Journal of Pharmaceutical and Biomedical Analysis xxx (2011) xxx–xxx

Contents lists available at ScienceDirect

Journal of Pharmaceutical and Biomedical Analysis

journal homepage: www.elsevier.com/locate/jpba

Review

Proteomics by mass spectrometry—Go big or go home?

Morgan F. Khan, Melissa J. Bennett, Chanelle C. Jumper, Andrew J. Percy, Leslie P. Silva, David C. Schriemer*

University of Calgary, Department of Biochemistry and Molecular Biology, 3330 Hospital Drive NW, Calgary, Alberta, Canada T2N 4N1

ARTICLE INFO

Article history:
Received 4 November 2010
Received in revised form 3 February 2011
Accepted 10 February 2011
Available online xxx

Keywords:
Proteomics
Mass spectrometry
Selected reaction monitoring
Chromatography
Bioinformatics

ABSTRACT

Mass spectrometry is an important technology for mapping composition and flux in whole proteomes. Over the last 5 years in particular, impressive gains in the depth of proteome coverage have been realized, particularly for model organisms. This review will provide an update on advancements in the key analytical techniques, methods and informatics directed towards whole proteome analysis by mass spectrometry. Practical issues involving sample requirements, analysis time and depth of coverage will be addressed, to gauge how useful data-driven approaches are for solving biological problems. Targeted mass spectrometric methods, based on selected reaction monitoring, are presented as a powerful alternative to data-driven methods. They offer robust, transferable protocols for hypothesis-directed monitoring of limited yet biologically significant tracts of any proteome.

Contents

1. Introduction

Proteomics as a discipline may be defined as the monitoring of all proteins within an organism, in both temporal and spatial terms. That is, at any given point in time, what proteins are expressed and where are they? While this sort of question defines the core technical issue for many endeavors in molecular biology, proteomics differentiates itself on the basis of the number of proteins monitored—all vs. a select few. A comprehensive analysis would have the advantage of avoiding bias when monitoring a disease state or a biological mechanism, and thus has considerable appeal.

As the field has existed for approximately 15 years, it is reasonable to evaluate how close we are to providing reliable methods for proteome analysis. Can established methods be placed into individual labs to deliver proteome characterization within a rea-

* Corresponding author. Tel.: +1 403 210 3811; fax: +1 403 283 8727.
E-mail address: dschriem@ucalgary.ca (D.C. Schriemer).

0731-7085/$ – see front matter © 2011 Published by Elsevier B.V.
doi:10.1016/j.jpba.2011.02.012

