From primary structure to function: biological insights from large-molecule mass spectra

Neil L Kelleher

As increasing information is available from genomic databases, mass spectrometry has begun to be used to identify and/or assess regions of predicted DNA or protein sequence. Mass spectrometry performance limits, together with experiments designed for genomic interplay, are being extended to allow accurate genotyping and protein profiling of cells at rates commensurate with the data-intensive future of biology.

Address: Department of Chemistry, 600 S. Mathews Avenue,
University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA.

E-mail: kelleher@scs.uiuc.edu

Chemistry & Biology 2000, **7**:R37–R45

1074-5521/00/$ – see front matter

Introduction

Mass spectrometry (MS) is now an essential complement to the more classical methods used for structural characterization of large biomolecules. Although molecules > 10 kDa have been fairly accessible to X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy for > 20 years, only in the last ten years have mass spectrometric techniques become available for characterizing such large molecules. The ionization techniques electrospray ionization (ESI) [1] and matrix-assisted laser-desorption ionization (MALDI) [2] have triggered the explosive growth in the use of MS to analyze biopolymers. This has been temporally coupled with an equally large surge in (shotgun) genome sequencing. With substantial improvements in the mass analyzers used to measure larger ions produced by ESI and MALDI, high resolving power ($> 10^4$) and increased tandem MS (MS/MS) capabilities have enabled older methods used for small molecules to be extended to molecules of higher relative molecular weight (M_r) and new approaches to be developed to increase sample throughput. MS is a complementary technique to X-ray and NMR in structural biology; it is sensitive, has a large number of data channels and is fast—characteristics that are especially advantageous for analysis of biomolecular primary structure of largely predicted sequence. New methodologies have been developed for determining 50,000 genotypes per day with ~100% accuracy, accomplishing reverse genetics 10^3 times per day, and increasing the quality and efficiency of protein and proteome analysis.

MS-based genotyping

Although large-molecule MS is advancing diverse research in the life sciences, its contribution to genetics has been limited. This situation should change, largely because MS offers an alternative to hybridization-based methods for assessing the mutational state of targeted single-nucleotide polymorphisms (SNPs). MS detection decouples any hybridization steps from the assessment of a SNP site. For example, the technology developed by Sequenom involves four main steps: PCR amplification of a genomic region of interest (a few hundred base pairs); annealing of a separate post-PCR primer adjacent to the SNP site; initiation of a high fidelity extension reaction with one or more pre-selected dideoxy bases [3]; and acquisition of MALDI/time-of-flight (TOF) mass spectra from nanoliter volumes of the extended primer products with automated data processing [4]. In general, knowledge about the DNA sequence flanking the SNP site allows judicious choice of post-PCR primer size, primer location, and dideoxy-nucleotide(s); this enables the assay to be designed so as to extract complete information about the SNP regardless of its state. Additionally, assay 'multiplexing' can be achieved by choosing post-PCR primers of varying lengths to dedicate predetermined *m/z* regions of the mass spectrum to specific SNPs (refer to *m/z* regions below SNP 1, 2, and so on in Figure 1). Such a procedure on patients' DNA to be evaluated for their propensity for heart disease assesses five SNP sites in three genes (Figure 1, top) in one mass spectrum. The post-PCR primers before and after extension are visible and their mass difference genotypes the SNP site. For example, the SNP in the coding region of the ApoB gene (Figure 1, far right) encodes a homozygous arginine in patient 1 and a heterozygous arginine/glutamic acid genotype in patient 2.

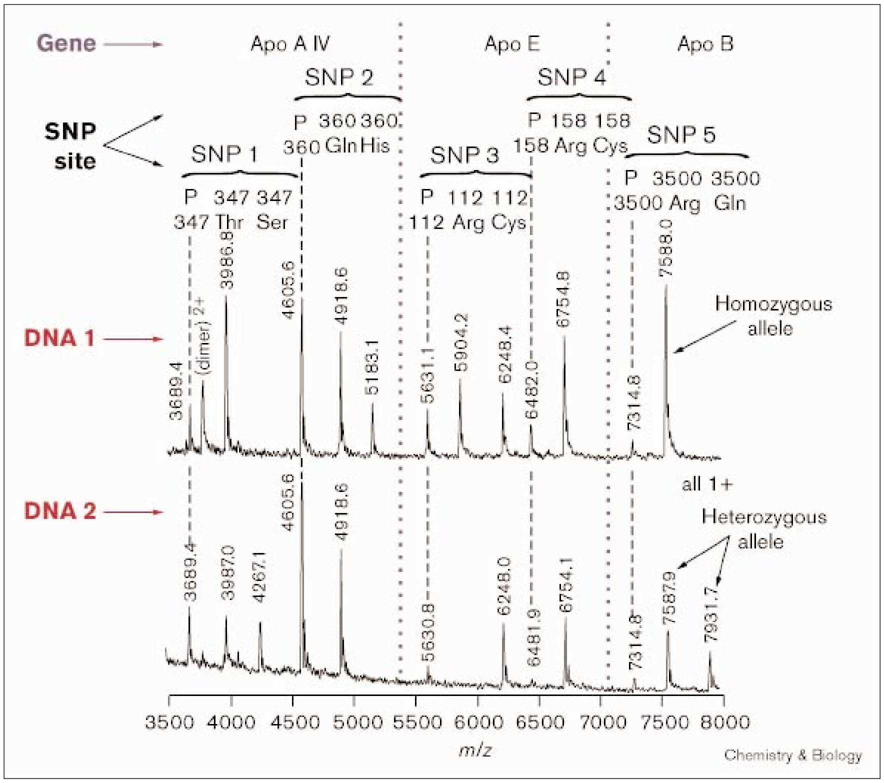
This technology for comparative DNA analysis can be applied to a broad spectrum of biology [5]. In addition to being used in human disease risk assessment [6] and pharmacogenomics, the technique could readily be used in the agriculture sciences to correlate genotypic information with traits of economic importance. For plants in particular, this can be done with a high degree of experimental control for more definitive genotype–phenotype correlations. With the 'pentaplexing' now being accomplished as in Figure 1 (five SNP assessments in one mass spectrum) combined with automated sample generation and transfer, a throughput of 50,000 genotypic assessments in 24 hours is attainable with an accuracy (assured by MS detection) well above 99%.

