

REVIEW ARTICLE

The proteomic future: where mass spectrometry should be taking us

Jay J. THELEN*† and Ján A. MIERNYK*†‡¹

*Department of Biochemistry, Plant Genetics Research Unit, University of Missouri, Columbia, MO 65211, U.S.A., †Interdisciplinary Plant Group, Plant Genetics Research Unit, University of Missouri, Columbia, MO 65211, U.S.A., and ‡USDA, Agricultural Research Service, Plant Genetics Research Unit, University of Missouri, Columbia, MO 65211, U.S.A.

A newcomer to the -omics era, proteomics, is a broad instrument-intensive research area that has advanced rapidly since its inception less than 20 years ago. Although the ‘wet-bench’ aspects of proteomics have undergone a renaissance with the improvement in protein and peptide separation techniques, including various improvements in two-dimensional gel electrophoresis and gel-free or off-gel protein focusing, it has been the seminal advances in MS that have led to the ascension of this field. Recent improvements in sensitivity, mass accuracy and fragmentation have led to achievements previously only dreamed of, including whole-proteome identification, and quantification and extensive mapping of specific PTMs (post-translational modifications). With such capabilities at present, one might conclude that

proteomics has already reached its zenith; however, ‘capability’ indicates that the envisioned goals have not yet been achieved. In the present review we focus on what we perceive as the areas requiring more attention to achieve the improvements in workflow and instrumentation that will bridge the gap between capability and achievement for at least most proteomes and PTMs. Additionally, it is essential that we extend our ability to understand protein structures, interactions and localizations. Towards these ends, we briefly focus on selected methods and research areas where we anticipate the next wave of proteomic advances.

Key words: label-free, mass spectrometry, post-translational modification, protein interaction, proteogenomics, proteomics.

BACKGROUND

The introduction of proteomics as ‘the next big thing’ was met initially with enthusiasm by the community of biologists. More recently, however, skepticism and even disappointment have arisen as it has become increasingly clear that the vast majority of proteomics publications are fractional in content or descriptive in scope (for example see [1,2]). Understandably all new fields go through developmental phases, most recently molecular biology. How many readers have at least one ‘gene sequence report’ on their CV? The development of proteomics has been no exception. However, it is somewhat disturbing that although proteomics has existed as a discipline for approximately 18 years, an unscientific comparison of the current literature with publications from the early days reveals that what appears to be an increasing majority remain fractional and descriptive. How can this be? One possible explanation is that the increase in the number of new ‘speciality’ proteomics journals, established in response to the growing demand and popularity of the field, has been accompanied by a concomitant lowering of the bar relative to research quality? Whatever the basis, the impact factors of the three major proteomics journals (*Journal of Proteome Research*, *Molecular and Cellular Proteomics* and *Proteomics*) are, as of this writing, all above 6.0. This indicates a devoted readership and bodes well for the continued growth of proteomics as a scientific discipline (Figure 1). We believe, however, that the impact of the discipline cannot be sustained without a transition to answering meaningful questions rather than continuing to accumulate and archive evermore descriptive information.

To achieve the transition, proteomics studies need to become more conclusive, and at the same time more global in scope in

order to reduce the ever-popular ‘follow-up study’ that consists of merely an expanded proteome dataset [3]. This will require improved research approaches, better methods of statistical analysis and, ultimately, editors to establish more stringent publication standards and reviewers to enforce the standards. Furthermore, as we learn more about the limitations of old technology it is important that we push the community into newer approaches and instrumentation. Although unlikely to be popular, we should harken to the lessons learned from other contemporary scientific disciplines and push forward. It is important for the participants to think about the future of proteomics and where the field should be going. With 18 years of hindsight, in the present review we attempt to project the future of proteomics and describe several areas that deserve more attention from the community.

STRATEGIES AND METHODS AND INSTRUMENTS! OH MY!

More, better, peptides

In bottom-up analyses, proteins are converted into peptides prior to MS analyses. The ideal peptide for analysis by collisional-activation MS/MS (tandem MS) should be 7–35 residues long, be protonated, and have a low charge state (z) and a high mass-to-charge ratio (m/z). Tryptic peptides from *Saccharomyces cerevisiae* have an average length of 8.4 residues and a C-terminal basic residue [4]. The success in using trypsin for analysis of the yeast proteome has led to a widespread adoption. In a non-scientific survey conducted on 11 February 2011, the PubMed database (<http://www.ncbi.nlm.nih.gov/pubmed/>) was queried using ‘proteomics’ as the search term. Of 100 primary research publications, 100 (100%) reported the use of trypsin to fragment

Abbreviations used: EM, electron microscopy; EST, expressed sequence tag; ETD, electron transfer dissociation; FDR, false-discovery rate; FT-ICR, Fourier-transformed ion cyclotron resonance; IM, ion mobility; IMS, imaging MS; LC, liquid chromatography; LTQ, linear trap quadrupole; MALDI, matrix-assisted laser-desorption ionization; MS/MS, tandem MS; Nedd, neural-precursor-cell-expressed developmentally down-regulated; NIMS, nanostructure initiator MS; nLC, nano-LC; ORF, open reading frame; PTM, post-translational modification; SUMO, small ubiquitin-related modifier; TOF, time-of-flight.

¹ To whom correspondence should be addressed (email MiernykJ@ars.usda.gov or MiernykJ@missouri.edu).

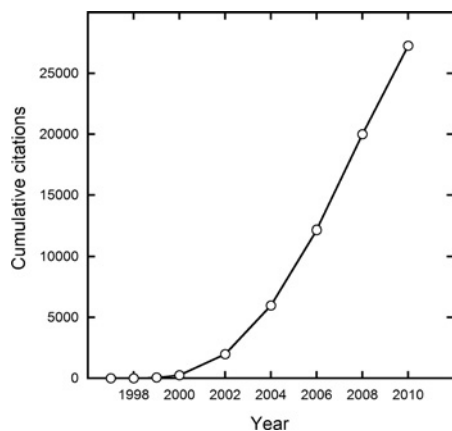


Figure 1 The increase in proteomics publications

The NCBI database PubMed (<http://www.ncbi.nlm.nih.gov/pubmed/>) was searched using 'proteomics' as the query term.

target proteins prior to MS analysis. Although this result might be slightly surprising, there is no doubt that a large majority of proteomic analyses employ trypsin [5]. Recombinant proteomics-grade trypsin is widely available, relatively inexpensive and easy to use. Trypsin cleaves polypeptide chains exclusively C-terminal to arginine and lysine residues, yielding peptides with basic residues at the C-terminus which typically give informative high mass y -ion series and easy to interpret spectra.

In addition to being a time-intensive process not well suited to automation, there are other disadvantages to using trypsin for target protein fragmentation; thermostability is relatively poor and it is labile to rapid autolysis at alkaline pH values. More importantly, lysine and arginine residues are not uniformly distributed throughout the proteome, especially in membrane proteins [6], and 100% sequence coverage cannot be achieved. Some peptides will be 'too short' or 'too long' for MS analysis, or cleavage will be blocked, or the results will be obscured by PTMs (post-translational modifications). Clearly trypsin alone is not enough. By analogy, where would we be today if EcoRI [7] were the only restriction enzyme used for DNA analysis or manipulation? Manipulation of DNA became progressively easier as more and varied restriction enzymes became readily available [8].

Refinements to the methods for trypsin digestion (microwave acceleration, ultrasonic assistance, use of nanobiocatalysts or quantum discs) [9–11] fail to address the essential problem of sequence coverage. Will solution to the 'tryptic flaws' come from a broader use of other proteases, discovery of new proteases or perhaps a more widespread use of chemical fragmentation? Yes. Or, more correctly, yes to all of these and more. Sequencing-grade chymotrypsin is commercially available and can be used to fragment proteins, cleaving C-terminal to phenylalanine, tyrosine, tryptophan and leucine [12]. Target-protein fragmentation by chymotrypsin alone is seldom an improvement over trypsin, but using the two proteases together often leads to significantly improved sequence coverage (for example [13]).

There are additionally the so-called 'sequencing endoproteases' [14], which include Lys-C, which cleaves only C-terminal to lysine residues, Arg-C, Glu-C and Asp-N. Used in conjunction with trypsin, these proteases also yield improved sequence coverage [4]. Disadvantages of using these endoproteases include their relatively high cost, catalytic inefficiency and that digestion with the alternative proteases can result in peptides that lack the C-terminal basic residue typically responsible for the intense y -ion

series. It has been recently demonstrated that the latter problem can be overcome by application of post-proteolysis N-terminal guanidination [15].

Chemical cleavage can be a useful adjunct to protease treatment. Methionine residues comprise less than 2% of protein amino acids, and treatment with cyanogen bromide cleaves the Met-X bond [16]. By itself, cyanogen bromide cleavage is not adequate for fragmentation of most proteins, but it can be easily used in conjunction with trypsin digestion or with other methods of chemical fragmentation, such as formic acid treatment, which cleaves aspartate–proline bonds [17]. Other methods of chemical fragmentation specifically target tryptophan or histidine residues [18].

Just as recombinant DNA-based research accelerated with the availability of numerous restriction enzymes, proteomics as a field would benefit greatly from the availability of additional defined-specificity endoproteases. As with endonucleases, these proteases might come from a survey of micro-organisms. Alternatively, it might be possible to use *in vitro* evolution to modify the specificity of extant proteases [19]. Whether it involves new proteases and new methods for chemical fragmentation, or simply new combinations of proven methods, the goal of 100% sequence coverage should be aggressively pursued.

Reduced chromatography times are necessary for LC (liquid chromatography)-MS/MS analyses

Improvements to mass analyser scanning frequency without equal advances in sensitivity will place a greater burden upon pre-fractionation prior to MS. Wet-bench manipulations are the antithesis of both sample throughput and reproducibility. Although proteomics began with two-dimensional gel profiling, it has slowly been shifting away from this approach in favour of gel-free MS-based profiling. One of the reasons is the serial, and therefore time-consuming, nature of protein identification with two-dimensional gels. Another reason is the biased nature of two-dimensional gels, with membrane and 'extreme' proteins (high/low mass or pI proteins) being under-represented [20]. But the use of electrophoresis-based techniques is widespread, and improvements, including immobilized pH gradients [21] and multiplex-staining of gels using specific fluorescent dyes [22,23], encourage applications. It is important, however, to consider the low-throughput predicament that accompanies protein identification after electrophoresis.

As an example, consider the identification of 200 protein spots excised from two-dimensional gels. Analysis of all 200 samples by LC-MS/MS would require approximately 2 weeks with 30 min chromatographic separations and at least one blank chromatographic run between samples to minimize peptide carryover. The poor throughput and high carryover associated with LC-MS/MS is a problem that can be addressed by reducing run-time, and testing and developing C-18 matrices with reduced 'memory effects'. A progressive shift from ion spray (millilitre flow rates) to capillary (microlitre flow rates) to nanospray (nanolitre flow rates) chromatography has been afoot since the advent of proteomics. Higher-pressure lower-flow chromatography for online nLC (nano-LC)-MS/MS [sometimes referred to as UPLC (ultra-performance liquid chromatography)] improves sensitivity and also reduces chromatography time by reducing peak width, particularly for targeted proteomics [24,25]. Nanoflow LC has the advantage of near-zero dead volume coupled to ultra-narrow peak widths (5 s) and is capable of reproducible separation of complex peptide samples in the 10–100 ng range. But although high sensitivity can be achieved, separation times are long and capacity is limited [26]. More recent developments include

two-dimensional nLC separations [27–29], and the introduction of nanospray microfluidics which use laser-etched column paths within an inert polymer chip [30]. Reducing both analysis and sample loading time will help to shorten the throughput and carry-over concerns associated with LC-MS, as this strategy continues to supplant use of electrophoresis-based methods.