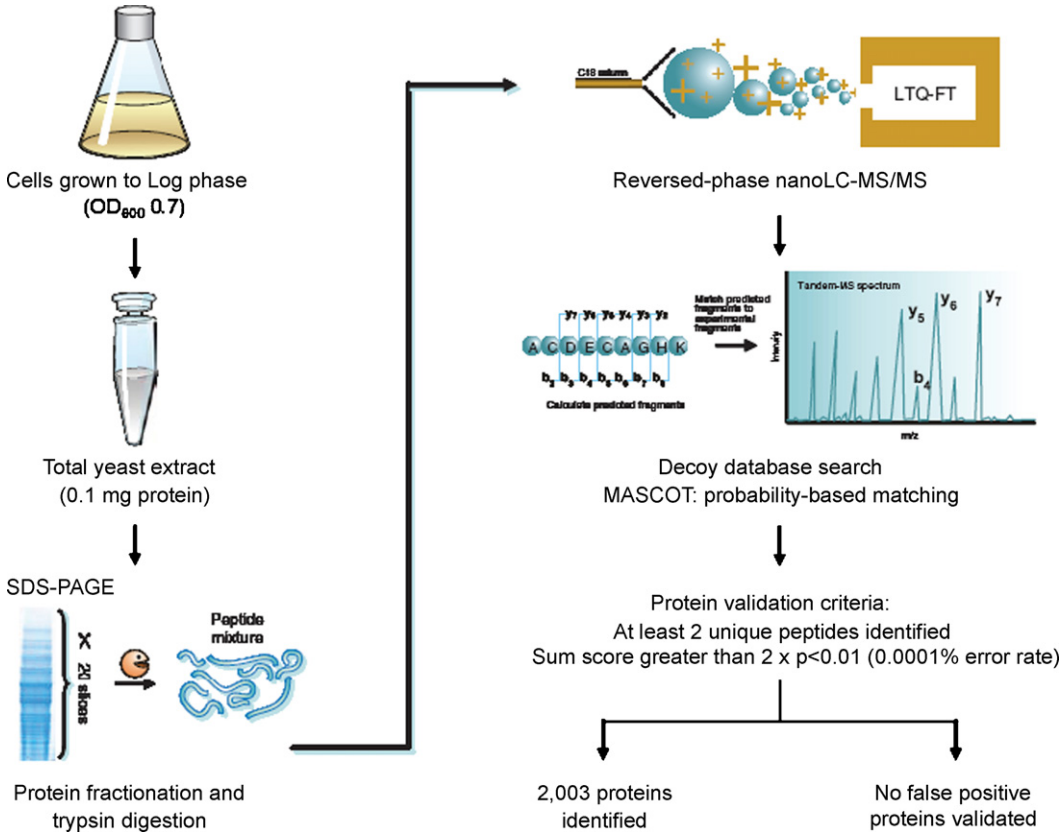

Fig. 1. Workflow for conventional data-driven, bottom-up proteomics experiment, here demonstrated with yeast analysis [29]. Reprinted with permission from Genome Biology.

sonable budget, useful for addressing unmet problems in basic or clinical science? This review will approach such questions from the perspective of the analytical scientist engaged in bioanalysis. The technical objectives of proteomics are not unlike those found in pharmaceutical analysis. An inspection of the FDA's guidance for industry, on the validation of analytical procedures for drug substance monitoring, discusses standards for detection and quantitation that really are universal. Specificity, detection limit, precision, accuracy, repeatability and robustness require consideration for any instrumental method applied to bioanalysis and we should at least consider how modern methods in proteomics measures up.

It could be argued that hoisting standards from pharmaceutical and clinical industries upon proteomics is modestly unfair, but two reasons justify this approach. In the first place, clinical analysis remains a strong driver for proteomics thus the standards of clinical lab testing seem to offer a reasonable perspective. In the second place, the readers of this journal may appreciate a perspective couched in familiar terms. Should the current methodological approaches measure up, analytical scientists within the disciplines targeted by this journal will find themselves in a position where such methods require adoption in their respective labs. Their engagement is therefore essential.

In this review, we will first consider recent exemplary work in whole proteome analysis, specifically those focused upon eukaryotic organisms of limited-to-high proteome complexity. We will provide an overview of technical solutions to the issue of sample complexity and data analysis. To avoid duplicating excellent review articles in the field, we will focus upon novel recent additions to the panel of methods, considering both the analytical and the informatics aspects of the workflow, and address practical matters related to sample amount and analysis time. We suggest that selective, targeted proteomics methods have a greater likelihood of extrapolation to a range of clinical and biological problems, and present a technical justification of this viewpoint, based on recent developments in the area.

2. Whole proteome analysis

2.1. The basic method

The current analytical *modus operandi* in proteomics is built upon a bottom-up approach, in which proteins are harvested from an organism and then digested with specific proteases. This rendering of protein fractions produces smaller peptides that are ideal for mass spectrometric analysis, in a data-driven process where peptides are dynamically selected and then sequenced using tandem MS methods (Fig. 1 and numerous reviews [1–3]). A wide variety of organisms, tissue types and biofluids have been interrogated with such bottom-up proteomics methods. In most cases, such interrogations represent "milestones" on the path to a complete proteome analysis [4]. That is, advancements are applied to a system of interest, in order to gauge the comprehensiveness of such analyses. Solving specific research problems with complete proteomics datasets actually remains an infrequent event, as it is pending a heightened confidence in the research community that depth of coverage is sufficiently exhaustive and reproducible.

2.2. Proteomics of simple model organisms

To this end, certain model organisms have been very useful in gauging the progress of MS-driven whole proteome analysis, but none more so than yeast. Early analyses based on MudPit-style identification strategies applied to yeast established the promise

Please cite this article in press as: M.F. Khan, et al., Proteomics by mass spectrometry—Go big or go home? J. Pharm. Biomed. Anal. (2011), doi:10.1016/j.jpba.2011.02.012

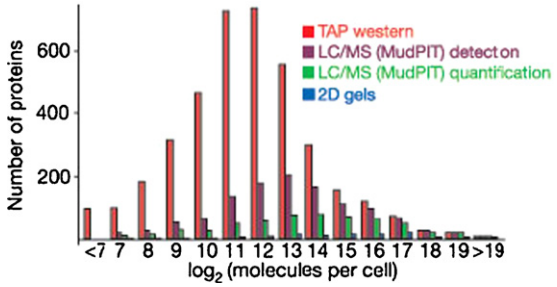

Fig. 2. Distribution of yeast proteins observed by TAP-based western blots, LC–MS/MS using the MudPIT strategies and 2D gel electrophoresis [6].
Reprinted with permission from Nature Publishing Group.

of MS-driven proteomics [5] and ever since, proteome characterization of this organism has served as a bellwether for performance. Yeast is a useful test case, in part because the total number of expressed proteins is known. Fusion libraries for every open reading frame have been generated, incorporating a high-affinity tag used to detect expression through semi-quantitative immunoassays [6]. This census has shown that approximately 80% of the proteome is expressed during log-phase growth (4251 proteins through this strategy). MS-driven approaches, circa 2002, were not strongly representative (Fig. 2) [6] however recently, this has improved to the extent that both the tag-based approach and the MS-driven approaches equivalently represent the yeast proteome, within error (Fig. 3) [7].