Selective detection: one peptide among many

MS is often used to detect and localize covalent modifications to proteins. For 'real-life' applications in cell biology and biochemistry, however, this process can become inefficient and even unsuccessful because of low (femtomol) levels of analyte. Parent ion scanning, developed on an ESI/triple quadrupole instrument (Table 1, row 2) [7], can make the analysis of modified peptides faster and more sensitive by ejecting a small ion characteristic of a targeted modification. Examples of distinctive ions include 79 *m/z* for phosphorylation (PO_3^-, a negative ion) and 204 *m/z* for most types of glycosylation ($HexNAc^+$). When scanning the low *m/z* region of the mass spectrum, such distinctive ions are ejected for detection, and while scanning the high *m/z* region (milliseconds later), the intact, modified peptide is detected. (An 86 *m/z* ion ejected from internal leucine or isoleucine residues was used to adapt parent ion scanning for proteomics [8]; detection limits were improved by up to two orders of magnitude by selective detection of peptides versus chemical noise after gel extraction.)

For situations in which stable isotopes can be partially incorporated into a covalently attaching moiety, selective

Figure 1



MALDI–TOF genotyping of five SNP sites in three related genes from two DNA samples; P, primer before base extension reaction; all ions are singly charged except $(dimer)^{2+}$, a noncovalent dimer of the 7588.0 peak. Peaks below each SNP site (indicated at top) indicate the state of SNP. For example, the 7314.8 Da primer for the ApoB polymorphism (SNP 5, far right) has two possible extension products: 7588 Da and 7932 Da, corresponding to arginine and glutamic acid codons, respectively. The sole extension product at 7588.0 Da for patient 1 (middle right) indicates a homozygous allele with an arginine codon. In contrast, the two extension products at 7587.9 and 7931.7 Da for patient 2 (bottom right) reveal a heterozygous allele at this position (arginine and glutamic acid codons).

Table 1

Common MS configurations and some applications*.

Ionization method	*m/z* Analyzer	An example application	Attainable resolving power[†]	Notes
MALDI	TOF	Protein identification by database retrieval	10^4	Peptide mapping can retrieve >70% of proteins[‡]
ESI	Quadrupole	Parent ion scanning	10^3	Triple quadrupole required for parent ion scanning and MS/MS
ESI	Ion trap	Miniaturized ESI and MS/MS for database retrieval	10^4	Lower cost; retrieval often more reliable
ESI	Quadrupole-TOF (a hybrid)	Noncovalent interactions[§]	10^4	High *m/z* range (>20,000 *m/z*)
ESI	FTMS	Top-down protein analysis	10^6	MS/MS above 10 kDa

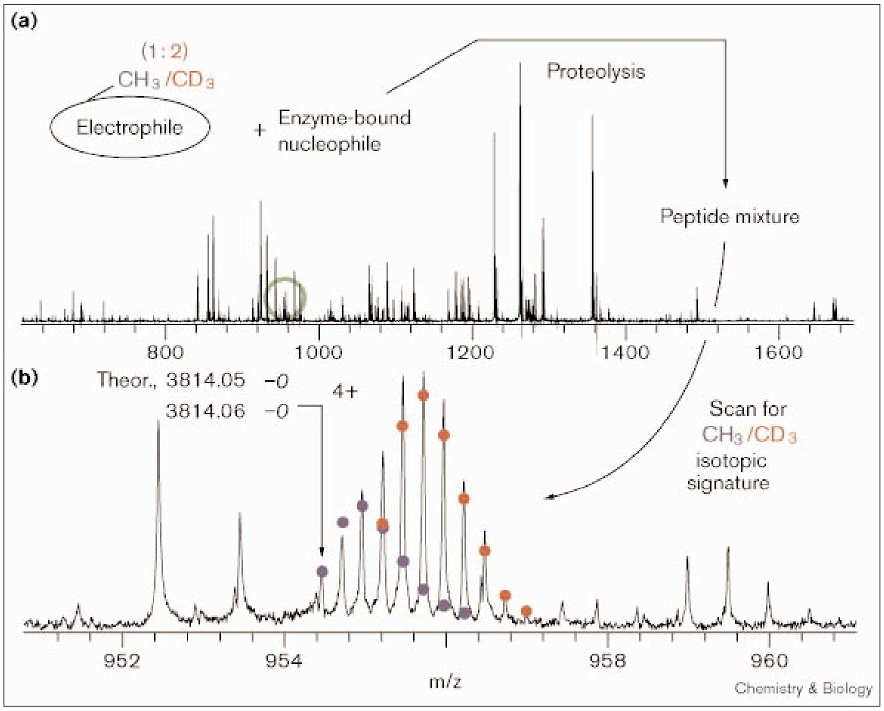
*Meant neither as exhaustive or exclusive; several more configurations are in use and some applications noted can be performed on a variety of instruments. [†]Resolving power is $m/\Delta m$ (analyte mass / resolvable mass difference). [‡]Study was performed in yeast [26]. [§]For a review see [45].

detection in a proteolytic mixture is afforded by a distinctive isotopic signature of peptides carrying the moiety [9]. For instance, engineering a CH_3/CD_3 mixture (1:2 ratio) of a suicide substrate followed by proteolytic digestion of the labeled protein enabled the selective detection of the modified peptide in a 40 component mixture (Figure 2a, green circle). Similarly, this procedure allowed the identification (now automated) of a covalently labeled peptide from a mixture of over 500 peptides from 2 to 30 kDa, measured at isotopic resolution and low part-per-million mass accuracy without chromatographic fractionation [10]. Such methods for specific detection of modified peptides will probably be extended to far more complex systems (e.g., a supramolecular assembly or an entire proteome) to identify quickly molecular targets of covalently binding inhibitors when affinity tags are untenable [11] or substrates of cellular modification enzymes.

