Advances in proteome analysis by mass spectrometry
Timothy J G Griffin*, David R Goodlett† and Ruedi Aebersold‡

Proteome characterization using mass spectrometry is essential for the systematic investigation of biological systems and for the study of gene function. Recent advances in this multifaceted field have occurred in four general areas: protein and peptide separation methodologies; selective labeling chemistries for quantitative measurement of peptide and protein abundances; characterization of post-translational protein modifications; and instrumentation.

Addresses
Institute for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA 98105, USA
*E-mail: tgriffin@systemsbiology.org
†E-mail: goodlett@systemsbiology.org
‡E-mail: raebersold@systemsbiology.org


958-1669/01/$ — see front matter © 2001 Elsevier Science Ltd. All rights reserved.

Abbreviations
2DE two-dimensional gel electrophoresis
CID collision-induced dissociation
ECD electron-capture dissociation
ESI electrospray ionization
FTICR Fourier transform ion cyclotron resonance
HPLC high-performance liquid chromatography
ICAT™ isotope-coded affinity tag
MALDI matrix-assisted laser desorption/ionization
MS mass spectrometry
MS/MS tandem mass spectrometry
MudPIT multidimensional protein identification technology
QTOF quadrupole time-of-flight
RP-µLC reverse-phase micropipillary liquid chromatography
TOF time-of-flight

Introduction
The recent publication of the human genome sequence [1,2] has moved the focus of biological research towards the functional analysis of the discovered genes and has catalyzed the emergence of discovery science, a research approach that is characterized by the systematic analysis of the components that constitute a biological system. Because proteins are immediately relevant for biological function, proteomics — the branch of discovery science focusing on proteins — has enjoyed a rapidly increasing level of attention. Mass spectrometry (MS) is an essential core technology for many types of proteomic studies, especially those that require the identification of proteins or the analysis of their covalent structure [3,4]. This review will concentrate on the most recent advances in MS-based methodologies along with some of the challenges and prospects for the future of proteomics research.

Separation strategies
Traditionally, proteomics research has been based upon the initial separation of a complex protein mixture by two-dimensional gel electrophoresis (2DE) followed by the identification of the separated proteins. When MS is used to identify the proteins, the separated species are proteolyzed and the extracted peptide products analyzed using either matrix-assisted laser desorption/ionization (MALDI) MS or electrospray ionization (ESI) MS. The combination of ESI, tandem mass spectrometry (MS/MS), on-line peptide separation by reverse-phase micropipillary liquid chromatography (RP-µLC), and sequence database searching using collision-induced dissociation (CID) spectra [5,6] has proven to be a particularly powerful method for the identification of proteins separated by 2DE or those contained in complex mixtures. Recent improvements in 2DE technology have included the development of a two-color fluorescence method for differential screening of protein expression [7], the development of ‘zoom gels’ which offer more sensitivity and resolving power [8], and continued progress in approaches to separate difficult to analyze membrane proteins [9].

Despite these continued incremental improvements, 2DE has some inherent limitations as a general strategy for proteome analysis. Using the yeast proteome as a model system, Gygi et al. [10••] investigated the effectiveness of 2DE in conjunction with RP-µLC and ESI MS/MS as a tool for the analysis of unfractionated proteomes. They concluded that even under conditions of maximum sample loading and extended electrophoretic separation, proteins of low-abundance, which comprise approximately one-half of the yeast proteome, were not detected. In fact, combining all of the proteins identified at the time from published, large-scale studies in yeast using 2DE and MS analysis showed that almost no low-abundance proteins had been identified in these studies [11].

Given the limitations of 2DE, alternative methodologies employing multidimensional chromatography for the separation of complex peptide mixtures prior to analysis by MS have seen increased use in proteomics studies. One powerful approach involves the digestion of protein mixtures followed by the initial separation of the resulting peptides by their electrostatic charge using strong cation-exchange liquid chromatography; the peptides are then further separated by their hydrophobicity using RP-µLC (Figure 1). Link et al. [12] originally described this technique coupled with MS/MS analysis, in an approach termed multidimensional protein identification technology (MudPIT). More recently, the MudPIT approach has been used for the large-scale analysis of the yeast proteome [13••], in which the largest number of proteins to date (1484) from the yeast proteome was identified. The authors concluded that this approach was largely unbiased, allowing for the analysis of all extremes of protein types including low-abundance and hydrophobic proteins.
Other alternatives to 2DE-based separations have also been developed, including the use of high-pressure, high-capacity RP-µLC for the separation and direct analysis by MS of very complex peptide mixtures [14]. Shen et al. [15] reported the detection of over one million peptides using high-capacity RP-µLC and ESI Fourier transform ion cyclotron resonance (FTICR) MS. The fractionation of whole protein mixtures by non-porous reverse-phase high-performance liquid chromatography (RP-HPLC), either as a stand-alone separation strategy [16] or as a second dimension of separation after initial separation by isoelectric focusing [17], has been shown to be effective in the screening by MS of differentially expressed proteins from cancerous cell lines. Methods such as the characterization of proteins selectively isolated using protein-chip arrays [18,19,20•] and the analysis of proteins directly from tissues by MALDI MS [21] obviate the need for up-front separation steps. In the future, it is expected that the development of robust, reproducible and highly resolving peptide separation tools will be a high priority for proteomics technology.