Going global will require improvements in instrument sensitivity

The overall efficiency of a mass spectrometer can be separated into three components: (i) efficiency of ionization (ions/molecule); (ii) transmission of the ion optical system (ions out/ions in); and (iii) detection efficiency (detected pulses out/ions in) [26]. As more biological questions are asked at the tissue and cellular level (e.g. laser-capture microdissection), it will be increasingly challenging to isolate sufficient protein for comprehensive analysis using extant instruments. Pursuit of improved instrument efficiency/sensitivity is ongoing and has been achieved through refinements in ionization efficiency and ion detection (Table 1). For example, the recent development of ‘S-lens’ technology, which captures and focuses ions more efficiently than previous generation ion transfer tubes, has led to a reported 10-fold increase in ion-capture efficiency [31]. It is likely that ongoing improvements in instruments and strategies will continue to improve MS sensitivity.

In contrast with bottom-up analysis, in top-down proteomics intact protein molecular ions are introduced into the mass analyser and are subjected to gas-phase fragmentation. This requires both the use of instruments with high mass accuracy, and the use of deconvolution algorithms [32–34]. The two major advantages of the top-down strategy are the potential access to complete protein sequences and the ability to localize and characterize PTMs. In addition, the time-consuming protein digestion required for bottom-up methods is eliminated. The mass accuracy possible with high magnetic fields has made FT-ICR (Fourier-transformed ion cyclotron resonance) mass spectrometers the instruments of choice for top-down proteomics [35,36]. These instruments are capable of mass accuracies of <2 p.p.m. and a resolution to 10^6 . Furthermore, Orbitrap instruments have a reported p.p.b. mass accuracy when internally calibrated [37].

The results obtained by top-down and bottom-up strategies are complementary, and in the foreseeable future both will be used in proteomic analyses. The top-down approach is newer to proteomics applications, and is probably in the ‘lag phase’ of methods development. It is of note that the entire June 2010 issue of the *Journal of the American Society for Mass Spectrometry* (volume 21, number 6) was devoted to recent developments in top-down proteomics. We can reasonably expect substantial improvements in both strategies and methods to be developed in the short-term future.

A hybrid of the two strategies, ‘middle-down’ proteomics, has already emerged [38,39]. In this approach, large proteins are subjected to limited fragmentation, yielding peptides in the 5–20 kDa range. These more manageable peptides are sequenced using the top-down strategy, maintaining the advantages of a high-percentage sequence coverage and retention of PTM information.

BETTER TARGETS

Quantitative proteomics . . . don’t put a label on it

It is absolutely critical that proteomics transcend qualitative protein identification, and move on to quantitative analysis. Ideally this would involve the universal adoption of a single inexpensive, and sensitive yet rigorous, method. Sadly, few

of our laboratories are located in Camelot, and until there is development/adoption of methods which allow facile direct comparisons within/between/among datasets it will remain impossible to directly compare results [3]. Presently, proteomics broadly encompasses relative (or differential) and absolute quantification, within the constraints of gel-based compared with gel-free sample preparation. Gel-free proteomics is typically separated into quantitative sub-categories based upon: (i) metabolic labelling, (ii) chemical labelling or (iii) label-free [40,41]. Each approach has advantages and limitations that must be considered during experimental design. Metabolic labelling requires the quantitative incorporation of stable isotope-labelled substrate (typically amino acids) into proteins *in vivo*, which is best accomplished using cell cultures. The QconCAT strategy is a fascinating elaboration of the stable isotope-labelled absolute quantification peptide-based approach. However, descriptions of recent applications (for example [42]) clearly indicate that the method is not yet ready for ‘prime-time’ adoption. Substantial problems remain relative to design of the synthetic gene that comprises the basis for quantification, and even in the detection of both the native and quantotypic peptides [42]. Chemical labelling is a post-protein isolation technique that typically involves primary-amine reactive chemistry through the use of commercially available activated isobaric tags that fragment during MS/MS analysis [20]. This method, typically referred to as isobaric tagging, allows simultaneous analysis of four to eight samples, on the basis of 1 amu (atomic mass unit)-resolved low-mass fragments. There is the advantage that the method is system-independent since labelling is done post-isolation. Sensitivity has been an issue, owing to the incomplete fragmentation of parent ions well-known to those studying protein PTMs using ion-trap instruments [20,43]. However, with the development of high-energy collision dissociation-based methods, quantitative fragmentation could become routine [44].

The third quantitative approach, label-free quantification, is the most facile and inexpensive quantification strategy. It can be further refined into spectral counting and peak integration strategies, with variations therein [45–49]. Spectral counting can be thought of as the MS equivalent to estimating mRNA abundance by counting the number of ESTs (expressed sequence tags) in an EST library, the so-called ‘digital Northern’ [50]. The accuracy and sensitivity of any sampling technique is dependent upon sample size. Like mRNA quantification, peptide/protein quantification by spectral counting is instrument-dependent. Owing to their fast-scanning capabilities, ion-trap MS instruments are generally considered to be the most proficient for spectral counting. Current linear ion traps routinely acquire 5–10 scans/s, or approximately 18000–36000 scans during a 60 min LC-MS/MS analysis. Improvements to even this instrumentation scanning speed are possible, ensuring that spectral counting will remain a viable quantification approach in the future. Given the ease of this approach compared with peak integration and the label-based approaches [45–49,51], it is probable that this will become the preferred strategy for discovery-based comparative proteomics, whereas peak integration will remain the more rigorous strategy for comparative (and absolute) quantification for targeted samples [45,52].

Protein interactions/structural proteomics

The functional complexity of an organism cannot be solely defined by the number of proteins that are present or their suite of PTMs; at least equally important is the number of biologically relevant multi-protein complexes [53–55]. It has been inferred that the interactome of the relatively simple eukaryote *S. cerevisiae*

Table 1 Overview of mass spectrometers available for protein analysis

The sensitivity of each instrument was determined using (1) reserpine, (2) [Glu1]-fibrinopeptide B, (3) ubiquitin, (4) BSA digest or (5) verapamil. All data were obtained from brochures available at the manufacturer's website (<http://www.thermoscientific.com>, <http://www.waters.com>, <http://www.bdal.com>, <http://www.agilent.com> and <http://www.absciex.com>) or through a sales representative. APCI, atmospheric pressure chemical ionization; API, atmospheric pressure ionization; APPI, atmospheric pressure photoionization; ASAP, atmospheric solids analysis probe; CE/MS, capillary electrophoresis MS; CID, collision-induced dissociation; ESCi, electrospray atmospheric pressure chemical ionization; ESI, electrospray ionization; FTMS, Fourier transform MS; HCD, higher-energy C-trap dissociation; HESI, heated electrospray ionization; MMI, multi-mode ionization; MRM, multiple reaction monitoring; nano, nanospray; photo, photoionization; PQD, pulsed Q collision-induced dissociation; PTR, proton transfer reaction; SORI, sustained off-resonance irradiation.

Model	Manufacturer	Analyser	Accuracy (p.p.m.)	Sensitivity	Range (<i>m/z</i>)	Resolution (FWHM)	Scan rate	Fragmentation	Ionization
Orbitrap (with LTQ XL or Velos)	Thermo Scientific	Hybrid	<1	Attomole ¹	50–4000	>100 000	1 spectra/s	CID, HCD, ETD, PQD	MALDI, HESI, ESI, nano, API, APCI, APCI/APPI
TSQ Vantage	Thermo Scientific	Triple quadrupole	50*	Attomole ¹	10–1500	7500	5000 units/s	CID	HESI, APCI/APPI, nano
Velos	Thermo Scientific	Linear ion trap	50*	Attomole ¹	50–4000	>25 000	33 333 units/s	CID, ETD, PQD	ESI, APCI, APCI/APPI, nano
LTQ XL	Thermo Scientific	Linear ion trap	50*	Attomole ¹	50–4000	>25 000	16 000 units/s	CID, ETD, PQD	ESI, APCI, APCI/APPI, nano
Xevo G2 Qtof	Waters	Quadrupole TOF	<1	Femtomole ²	<100 000	>22 500	30 spectra/s	CID	ESI/APCI/ESCI, APCI, APPI/APCI, nano, ASAP, APGC, TRIZACTM
Xevo TQ-S	Waters	Tandem quadrupole	50*	Attomole ¹	2–2048	Not provided	10 000 units/s	CID	ESI/APCI/ESCI, APCI, APPI/APCI, nano, ASAP, APGC, TRIZACTM
Synapt G2 HDMS	Waters	Quadrupole TOF	<1	Femtomole ²	<100 000	40 000	20 spectra/s	CID	ESI/APCI/ESCI, APCI, APPI/APCI, nano, MALDI, ASAP, APGC, TRIZACTM
SolariX FTMS	Bruker Daltonics	FTMS	<1	Attomole ³	<100 000	>1 000 000	Not provided	CID, ETD, ECD, (SORI)-CID	ESI, nano, APCI, APPI
Amazon ETD	Bruker Daltonics	Linear ion trap	50*	Attomole ¹	50–3000	20 000	52 000 units/s	CID, ETD/PTR	ESI, APCI, APPI, nano, HPLC-Chip, ESI/MALDI
micrOTOF-Q II	Bruker Daltonics	Quadrupole TOF	<2	Attomole ⁴	50–20 000	20 000	Not provided	CID	ESI, APCI, ESI/APCI, APPI, nano, CE/MS
6490 Triple Quadrupole	Agilent	Triple quadrupole	50*	Zeptomole ⁵	10–2000	Not provided	150 MRM/s	CID	HPLC-chip, ESI, APCI, APPI, MMI
6500 Q-TOF	Agilent	Quadrupole TOF	<2	Femtomole ¹	20–20 000	40 000	20 spectra/s	CID	ESI, APCI, ESI/APCI, APPI, MALDI, HPLC-Chip
TripleTOF 5600	ABSciex	Triple TOF	<1	Femtomole ¹	<40 000	40 000	100 spectra/s	CID	ESI/APCI, Turbo V, nano
Qtrap 5500	ABSciex	Triple quadrupole	100*	Femtomole ¹	5–1250	Not provided	12 000 units/s	CID/ETD	APCI, Turbo V, nano, ESI/APCI, photo
		Linear ion trap	100*	Femtomole ¹	50–1000	9200	20 000 units/s		

*Instruments with mass accuracy reported in Daltons were converted into p.p.m. using an *m/z* of 1000.