Electron capture dissociation: a new ion fragmentation method for tandem MS

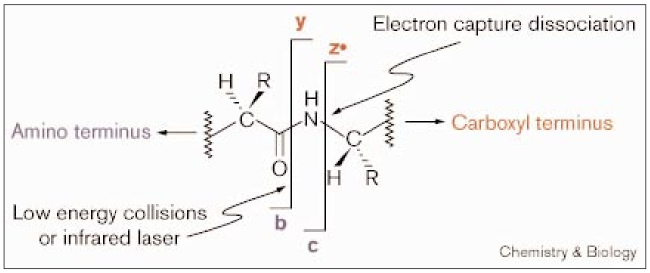
After acquisition of a mass spectrum of a mixture, ions of interest can be 'purified' (automatically in some MS instruments), dissociated, and a mass spectrum of the fragment ions recorded [12]. This tandem use of MS (i.e., MS/MS) often employs low energy collisions with gas for the ion dissociation step. For peptides, fragmentation of

Figure 2



Specific detection of a covalently modified peptide by engineering isotopic heterogeneity. **(a)** ESI–FT mass spectrum (6.1 Tesla, single scan) of an unfractionated proteolytic digest of a 42 kDa protein labeled with a d_0/d_3 mixture of suicide substrates. **(b)** Selective detection of a distinctive isotopic distribution of the covalently-labeled peptide recognized from the mixture of 40 peptides in (a); blue and red dots, theoretical isotopic distributions for the 3.8 kDa peptide labeled with the d_0- and d_3-forms of the suicide inhibitor, respectively. The M_r value 3814.06 Da is for the monoisotopic peak; (There are three M_r values for large molecules: average, monoisotopic, and most abundant isotopic peak; values for carbon used for calculation of each are 12.011, 12.000 and 13.003, respectively). For proteins and peptides, isotopic peaks arise mainly from the 1.1% of ^{13}C in nature, but also from ^{2}H, ^{15}N, ^{18}O, ^{34}S, etc.) for spectra with isotopic resolution, a convenient nomenclature for isotopic bookkeeping is to indicate the mass difference (in units of 1.0034 Da) between the most abundant isotopic peak and the monoisotopic peak in italics after each M_r value (see Figure 5).

Figure 3

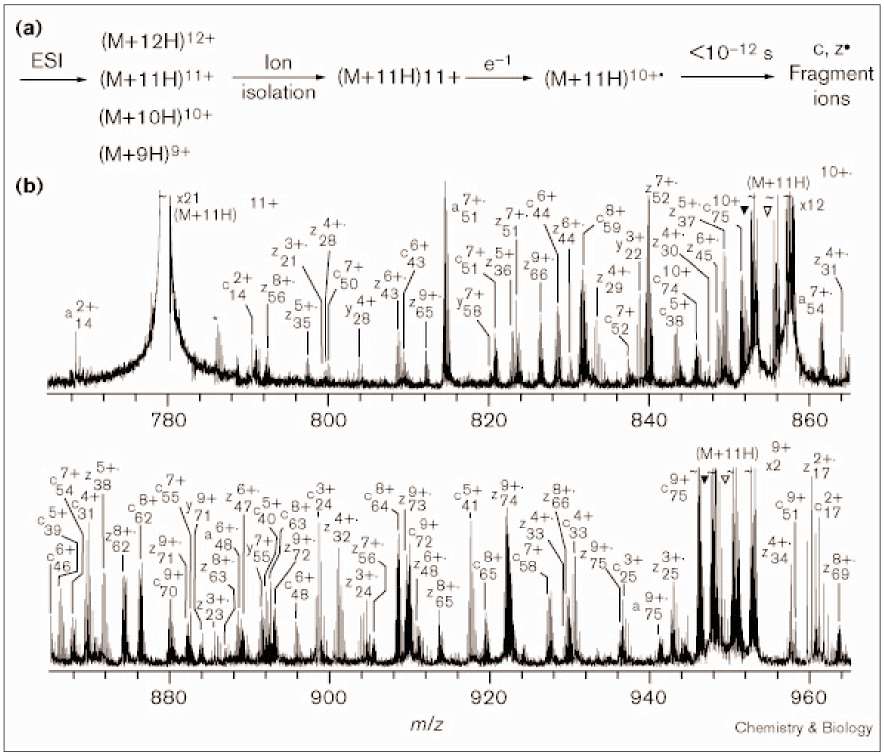


Nomenclature for fragmentation of peptide and protein ions.

the amide bond is most common (Figure 3). (Mass spectrometrists refer to the resulting fragment ions as *b*- and *y*-type ions for those that contain the amino- or carboxy-terminal portion of the peptide, respectively, after fragmentation [13].) Any single ion usually fragments only once; collections of ions fragmenting at every amide bond produce fragment ion ladders used for *de novo* sequencing. For larger peptides (> 25 residues) and proteins, fragment ions usually come in nested sets (from cleavage of adjacent amide bonds) and other conventional techniques for MS/MS fragmentation (collisions with surfaces and infrared laser dissociation) give similar fragmentation patterns via cleavage of identical amide sites. Recently, capture of near-zero energy electrons by multiply charged cations from ESI inside a Fourier-transform (FT) MS instrument has resulted in a unique fragmentation of N–C_α bonds in peptides and proteins (Figure 3) [14]. Such electron capture dissociation (ECD) now provides information complementary to collisional and photo dissociation, enabling the *de novo* sequencing of a 8.6 kDa protein using only MS/MS (no protease) [10]. Unlike conventional MS/MS methods, ECD shows little preference for backbone fragmentation sites, thus providing information-rich MS/MS spectra (Figure 4). ECD has not been demonstrated for proteins > 20 kDa using ESI with FTMS.