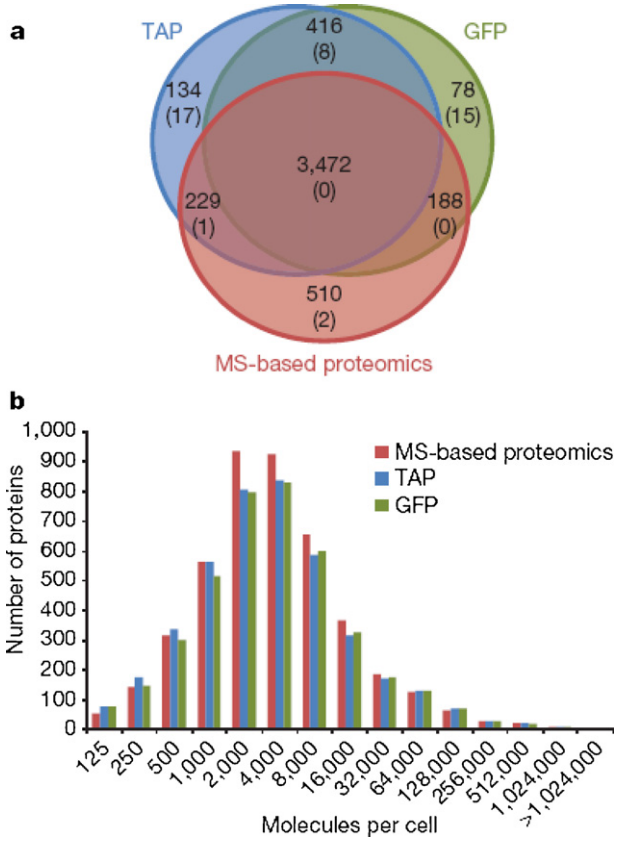

Fig. 3. Yeast proteome coverage, comparing MS-based proteomics with GFP and TAP tagging methods (a) on the basis of identification overlap, where numbers are the identified proteins by each method and in parentheses the number of suspect open reading frames, and (b) identified proteins, as a function of expression levels in the cell, in terms of copy number [7].
Reprinted with permission from Nature Publishing Group.

2.3. Proteomics of complex organisms

This level of in performance has not been achieved in higher organisms, although impressive gains have been realized. For example, an initiative in shotgun proteomics has cataloged large fractions of the *Caenorhabditis elegans* proteome, as many as 10,977 proteins or 54% of the predicted genome [8]. This represents a near-doubling of the number of proteins identified through studies 1–2 years earlier [9], highlighting the pace of improvements in the field.

These efforts provide more value than simply generating "catalogs" of protein identifications. The product of such exercises usually includes quantitative data, permitting a systems-wide perspective of the response to a specific perturbation, and comparisons between organisms. For example, comparing the proteome of *C. elegans* with that of *Drosophila melanogaster* reveals a remarkable correlation between individual protein expression levels, the latter covering approximately 65% of the predicted open reading frames [8]. Such data also informs, in a global way, on mechanisms of transcriptional regulation and provides a certain proofreading capability for genome annotation [9–11]. Datasets arising from these global comparisons have been instrumental in refining the assumption that transcript levels correlate with protein levels—there is a correlation but it consistently is demonstrated to be weak [7,12,13]. Finally, these studies highlight errors in transcriptomic data – often assumed to be quantitative and robust measures of global transcript levels – indicating that data quality issues in any large-scale initiative require careful consideration [14,15].

No other complex, multicellular organisms (with the exception of *Arabidopsis thaliana*) have been probed to this depth in singular experimental excursions, and with the level of rigor presented in the above studies. However, as the obvious appeal in proteomics is to profile species more directly linked to drug development and human health, proteome characterizations of tissues derived from mouse, rat and humans abound [16–21]. Such studies rapidly become mired within the technical challenges associated with protein extraction and sample prefractionation, and to date no proteomic surveys approach the comprehensiveness of the simpler organisms.