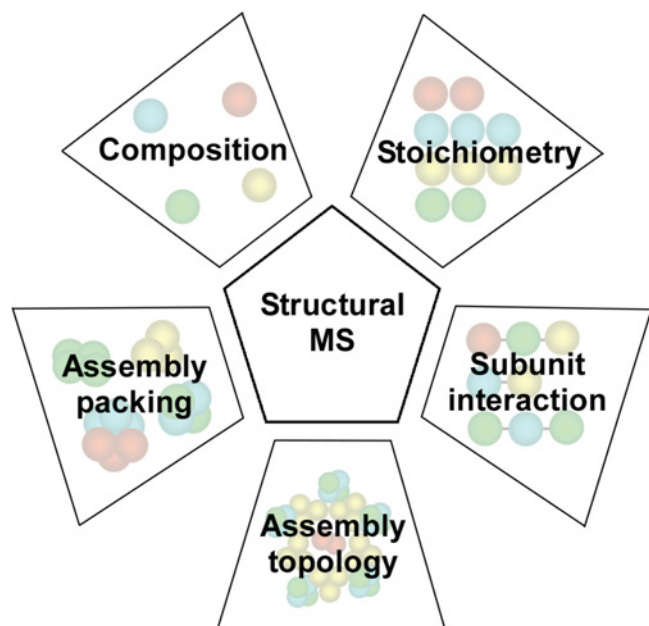


Figure 2 Interacting strategies used to study protein interactions/structural MS

Stoichiometry can be determined under conditions that maintain non-covalent interactions, and thus measure the mass of the intact complex. MS/MS at increasing collision energy can be used to distinguish core from peripheral components, whereas dissociation in the gas phase can be used to identify subcomplexes and assembly packaging. The subunit copy number can be determined by using the summing of masses for interaction topology algorithm [55].

comprises 35 000 protein complexes, and that each of the proteins interacts with six non-redundant partners [56,57]. In the overall picture, it is essential that we understand how the function of multi-protein complexes differs from the functions of the individual components.

Early descriptions of protein interactions were for the most part based upon results from synthetic genetic screens, such as the yeast two-hybrid system [58], affinity tagging [59,60] or co-precipitation studies [61]. The surprisingly poor initial agreement of results using these strategies [62] indicated both the difficulty of the questions being asked and the extent to which improved methods were, and continue to be, needed.

Extant protein-interaction maps of complex multicellular organisms, such as *Caenorhabditis elegans* and *Drosophila melanogaster* [63,64], are based upon results from the application of a variety of computational [65], evolutionary [66], synthetic genetic [58] and even transcript co-expression screens [67]. Although there have been proteomics-based studies of these models [56,61,68], they have not yet contributed as much as they should and, ultimately, will (Figure 2). A description of any organism-scale protein-interaction network must accommodate the sub-networks arising from studies of organellar protein interactions. A relatively straightforward strategy for analysis of Megadalton-sized *Arabidopsis thaliana* chloroplastidial multi-protein complexes has been described by Olinares et al. [69].

The utility of MS in the study of protein interactions and, simultaneously, the genesis of structural MS, are based upon the observation that non-covalent interactions can be maintained in the gas phase [53]. Advantages of using MS to study multi-protein complexes include the ability to study both symmetric and asymmetric heterogeneous complexes at picomolar concentrations, and to do this in real time (Figure 2). An early application of structural MS involved determining the

stoichiometry of individual proteins in a complex [70] (Figure 2). A previously isolated complex was dissociated into individual components, which were then separated by LC prior to MS analysis. The m/z results can be converted into actual mass values by the use of maximum-entropy deconvolution [71,72]. More recently, a hybrid MALDI (matrix-assisted laser-desorption ionization) LTQ (linear trap quadrupole) Orbitrap instrument was used to resolve the composition of a nuclear pore subcomplex without prior dissociation [73].

The most common strategy used for determining the existence of a 'complex' is some form of co-precipitation. This method can employ antibodies, engineered proteins or might simply target complexes that are so large that they can be isolated by rate-zonal sedimentation (e.g. the pyruvate dehydrogenase complex [74]). Co-precipitated proteins can be digested and analysed by MS, either with or without including an electrophoretic or LC separation step. If the target protein can be engineered to include a purification aid, such as a His₆ tag, then potential complexes can be easily isolated from transformed sources. The state-of-the-art version of this strategy is referred to as tandem-affinity purification. In this case, two distinct affinity tags are employed in order to increase the stringency of washing and reduce artefactual associations that are formed only after cell disruption [64,68]. It remains critically important to independently validate any co-precipitation data before assuming concurrence in a complex. Co-precipitation can be conducted in either the presence or absence of chemical cross-linkers [75]. The cross-linking will stabilize ephemeral interactions, and, especially if cleavable cross-linkers are used, need not interfere with downstream analyses [76]. Furthermore, substantial information about the protein surfaces interacting and the specific amino acids involved can be obtained through using cross-linkers of differing lengths and functional-group targets [77].

Once the composition, stoichiometry and mass of the subunits have been determined, the next step is to identify which subunits comprise the core and which are peripheral to it (Figure 2). The peripheral subunits are more exposed, more easily unfolded and the first to be released. Thus step-wise increases in accelerating voltage can be used to identify them [55,78]. Alternatively, a surface-induced dissociation strategy could be employed for identification of peripheral components. Very large or polydisperse protein complexes can result from the transient, but still functional, association of smaller subcomplexes, complicating the generation of any sort of connectivity map (Figure 2). Furthermore, protein associations can be transient, formed in response to specific biochemical or environmental cues, further complicating analysis [57]. Finally, considering the multiplicity of potential spatial conformations, preparing a connectivity map for a multi-component complex is not trivial, but ultimately rewarding and useful in terms of functional interpretation.

Although the use of MS is the best extant method to identify the protein composition of a complex on the basis of the ability to provide an exact molecular mass, the reliance on genomic databases can at the same time be problematic. Essentially all cellular proteins are subject to multiple PTMs, which result in both increases and decreases in protein mass. Furthermore the PTMs themselves can be responsible for protein associations and complex formation. This subject is more fully addressed in a subsequent section, and has implications regarding all aspects of proteomic analysis.

The recent combination of IM (ion mobility) with MS has provided a new dimension in analysis of protein and protein complexes [55,79,80]. A travelling voltage wave propels the ions through the IM cell, reducing their transit time, which serves to

increase both sensitivity and analysis speed. In IM-MS, ions are separated on the basis of differences in charge, size and shape, and can provide information on the stoichiometry, topology and cross-section of both protein complexes and their composite subunits (Figure 2). Thus ions can be identified by their mass, whereas their overall structure can be simultaneously determined, providing insight into functional protein complexes including Megadalton-sized molecular machines such as the proteasome [81].

Use of IM-MS can increase our understanding of not only the shape of an entire multi-protein complex, but also that of individual components if analysed separately. When high-resolution protein structures are not available, it is possible to combine results from IM-MS with those from structural EM (electron microscopy) and computational modelling [80]. Unfortunately, to date only a single commercial IM-MS instrument capable of analysing large protein complexes is available [82], so widespread application of IM-MS will require either modification of extant instruments or new developments by instrument vendors.

A fascinating hybrid strategy was recently described by Richter et al. [83]. The authors describe the straightforward isolation of multi-protein complexes by cell lysis followed by rate-zonal sedimentation. By employing GraFix (gradient fixation) during centrifugation, even relatively labile complexes remained intact. After gradient fractionation, single-particle EM was used to visualize the complexes that had been adsorbed on to EM carbon films. In parallel, identically prepared specimens were digested and analysed by MS (EM carbon film-assisted digestion). The EM and MS results can be directly correlated. In their first full description of the method, it was reported that as little as 50 fmol were sufficient for a comprehensive protein description of two model complexes [83]. Given the availability of structural EM plus MS data, it would be feasible to then model protein-binding interfaces [84]. This and many other aspects of MS require a partnership with computational analysis in order to provide both a comprehensive and comprehensible 'holo-picture' (for example [85–88]) (Figure 2). Coverage of the roles of several important computational methods can be found in the excellent recent review by Sharon [55].

IMS (imaging MS)

The first description of IMS appeared in 1997, and application of this technique has increased rapidly since then. By December 2011 there were 583 entries in the PubMed (<http://www.ncbi.nlm.nih.gov/pubmed/>) database that included IMS in the title and/or abstract. IMS is often referred to as MALDI imaging, on the basis of the type of instrument used [89], or occasionally as 'mass microscopy' [90,91].

All of the early research using IMS addressed analysis of low-molecular-mass compounds. Even today the majority of IMS-based publications are from research that has focused on metabolomic profiling or pharmaceutical drug analysis. There is, however, an increasing enthusiasm for developing improved methods for analysis of proteins/peptides [92–96].

The use of IMS for direct analysis of proteins has been most successful in studies of the low- to mid-molecular-mass proteome. Because this method analyses intact tissue or tissue slices, it avoids homogenization and separation steps, and the spatial distribution of proteins within the tissue is preserved. The process is relatively simple; a matrix is deposited on the sample followed by irradiation with a laser which desorbs and ionizes the peptides. The use of MALDI is typically coupled with TOF (time-of-flight) mass analysers where the ions are accelerated at a fixed potential, traverse a field-free flight tube where they are separated

based on their m/z ratio, and are subsequently detected. The mass range of TOF analysis is virtually unlimited and capable of measuring analytes >200 kDa [97]; however, IMS studies are carried out using TOF mass analysers that have a resolving power of approximately 15 000 for peptides in the range ~ 1500 Da. Use of FT-ICR instruments allows resolution in excess of 1 000 000, which would allow baseline resolution of species differing in millimass units. When analysing samples that have the same nominal mass but different exact masses, TOF-based methods cannot differentiate individual spatial distributions, whereas this would be possible using FT-ICR-based methods.

The spatial localization of discrete proteins by IMS has a lateral resolution of 10–100 μm . A thin (~ 10 μm) sample can be mounted on a target plate and a matrix applied to the surface using a pneumatic nebulizer (Figure 3) [98]. Spectra are recorded in a systematic fashion by moving the sample stage beneath a fixed laser position. The resulting spot array comprises an image dataset analogous to pixels in a digital photograph. Each laser-irradiated spot (pixel) gives rise to a mass spectrum correlated with the discrete X/Y co-ordinate. Thus each spot contains a dataset with thousands of channels (m/z values), each of which has its own intensity. The intensity values can be expressed in a whole-array context as a two-dimensional ion-density map. These data can then be used to generate images depicting the localization and relative intensities of hundreds of ions from the sample (Figure 3).

The current state-of-the-art requires independent verification of peptide/protein identification, typically achieved by homogenizing the sample and pre-fractionating the proteins by either electrophoresis or LC prior to MS analysis. Newly developed methods should obviate this necessity. Debois et al. [99] recently described the use of MALDI ISD (in-source decay) to fragment ions directly in the MS ion source by using specific matrices (2,5-dihydroxybenzoic acid or 1,5-diaminonaphthalene). This application has led to robust direct protein identification, and there is additionally the possibility of *de novo* sequence analysis.

A consistent problem with IMS is the inconsistency of the methods used for matrix application. Although there is no doubt that improved methods for matrix application will be developed, matrix-free methods are also being developed [100]. One promising matrix-free method has been termed NIMS (nanostructure initiator MS). Instead of coating the surface of the sample with matrix, in NIMS samples are placed on top of a nanostructured matrix [101–103]. A sort of 'Western blot' IMS hybrid has been proposed recently [104], and an interesting and potentially very useful variant of NIMS involves incorporation of the substrate for a given enzyme into the matrix [105]. Conceptually related to enzyme cytochemistry, it would be the product produced during enzyme catalysis that is detected.

