establish the site of sugar attachment. For N-linked glycans, the conversion of asparagine to aspartate leads to an increase of 1 mass unit, facilitating the identification of the glycan attachment site at the deglycosylated peptide. (2) Glycoprotein/N-linked glycopeptide enrichment by lectin affinity chromatography followed by LC/MS analysis [45]. In addition, some LC-based strategies, such as the use of graphitized columns, improve the identification process [46,47].

In spite of the developments in proteomics software for the automatic assignment of unknown peptides, the analysis of glycopeptides by MS methods still remains a challenge in terms of assignment and interpretation of spectra. Further development of suitable software tools will greatly enhance throughput in glycoproteomics [48].

Phosphoproteomics

Protein phosphorylation is a reversible post-translational modification that is involved in the majority of cellular processes. Abnormal phosphorylation activity has been linked to various disease states [49]. The elucidation of protein phosphorylation sites may reveal potential drug targets as well as contribute to our understanding of cell-signaling mechanisms. A general approach for the characterization of phosphorylated proteins is to detect the phosphor moiety of phosphorylated peptides under CID conditions in MS/MS experiments. Neutral loss scanning (98 mass units, H_3PO_4 or HPO_3 and H_2O) in positive ion mode and precursor ion scanning (−79 mass units, PO_3^-) in negative ion mode detect the loss of the phosphate group to identify phosphopeptides from complicated samples using a quadrupole-linear ion trap mass spectrometer [50]. Multiple reaction monitoring (MRM) has also been used in phosphoproteomic experiments. A transition of 98 mass units has been employed to monitor for the loss of a phosphate group on a specific phosphopeptide. This procedure has resulted in increased sensitivity because of the use of only a few specific transitions during the experiments [51]. Phosphopeptide sequences can be identified by MS/MS experiments, although commercially available software packages may not work well owing to complexity of the fragmentation or inefficiency of CID process. A new probability-based statistical approach has been reported to improve the likelihood of determining correct phosphorylation sites based on the presence and intensity of certain ions in the MS/MS data [52].

There are several challenges in the MS analysis of phosphorylated proteins/peptides. The phosphorylation process may have low stoichiometry because of the transient nature of this post-translational modification. Phosphorylated proteins may be at relatively low abundances. Sensitive and robust enrichment strategies are often required before MS analysis on complicated samples. Immobilized metal affinity chromatography (IMAC) is a relatively mature technique for the selective enrichment of phosphopeptides [53]. Recently, the use of titanium dioxide, zirconium dioxide and aluminum hydroxide enhances the enrichment of phosphopeptides by a similar mechanism to IMAC [54–56]. Affinity-based enrichment of phosphotyrosine-containing proteins employs antiphosphotyrosine antibodies to immunoprecipitate and isolate phosphotyrosine-containing proteins [57]. Following affinity enrichment, phosphopeptides are usually transferred to a C18 trapping column for separation and analysis.

Future directions of LC/MS in proteomics studies

As illustrated in this article, LC/MS plays important roles in protein structural identification and quantitative measurements in proteomics research, an integral part of the drug discovery process. Its broad applications include target protein characterization and biomarker discovery in addressing challenging issues related to drug efficacy and safety in drug discovery environment. New technological advances in this area can provide additional capabilities in using LC/MS for proteomics studies.

One of the latest advancements in separation science is UHPLC [58] technology. It utilizes small particles in a packed column, resulting in increased speed of analysis with excellent resolution, sensitivity and peak capacity. Coupling UHPLC to MS greatly facilitates analysis of complex mixtures, further improving the bottom-up method, as illustrated in a recent study [59]. An automated ultra-high-pressure multidimensional liquid chromatography system operating up to approximately 20 kpsi was developed to improve separations and increase protein coverage from limited amounts of samples. In this study, a reversed-phase gradient was