A survey of plasma proteomics – perhaps the most complex sample type in the field for its wide dynamic range of protein concentration – provides a useful means of evaluating the challenges within bottom-up proteomics. To date, no comprehensive analysis of plasma has been achieved, in spite of impressive efforts [22,23]. A recent review of blood proteomics in general surveys the numerous attempts [24]. The plasma proteome project cataloged just over 3000 proteins using data collected from 35 member laboratories, with only 889 considered to pass stringent identification criteria [25]. One particular study illustrates the technical challenges ahead [26,27]. Using multiple protein fractionation concepts and LC–MS/MS of the resulting peptides resulted in the identification of 2254 proteins. The plasma proteome is of indeterminate size, because it also represents dynamic processes that shed cellular proteins into circulation, so it is difficult to gauge completion [22]. However, this analysis required significant amounts of serum protein to achieve this level of characterization (146 mg) and greater than 26 days of instrument time. Interestingly, parallel efforts in the analysis of mouse plasma have been able to achieve moderately higher identification rates than from human samples [17].

Under the guidance of the Human Proteome Organization (HUPO), a distributed effort has been mounted to catalog proteomes in key human tissues, which has the advantage of accumulating data drawn from numerous labs, all following standardized protocols. For example, the Human Liver Proteome Project has amassed a database of 13,222 protein entries, containing an

Please cite this article in press as: M.F. Khan, et al., Proteomics by mass spectrometry—Go big or go home? J. Pharm. Biomed. Anal. (2011), doi:10.1016/j.jpba.2011.02.012

index of the underlying peptide identifications supporting this list [28]. Organellar and protein-level fractionation has proven to be one of the key methods by which to increase the depth of coverage for the liver proteome. This finding is holding true in most cases of tissue proteome profiling.

2.4. On the discrepancy between simple and complex organisms

Intriguingly, a variety of rigorous individual studies on human-derived samples offer identification lists rarely exceeding 2000 proteins, significantly less than the totals seen from the analysis of lower organisms. This highlights that increased complexity reduces the probability of identification, by a factor only partly dependent on the proteome size [29]. Reduced probability arises in part from the greater dynamic range of protein expression within human samples, and its effect on LC/MS-based protein identification ([30] and see below). Existing strategies are simply better adapted to the dynamic range of expression for simple systems ($\sim 10^4$ for yeast [7]) than they are for higher organism (e.g. $>10^{11}$ for human plasma [31]). Probabilities are further reduced by the greater heterogeneity within individual proteins, at the level of isoforms and post-translational modifications, to the extent that "signal splitting" conspires to reduce identification rates in unusual ways. For example, variable glycosylation can affect the fractionation of proteins and peptides by either size, charge or hydrophobicity, diffusing protein levels across multiple fractions in separation systems based on these principles.

It is useful to dissect the capacity issues associated with modern bottom-up methods in some additional detail, prior to considering new developments in the area of complex protein mixture analysis. If we define "identification probability" as the likelihood of generating a unique protein identification within a sample of a given complexity, then we see that there is a relationship between this probability, protein dynamic range and the number of unique proteins present in the sample. This can be represented by Fig. 4. Increasing the amount of sample loaded into any given proteome analysis "engine" reaches a point where the identification probability saturates [32–34], beyond which further sample increases are generally wasteful (Fig. 4a).

However, digests of protein mixtures generate pools of peptides that tax the sequencing speed of modern instruments, such that any single analysis usually misses "sequenceable" peptides. It has been demonstrated that replicate sample analysis can increase the number of proteins identified in a sample of a fixed mass, reflecting a quasi-stochastic quality to the peptide selection process during mass analysis [35]. However, the full benefit of this replication is really only seen when the dynamic range of expression is lower (Fig. 4b), because the selection process is typically driven by peptide intensity.

Issues of high sample dynamic range can be addressed in part by fractionating the peptide pool using two or more dimensions of separation. This may also be implemented at the protein level, and is perhaps the most successful way to do so [36]. However, fractionation by chromatography or electrophoresis does not discriminate based on peptide abundance levels but rather simply creates a series of less complex mixtures. Here, we use the term "complexity" to simply indicate the number of unique peptides present in the sample. Although increasing the dimensions of sample fractionation generally increases the number of proteins identified in a sample of a given size (Fig. 4c), the greatest impact on the depth of coverage in a sample should be seen in samples that are less complex to begin with. This may seem odd at first, but the primary function of any fractionation strategy is to reduce the complexity of the pools of peptides, so that the mass spectrometer has sufficient time to fully interrogate their contents and so that ion suppression is eased [29]. For any given multidimensional fractionation scheme, this will have its greatest impact in samples with fewer numbers of proteins where the total pool of peptides is smaller [37]. The reader is directed to thoughtful discussions of issues related to sample complexity, fractionation, dynamic range and sequencing rates by de Godoy et al. [7,29].