Accumulating data indicate that dissociation by ECD is nonergodic [15], so backbone bonds in polypeptides are cleaved without significant randomization of vibrational energy before ion fragmentation (Figure 4a, far right). Labile covalent modifications not stable to any other MS/MS methods (*e.g.*, γ-carboxylation of glutamate) can be localized by ECD [16]. ECD does not appear to be confined to FTMS instruments, as in-source decay in MALDI/TOF apparently causes ECD of 2+ MALDI ions; many N–C_α bonds are cleaved, which allow over 50 contiguous residues to be sequenced [17]. (Peptide sequencing is therefore on a par with the MS/MS-based sequencing of oligonucleotides, in which a 50-mer DNA can be completely sequenced using ESI/FTMS [18].) These aspects of ECD may prove especially useful for efficient mapping of H/D exchange sites in solution that

Figure 4



Electron capture dissociation (ECD) of ubiquitin (76 residues). **(a)** Process of ECD of ions produced by ESI and trapped inside a FTMS. **(b)** Partial ECD MS/MS spectrum of (ubiquitin)$^{11+}$ ions, 8.6 kDa, collected on the Cornell ESI–FT instrument (6.1 Tesla, 70 scans); *c* and *z*• ions are defined in Figure 3. Data were analyzed by the Thrash Algorithm recently developed for automated FTMS analysis [10].

report on noncovalent interactions [19]. ECD of partially deuterated peptides (produced by acidic proteolysis of proteins engaging in noncovalent binding) could allow NMR-level determination of amide exchange sites and rates; use of MS for this would be far more efficient (faster with 10^3–10^6 lower sample requirements) even for proteins >40 kDa, but has been hampered to date by H/D scrambling during the MS/MS process [20]. Combining ECD with conventional MS/MS fragmentation methods results in a diverse tool set for protein microcharacterization and *de novo* sequencing—vital roles for MS in biochemistry and immunology.

Proteomics without the gel

Identification of stained proteins on one-dimensional (1-D) or 2-D polyacrylamide gels by MS has surpassed Edman degradation in speed, sensitivity and reliability. MS identification by database retrieval is based on either tryptic peptide mass values alone (i.e., peptide mapping) [21], or peptide mass values and those of fragment ions produced during MS/MS. The MS/MS approach resulted mainly from the successful miniaturization of ESI through microelectrospray and nanospray. (Microelectrospray is when pumped flow is used (in direct infusion or capillary chromatography) [22], whereas in nanospray, the flow is caused solely by the ESI voltage; flow rates are typically 200–700 nl/min and 20–70 nl/min [23], respectively.) Peptide MS/MS data are formulated into a peptide sequence tag [24] or the raw data pattern is used directly for database retrieval [25]. The proteomic strategy for many labs is to use ~10% of a gel-extracted, tryptic digest for peptide mapping by MALDI/TOF. This identifies ~70% of proteins (in bacteria and yeast) and, if required, the remaining material is used for microcapillary liquid chromatography (μLC)- or nanospray-MS/MS [26].

Recently, direct analysis of even the largest protein complexes [27] by proteolytic digestion *en masse* has allowed identification of the protein components in the complex without prior gel separation as before [28]. Proteolytic mixtures with up to 1000 peptides are brought to an ion trap mass spectrometer via μLC online with ESI and automated MS/MS. For example, over 90 proteins in the yeast 80S ribosome were identified in one μLC-MS/MS run [27]. Such an approach to proteomics would provide protein identity, but without densitometric gel analysis no relative abundance information would be obtained for comparison with another cellular state.