As with all other aspects of MS-based proteomic analyses, IMS will continue to benefit from advances in instrument capability, and from application of new and improved methods [106,107]. Intriguing recent developments include affinity-imaging MS, and the potential to use specific chemical barcodes. Imanishi et al. [108] have described a method for phospho-peptide enrichment with claimed femtomolar sensitivity. As described, the method uses glass slides rather than intact tissue, but if one envisions adoption of a liposome-mediated delivery system, such as that described in [109], then future translation of this method to tissues or tissue slices does not seem impossible. The use of liposomes to deliver specific probes (chemical barcodes), coupled with analysis by TOF-SIMS (secondary ion MS) has the potential to localize either specific domains or individual proteins at the single-molecule level [109]. Hopefully there will be a future synergism among researchers in MS, proteomics and cell biology

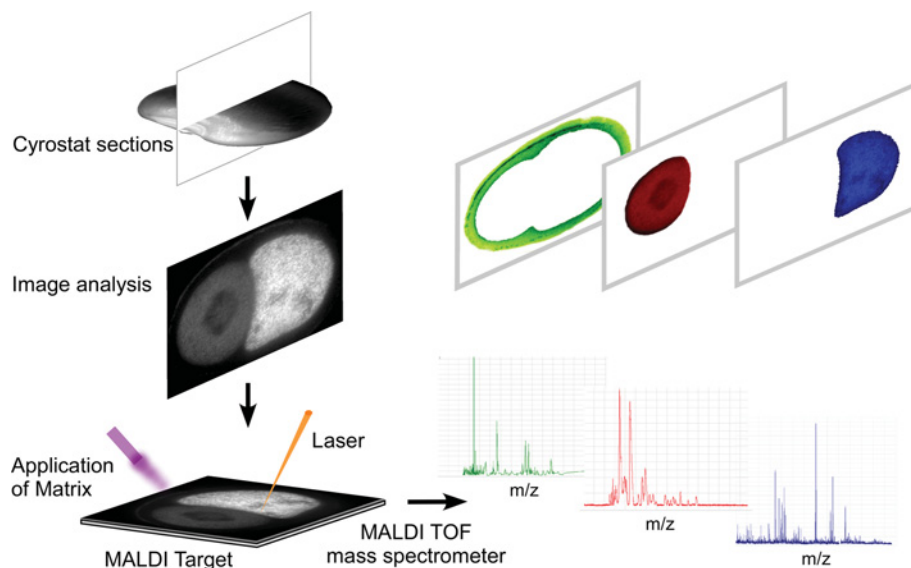


Figure 3 IMS

Thick samples can be sectioned, subjected to imaging and then positioned on the target, followed by matrix deposition using a pneumatic nebulizer. The sample is then irradiated with a laser which desorbs and ionizes the peptides. The use of MALDI is typically coupled with TOF mass analysers where the ions are accelerated at a fixed potential, traverse a field-free flight tube where they are separated by their m/z ratio, and subsequently detected. Spectra are systematically recorded by moving the sample stage beneath a fixed laser position. The resulting spot array comprises a dataset analogous to pixels in a digital image. Each laser-irradiated spot (pixel) gives rise to a mass spectrum correlated to the discrete X/Y co-ordinate. The data can then be used to generate images depicting the localization and relative intensities of ions from the sample.

that will push IMS towards the kind of quantum advances made in light microscopy in recent years (for example [110]).

The (whole) proteome

Although the precise number of genes that comprise the human genome remains elusive, the most commonly stated number is approximately 23 000. Although this is more than enough to generate a complex jigsaw puzzle, it pales in comparison with the size and complexity of the corresponding proteome! Proteome diversity has been estimated as two to three orders of magnitude greater than predicted by the encoding genome (>1 000 000 molecular species of proteins) [111].

A paradox has been developing in recent years relative to proteome complexity. On the one hand, it is common knowledge that proteins are post-translationally modified and occur in different 'isoforms'. On the other hand, biology disclaims PTMs by linking protein names directly with functions. Optimal sequence coverage today is achieved through a combination of bottom-up plus top-down proteomics strategies. Simple identification of a peptide, regardless of how technically difficult this might be, fails to consider that the original primary gene product is sure to exist as the combinatorial sum of several differentially modified forms [112]. Especially with the concept of systems biology looming on the horizon, it is essential that quantitative proteomic analyses address the range of protein species arising from PTMs.

It is generally accepted that there are >200 distinct PTMs. As previously noted, the combinatorial consequences are staggering. Mining of large MS datasets to discover PTM peptides requires careful quality control using techniques such as decoy databases (reversed or randomized forms of the database used for querying). By using such an approach the FDR (false-discovery rate) for any database search can be determined, and it has been determined

that accurate precursor mass (<5 p.p.m.) is essential to keep FDR values below 1%. Thus it will always be an advantage to employ workflows and instruments that offer the best practical mass accuracy. One strategy improved accuracy to the p.p.b. range by internal calibration with the residual electron donor analyte from ETD (electron transfer dissociation) analysis [31]. Of course, this approach requires a high mass accuracy instrument coupled to ETD. Improvements in mass accuracy will continue to be important as this area moves forward.

The most studied [113,114], and to some extent most understood [115], PTM is reversible protein phosphorylation. With the arrival of phosphoproteomics as a popular subdiscipline, novel enrichment and MS analysis approaches have been developed and we anticipate the rapid future expansion of phosphorylation-site mapping. Most phosphopeptide-enrichment strategies utilize metal-affinity chromatography, either transition metals (Fe, Ga and Co) or metal oxides (titanium and zirconium). There are an increasing number of indications that the transition metal and metal oxide approaches are more complementary than redundant in terms of their phosphopeptide selectivity, and are most productively used in a parallel or integrated fashion [116]. Strong cation-exchangers are an alternative to metal-affinity chromatography [117,118], and immunoprecipitation seems to work well with phosphotyrosine-containing peptides [119]. Other approaches have recently been demonstrated, including a variant of hydrophilic-interaction chromatography, termed electrostatic-repulsion hydrophilic-interaction chromatography [120]. Improvements in phosphopeptide enrichment will certainly continue in the coming years, and it is our hope that the reproducibility of using these techniques will be one of the developments.

Improvements in MS instrument sensitivity and speed have greatly aided discovery-based PTM mapping, most notably the release of the dual-pressure linear ion trap that can function either as a stand-alone [29] or when coupled to an Orbitrap mass

analyser [121]. The dual pressure instrument offers improved ion-source transmission and more efficient ion extraction for the higher-energy collision dissociation cell. The improvements in instrument design and performance allowed a 10-fold increase in ion transmission and a 2-fold increase in scanning speed, the combination of which translated to near routine identification of thousands of phosphorylation sites from a single biological sample.

It now appears that the next big thing in the field of PTM will be protein lysine acetylation [122]. Unlike N-terminal residue acetylation, lysine acetylation is reversible and can thus play a regulatory role. Lysine acetylation was initially discovered as a PTM of histone proteins, but it now seems as if the occurrence and diversity of this PTM increases with each new issue of the biochemistry/proteomics/MS journals (for example [123–125]). It has been suggested that protein lysine acetylation will ultimately be as common as phosphorylation. This remains to be seen, but it is noteworthy that there have already been more lysine-acetylated proteins identified in *Escherichia coli* than phosphoproteins [122]. For bottom-up analyses, proteases other than trypsin should be considered. Once the proteins have been fragmented, specific methods of lysine-acetylated peptide enrichment can be used, which will increase discovery rates. Lysine-acetylated peptides do not seem to fragment very differently from their unmodified lysine counterparts in electrospray MS/MS [126].

Protein lysine residues can also be reversibly N-methylated [127] or modified by formation of an isopeptide bond during post-translational conjugation with the small peptide modifiers ubiquitin/SUMO (small ubiquitin-related modifier)/Nedd (neural-precursor-cell-expressed developmentally down-regulated) 8 [128–130]. Although the small peptide modifiers are typically considered as a component of targeted proteostasis, there is also evidence for a role in the control of catalytic activity [131]. In some instances protein purification under denaturing conditions is necessary to preserve protein ubiquitination/SUMOylation/Neddylolation [132].

There are relatively few instances where the complete suite of PTMs has been mapped for a specific protein. One case, however, where this has been achieved is the lens beaded filament protein filensin [133]. A total of nine phosphorylation sites were mapped. Additionally, filensin is proteolytically processed at Asp⁴³¹ and Leu³⁹, and the resulting new N-termini are N-myristoylated and N-acetylated respectively. Finally, aspartic acid isomerization to isoaspartic acid occurs at Asp⁴³¹, yielding a total of 14 distinct PTMs. The number of possible combinations of this number of PTMs (2^{14}) is 16384, well beyond our abilities to separate and/or detect! But even the entirely reasonable number of three PTMs would result in a complex mixture of eight peptides. A virtual comparison of a parent profile with that of the same peptide plus three PTMs is presented in Figure 4. Although it is unlikely that all proteins are as extensively modified as filensin, it seems equally unlikely that there are many proteins with no PTMs. Many aspects of the new biology are fusions between experimental and computational analyses (for example [134,135]). Efforts towards defining the complete proteome clearly represent an area that will benefit from such a hybrid approach.

Cultivate a computer science friend

As an example of the potential value of a computational/experimental partnership for proteome analysis, we will briefly address recent developments in the study of protein phosphorylation. Prediction algorithms can be either general or kinase-specific. As this is written there are more than a dozen widely available phosphorylation-site prediction programs,

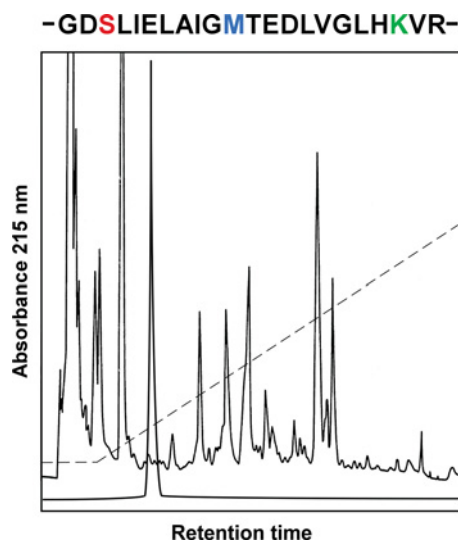


Figure 4 Virtual LC profile of the eight permutations of a peptide with zero, one, two or three potential PTMs

S, serine or phosphoserine; M, methionine or methionine sulfoxide; K, lysine or N-acetylated lysine. The native peptide is a single symmetrical peak, in contrast with the complexity of the virtual profile of a protein with three PTMs.

including the general tools DISPHOS [136], NetPhos [137] and scan-x [138]. Recently, a new program called MUSite, that incorporates disorder prediction as one of three parameters, was shown to outperform the three previous prediction tools on the basis of sensitivity and specificity [139]. When applied to the TAIR9 *Arabidopsis* proteome (ftp://ftp.arabidopsis.org/Genes/TAIR9_genome_release/readme_TAIR9.txt), approximately 18000 potential phosphorylation sites were predicted at the 99% confidence interval. More recent algorithms have validated the importance of protein disorder [140]. Using MS-based approaches approximately 2000 phosphorylation events have been identified from *Arabidopsis* [141,142] and archived in the Plant Protein Phosphorylation Database (P³DB; <http://www.p3db.org>) [143]. These results suggest that although MS-based approaches are accurate, they are not sensitive [144]. The converse could be appropriately stated for the prediction algorithms. In a similar manner to protein structure determination or subcellular localization predictions, however, as we learn more about the rules for protein phosphorylation prediction the tools will become more accurate, a Yin–Yang relationship. Although those using computational tools to study protein phosphorylation have taken an early lead in the application of bioinformatic methods to analysis of the proteome, we assume that those more interested in or intrigued by other PTMs will not be far behind [145].

Proteogenomics

The term proteogenomics was coined relatively recently [146,147], and refers to the use of MS-based proteomic analyses to correct/improve annotation of the genome-based proteome [148,149]. As the post-genomic era advances, it has become increasingly clear that the existence of a DNA sequence is insufficient to provide an understanding of complex biological processes. Despite this, the bulk of current validation methods use information derived solely from an annotated genome. The genome sequence does not, however, unambiguously demonstrate that a predicted ORF (open reading frame) is translated into a protein. This is an untenable situation for proteome annotation.