**Stable-isotope labeling**

The application of stable-isotope dilution and MS to the analysis of peptides (Figure 2) has proven effective for comparing proteome-wide changes in protein abundance, thus extending the reach of MS-based analyses beyond descriptive protein cataloging. The introduction of the isotope-coded affinity tag (ICAT™) reagent [22] has led the way in enabling the quantitative measurement of protein expression. In a seminal study, the galactose utilization pathway in yeast, which has been studied for decades using genetic and biochemical methods, was analyzed...
using a combination of discovery science tools. Quantitative proteomic data were obtained using the ICAT™ reagent technology, transcript profiles were obtained from full genome cDNA microarrays, and protein interaction data were obtained with the yeast two-hybrid technique. The results of the studies were integrated into a model of the process [23*]. The resulting model was consistent with key aspects of the biochemical model and suggested numerous new testable hypotheses.

A wide variety of stable-isotope labeling methodologies for proteins and peptides have been recently described. Metabolic labeling through the use of stable-isotope-enriched media was used in conjunction with the isolation of cysteine-containing peptides and analysis by RP-μLC and FTICR MS for quantitative protein analysis in both bacterial and mammalian cell lines [24]. Munchbach et al. [25] developed a chemical method for labeling the N termini of peptides with an isotopically coded reagent. This type of tagging facilitates not only peptide quantification, but also de novo sequencing of peptides by MS/MS. Proteolysis-catalyzed labeling of peptides at their C termini using $^{18}$O has also proven effective for comparative measurements of protein abundance within complex protein mixtures [26]. Goodlett and co-workers [27] have developed an alternative isotopic labeling method that involves the labeling of carboxylic acid residues with isotopically normal or heavy methanol. This method allows
for the relative quantification of protein abundance and facilitates de novo peptide sequencing via a novel algorithm that uses the MS/MS spectra of differentially isotopically labeled peptides of identical sequence [27]. In an alternate application of stable-isotope protein labeling, Müller et al. [28] designed isotopically tagged cross-linking reagents for the investigation of protein structure. By adding equal amounts of the two isotopically coded, chemically identical cross-linking reagents to either a single protein or a protein complex the authors showed the effective and straightforward assignment of amino acid cross-links. Protease digests of the cross-linked proteins were analyzed by MALDI MS. It is apparent that stable-isotope labeling and MS will continue to see use as a valuable tool for protein quantification and other applications.

**Functional proteomics**

MS continues to evolve as the most significant technique for the analysis of protein modifications, many of which are critical for maintaining or controlling biological function. Particularly interesting were initial attempts to analyze protein modifications in complex protein samples, if not on a proteome-wide scale. Several improvements to the analysis of phosphorylated proteins have been described that enable the selective enrichment of phosphorylated proteins or peptides from a mixture of phosphorylated and non-phosphorylated peptides. The phosphorylated proteins or peptides are then analyzed by RP-µLC MS and RP-µLC MS/MS. Zhou et al. [29••] described an approach for the systematic analysis of protein phosphorylation. The approach employs a chemical procedure based on the formation of reversible phosphoramidates at the sidechains of the phosphorylated hydroxycamino acids. Following treatment, serine-, threonine- and tyrosine-phosphorylated peptides can be isolated from complex mixtures using selective solid-phase isolation methods. The isolated phosphopeptides were analyzed by RP-µLC MS/MS. In the study, 24 phosphopeptides from 13 phosphoproteins from a total yeast lysate were identified, the majority of which were not previously known to be phosphorylated. Oda et al. [30••] have described a method for the selective labeling and affinity enrichment of phosphoseryl and phosphothreonyl peptides based on a β-elimination reaction. Smith and co-workers [31] extended this approach through the development of a class of isotopically coded affinity reagents that enable the quantitative measurement of these peptides. In a related method, Weckwerth et al. [32] used the same chemical reaction to isotopically label phosphopeptides for quantification. However, this method lacked provision for selective phosphopeptide enrichment and is, therefore, less well suited to the analysis of protein mixtures.