operated in the constant-flow mode so that the gradient shape was fully controllable and could be optimized for the type of samples to be analyzed. This approach was validated on a soluble fraction from yeast lysate where approximately 30% more protein identifications were obtained using a 60-cm-long triphasic capillary column than with the traditional method. Another important advance in LC is the use of microfluidic systems to improve peptide separations [60–62]. In one study, a new, integrated microfluidic device (HPLC-Chip) was coupled to ion trap mass spectrometry for the identification of proteins obtained from 2D gel electrophoresis or chromatography [62]. A sensitivity increase of a factor of at least fivefold was obtained compared to the results obtained previously by conventional nano-LC/MS/MS. The use of microfluidic devices or miniaturized systems can be a valuable tool in identifying low-abundance proteins in proteomics studies.

To analyze complex protein samples better, a different type of mass spectrometer, incorporating high-efficiency gas phase ion mobility separation, has been described which improves speed, reproducibility and coverage for the analysis of protein mixtures [63]. Ion mobility MS differentiates protein samples based on size, shape and charge as well as mass. As reported in the literature, a combination of strong cation-exchange and reversed-phase LC with ion mobility MS has been used to characterize complex mixtures of proteins associated with the human plasma proteome [64]. The increase in separation capacity associated with the inclusion of ion mobility separation led to the generation of one of the most extensive proteome maps to date with a preliminary identification of 9087 proteins from 37,842 unique peptide assignments. An analysis of expected false-positive rates resulted in a high-confidence identification of 2928 proteins. The results were cataloged in a fashion that includes positions and intensities of assigned features observed in the datasets, as well as related identification information such as protein accession number, mass and homology score/confidence indicators.

Another emerging field in proteomics research is top-down proteomics. In contrast to bottom-up experiments, intact proteins are analyzed directly using high-resolution MS and dissociated subsequently by a tandem mass spectrometer [65]. Dissociation techniques used in top-down proteomics include CID, electron-capture dissociation (ECD) [66] and electron-transfer dissociation (ETD) [67]. In ECD, low-energy electrons are captured by multiply charged protein ions and fragmentation takes place by the cleavage of N–C(R) bonds to generate mainly 'c' and 'z' ions. In ETD, singly charged anions transfer an electron to multiply charged protein ions, resulting in fragmentation similar to ECD. Advantages of top-down proteomics include, potentially, 100% coverage for an entire protein sequence in protein identification and post-translational modifications largely remaining intact during dissociation at the protein level. Several studies have been carried out using this method and have illustrated the benefits of the technique in protein identification [68]. The use of LC/MS in a top-down approach has been reported in the literature, demonstrating feasibility of this method in obtaining data on a chromatographic time scale for the first time [69]. In a single LC–MS/MS run, 22 yeast proteins with molecular weights ranging from 14 to 35 kDa have been successfully identified. Additional 39 proteins from LC/MS/MS of selected anion exchange fractions were also identified after fractionating a whole-cell lysate using anion exchange chromatography before on-line LC–MS/MS. Although challenges remain for this approach with respect to its robustness for routine proteomics studies, the use of LC/MS in top-down proteomics can be a powerful technique, affording better understanding of protein fragmentation patterns, and allowing the incorporation of advances in MS activation methods/instrumentation and development of bioinformatics tools for database searching. Both bottom-up and top-down methods can be combined to offer complimentary approaches to achieve the best results for proteomics studies in complex biological systems.

Acknowledgement

The authors would like to thank Dr Malcolm MacCoss for his support on the projects.