However, there are now several studies in the literature where predicted genes have been validated at the protein level [150,151].

In addition to the validation of predicted genes and detection of novel genes, results from proteogenomic analyses provide validation of hypothetical ORFs [152], allow accurate determination of protein initiation and termination sites [153], and allow identification of splice variants at the protein level [154].

The need for improved data mining and informatics tools are a key challenge confronting the application of proteogenomic methods to genome annotation. In order for genome sequencing and annotation projects to include high-throughput LC-MS/MS datasets as an essential complement to results from gene-prediction programs, it is critical that new algorithms be developed. These new data mining and informatics tools must be able to seamlessly incorporate the information provided by LC-MS/MS datasets [145,146,148].

Although the principle of searching MS/MS spectra data against six-frame translated genomes to experimentally validate predicted protein-coding genes has been demonstrated in both prokaryotes [155,156] and eukaryotes [152,157], extant techniques are technically limited when contemplating analyses of complex genomes. An obvious disadvantage is the enormous size of a six-frame translated genome, which might be as large as 6 Gb. Simply scaling-up the use of current proteomic search strategies would be impractical, in addition to which the FDR increases in parallel with the database size, decreasing reliability. Improvements in speed and reliability will be necessary in order to develop search routines that are less problematic and more widely applicable. Ideally, any genome-sequencing project should be complemented by nLC-MS/MS-based proteomic profiling [157]. Unfortunately we are not yet at the point where this would be practical.

Biomarkers?

Although much of the interest in MS-based proteomics research has been focused on the intersection of instrumentation and protein (bio)chemistry, there is also an increasing interest in high-throughput biomarker analysis or clinical proteomics as it is sometimes called [158]. For example, in the conference programme for the 59th Annual Meeting of the American Society for Mass Spectrometry (2011) the number of abstracts for presentations on 'Biomarkers' was substantially larger than any other category. Unfortunately, for various biological and technical reasons, researchers have not yet been able to translate this interest with effectiveness to the degree of accuracy necessary for prognostic or monitoring assays. Can proteomics-based strategies actually be useful in a clinical context? Large-scale scepticism has now been expressed by both clinicians and regulatory agencies. It is now accepted that individual biomarkers are not likely to exist or to be widely useful, because the complexity of disease can seldom be captured by a single protein. Although it might be considered an extreme example, it was recently reported that researchers identified a set of 81 disease-associated proteins [159]. Thus a strategy more likely to be successful would employ a panel of biomarkers [160,161].

Unfortunately, the problems inherent to validation of a single biomarker increase logarithmically when considering the use of multiple interacting-markers to provide specific valid signatures [161,162]. How does one test the validity? Can a change in occurrence or abundance of a protein be confidently associated with the disease? Or might it only be an artefact resulting from technical variability? Is the source of the technical variability in the isolation/separation/pre-analytical steps, or during the analysis itself [163]? And finally, do we even know enough

about biological variability or experimental design to realistically consider the possibility of clinical proteomics?

It has been noted that although many biomarkers are proposed in highly cited studies as determinants of disease risk, prognosis or response to treatment, few ultimately reach clinical practice [164]. Thus far the highly cited biomarker studies often report larger effect estimates for associations than are reported in subsequent meta-analyses evaluating the same associations. How many samples need to be analysed in order to define a valid biomarker? How do we design a multi-biomarker clinical platform, and, having designed one, how should it be validated [165]? Are the statistical criteria accepted and acceptable for high-throughput clinical chemistry adequate or adaptable to clinical proteomics? Early results suggest that the increasingly robust statistical methods being developed for and used in bioinformatic analyses will be more appropriate. Results described in recent reports suggest that valid proteomic biomarkers for diagnosis and prognosis can only be developed by applying statistical data mining procedures [166,167]. The authors suggest that multiple testing is necessary, but that sample-size estimation can be performed on the basis of a smaller number of observations via re-sampling from pilot data, and conclude that the sorts of machine-learning algorithms used to predict protein PTMs [168] appear ideally suited to generate the desired signatures [169].

CONCLUDING REMARKS

In the present paper, on the basis of our training, background and experience, we provide a brief review of the state of proteomics research. This is set in the context of an admittedly biased, but not unreasonable, set of suggestions for the short-term future of the field. We touch upon the need for improved tools, strategies and targets. We champion a shift away from qualitative, incomplete and descriptive studies. If a question can be answered effectively by transcript profiling, then why apply a MS-based proteomics strategy? It is the questions that cannot be addressed by using the other omics-based tools, such as protein interactions and the structures of multi-protein complexes, that deserve more of our attention.

ACKNOWLEDGEMENTS

We thank M.L. Johnston who prepared the Figures and L. Meyer who prepared Table 1.

FUNDING

Work in the laboratory of J.J.T. is supported by the NSF and ILSI-Health Environmental Science Institute. Work in the J.A.M. laboratory is supported by the USDA, Agricultural Research Service, NSF, and the Nichols Foundation.

REFERENCES

- 1 Miernyk, J. A., Pret'ová, A., Olmedilla, A., Klubicova, K., Obert, B. and Hajduch, M. (2011) Using proteomics to study sexual reproduction in angiosperms. *Sexual Plant Reprod.* **24**, 9–22
- 2 Miernyk, J. A. and Hajduch, M. (2011) Seed proteomics. *J. Proteomics* **74**, 389–400
- 3 Nilsson, T., Mann, M., Aebersold, R., Yates, III, J.R., Bairock, A. and Bergeron, J.J.M. (2010) Mass spectrometry in high-throughput proteomics: ready for the big time. *Nat. Methods* **7**, 681–685
- 4 Swaney, D. L., Wenger, C. D. and Coon, J. J. (2010) Value of using multiple proteases for large-scale mass spectrometry-based proteomics. *J. Proteome Res.* **9**, 1323–1329