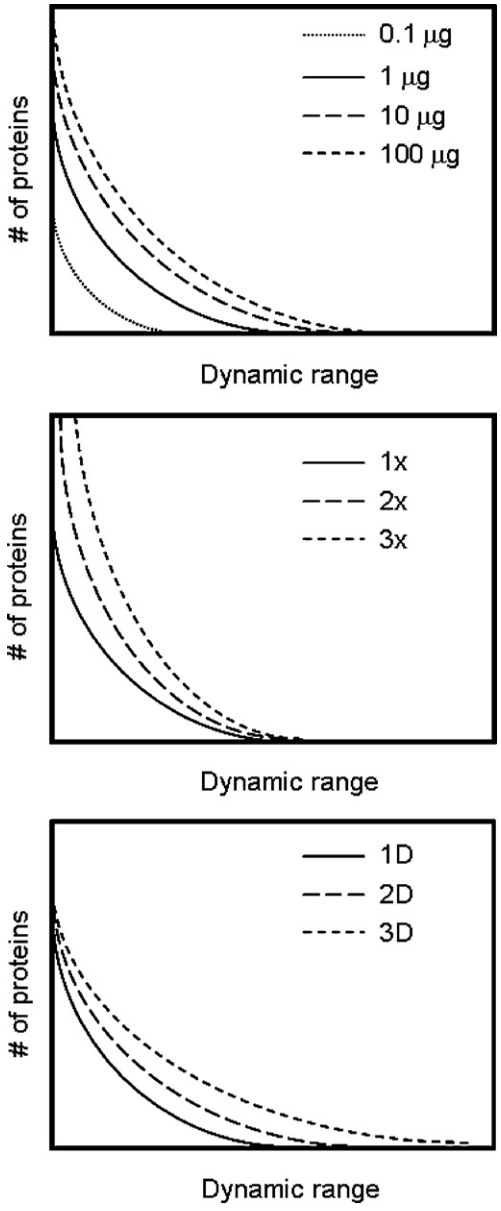

Fig. 4. General trends in how key analytical variables influence the probability of detecting a protein in a sample of specified protein complexity (# of proteins) and sample dynamic range. (a) Increasing the sample load in a single analysis for any given proteomics method (e.g. 1D LC–MS/MS). (b) For a given sample load, increasing the number of replicate analyses. (c) For a given sample load, increasing dimensionality in the fractionation of protein or peptide pools. In all cases, lines represent a given probability that any one protein would be correctly identified, at the specified sample composition. Numbers are to provide context to the trends and though reasonable, are not intended to be rigorous applied.

2.5. An assessment

The impact of these considerations on sample requirements and analysis time is best evaluated by assessing the progress towards full characterization of the yeast proteome. Approximately 2000 proteins were identified from 100 µg of protein extract, representing almost 2 days of LC–MS/MS instrument time [29]. The sample

Please cite this article in press as: M.F. Khan, et al., Proteomics by mass spectrometry—Go big or go home? J. Pharm. Biomed. Anal. (2011), doi:10.1016/j.jpba.2011.02.012

was fractionated at the protein level using SDS–PAGE and no replicates were conducted. Full coverage of the proteome (~4400 proteins) required 2.2 mg of protein and 38 days of instrument time, requiring an extensive fractionation scheme [7]. This highlights the declining benefits of such extra effort, as we have attempted to capture in Fig. 4.

In summary, while very impressive depths of coverage for the proteomes of simpler organisms have been achieved, and clear progress is being made towards similar coverage of human proteomes through community-based efforts, a set of analytical realities become clear. Current efforts directed towards deep proteome coverage remain large-scale endeavors requiring significant investment in equipment, informatics and sample handling. Although there are similarities to shotgun genome sequencing in that short "reads" of sequence are generated from which identity is established, bottom-up proteomics methods differ because individual peptide reads rarely overlap, nor are all peptides detected. In other words, identity is established with a minimum of sequence data in most cases, relying heavily on the detection of truly unique peptides for any given identification. Such determinations require on the order of 100 μg of protein per replicate for even fractional coverage of smaller proteomes.

3. Technological developments in whole proteome analysis

The core component in any bottom-up proteomics engine remains an LC–MS/MS instrument, where reversed-phase chromatography enriches and separates peptides for mass analysis and sequencing [3,38,39]. The most significant recent development in this component involves the proliferation of high resolution mass spectrometers such as the Orbitrap, allowing for routine measurement of peptides at consistently high mass measurement accuracy (<1 ppm) and higher quality ion selection in the data-dependent experiment. This alone has improved the quality of protein identifications as well as the quantity [40,41].