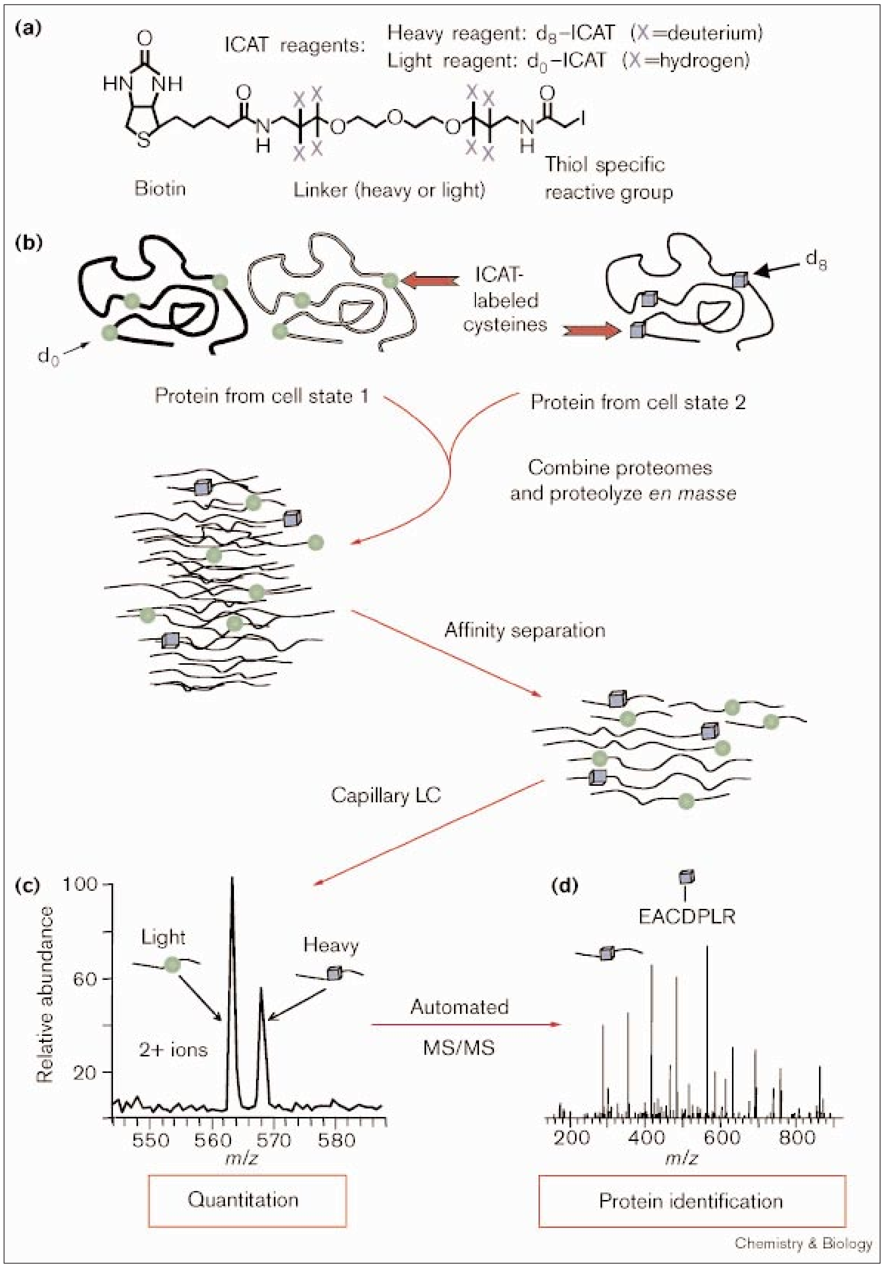
Several groups have begun bypassing the use of 2-D gels for quantitation of protein expression using a derivative of the isotope dilution method. (An isotopically different form of an analyte has identical chromatographic (and other) properties to the natural analyte, but they have different M_r values, allowing determination of their relative abundance in a mass spectrum.) Metabolic incorporation or depletion of stable isotopes during growth has allowed determination of protein expression rations in bacteria and yeast using the proteins themselves [29] or tryptic peptides [30] relative to control cells. The most elegant and broadly applicable method is that of Turecek, Gelb, Aebersold and coworkers [31] in which a post-growth biotin tag (Figure 5a) is introduced onto cysteine residues via iodoacetamide chemistry in one of two isotopic forms (d_0 or d_8); proteome samples from two cellular states are mixed, trypsinized *en masse*, and the peptide mixture simplified by affinity capture prior to μLC-ion trap-MS/MS (Figure 5b). A peptide is identified by its MS/MS signature (Figure 5d) from a database limited to those containing cysteine and an expression ratio for the peptide (and therefore its parent protein) is determined by the d_0/d_8 ratio (Figure 5c). This method is applicable to human tissue and alleviates much of the dynamic range problem in proteomics. (Proteome analyses using 2-D gels measure mainly abundant proteins because most of the protein mass of a cellular lysate applied to a gel (maximum of ~300 μg) is comprised of relatively few types of proteins. The affinity separation enabled by the biotin tag allows a far higher amount of starting material to be used and peptides from lower copy number proteins to be detected.) The authors note that 92% of yeast gene products contain cysteine and the determination of expression ratios is accurate to 2–12%. In a novel extension of the isotope dilution strategy, Chait and coworkers [30] have demonstrated differential abundance measurements not just for whole proteins, but also for the levels of post-translational modification (phosphorylation) at individual sites within a targeted protein.

Considering these advances, the prospect of correlating hundreds of protein expression ratios with those of their transcripts (determined by DNA microarrays) during development or disease is in reach. Do the levels of protein and RNA correlate [32]? How well? What families of proteins have a tight correlation between their gene expression and gene products? How much do correlation coefficients differ in the bacteria, archaea and eukarya? Although proteomic analysis lags DNA microarrays by at least 10^2 in throughput and 10^3 in sensitivity (no RNA amplification), protein profiling has just witnessed a major advance.

Top-down versus bottom-up protein analysis

The proteomic detector of the future should provide comprehensive information on identity, abundance and covalent state of gene products above single copy from 10^5 cells. Of these goals for abundant proteins from 10^9 cells today, detection of post-translational modifications is ‘largely an unsolved problem in high-throughput proteomics’ [33]. In principle, MS is best suited to detect biological events that generate a protein of different M_r value than that predicted from its open reading frame. However, the < 3 kDa peptides produced by ‘bottom-up’ proteolysis (Figure 6a) that allow unambiguous protein identification

Figure 5

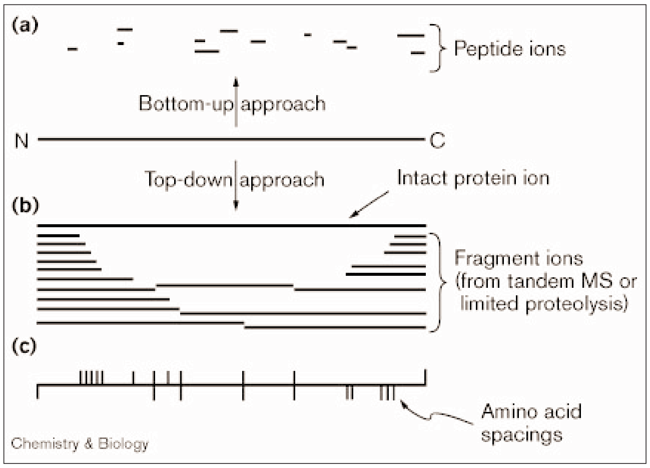


Summary of the isotope-coded affinity tag (ICAT) strategy for identification and relative quantitation of proteins from two cellular states. **(a)** Structure of ICAT with its three functional components: a biotin affinity tag, a linker with or without stable isotopic enrichment, and a cysteine-specific alkylation site. **(b)** A single kind of protein is shown to be labeled with either the d_0-ICAT ('light', left) or d_8-ICAT ('heavy', right). After combination of two proteomes and typsinolysis, affinity capture followed by µLC fractionation submits ~10 peptides per minute to an ESI/ion trap instrument. **(c)** Peptide M_r measurement and determination of relative abundance follows, with **(d)** automated MS/MS allowing protein identification by database retrieval. The peptide's MS/MS spectrum and its d_0/d_8 ratio report the identity and the expression ratio, respectively, of the intact protein from which it originated.