Advances in MS instrumentation have also enabled more sensitive and accurate analysis of protein phosphorylation. The use of precursor ion scanning using an ESI quadrupole time-of-flight (QTOF) mass spectrometer was described for the selective characterization of tyrosine-phosphorylated peptides contained in mixtures of phosphorylated and non-phosphorylated peptides [33]. The use of electron-capture dissociation (ECD) MS was shown to give more sequence-specific information on phosphopeptides than the traditional methodology using CID [34], as the phosphate moiety is not as labile during ECD as compared with CID. ECD MS was also shown to be capable of determining the sites of phosphorylation in an intact protein, without prior proteolysis. Annan et al. [35] described a multistep method for mapping the phosphorylation sites on purified proteins using RP-µLC MS and RP-µLC MS/MS.

Several recent advances have also been made using MS for the analysis of glycoproteins. These include gel-electrophoresis-based methods for the characterization of both the glycoprotein and the linked oligosaccharides [36,37], a novel affinity method for the selective capture of glycoproteins followed by analysis with MS [38], and specific procedures for the analysis of O-linked or N-linked glycosylation sites using MS [39–41]. Despite the steady progress that has been made, the continued development of more rapid, accurate and sensitive MS-based approaches for the analysis of these complex molecules is needed.

**Instrumentation**

Rapid advances in MS instrumentation, both in terms of hardware and software, are enhancing current proteomics techniques and have catalyzed the development of new approaches. The MALDI-TOF/TOF MS [42] is an emerging instrument that is predicted to be capable of conducting MS/MS scans an order of magnitude faster than current ion trap (IT) or QTOF instruments. This type of instrument offers the possibility of obtaining higher coverage of a proteome in a shorter time, although its full potential remains to be realized. The more mature MALDI-QTOF instruments have proven effective in the analysis of complex peptide mixtures [43–45] and in the comprehensive determination of protein sequences [46••]. The combination of stable-isotope labeling, peptide fractionation by RP-µLC, and MALDI-QTOF MS allow for the abundance-dependent selection of peptides for sequencing. By focusing specifically on those peptides that show differential expression in a particular biological context, the throughput of quantitative protein profiling is potentially dramatically increased. Furthermore, the data obtained are immediately relevant for the biological system studied [43]. Developments in FTICR-MS suggest that the need for ‘serial’ MS/MS to identify proteins from species with completely sequenced genomes will eventually be obviated. For example, it has been demonstrated [47•] that simultaneous or ‘parallel’ trapping of a group of unrelated peptides (i.e. not derived from the same parent protein) followed by concurrent fragmentation of these peptides in the ion cyclotron resonance cell can be used to identify the proteins from which each peptide was derived without MS/MS analysis (i.e. selection of a single ion followed by fragmentation). The promise of FTICR-MS for proteomics remains extraordinarily high due to the high-mass accuracy, high-resolution, and outstanding sensitivity of the method [48,49]. For now, however, the
Advances in proteome analysis by mass spectrometry

Giffin, Goodlett and Aebersold 611

instruments remain in the laboratories of MS experts, owing to the increased difficulty of operation compared with IT and QTOF instruments. Other novel instrumentation that has potential for the general, sensitive sequencing of intact proteins includes accelerator MS, which is capable of protein sequence determination at the attomole level [50•], and comprehensive, automated protein sequencing using ECD MS [51•]. Advances in MS instrumentation will continue to increase the sensitivity, throughput and depth of attainable proteome characterization.

Conclusions

The recent advances described in this overview have dramatically increased the power and scope of proteomic studies by MS, but many challenges remain. Routine analysis over the full range of protein expression, which for human cells and tissues spans at least six orders of magnitude, remains difficult, although the continued development of more sensitive instrumentation is making this goal more attainable. The use of stable-isotope labeling, such as the ICAT™ reagent labeling approach, has greatly improved our ability to analyze the proteome in a quantitative manner. The ICAT™ reagent approach is not without disadvantages, however. Limitations include low protein sequence coverage, as only a subset of labeled peptides are analyzed (in many cases only one peptide is identified per protein), and a lack of data on protein modifications. The first of these limitations might be addressed by the further analysis and identification of non-labeled peptides, which will not give quantitative information but can improve the confidence in the identification of proteins. The second limitation might be overcome by the complementary use of some of the methods described here to analyze modified proteins or by more novel approaches, such as the coupling of specialized protein chips [20•] with MS for functional investigations.

The power of proteomic information, combined with other levels of biological data, is beginning to be realized for the characterization of molecular circuitry on a system-wide level [25•]. Continued advances in the areas described here will increase the scope and depth of the data that are attainable in proteomic analyses by MS, and will greatly aid in the level of understanding of biological systems.

Acknowledgements

We apologize to those authors whose work could not be included here due to the constraints of space. We would also like to thank Sam Donohoe and Michael Wright for their assistance with the manuscript. TJG was funded by an NIH Postdoctoral Genome Training Grant fellowship. This work was also supported by grants from the National (USA) Cancer Institute (1R33CA84698) and from the NIH (RO1 AI41109-01).

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

• of special interest
•• of outstanding interest


studies. The chip is capable of determining kinase-substrate interactions in a highly parallel and specific manner.