3.1. Improving LC–MS/MS performance

This core component has been rendered more effective by the gradual standardization of sample introduction methods, and the implementation of new approaches. Most whole proteome analyses incorporate at least one dimension of protein separation now, and often more. GeLC–MS is the most widely used approach (see Fig. 1). A fusion of SDS–PAGE with LC–MS/MS, this involves the separation of protein in one dimension of a gel, followed by segmenting the entire lane into a series of fractions for in-gel tryptic digestion, and then reversed-phase LC–MS/MS [42]. This has been particularly effective for smaller proteomes. Samples of higher complexity are now often fractionated with 2D-LC of proteins, involving either ion exchange or isoelectric focusing in the first dimension, followed by reversed phase (e.g. IPAS) [26,43]. If the samples contain a number of proteins at very high abundance, as in plasma, this protein fractionation step is usually fronted by immunodepletion for their removal. For example, blended antibody columns are available for extracting albumin, IgGs, transferrin, fibrinogen and 16 other abundant proteins from plasma [44]. This enriches the remaining protein and is a very effective way of reducing the sample dynamic range. Such reagents are becoming available for a wider variety of organisms, including plants for the removal of Rubisco [45].

3.2. Recent developments in peptide and protein fractionation

One of the more successful additions to sample fractionation methods involves *peptide-level* isoelectric focusing (Fig. 5) [46]. With the addition of devices facilitating the recovery of peptides from IPG strips, it has been shown that this means of separation

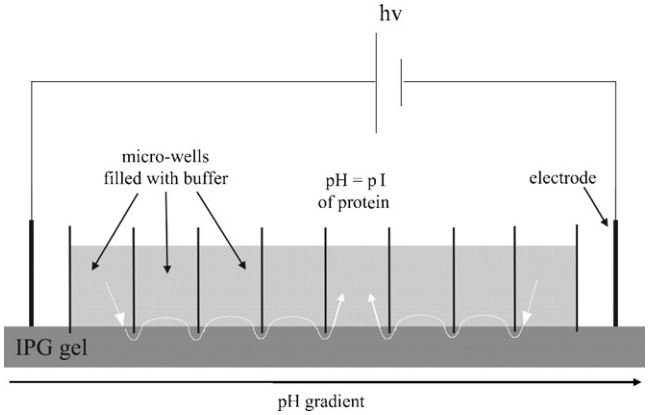

Fig. 5. Schematic of the setup used for OFFGEL electrophoresis [46]. Reprinted with permission from Molecular & Cellular Proteomics.

provides complementary coverage to the conventional GeLC–MS method and higher identification rates overall [37]. It has become a key approach in the full proteome analyses of simpler organisms and will likely find considerable use going forward. However, it is important to stress that complementarity with GeLC–MS means that both should be used, if one expects to maximize proteome coverage.

Whatever the precise protocol, replicate sample analysis in GeLC–MS has become a fixture in whole proteome experiments, not to test for reproducibility as would usually be the case, but as a means to increase the number of peptides identified as discussed above. This is a simple approach, but one that delivers declining benefits with each additional replicate; three or four runs are usually sufficient to maximize the benefit of this approach [13,47,48]. This is demonstrated in the recent study by Wang et al. [36]. Although it showed that inclusion of a *protein-level* IEF separation can compensate for replicate analysis, with an overall equivalency in sample consumption and instrument time (Fig. 6), replication would presumably be beneficial here as well.

Other notable fractionation concepts that have improved depth of proteome coverage include peptide ion mobility in the gas phase prior to MS detection [49], and "replay" analyses that analyzes a sample twice by collecting undersampled fractions for further

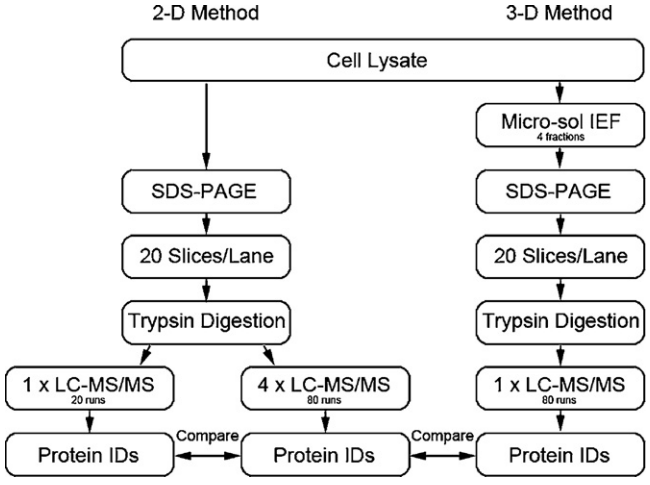

Fig. 6. Experimental outline for a 2D and a 3D proteome processing method, using the same amount of protein. Authors show that incorporating a simple protein fractionation step (micro-sol IEF) can largely compensate for replicate gel band analysis [36].
Reprinted with permission from the American Chemical Society.