(and now quantitation) obviate analysis of the entire primary structure. The intact protein mass is never obtained either because the proteome (or protein complex) is digested directly or extraction from a gel is inefficient (gel imaging by MALDI/TOF of intact proteins is being explored [34]); sequence coverage is usually 5–40%. This peptide-mapping approach verifies small portions of a DNA-derived sequence and does not involve direct determination of the amino acid sequence. For example, if a 2000.0 Da M_r value matches a predicted peptide of 2000.1 Da, all 18 of its residues are inferred to be correct. The confidence in the protein identification comes from the mass accuracy of the spectrometer (now usually < 25 ppm) [21] and other peptide matches in the spectrum. Complete (or 100%) sequence coverage means that the whole DNA-predicted composition of the protein has been verified within experimental error; together with the high fidelity of transcription and translation, the sequence can be said to be correct.

Efficient surveying of the complete DNA-predicted sequence of a protein is possible using a 'top-down' strategy (Figure 6b) [35]. The top-down approach using ESI/FTMS (Table 1, row 5) generates larger protein

Figure 6



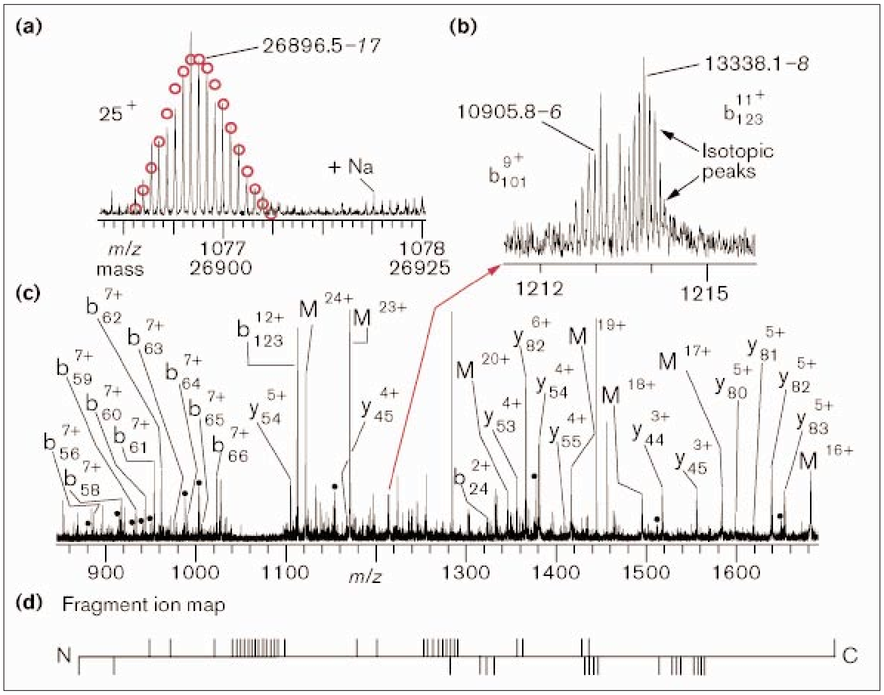
Strategies for DNA-predicted protein sequence analysis. **(a)** Bottom-up versus **(b)** top-down analysis. **(c)** A short hand representation of the hypothetical data in (b).

fragments (5–40 kDa) from direct MS/MS of intact protein ions. This results in more sequence coverage from fewer protein pieces [36] and usually enables 100% sequence coverage below 50 kDa using a 6 Tesla instrument. Such MS/MS data allow protein identification by database retrieval [37], verification of the amino- and carboxy-terminal positions, confirmation of large sections of DNA-predicted sequence, and localization of regions of error or modification [38]. For example, application of the top-down approach to the gene products of an *Escherichia coli* operon efficiently uncovered reading frame errors masking the presence of a small unidentified reading frame whose protein product (7.3 kDa) covalently sequesters a sulfur atom for eventual incorporation into thiamin [36]. For a 27 kDa protein in this operon, an MS/MS spectrum (Figure 7b) of the intact protein ions (Figure 7a) yielded 51 fragment ions, only a few of which are required for database retrieval and resolution of a start codon ambiguity [36].

Recently, work to establish efficient sampling methods devoted to intact proteins has begun [34,39–42]. Although identification can become easier at high mass (as there are fewer proteins in the database), the goal of complete sequence coverage becomes more difficult. Although it is possible to measure proteins as large as 112 kDa with isotopic resolution [43], sensitivity decreases as protein size increases, which is due to a larger number of charge states (for ESI) and isotopic peaks for larger proteins, and MS/MS is less routine. Extension of the top-down approach to proteins > 50 kDa requires a robust method for production of 10–40 kDa peptides and/or increased (FT) MS capability at high mass. Just as hybridizing a quadrupole with TOF MS has delivered an enhanced performance instrument (the Q-TOF; Table 1, row 4), a hybrid ‘Q-FTMS’ now in development should extend the routine mass range for top-down protein analysis and dramatically increase the dynamic range for MS/MS experiments in FTMS [44].

Figure 7

Top-down sequence analysis of a 27 kDa protein by ESI/FTMS. **(a)** Partial ESI/FT mass spectrum of an intact 27 kDa protein; 25^+ ions; +Na, sodium adduct (+22 Da); red open circles, theoretical isotopic distribution generated from the empirical formula $C_{1196}H_{1935}N_{327}O_{355}S_{10}$; italicized number after the M_r value indicates the heavy isotope peak predicted to be most abundant (see Figure 2 legend). **(b)** MS/MS spectrum resulting from low energy collisions of the 25+ ions in (a) with Argon gas (6.1 Tesla, 10 scans); dots, peaks from neutral losses of ammonia or water. **(c)** Expansion of a small *m/z* range containing two fragment ions (the 1 Da spacing of the isotopic peaks are critical for direct determination of ion charge state (*z*) and therefore M_r value). **(d)** Summary of fragmentation data in (b) by mapping fragment ions to the DNA-predicted sequence.