References

1 Pharmaceutical Research and Manufacturers of America. *New Drug Approvals Series (1994–2003), New Drugs in Development Series 2001–2003*, PhRMA
2 Kola, I. and Landis, J. (2004) Can the pharmaceutical industry reduce attrition rates? *Nat. Rev. Drug Discov.* 3, 711–716
3 MacCoss, M. and Baillie, T.A. (2004) Organic chemistry in drug discovery. *Science* 303, 1810–1813
4 Rifai, N. *et al.* (2006) Protein biomarker discovery and validation: the long and uncertain path to clinical utility. *Nat. Biotechnol.* 24, 971–983
5 Smith, J.C. *et al.* (2007) Proteomics in 2005/2006: developments, applications and challenges. *Anal. Chem.* 79, 4325–4344
6 Lescuyer, P. *et al.* (2007) How shall we use the proteomics toolbox for biomarker discovery? *J. Proteome Res.* 6, 3371–3376
7 Reyzer, M.L. and Caprioli, R.M. (2007) MALDI-MS-based imaging of small molecules and proteins in tissues. *Curr. Opin. Chem. Biol.* 11, 29–35
8 Wilkins, M.R. *et al.* (1996) Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. *Biotechnol. Genet. Eng. Rev.* 13, 19–50
9 Tyers, M. and Mann, M. (2003) From genomics to proteomics. *Nature* 422, 193–197
10 Chen, G. *et al.* (2007) Applications of LC/MS in structure identifications of small molecules and proteins in drug discovery. *J. Mass Spectrom.* 42, 279–287
11 Chen, G. *et al.* (2007) LC/MS analysis of proteins and peptides in drug discovery. In *HPLC for Pharmaceutical Scientists* (Kazakevich, Y. and LoBrutto, R., eds), pp. 837–899, Wiley
12 Chen, G. and Pramanik, B.N. (2008) LC–MS for protein characterization: current capabilities and future trends. *Expert Rev. Proteomics* 5, 435–444
13 Fenn, J.B. *et al.* (1989) Electrospray ionization for mass spectrometry of large biomolecules. *Science* 246, 64–71
14 Karas, M. *et al.* (1987) Matrix-assisted ultraviolet laser desorption of non-volatile compounds. *Int. J. Mass Spectrom. Ion Process.* 78, 53–68
15 Henzel, W.J. *et al.* (1993) Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases. *Proc. Natl. Acad. Sci. U. S. A.* 90, 5011–5015
16 Hunt, D.F. *et al.* (1986) Protein sequencing by tandem mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* 83, 6233–6237
17 Cooks, R.G. *et al.* (1997) Quadrupole mass filters and quadrupole ion traps. In *Selected Topics in Mass Spectrometry in the Biomolecular Sciences*, (Vol. 504) (Caprioli, R.M. *et al.* eds), pp. 213–238, Kluwer Academic Publishers
18 Schwartz, J.C. *et al.* (2002) A two-dimensional quadrupole ion trap mass spectrometer. *J. Am. Soc. Mass Spectrom.* 13, 659–669
19 Chernushevich, I.V. *et al.* (2001) An introduction to quadrupole-time-of-flight mass spectrometry. *J. Mass Spectrom.* 36, 849–865
20 Hu, Q. *et al.* (2005) The Orbitrap: a new mass spectrometer. *J. Mass Spectrom.* 40, 430–443
21 Bogdanov, B. and Smith, R.D. (2005) Proteomics by FTICR mass spectrometry: top down and bottom up. *Mass Spectrom. Rev.* 24, 168–200