Please cite this article in press as: M.F. Khan, et al., Proteomics by mass spectrometry—Go big or go home? J. Pharm. Biomed. Anal. (2011), doi:10.1016/j.jpba.2011.02.012

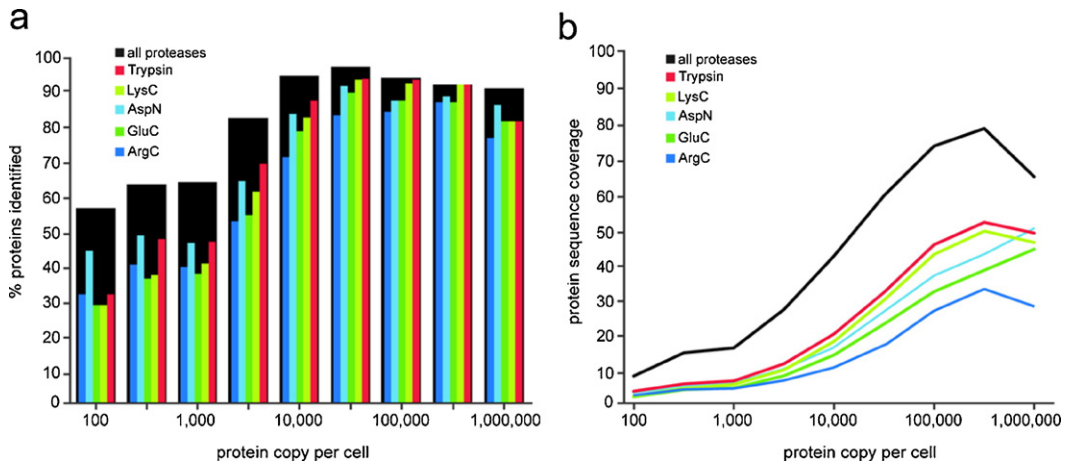

Fig. 7. Effect of individual and multiple enzymes on (a) the probability of protein identification and on (b) protein sequence coverage as a function of protein abundance [55]. Reprinted with permission from the American Chemical Society.

inspection [50]. It should be clear at this point that depth of proteome coverage using processes that feed a data-dependent LC–MS/MS core are close to "hitting a wall". While there is always the possibility for an unseen disruptive technology, increased sample processing is not likely to deliver whole proteome analyses on large numbers of samples, except in highly specialized laboratories.

3.3. Improvements in peptide ion fragmentation

Looking back, the greatest impact in depth of coverage has been achieved through advancements in MS methods and technology, with the associated development of computational tools [1]. One activity in this area that is worth highlighting is the development of new ion fragmentation technology, involving electron capture or electron transfer dissociation [51–53]. When compared to conventional collisionally-induced dissociation (CID), this fragmentation technology provides longer sequence "reads" for larger peptides but no real benefit for smaller peptides, where CID excels. The two fragmentation methods therefore complement each other. Swaney et al. have developed a more sophisticated approach to a data-dependent experiment on this basis [54]. Peptide ions detected at a given point in chromatography time are assessed for the greatest likelihood of a successful sequence event, and then metered out to CID or ETD accordingly. This comes with no extra sample fractionation, and generates almost 40% more peptide identifications compared to CID alone.

A very recent work by the same lab incorporated five different proteases to obtain near-complete coverage of the yeast proteome, in concert with this "decision-tree" approach [55]. This is impressive, in part because it only required a conventional 2D peptide separation front-end. Each digest required only 12 LC–MS/MS runs, for a total of approximately 5 days instrument time. Perhaps most importantly, there was a >2-fold increase in average sequence coverage using this method (Fig. 7), and overlapping peptides should return a true "shotgun" style quality to the bottom up method, as in genome sequencing. This would aid in sequence-building exercises, something that a purely tryptic digestion does not deliver.

3.4. Departing from the data-driven experiment

Various other strategies have emerged, designed to utilize the MS data more effectively and address the quasi-stochastic nature of conventional data-dependent experiments. Accurate mass tags (AMTs) rely upon prior identification of peptides, after which they can be identified in subsequent analyses based solely upon accurate mass measurements of the peptide, as well as their retention time in a standardized chromatographic method [56,57]. This can work very well, as the number of peptides "recognized" by the instrument is no longer determined by a selection and sequencing event. The most rigorous approach requires the mining of MS/MS data in the form of a reference database, after which the method becomes very effective in subsequent analyses [58]. Accurate retention time prediction software may have a role in building these reference databases as well [59,60]. In general, the idea of mining pre-existing datasets, in the form of spectral searching or otherwise, represents an important trend in proteomics [61].