Extending the top-down approach to an entire proteome would establish a basis set of expressed protein primary structures. However, varying amounts of effort would be required to identify the cause of mass discrepancies between DNA prediction and protein reality detected in top-down proteome scanning. For instance, database sequence errors could slow efforts to find covalent modifications important in signaling, *in vivo* stability, subcellular localization, and formation of catalytically competent enzymes, some of which are not reliably predicted from a genomic sequence or are yet unknown. MS must therefore continue evolving to enable characterization of proteins as efficiently as peptides are processed today; accomplishing and combining this with isotope dilution strategies would be a major step towards establishing a futuristic proteomic detector.

Conclusions

Protein function can be effected at many levels including mutation, transcription, mRNA processing, translation, subcellular localization, complexation, post-translational modification and degradation. MS-based techniques can help to deconvolute and assess these levels of functional regulation by analysis of DNA and protein primary structures. Most critical for continued MS advancements in the near future is further development of front-end procedures and automation; these, coupled with upcoming MS instrumental improvements, will increase sample sensitivity and throughput to conceptual limits. Such efficiency will short-circuit cell biology with protein microcharacterization and allow extensive correlations to RNA- and DNA-level information to provide a fairly comprehensive, temporally resolved picture of changes in developing, diseased, and normal cells and tissues. The interplay between MS and a genomic sequence is prototypical for how a genome sequence merely establishes a basis for faster data production, the understanding and contextualization of which should help fulfill some postgenomic expectations.

Acknowledgements

The author thanks the National Institute of Allergy and Infectious Diseases for postdoctoral (F32 AI10087) and likely transitional career support (K22 AI01748). Figures 1, 3 and 4 were adapted from those provided by Daniel Little (Sequenom), Dave Horn (McLafferty research group), and Steven Gygi (Aebersold research group), respectively; the author thanks these researchers and Peter Belshaw, Chris Hendrickson, Jim Vath and Christopher Walsh for helpful comments and discussions.

References

1. Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F. & Whitehouse, C.M. (1989). Electrospray ionization for mass spectrometry of large biomolecules. *Science* **246**, 64-71.
2. Hillenkamp, F., Karas, M., Beavis, R.C. & Chait, B.T. (1991). Matrix-assisted laser desorption/ionization mass spectrometry of biopolymers. *Anal. Chem.* **63**, 1193A-1203A.
3. Braun, A., Little, D.P. & Köster, H.(1997). Detecting CFTR gene mutations using primer oligo base extension and mass spectrometry. *Clin. Chem.* **43**, 1151-1158.
4. Little, D.P., Cornish, T.J., ODonnell, M.J., Braun, A., Cotter, R.J. & Köster, H. (1997). MALDI on a chip: analysis of arrays of low- to sub-femtomole quantities of DNA diagnostic products dispensed by a piezoelectric pipette. *Anal. Chem.* **69**, 4540-4546.
5. Cantor, C.R. & Little, D.P. (1998). Massive attack on high-throughput biology. *Nat. Genet.* **20**, 5-6.
6. Wang, D.G., *et al.*, & Al, E. (1998). Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* **280**, 1077-1082.
7. Annan, R.S. & Carr, S.A. (1997). The essential role of mass spectrometry in characterizing protein structure: mapping posttranslational modifications. *J. Protein Chem.* **16**, 391-402.
8. Wilm, M., Neubauer, G. & Mann, M. (1996). Parent ion scans of unseparated peptide mixtures. *Anal. Chem.* **68**, 527-533.
9. Kelleher, N.L., Nicewonger, R.B., Begley, T.P. & McLafferty, F.W. (1997). Identification of modification sites in large biomolecules by stable isotope labeling and tandem high resolution mass spectrometry: the active site nucleophile of thiaminase I. *J. Biol. Chem.* **272**, 32215-32220.
10. McLafferty, F.W., Fridriksson, E.K., Horn, D.M., Lewis, M.A. & Zubarev, R.A. (1999). Biomolecule mass spectrometry. *Science* **284**, 1289-1290.
11. Taunton, J. (1997). How to starve a tumor. *Chem. Biol.* **4**, 493-496.
12. McLafferty, F.W. (1981). Tandem mass spectrometry. *Science* **214**, 280-287.
13. Roepstorff, P. & Fohlman, J. (1984). Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed. Mass Spectrom.* **11**, 601.
14. Zubarev, R.A., Kelleher, N.L. & McLafferty, F.W. (1998). Electron capture dissociation of multiply-charged protein cations. A nonergodic process. *J. Am. Chem. Soc.* **120**, 3265-3266.
15. Zubarev, R.A., *et al.*, & McLafferty, F.W. (1999). Electron capture dissociation of gaseous multiply-charged proteins is favored at disulfide bonds and other sites of high hydrogen atom affinity *J. Am. Chem. Soc.* **121**, 2857-2862.
16. Kelleher, N.L., *et al.*, & Walsh, C.T. (1999). Localization of labile posttranslational modifications by electron capture dissociation: the case of γ-carboxyglutamic acid. *Anal. Chem.* **71**, 4250-4253.
17. Katta, V., Chow, D.T. & Rohde, M.F. (1998). Applications of in-source fragmentation of protein ions for direct sequence analysis by delayed extraction MALDI-TOF mass spectrometry. *Anal. Chem.* **70**, 4410-4416.
18. Little, D.P., Aaserud, D.J., Valaskovic, G.A. & McLafferty, F.W. (1996). Sequence information from 42–108-mer DNAs (complete for a 50-mer) by tandem mass spectrometry. *J. Am. Chem. Soc.* **118**, 9352-9359.
19. Wang, F., *et al.*, & Zhang, Z.Y. (1998). Conformational and dynamic changes of yersinia protein tyrosine phosphatase induced by ligand binding and active site mutation and revealed by H/D exchange and electrospray ionization fourier transform ion cyclotron resonance mass spectrometry. *Biochemistry* **37**, 15289-15299.
20. Deng, Y., Pan, H. & Smith, D.L. (1999). Selective isotope labeling demonstrates that hydrogen exchange at individual peptide amide linkages can be determined by collision-induced dissocation mass spectrometry. *J. Am. Chem. Soc.* **121**, 1966-1967.
21. Clauser, K.R., Baker, P. & Burlingame, A.L. (1999). Role of accurate mass measurement (± 10 ppm) in protein identification strategies employing MS or MS/MS and database searching. *Anal. Chem.* **71**, 2871-2882.
22. Emmett, M.R. & Caprioli, R.M. (1994). Micro-electrospray mass spectrometry: ultra-high-sensitivity analysis of peptides and proteins. *J. Am. Soc. Mass Spectrom.* **5**, 605-613.
23. Wilm, M. & Mann, M. (1996). Analytical properties of the nanoelectrospray ion source. *Anal. Chem.* **68**, 1-6.
24. Mann, M. & Wilm, M. (1994). Error-tolerant identification of peptides in sequence databases by peptide sequence tags. *Anal. Chem.* **66**, 4390-4399.
25. Yates, J.R., Morgan, S.F., Gatlin, C.L., Griffin, P.R. & Eng, J.K. (1998). Method to compare collision-induced dissociation spectra of peptides: potential for library searching and subtractive analysis. *Anal. Chem.* **70**, 3557-3565.
26. Shevchenko, A., *et al.*, & Mann, M. (1996). Linking genome and proteome by mass spectrometry: large-scale identification of yeast proteins from two dimensional gels. *Proc. Natl Acad. Sci. USA* **93**, 14440-14445.
27. Link, A.J., *et al.*, & Yates, J.R. (1999). Direct analysis of protein complexes using mass spectrometry. *Nat. Biotechnol.* **17**, 676-682.
28. Neubauer, G., Gottschalk, A., Fabrizio, P., Seraphin, B., Luhrmann, R., & Mann, M. (1997). Identification of the proteins of the yeast U1 small nuclear ribonucleoprotein complex by mass spectrometry. *Proc. Natl Acad. Sci. USA* **94**, 385-390.
29. Pasa-Tolic, L., *et al.*, & Smith, R.D. (1999). High throughput proteome-wide precision measurements of protein expression using mass spectrometry. *J. Am. Chem. Soc.* **121**, 7949-7950.