22 Makarov, A. *et al.* (2006) Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer. *Anal. Chem.* 78, 2113–2120
23 May, D. *et al.* (2007) A platform for accurate mass and time analyses of mass spectrometry data. *J. Proteome Res.* 6, 2685–2694
24 Wysocki, V.H. *et al.* (2005) Mass spectrometry of peptides and proteins. *Methods* 35, 211–222
25 Domon, B. and Aebersold, R. (2006) Mass spectrometry and protein analysis. *Science* 312, 212–217
26 Norbeck, A.D. *et al.* (2005) The utility of accurate mass and LC elution time information in the analysis of complex proteomes. *J. Am. Soc. Mass Spectrom.* 16, 1239–1249
27 Berg, M. *et al.* (2006) Reproducibility of LC–MS-based protein identification. *J. Exp. Bot.* 57, 1509–1514
28 Zimmer, J.S.D. *et al.* (2006) Advances in proteomics data analysis and display using an accurate mass and time tag approach. *Mass Spectrom. Rev.* 25, 450–482
29 Hardouin, J. *et al.* (2007) HPLC-chip–mass spectrometry for protein signature identifications. *J. Sep. Sci.* 30, 1482–1487
30 Righetti, P.G. *et al.* (2005) Prefractionation techniques in proteome analysis: the mining tools of the third millennium. *Electrophoresis* 26, 297–319
31 McDonald, W.H. and Yates, J.R., III (2002) Shotgun proteomics and biomarker discovery. *Dis. Markers* 18, 99–105
32 Florens, L. and Washburn, M.P. (2006) Proteomic analysis by multidimensional protein identification technology. *Methods Mol. Biol.* 328, 159–175
33 Panchaud, A. *et al.* (2008) Experimental and computational approaches to quantitative proteomics: status quo and outlook. *J. Proteomics* 71, 19–33
34 Berkelman, T. (2008) Quantitation of protein in samples prepared for 2-D electrophoresis. *Methods Mol. Biol.* 424, 43–49
35 Wu, C.C. and MacCoss, M.J. (2007) Quantitative proteomic analysis of mammalian organisms using metabolically labeled tissues. *Methods Mol. Biol.* 359, 191–201
36 Mann, M. (2006) Functional and quantitative proteomics using SILAC. *Nat. Rev. Mol. Cell Biol.* 7, 952–958
37 Gygi, S.P. *et al.* (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* 17, 994–999
38 Stewart, I.I. *et al.* (2001) ^{18}O labeling: a tool for proteomics. *Rapid Commun. Mass Spectrom.* 15, 2456–2465
39 Aggarwal, K. *et al.* (2006) Shotgun proteomics using the iTRAQ isobaric tags. *Brief Funct. Genomic. Proteomic.* 5, 112–120
40 Old, W.M. *et al.* (2005) Comparison of label-free methods for quantifying human proteins by shotgun proteomics. *Mol. Cell. Proteomics* 4, 1487–1502
41 Wiener, M.C. *et al.* (2004) Differential mass spectrometry: a label-free LC–MS method for finding significant differences in complex peptide and protein mixtures. *Anal. Chem.* 76, 6085–6096
42 Meng, F. *et al.* (2007) Quantitative analysis of complex peptide mixtures using FTMS and differential mass spectrometry. *J. Am. Soc. Mass Spectrom.* 18, 226–233
43 Wuhrer, M. *et al.* (2007) Glycoproteomics based on tandem mass spectrometry of glycopeptides. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* 849, 115–128
44 Bond, M.R. and Kohler, J.J. (2007) Chemical methods for glycoprotein discovery. *Curr. Opin. Chem. Biol.* 11, 52–58
45 Mechref, Y. *et al.* (2008) Glycoprotein enrichment through lectin affinity techniques. *Methods Mol. Biol.* 424, 373–396
46 Wagner-Rousset, E. *et al.* (2008) The way forward, enhanced characterization of therapeutic antibody glycosylation: comparison of three level mass spectrometry-based strategies. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* 872, 23–37
47 Fauquenoy, S. *et al.* (2008) Proteomics and glycomics analyses of N-glycosylated structures involved in *Toxoplasma gondii*–host cell interactions. *Mol. Cell Proteomics* 7, 891–910