- 5 Olsen, J. V., Ong, S. E. and Mann, M. (2004) Trypsin cleaves exclusively C-terminal to arginine and lysine residues. *Mol. Cell. Proteomics* **3**, 608–614
- 6 Baud, F. and Karlin, S. (1999) Measures of residue density in protein structures. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 12494–12499
- 7 Dana, K. and Nathans, D. (1971) Specific cleavage of Simian virus 40 DNA by restriction endonuclease of *Hemophilus influenzae*. *Proc. Natl. Acad. Sci. U.S.A.* **68**, 2913–2917
- 8 Roberts, R. J. (2005) How restriction enzymes became the workhorses of molecular biology. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 5905–5908
- 9 Capelo, J. L., Carreira, R., Diniz, M., Fernandes, L., Galesio, M., Lodeiro, C., Santos, H. M. and Vale, G. (2009) Overview on modern approaches to speed up protein identification workflows relying on enzymatic cleavage and mass spectrometry-based techniques. *Anal. Chim. Acta* **650**, 151–159
- 10 Kim, J., Kim, B. C., Lopez-Ferrer, D., Petritis, K. and Smith, R. D. (2010) Nanobiocatalysis for protein digestion in proteomic analysis. *Proteomics* **10**, 687–699
- 11 Ray, S., Chandra, H. and Srivastava, S. (2010) Nanotechniques in proteomics: current status, promises and challenges. *Biosens. Bioelectron.* **25**, 2389–2401
- 12 Appel, W. (1986) Chymotrypsin: molecular and catalytic properties. *Clin. Biochem.* **19**, 317–322
- 13 Fischer, F. and Poetsch, A. (2006) Protein cleavage strategies for an improved analysis of the membrane proteome. *Proteome Sci.* **4**, 2
- 14 Wada, Y. and Kadoya, M. (2003) In-gel digestion with endoproteinase LysC. *J. Mass Spectrom.* **38**, 117–118
- 15 Carabetta, V. J., Li, T., Shakya, A., Greco, T. M. and Cristea, B. M. (2010) Integrating Lys-N proteolysis and N-terminal guanidination for improved fragmentation and relative quantification of singly-charged ions. *J. Am. Soc. Mass Spectrom.* **21**, 1050–1060
- 16 Skopp, R. N. and Lane, L. C. (1989) Fingerprinting of proteins cleaved in solution by cyanogen bromide. *Appl. Theor. Electrophor.* **1**, 61–64
- 17 Schroeder, W. A., Shelton, J. A. and Shelton, J. R. (1969) An examination of conditions for the cleavage of polypeptide chains with cyanogen bromide. *Arch. Biochem. Biophys.* **130**, 551–556
- 18 Zhang, H., Yan, W. and Aebersold, R. (2004) Chemical probes and tandem mass spectrometry: a strategy for the quantitative analysis of proteomes and subproteomes. *Curr. Opin. Chem. Biol.* **8**, 66–75
- 19 O'Loughlin, T. L., Greene, D. N. and Matsumura, I. (2006) Diversification and specialization of HIV protease function during *in vitro* evolution. *Mol. Biol. Evol.* **23**, 764–772
- 20 Brewis, I. A. and Brennan, P. (2010) Proteomics technologies for the global identification and quantification of proteins. *Adv. Protein Chem. Struct. Biol.* **80**, 1–44
- 21 Gianazza, E. and Righetti, P. G. (2009) Immobilized pH gradients. *Electrophoresis (Suppl. 1)*, S112–S121
- 22 Minden, J. S., Dowd, S. R., Meyer, H. E. and Stühler, K. (2009) Difference gel electrophoresis. *Electrophoresis (Suppl. 1)*, S156–S161
- 23 Lewandowska-Gnatowska, E., Johnston, M. L., Antoine, W., Szczegielniak, J., Muszyńska, G. and Miernyk, J. A. (2011) Using multiplex-staining to study changes in the maize leaf phosphoproteome in response to mechanical wounding. *Phytochemistry* **72**, 1285–1292
- 24 Pieters, S., Dejaeger, B. and Vander Heyden, Y. (2010) Emerging analytical separation techniques with high throughput potential for pharmaceutical analysis, part I: stationary phase and instrumental developments in LC. *Comb. Chem. High Throughput Screen.* **13**, 510–529
- 25 Hernández-Borges, J., Aturki, Z., Rocco, A. and Fanali, S. (2007) Recent applications in nanoliquid chromatography. *J. Sep. Sci.* **30**, 1589–1610
- 26 Vestal, M. L. (2011) The future of biological mass spectrometry. *J. Am. Soc. Mass Spectrom.* **22**, 953–959
- 27 Dumpala, P. R., Lawrence, M. L. and Karsi, A. (2009) Proteome analysis of *Edwardsiella ictaluri*. *Proteomics* **9**, 1353–1363
- 28 Carranza, P., Grunau, A., Schneider, T., Hartmann, I., Lehner, A., Stephan, R., Gehrig, P., Grossmann, J., Groebel, K., Hoelzle, L. E. et al. (2010) A gel-free quantitative proteomics approach to investigate temperature adaptation of the food-borne pathogen *Cronobacter turicensis*. *Proteomics* **10**, 3248–3261
- 29 Ficarro, S. B., Zhang, Y., Carrasco-Alfonso, M. J., Garg, B., Adelmant, G., Webber, J. T., Luckey, C. J. and Marto, J. A. (2011) Online nanoflow multidimensional fractionation for high efficiency phosphopeptide analysis. *Mol. Cell. Proteomics* **83**, 6996–7005
- 30 Freire, S. L., Yang, H. and Wheeler, A. R. (2008) A practical interface for microfluidics and nano-electrospray mass spectrometry. *Electrophoresis* **29**, 1836–1843
- 31 Second, T. P., Blethrow, J. D., Schwartz, J. C., Merrihew, G. E., MacCoss, M. J., Swaney, D. L., Russell, J. D., Coon, J. J. and Zabrouskov, V. (2009) A novel dual-pressure linear ion trap mass spectrometer improves the analysis of complex protein mixtures. *Anal. Chem.* **81**, 7757–7765
- 32 Garcia, B. A. (2010) What does the future hold for top down mass spectrometry? *J. Am. Soc. Mass Spectrom.* **21**, 193–202
- 33 Liu, X., Inbar, Y., Dorrestein, P. C., Wynne, C., Edwards, N., Souda, P., Whitelegge, J. P., Bafna, V. and Pevzner, P. A. (2010) Deconvolution and database search of complex tandem mass spectra of intact proteins: a combinatorial approach. *Mol. Cell. Proteomics* **9**, 2772–2782
- 34 Tipton, J. D., Tran, J. C., Catherman, A. D., Ahlf, D. R., Durban, K. R. and Kelleher, N. L. (2011) Analysis of intact protein isoforms by mass spectrometry. *J. Biol. Chem.* **286**, 25451–25458
- 35 Tolmachev, A. V., Robinson, E. W., Wu, S., Paša-Tolić, L. and Smith, R. D. (2009) FT-ICR MS optimization for the analysis of intact proteins. *Int. J. Mass Spectrom.* **281**, 32–38
- 36 Shah, B., Kozłowski, R. L., Han, J. and Borchers, C. H. (2011) Emerging mass spectrometry-based technologies for analyses of chromatin changes: analysis of histones and histone modifications. *Methods Mol. Biol.* **773**, 259–303
- 37 Wenger, C. D., McAlister, G. C., Xia, Q. and Coon, J. J. (2010) Sub-part-per-million precursor and product mass accuracy for high-throughput proteomics on an electron transfer dissociation-enabled orbitrap mass spectrometer. *Mol. Cell. Proteomics* **9**, 754–763
- 38 Meyer, B., Papatotiriou, D. G. and Karas, M. (2011) 100% protein sequence coverage: a modern form of surrealism in proteomics. *Amino Acids* **41**, 291–310
- 39 Cannon, J., Lohnes, K., Wynne, C., Wang, Y., Edwards, N. and Fenselau, C. (2010) High-throughput middle-down analysis using an orbitrap. *J. Proteome Res.* **9**, 3886–3890
- 40 Thelen, J. J. and Peck, S. C. (2007) Quantitative proteomics in plants: choices in abundance. *Plant Cell* **19**, 3339–3346
- 41 Schulze, W. X. and Usadel, B. (2010) Quantitation in mass-spectrometry-based proteomics. *Annu. Rev. Plant Biol.* **61**, 491–516
- 42 Carroll, K. M., Simpson, D. M., Eyers, C. E., Knight, C. G., Brownbridge, P., Dunn, W. B., Winder, C. L., Lanthaler, K., Pir, P., Malys, N. et al. (2011) Absolute quantification of the glycolytic pathway in yeast: deployment of a complete QconCAT approach. *Mol. Cell. Proteomics* **10**, M111.007633
- 43 Treumann, A. and Thiede, B. (2010) Isobaric protein and peptide quantification: perspectives and issues. *Expert Rev. Proteomics* **7**, 647–653
- 44 Dayon, L., Pasquarello, C., Hoogland, C., Sanchez, J. C. and Scherl, A. (2010) Combining low- and high-energy tandem mass spectra for optimized peptide quantification with isobaric tags. *J. Proteomics* **73**, 769–777
- 45 Stevenson, S. E., Chu, Y., Ozias-Akins, P. and Thelen, J. J. (2009) Validation of gel-free, label-free quantitative proteomics approaches: applications for seed allergen profiling. *J. Proteomics* **72**, 555–566
- 46 Collier, T. S., Sarkar, P., Franck, W. L., Rao, B. M., Dean, R. A. and Muddiman, D. C. (2010) Direct comparison of stable isotope labeling by amino acids in cell culture and spectral counting for quantitative proteomics. *Anal. Chem.* **82**, 8696–8702
- 47 Neilson, K. A., Ali, N. A., Muralidharan, S., Mirzaei, M., Mariani, M., Assadourian, G., Lee, A., van Sluyter, S. C. and Haynes, P. A. (2011) Less label, more free: approaches in label-free quantitative mass spectrometry. *Proteomics* **11**, 535–553
- 48 Lee, S., Kwon, M., Lee, H., Paik, Y., Tang, H., Lee, J. K. and Park, T. (2011) Enhanced peptide quantification using spectral count clustering and cluster abundance. *BMC Bioinformatics* **12**, 423
- 49 Gokce, E., Shuford, C. M., Franck, W. L., Dean, R. A. and Muddiman, D. C. (2011) Evaluation of normalization methods on GeLC-MS/MS label-free spectral counting data to correct for variation during proteomic workflows. *J. Am. Soc. Mass Spectrom.* **22**, 2199–2208
- 50 Mekhedov, S., de Ilárduya, O. M. and Ohlrogge, J. B. (2000) Toward a functional catalog of the plant genome. A survey of genes for lipid biosynthesis. *Plant Physiol.* **122**, 389–402
- 51 Dicker, L., Lin, X. and Ivanov, A. R. (2010) Increased power for the analysis of label-free LC-MS/MS proteomics data by combining spectral counts and peptide peak attributes. *Mol. Cell. Proteomics* **9**, 2704–2718
- 52 Houston, N. L., Lee, D. G., Stevenson, S. E., Ladics, G. S., Bannon, G. A., McClain, S., Privalle, L., Stagg, N., Herouet-Guicheney, C., Macintosh, S. C. and Thelen, J. J. (2011) Quantitation of soybean allergens using tandem mass spectrometry. *J. Proteome Res.* **10**, 763–773
- 53 Alberts, B. (1998) The cell as a collection of protein machines: preparing the next generation of molecular biologists. *Cell* **92**, 291–294
- 54 Oeljeklaus, S., Meyer, H. E. and Warscheid, B. (2009) New dimensions in the study of protein complexes using quantitative mass spectrometry. *FEBS Lett.* **583**, 1674–1683
- 55 Sharon, M. (2010) How far can we go with structural mass spectrometry of protein complexes? *J. Am. Soc. Mass Spectrom.* **21**, 487–500
- 56 Hart, G. T., Ramani, A. K. and Marcotte, E. M. (2006) How complete are current yeast and human protein-interaction networks? *Genome Biol.* **7**, 120
- 57 Reguly, T., Breitkreutz, A., Boucher, L., Breitkreutz, B. J., Hon, G. C., Myers, C. L., Parsons, A., Friesen, H., Oughtred, R., Tong, A. et al. (2006) Comprehensive curation and analysis of global interaction networks in *Saccharomyces cerevisiae*. *J. Biol.* **5**, 11