30. Oda, Y., Huang, K., Cross, F.R., Cowburn, D. & Chait, B.T. (1999). Accurate quantitation of protein expression and site-specific phosphorylation. *Proc. Natl Acad. Sci. USA* **96**, 6591-6596.
31. Gygi, S.P., Rist, B., Gerber, S.A., Turecek, F., Gelb, M.H. & Aebersold, R. (1999). Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* **17**, 994-999.
32. Haynes, P.A., Gygi, S.P., Figeys, D. & Aebersold, R. (1998). Proteome analysis: biological assay or data archive. *Electrophoresis* **19**, 1862-1871.
33. Mann, M. (1999). Quantitative proteomics? *Nat. Biotechnol.* **17**, 954-955.
34. Loo, J.A., Brown, J., Critchley, G., Mitchell, C., Andrews, P.C. & Ogorzalek Loo, R.R. (1999). High sensitivity mass spectrometric methods for obtaining intact molecular weights from gel-separated proteins. *Electrophoresis* **20**, 743-748.
35. Kelleher, N.L., Lin, H.Y., Valaskovic, G.A., Aaseruud, D.J., Fridriksson, E.K. & McLafferty, F.W. (1999). Top down versus bottom up protein characterization by tandem high-resolution mass spectrometry. *J. Am. Chem. Soc.* **121**, 806-812.
36. Kelleher, N.L., *et al.*, & McLafferty, F.W. (1998). Efficient sequence analysis of the six gene products (7–74 kDa) from the *Escherichia coli* thiamin biosynthetic operon by tandem high-resolution mass spectrometry. *Protein Sci.* **7**, 1796-1801.
37. Mörtz, E., *et al.*, & Mann, M. (1996). Sequence tag identification of intact proteins by matching tandem mass spectral data against sequence data bases. *Proc. Natl Acad. Sci. USA* **93**, 8264-8267.
38. Kelleher, N.L., Costello, C.A., Begley, T.P. & McLafferty, F.W. (1995). Thiaminase I (42 kDa) heterogeneity, sequence refinement, and active site location from high-resolution tandem mass spectrometry. *J. Am. Soc. Mass Spectrom.* **6**, 981-984.
39. Valaskovic, G.A., Kelleher, N.L. & McLafferty, F.W. (1996). Attomole protein characterization by capillary electrophoresis-mass spectrometry. *Science* **273**, 1199-1202.
40. Yang, L., Lee, C.S., Hofstadler, S.A., Pasa-Tolic, L. & Smith, R.D. (1998). Capillary isoelectric focusing-electrospray ionization fourier transform ion cyclotron resonance mass spectrometry for protein characterization. *Anal. Chem.* **70**, 3235-3241.
41. Li, W., Hendrickson, C.L., Emmett, M.R. & Marshall, A.G. (1999). Identification of intact proteins in mixtures by alternated capillary liquid chromatography electrospray ionization and LC ESI infrared multiphoton dissociation fourier transform ion cyclotron resonance mass spectrometry. *Anal. Chem.* **71**, 4397-4402.
42. Opiteck, G.J., Ramirez, S.M., Jorgenson, J.W. & Moseley, M.A. (1998). Comprehensive two-dimensional high-performance liquid chromatography for the isolation of overexpressed proteins and proteome mapping. *Anal. Biochem.* **258**, 349-361.
43. Kelleher, N.L., Senko, M.W., Siegel, M.M. & McLafferty, F.W. (1997). Unit resolution mass spectra of 112 kDa molecules with 3 Da accuracy. *J. Am. Soc. Mass Spectrom.* **8**, 380-383.
44. Shi, S., Hendrickson, C.L., Quinn, J. P. & Marshall, A.G. (1999). Optimized multipole ion injection system for external accumulation electrospray FT-ICR mass spectrometry. *Proceedings of the 47th ASMS*, pp. 2417-2418, Dallas, TX.
45. Rostom, A.A., & Robinson, C.V. (1999). Disassembly of intact multiprotein complexes in the gas phase. *Curr. Opin. Struct. Biol.* **9**, 135-141.