48 Ashline, D.J. *et al.* (2007) Carbohydrate structural isomers analyzed by sequential mass spectrometry. *Anal. Chem.* 79, 3830–3842
49 Blume-Jensen, P. and Hunter, T. (2001) Oncogenic kinase signaling. *Nature* 411, 355–365
50 Le Blanc, J.C. *et al.* (2003) Unique scanning capabilities of a new hybrid linear ion trap mass spectrometer (Q TRAP) used for high sensitivity proteomics applications. *Proteomics* 3, 859–869
51 Unwin, R.D. *et al.* (2005) Multiple reaction monitoring to identify sites of protein phosphorylation with high sensitivity. *Mol. Cell Proteomics* 4, 1134–1144
52 Beausoleil, S.A. *et al.* (2006) A probability-based approach for high-throughput protein phosphorylation analysis and site localization. *Nat. Biotechnol.* 24, 1285–1292
53 Corthals, G.L. *et al.* (2005) Identification of phosphorylation sites using microimmobilized metal affinity chromatography. *Methods Enzymol.* 405, 66–81
54 Kweon, H.K. and Håkansson, K. (2008) Metal oxide-based enrichment combined with gas-phase ion-electron reactions for improved mass spectrometric characterization of protein phosphorylation. *J. Proteome Res.* 7, 749–755
55 Kweon, H.K. and Håkansson, K. (2006) Selective zirconium dioxide-based enrichment of phosphorylated peptides for mass spectrometric analysis. *Anal. Chem.* 78, 1743–1749
56 Cantin, G.T. *et al.* (2007) Optimizing TiO_2-based phosphopeptide enrichment for automated multidimensional liquid chromatography coupled to tandem mass spectrometry. *Anal. Chem.* 79, 4666–4673
57 Zheng, H. *et al.* (2005) Phosphotyrosine proteomic study of interferon alpha signaling pathway using a combination of immunoprecipitation and immobilized metal affinity chromatography. *Mol. Cell Proteomics* 4, 721–730
58 MacNair, J.E. *et al.* (1999) Ultrahigh-pressure reversed-phase capillary liquid chromatography: isoatic and gradient elution using columns packed with 1.0-micron particles. *Anal. Chem.* 71, 700–708
59 Motoyama, A. *et al.* (2006) Automated ultra-high-pressure multidimensional protein identification technology (UHP-MudPIT) for improved peptide identification of proteomic samples. *Anal. Chem.* 78, 5109–5118
60 Fortier, M.H. *et al.* (2005) Integrated microfluidic device for mass spectrometry-based proteomics and its application to biomarker discovery programs. *Anal. Chem.* 77, 1631–1640
61 Yin, H. *et al.* (2005) Microfluidic chip for peptide analysis with an integrated HPLC column, sample enrichment column, and nanoelectrospray tip. *Anal. Chem.* 77, 527–533
62 Hardouin, J. *et al.* (2006) Usefulness of an integrated microfluidic device (HPLC-Chip–MS) to enhance confidence in protein identification by proteomics. *Rapid Commun. Mass Spectrom.* 20, 3236–3244
63 Liu, X. *et al.* (2004) Development of high throughput dispersive LC-ion mobility-TOFMS techniques for analysing the human plasma proteome. *Brief Funct. Genomic. Proteomic.* 3, 177–186
64 Liu, X. *et al.* (2007) Mapping the human plasma proteome by SCX-LC–IMS–MS. *J. Am. Soc. Mass Spectrom.* 18, 1249–1264
65 McLafferty, F.W. *et al.* (2007) Top-down MS, a powerful complement to the high capabilities of proteolysis proteomics. *FEBS J.* 274, 6256–6268
66 Ge, Y. *et al.* (2002) Top down characterization of larger proteins (45 kDa) by electron capture dissociation mass spectrometry. *J. Am. Chem. Soc.* 124, 672–678
67 Syka, J.E. *et al.* (2004) Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* 101, 9528–9533
68 Siuti, N. and Kelleher, N.L. (2007) Decoding protein modifications using top-down mass spectrometry. *Nat. Methods* 4, 817–821
69 Parks, B.A. *et al.* (2007) Top-down proteomics on a chromatographic time scale using linear ion trap Fourier transform hybrid mass spectrometers. *Anal. Chem.* 79, 7984–7991