- 58 Brückner, A., Polge, C., Lentze, N., Auerbach, D. and Schlattner, U. (2009) Yeast two-hybrid, a powerful tool for systems biology. *Int. J. Mol. Sci.* **10**, 2763–2788
- 59 Collins, M. O. and Choudhary, J. S. (2008) Mapping multiprotein complexes by affinity purification and mass spectrometry. *Curr. Opin. Biotechnol.* **19**, 324–330
- 60 Kaake, R. M., Wang, X. and Huang, L. (2010) Profiling of protein interaction networks of protein complexes using affinity purification and quantitative mass spectrometry. *Mol. Cell. Proteomics* **9**, 1650–1665
- 61 Moresco, J. J., Carvalho, P. C. and Yates, III, J.R. (2010) Identifying components of protein complexes in *C. elegans* using co-immunoprecipitation and mass spectrometry. *J. Proteomics* **73**, 2198–2204
- 62 Ito, T., Ota, K., Kubota, H., Yamaguchi, Y., Chiba, T., Sakuraba, K. and Yoshida, M. (2002) Roles for the two-hybrid system in exploration of the yeast protein interactome. *Mol. Cell. Proteomics* **1**, 561–566
- 63 Giot, L., Bader, J., Brouwer, C., Chaudhuri, A., Kuang, B., Li, Y., Hao, Y., Ooi, C., Godwin, B., Vitols, E. et al. (2004) A protein interaction map of *Drosophila melanogaster*. *Science* **302**, 1727–1736
- 64 Li, S., Armstrong, C., Bertin, N., Ge, H., Milstein, S., Boxem, M., Vidalain, P. O., Han, J. D., Chesneau, A., Ha, T. et al. (2004) A map of the interactome network of the metazoan *C. elegans*. *Science* **303**, 540–543
- 65 Shin, C. J., Davis, M. J. and Ragan, M. A. (2009) Towards the mammalian interactome: inference of a core mammalian interaction set in mouse. *Proteomics* **9**, 5256–5266
- 66 Lewis, A. C., Saeed, R. and Deane, C. M. (2010) Predicting protein-protein interactions in the context of protein evolution. *Mol. Biosyst.* **6**, 55–64
- 67 Gomez, S. M. (2003) Prediction of protein-protein interaction networks. *Curr. Protoc. Bioinformatics* **8**, 8.2
- 68 Puig, O., Casparly, F., Rigaut, G., Rutz, B., Bouveret, E., Bragado-Nilsson, E., Wilm, M. and Séraphin, B. (2001) The tandem affinity purification (TAP) method: a general procedure of protein complex purification. *Methods* **24**, 218–229
- 69 Olinares, P.D.B., Ponnala, L. and van Wijk, K. J. (2010) Megadalton complexes in the chloroplast stroma of *Arabidopsis thaliana* characterized by size exclusion chromatography, mass spectrometry, and hierarchical clustering. *Mol. Cell. Proteomics* **9**, 1594–615
- 70 Smith, V. F., Schwartz, B. L., Randall, L. L. and Smith, R. R. (1996) Electrospray mass spectrometric investigation of the chaperone SecB. *Protein Sci.* **5**, 488–494
- 71 Ferrige, A. G., Seddon, M. J., Jarvis, S., Skilling, J. and Aplin, R. (1991) Maximum entropy deconvolution in electrospray mass spectrometry. *Rapid Commun. Mass Spectrom.* **5**, 374–377
- 72 Ferrige, A. G., Seddon, M. J., Green, B. N., Jarvis, S. A., Skilling, K. and Staunton, J. (1992) Disentangling electrospray spectra with maximum entropy. *Rapid Commun. Mass Spectrom.* **6**, 701–711
- 73 Luo, Y., Li, T., Yu, F., Kramer, T. and Cristea, I. M. (2010) Resolving the composition of protein complexes using a MALDI LTQ Orbitrap. *J. Am. Soc. Mass Spectrom.* **21**, 31–46
- 74 Miernyk, J. A. and Thelen, J. J. (2008) Biochemical approaches for discovering protein-protein interactions. *Plant J.* **53**, 597–609
- 75 Renzone, G., Salzano, A. M., Arena, S., D'Ambrosio, C. and Scaloni, A. (2007) Mass spectrometry-based approaches for structural studies on protein complexes at low-resolution. *Curr. Proteomics* **4**, 1–16
- 76 Götz, M., Pettelkau, J., Schaks, S., Bosse, K., Ihling, C. H., Krauth, F., Fritzsche, R., Kühn, U. and Sinz, A. (2012) StavroX-A software for analyzing crosslinked products in protein interaction studies. *J. Am. Soc. Mass Spectrom.* **23**, 76–87
- 77 Paramelle, D., Cantel, S., Enjalbal, C., Amblard, M., Forest, E., Heymann, M., Geourjon, C., Martinez, J. and Subra, G. (2009) A new generation of cross-linkers for selective detection by MALDI MS. *Proteomics* **9**, 5384–5388
- 78 Benesch, J.L.P., Aquilina, J. A., Ruotolo, B. T., Sobott, F. and Robinson, C. V. (2006) Tandem mass spectrometry reveals the quaternary organization of macromolecular assemblies. *Chem. Biol.* **13**, 597–605
- 79 Pukala, T. L., Ruotolo, B. T., Zhou, M., Politis, A., Stefanescu, R., Leary, J. A. and Robinson, C. V. (2009) Subunit architecture of multiprotein assemblies determined using restraints from gas-phase measurements. *Structure* **17**, 1235–1243
- 80 Utrecht, C., Rose, R. J., van Duijn, E., Lorenzen, K. and Heck, A. J. (2010) Ion mobility mass spectrometry of proteins and protein assemblies. *Chem. Soc. Rev.* **39**, 1633–1655
- 81 Cheng, Y. (2009) Toward an atomic model of the 26S proteasome. *Curr. Opin. Struct. Biol.* **19**, 203–208
- 82 Giles, K., Pringle, S. D., Worthington, K. R., Little, D., Wildgoose, J. L. and Bateman, R. H. (2004) Applications of a travelling wave-based radio-frequency-only stacked ring ion guide. *Rapid Commun. Mass Spectrom.* **18**, 2401–2414
- 83 Richter, F. M., Sander, B., Golas, M. M., Stark, H. and Urlaub, H. (2010) Merging molecular electron microscopy and mass spectrometry by carbon film-assisted endoprotease digestion. *Mol. Cell. Proteomics* **9**, 1729–1741
- 84 Kundrotas, P. J. and Vakser, I. A. (2010) Accuracy of protein-protein binding sites in high-throughput template-based modeling. *PLoS Comput. Biol.* **6**, e1000727
- 85 Guda, C., King, B. R., Pal, L. R. and Guda, P. (2009) A top-down approach to infer and compare domain-domain interactions across eight model organisms. *PLoS ONE* **4**, e5096
- 86 Hugo, W., Song, F., Aung, Z., Ng, S. K. and Sung, W. K. (2010) SLiM on diet: finding short linear motifs on domain interaction interfaces in Protein Data Bank. *Bioinformatics* **26**, 1036–1042
- 87 Blayney, M. J., Whitney, S. M. and Beck, J. L. (2011) NanoESI mass spectrometry of Rubisco and Rubisco activase structures and their interactions with nucleotides and sugar phosphates. *J. Am. Soc. Mass Spectrom.* **22**, 1588–1601
- 88 Hamdy, O. M. and Julian, R. R. (2012) Reflections on charge state distributions, protein structure, and the mystical mechanism of electrospray ionization. *J. Am. Soc. Mass Spectrom.* **23**, 1–6
- 89 Esquenazi, E., Coates, C., Simmons, L., Gonzalez, D., Gerwick, W. H. and Dorrestein, P. C. (2008) Visualizing the spatial distribution of secondary metabolites produced by marine cyanobacteria and sponges via MALDI-TOF imaging. *Mol. Biosyst.* **4**, 562–570
- 90 Setou, M., Shrivasa, K., Sroyraya, M., Yang, H., Sugiura, Y., Moribe, J., Kondo, A., Tsutsumi, K., Kimura, Y., Kurabe, N. et al. (2010) Developments and applications of mass microscopy. *Med. Mol. Morphol.* **43**, 1–5
- 91 Setou, M. and Kurabe, N. (2010) Mass microscopy: high-resolution imaging mass spectrometry. *J. Electron Microsc. (Tokyo)* **60**, 47–56
- 92 Seeley, E. H. and Caprioli, R. M. (2008) Mass spectrometry special feature: molecular imaging of proteins in tissues by mass spectrometry. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 18126–18131
- 93 Seeley, E. H., Oppenheimer, S. R., Mi, D., Chaurand, P. and Caprioli, R. M. (2008) Enhancement of protein sensitivity for MALDI imaging mass spectrometry after chemical treatment of tissue sections. *J. Am. Soc. Mass Spectrom.* **19**, 1069–1077
- 94 van Remoortere, A., van Zeijl, R.J.M., van den Oever, N., Franck, J., Longuespee, R., Wiszorski, M., Salzet, M., Deelder, A. M., Fournier, I. and McDonnell, L.A. (2010) MALDI imaging and profiling MS of higher mass proteins from tissue. *J. Am. Soc. Mass Spectrom.* **21**, 1922–1929
- 95 Pól, J., Strohal, M., Havlíček, V. and Volný, M. (2010) Molecular mass spectrometry imaging in biomedical and life science research. *Histochem. Cell Biol.* **134**, 423–443
- 96 Watrous, J. D., Alexandrov, T. and Dorrestein, P. C. (2011) The evolving field of imaging mass spectrometry and its impact on future biological research. *J. Mass Spectrom.* **46**, 209–222
- 97 Kurnosenko, S. and Moskovets, E. (2010) On the high-resolution mass analysis of the product ions in tandem time-of-flight (TOF/TOF) mass spectrometers using a time-dependent re-acceleration technique. *Rapid Commun. Mass Spectrom.* **24**, 63–74
- 98 Kaletas, B. K., van der Wiel, I. M., Stauber, J., Dekker, L. J., Güzel, C., Kros, J. M., Luider, T. M. and Heeren, R. M. (2009) Sample preparation issues for tissue imaging by imaging MS. *Proteomics* **9**, 2622–2633
- 99 Debois, D., Bertrand, V., Quinton, L., De Pauw-Gillet, M. C. and De Pauw, E. (2010) MALDI-in source decay applied to mass spectrometry imaging: a new tool for protein identification. *Anal. Chem.* **82**, 4036–4045
- 100 Vidová, V., Novák, P., Strohal, M., Pól, J., Havlíček, V. and Volný, M. (2010) Laser desorption-ionization of lipid transfers: tissue mass spectrometry imaging without MALDI matrix. *Anal. Chem.* **82**, 4994–4997
- 101 Northen, T. R., Yanes, O., Northen, M. T., Marrinucci, D., Uritboonthai, W., Apon, J., Golledge, S. L., Nordström, A. and Siuzdak, G. (2007) Clathrate nanostructures for mass spectrometry. *Nature* **449**, 1033–1036
- 102 Woo, H. K., Northen, T. R., Yanes, O. and Siuzdak, G. (2008) Nanostructure-initiator mass spectrometry: a protocol for preparing and applying NIMS surfaces for high-sensitivity mass analysis. *Nat. Protoc.* **3**, 1341–1349
- 103 Yanes, O., Woo, H. K., Northen, T. R., Oppenheimer, S. R., Shriver, L., Apon, J., Estrada, M. N., Potchoiba, M. J., Steenwyk, R., Manchester, M. and Siuzdak, G. (2009) Nanostructure initiator mass spectrometry: tissue imaging and direct biofluid analysis. *Anal. Chem.* **81**, 2969–2975
- 104 van Hove, E.R.A., Smith, D. F., Fornai, L., Glunde, K. and Heeren, R.M.A. (2011) An alternative paper based tissue washing method for mass spectrometry imaging: localized washing and fragile tissue analysis. *J. Am. Soc. Mass Spectrom.* **22**, 1885–1890
- 105 Northen, T. R., Lee, J. C., Hoang, L., Raymond, J., Hwang, D. R., Yannone, S. M., Wong, C. H. and Siuzdak, G. (2008) A nanostructure-initiator mass spectrometry-based enzyme activity assay. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 3678–3683
- 106 Stauber, J., MacAleese, L., Franck, J., Claude, E., Snel, M., Kaletas, B. K., Wiel, I. M., Wiszorski, M., Fournier, I. and Heeren, R. M. (2010) On-tissue protein identification and imaging by MALDI-ion mobility mass spectrometry. *J. Am. Soc. Mass Spectrom.* **21**, 338–47
- 107 Inutan, E. D., Richards, A. L., Wager-Miller, J., Mackie, K., McEwen, C. N. and Trimpin, S. (2011) Laserspray ionization, a new method for protein analysis directly from tissue at atmospheric pressure with ultrahigh mass resolution and electron transfer dissociation. *Mol. Cell. Proteomics* **10**, 1074

- 108 Imanishi, S. Y., Kouvonen, P., Småtå, J. H., Heikkilä, M., Peuhu, E., Mikhailov, A., Ritala, M., Lindén, M., Corthals, G. L. and Eriksson, J. E. (2009) Phosphopeptide enrichment with stable spatial coordination on a titanium dioxide coated glass slide. *Rapid Commun. Mass Spectrom.* **23**, 3661–3667
- 109 Gunnarsson, A., Sjövall, P. and Höök, F. (2010) Liposome-based chemical barcodes for single molecule DNA detection using imaging mass spectrometry. *Nano Lett.* **10**, 732–737
- 110 Mortensen, K. I., Churchman, L. S., Spudich, J. A. and Flyvbjerg, H. (2010) Optimized localization analysis for single-molecule tracking and super-resolution microscopy. *Nat. Methods* **7**, 377–381
- 111 Walsh, C. T., Garneau-Tsodikova, S. and Gatto, Jr, G. J. (2005) Protein posttranslational modifications: the chemistry of proteome diversifications. *Angew. Chem., Int. Ed. Engl.* **44**, 7342–7372
- 112 Seo, J., Jeong, J., Kim, Y. M., Hwang, N., Paek, E. and Lee, K. J. (2008) Strategy for comprehensive identification of post-translational modifications in cellular proteins, including low abundant modifications: application to glyceraldehyde-3-phosphate dehydrogenase. *J. Proteome Res.* **7**, 587–602
- 113 Kosako, H. and Nagano, K. (2011) Quantitative phosphoproteomics strategies for understanding protein kinase-mediated signal transduction pathways. *Expert Rev. Proteomics* **8**, 81–94
- 114 Ozlu, N., Akten, B., Timm, W., Haseley, N., Steen, H. and Steen, J. A. (2010) Phosphoproteomics. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **2**, 255–276
- 115 Cutillas, P. R. and Jorgensen, C. (2011) Biological signalling activity measurements using mass spectrometry. *Biochem. J.* **434**, 189–199
- 116 Wilson-Grady, J. T., Villen, J. and Gygi, S. P. (2008) Phosphoproteome analysis of fission yeast. *J. Proteome Res.* **7**, 1088–1097
- 117 Cantin, G., Yi, W., Lu, B., Park, S. K., Xu, T., Lee, J. D. and Yates, III, J.R. (2008) Combining protein-based IMAC, peptide-based IMAC, and MudPIT for efficient phosphoproteomic analysis. *J. Proteome Res.* **7**, 1346–1351
- 118 Villen, J. and Gygi, S. P. (2008) The SCX/IMAC enrichment approach for global phosphorylation analysis by mass spectrometry. *Nat. Protoc.* **3**, 1630–1638
- 119 Villen, J., Beausoleil, S. A., Gerber, S. A. and Gygi, S. P. (2007) Large-scale phosphorylation analysis of mouse liver. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 1488–1493
- 120 Alpert, A. J. (2008) Electrostatic repulsion hydrophilic interaction chromatography for isocratic separation of charged solutes and selective isolation of phosphopeptides. *Anal. Chem.* **80**, 62–76
- 121 Olsen, J. V., Schwartz, J. C., Griep-Raming, J., Nielsen, M. L., Damoc, E., Denisov, E., Lange, O., Remes, P., Taylor, D., Splendore, M. et al. (2009) A dual pressure linear ion trap orbitrap instrument with very high sequencing speed. *Mol. Cell. Proteomics* **8**, 2759–2769
- 122 Mischerikow, N. and Heck, A. J. (2011) Targeted large-scale analysis of protein acetylation. *Proteomics* **11**, 571–589
- 123 McEwan, D. G. and Dikic, I. (2011) The three musketeers of autophagy: phosphorylation, ubiquitylation and acetylation. *Trends Cell Biol.* **21**, 195–201
- 124 Wu, X., Oh, M. H., Schwarz, E., Larue, C. T., Sivaguru, M., Imai, B. S., Yau, P. M., Ort, D. R. and Huber, S. C. (2011) Lysine acetylation is a widespread protein modification for diverse proteins in *Arabidopsis*. *Plant Physiol.* **155**, 1769–1778
- 125 Finkemeier, I., Laxa, M., Miguet, L., Howden, A. and Sweetlove, L. (2011) Proteins of diverse function and subcellular location are lysine-acetylated in *Arabidopsis*. *Plant Physiol.* **155**, 1779–1790
- 126 Li, Y. and Ball, H. L. (2009) Study of acetylation on Ser/Thr/Tyr/Lys, and trimethylation on Lys using electrospray tandem mass spectrometry. *Int. J. Mass Spectrom.* **281**, 24–30
- 127 Villeneuve, L. M., Reddy, M. A. and Natarajan, R. (2011) Epigenetics: deciphering its role in diabetes and its chronic complications. *Clin. Exp. Pharmacol. Physiol.* **38**, 401–409
- 128 Komander, D. (2009) The emerging complexity of protein ubiquitination. *Biochem. Soc. Trans.* **37**, 937–953
- 129 Hannoun, Z., Greenhough, S., Jaffray, E., Hay, R. T. and Hay, D. C. (2010) Post-translational modification by SUMO. *Toxicology* **278**, 288–293
- 130 Rabut, G. and Peter, M. (2008) Function and regulation of protein neddylation. *EMBO Rep.* **9**, 969–976
- 131 Uhrig, R. G., She, Y. M., Leach, C. A. and Plaxton, W. C. (2008) Regulatory monoubiquitination of phosphoenolpyruvate carboxylase in germinating castor oil seeds. *J. Biol. Chem.* **283**, 29650–29657
- 132 Maine, G. N., Li, H., Zaidi, I. W., Basrur, V., Elenitoba-Johnson, K. S. and Burstein, E. (2010) A bimolecular affinity purification method under denaturing conditions for rapid isolation of a ubiquitinated protein for mass spectrometry analysis. *Nat. Protoc.* **5**, 1447–1459
- 133 Wang, Z., Obidike, J. E. and Schey, K.L. (2010) Posttranslational modifications of the bovine lens beaded filament proteins filensin and CP49. *Invest. Ophthalmol. Vis. Sci.* **51**, 1565–1574
- 134 Daga, P. R., Patel, R. Y. and Doerksen, R. J. (2010) Template-based protein modeling: recent methodological advances. *Curr. Top. Med. Chem.* **10**, 84–94
- 135 Rastogi, S. and Rost, B. (2010) Bioinformatics predictions of localization and targeting. *Methods Mol. Biol.* **619**, 285–305
- 136 Iakoucheva, L. M., Radivojac, P., Brown, C. J., O'Connor, T. R., Sikes, J. G., Obradovic, Z. and Dunker, A. K. (2004) The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res.* **32**, 1037–1049
- 137 Blom, N., Gammeltoft, S. and Brunak, S. (1999) Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J. Mol. Biol.* **294**, 1351–1362
- 138 Schwartz, D., Chou, M. F. and Church, G. M. (2009) Predicting protein post-translational modifications using meta-analysis of proteome scale data sets. *Mol. Cell Proteomics* **8**, 365–379
- 139 Gao, J., Thelen, J. J., Dunker, A. K. and Xu, D. (2010) MUsite: a tool for global prediction of general and kinase-specific phosphorylation sites. *Mol. Cell. Proteomics* **9**, 2586–2600
- 140 Rao, R. S. and Møller, I. M. (2012) Large-scale analysis of phosphorylation site occupancy in eukaryotic proteins. *Biochim. Biophys. Acta* **1824**, 405–412
- 141 Sugiyama, N., Nakagami, H., Mochida, K., Daudi, A., Tomita, M., Shirasu, K. and Ishihama, Y. (2008) Large-scale phosphorylation mapping reveals the extent of tyrosine phosphorylation in *Arabidopsis*. *Mol. Syst. Biol.* **4**, 193
- 142 Reiland, S., Messerli, G., Baerenfaller, K., Gerrits, B., Endler, A., Grossmann, J., Gruissem, W. and Baginsky, S. (2009) Large-scale *Arabidopsis* phosphoproteome profiling reveals novel chloroplast kinase substrates and phosphorylation networks. *Plant Physiol.* **150**, 889–903
- 143 Gao, J., Agrawal, G. K., Thelen, J. J. and Xu, D. (2009) P3DB: a plant protein phosphorylation database. *Nucleic Acids Res.* **37**, D960–D962
- 144 Brenton, A. G. and Godfrey, A. R. (2010) Accurate mass measurements: terminology and treatment of data. *J. Am. Soc. Mass Spectrom.* **21**, 1821–1835
- 145 Kim, M. S. and Pandey, A. (2012) Electron transfer dissociation mass spectrometry in proteomics. *Proteomics*, doi: 10.1002/pmic.201100517
- 146 Ansong, C., Purvine, S. O., Adkins, J. N., Lipton, M. S. and Smith, R. D. (2008) Proteogenomics: needs and roles to be filled by proteomics in genome annotation. *Briefings Funct. Genomics Proteomics* **7**, 50–62
- 147 Armengaud, J. (2009) A perfect genome annotation is within reach with the proteomics and genomics alliance. *Curr. Opin. Microbiol.* **12**, 292–300
- 148 Renuse, S., Chaerkady, R. and Pandey, A. (2011) Proteogenomics. *Proteomics* **11**, 620–630
- 149 Krug, K., Nahnsen, S. and Macek, B. (2011) Mass spectrometry at the interface of proteomics and genomics. *Mol. Biosyst.* **7**, 284–291
- 150 Castellana, N. E., Payne, S. H., Shen, Z., Stanke, M., Bafna, V. and Briggs, S. P. (2008) Discovery and revision of *Arabidopsis* genes by proteogenomics. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 21034–21038
- 151 Bringans, S., Hane, J. K., Casey, T., Tan, K. C., Lipscombe, R., Solomon, P. S. and Oliver, R. P. (2009) Deep proteogenomics: high throughput gene validation by multidimensional liquid chromatography and mass spectrometry of proteins from the fungal wheat pathogen *Stagonospora nodorum*. *BMC Bioinformatics* **10**, 301
- 152 Xing, X. B., Li, Q. R., Sun, H., Fu, X., Zhan, F., Huang, X., Li, J., Chen, C. L., Shyr, Y., Zeng, R. et al. (2011) The discovery of novel protein-coding features in mouse genome based on mass spectrometry data. *Genomics* **98**, 343–351
- 153 Castellana, N. and Bafna, V. (2010) Proteogenomics to discover the full coding content of genomes: a computational perspective. *J. Proteomics* **73**, 2124–2135
- 154 Armengaud, J. (2010) Proteogenomics and systems biology: quest for the ultimate missing parts. *Exp. Rev. Proteomics* **7**, 65–77
- 155 Christie-Oleza, J. A., Miotello, G. and Armengaud, J. (2012) High-throughput proteogenomics of *Ruegeria pomeroyi*: seeding a better genomic annotation for the whole marine *Roseobacter clade*. *BMC Genomics* **13**, 73
- 156 Venter, E., Smith, R. D. and Payne, S. H. (2011) Proteogenomic analysis of bacteria and archaea: a 46 organism case study. *PLoS ONE* **6**, e27587
- 157 Helmy, M., Tomita, M. and Ishihama, Y. (2011) OryzaDB: rice proteome database based on shotgun proteogenomics. *BMC Plant Biol.* **11**, 63
- 158 Dakna, M., Keith, H., Kalousis, A., Carpenter, S., Kolch, W., Schanstra, J., Haubitz, M., Vlahou, A., Mischak, H. and Girolami, M. (2010) Addressing the challenge of defining valid proteomic biomarkers and classifiers. *BMC Bioinformatics* **11**, 594
- 159 Schlatter, D. M., Dazard, J. E., Ewing, R. M., Ilchenko, S., Tomcheko, S. E., Eid, S., Ho, V., Yanik, G., Chance, M. R. and Cooke, K. R. (2012) Human biomarker discovery and predictive models for disease progression for idiopathic pneumonia syndrome following allogeneic stem cell transplantation. *Mol. Cell. Proteomics*, doi:10.1074/mcp.M111.015479
- 160 Tjalsma, H. (2010) Identification of biomarkers for colorectal cancer through proteomics-based approaches. *Exp. Rev. Proteomics* **7**, 879–895

- 161 Fu, Q., Schoenhoff, F. S., Savage, W. J., Zhang, P. and Van Eyk, J. E. (2010) Multiplex assays for biomarker research and clinical application: translational science coming of age. *Proteomics Clin. Appl.* **4**, 271–284
- 162 Imai, K., Koshiyama, A. and Nakata, K. (2011) Towards clinical proteomics analysis. *Biomed. Chromatogr.* **25**, 59–64
- 163 Walsh, G. M., Rogalski, J. C., Klockenbusch, C. and Kast, J. (2010) Mass spectrometry-based proteomics in biomedical research: emerging technologies and future strategies. *Exp. Rev. Mol. Med.* **12**, e30
- 164 Ioannidis, J. P. and Panagiotou, O. A. (2011) Comparison of effect sizes associated with biomarkers reported in highly cited individual articles and in subsequent meta-analyses. *JAMA, J. Am. Med. Assoc.* **305**, 2200–2211
- 165 Ioannidis, J. P. and Khoury, M. J. (2011) Improving validation practices in 'omics' research. *Science* **334**, 1230–1232
- 166 Rajalahti, T., Kroksveen, A. C., Arneberg, R., Berven, F. S., Vedeler, C. A., Myhr, K. M. and Kvalheim, O. M. (2010) A multivariate approach to reveal biomarker signatures for disease classification: application to mass spectral profiles of cerebrospinal fluid from patients with multiple sclerosis. *J. Proteome Res.* **9**, 3608–3620
- 167 Huillet, C., Adrait, A., Lebert, D., Picard, G., Trauchessec, M., Louwagie, M., Dupuis, A., Hittinger, L., Ghaleh, B., Le Corvoisier, P. et al. (2012) Accurate quantification of cardiovascular biomarkers in serum using protein standard absolute quantification (PSAQ) and selected reaction monitoring. *Mol. Cell. Proteomics* **11**, M111.008235
- 168 Chalkley, R. J. and Clauser, K. R. (2012) Modification site localization scoring: strategies and performance. *Mol. Cell. Proteomics*, doi:10.1074/mcp.R111.015305
- 169 Naegle, K. M., Welsch, R. E., Yaffe, M. B., White, F. M. and Lauffenburger, D. A. (2011) MCAM: multiple clustering analysis methodology for deriving hypotheses and insights from high-throughput proteomic datasets. *PLoS Comput. Biol.* **7**, e1002119

Received 25 February 2011/22 February 2012; accepted 1 March 2012
Published on the Internet 11 May 2012, doi:10.1042/BJ20110363