3.5. Developments in bioinformatics

Databases built on prior acquisitions of MS/MS data will only be valuable if the resulting peptide identifications are accurate, but evaluating accuracy is not a trivial undertaking. This informatics problem has been embraced by many labs engaged in whole proteome analysis, and centers upon statistically-based determinations of global false discovery rates (FDR's) or local false discovery rates (fdr's), also referred to as posterior error probabilities [62]. The former is a metric that returns an assessment of the error rate in the entire set of spectra searched, but does not inherently convey a quality assessment of individual spectra. The latter seeks to do this. Both methods have their place, but in large-scale undertakings such as described in this review, the FDR approach perhaps has the greatest utility [63]. It requires the use of decoy databases in order assess the null distribution for the dataset tested. This allows for a straightforward calculation of the global error rate.

To obtain greater confidence in any particular spectral match, a series of methods have been developed in order to tease out the null distribution from the datasets submitted to a search. Most base the probability of a match upon the scores that database search engines assign to all candidate peptides for any given MS/MS spectrum [64]. Originally built from training sets of data characterized by a high confidence in both correct and incorrect assignments, discriminating functions can be generated that essentially translate the scores (and other useful information) into probabilities for the correctness of each individual peptide assignment. Currently there is less reliance on training sets, as techniques are used to generate these discriminating functions from the distribution of scores returned from a search of any given large dataset. The reader is referred to seminal articles and reviews in the area for additional information [65–69].

Here, the most significant point we wish to emphasize relates to the threshold for successful *protein* identification. These are obvi-

Please cite this article in press as: M.F. Khan, et al., Proteomics by mass spectrometry—Go big or go home? J. Pharm. Biomed. Anal. (2011), doi:10.1016/j.jpba.2011.02.012

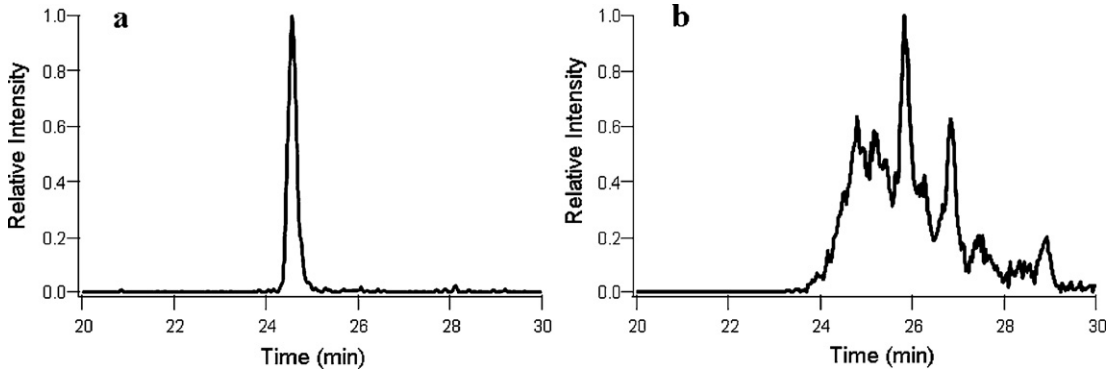

Fig. 8. Positive effect of immunoaffinity enrichment of a proteotypic peptide for galectin-1 (DGGAWGAEQR), monitoring the 523.7/230.1 transition in a targeted experiment. Extracted ion chromatogram of the proteotypic peptide (a) after and (b) before immunoaffinity isolation from 2.5 mg/ml serum protein digest [92].

ously based upon peptide data, but in large-scale analyses the association between peptides has been erased, because of the global digestion step. Referred to as the "protein inference problem", the methods used for significance determination are direct extensions of the methods used for peptides [70]. The "two-peptide" requirement for protein identification often seen in the literature has been shown to be overly conservative; a metric based on calculated error rates as described above in generally regarded as a better approach [69].

Returning to the notion of databases assembled from annotated peptide MS/MS spectra, we see that error evaluation is required. Resources such as PeptideAtlas [71] have applied error quantification to data generated from a variety of sources, such that error rates can be determined both for individual identifications and across the whole atlas [72,73]. This sort of community-driven approach to curation leads to new opportunities for spectral searching and has been pursued by other groups as well [61,74]. The massive amounts of peptide MS/MS data being generated has also supported growing sophistication in the theoretical prediction of peptide ion fragmentation spectra, including fragment abundance levels and fragmentation "rules" [75,76]. While the rules are not of the sort that are rigorously obeyed, the accumulated knowledge in gas-phase ion chemistry has allowed for surprising utility in the area of "simulated" spectral libraries, both for CID and for ETD fragmentation [77].

4. Targeted proteomics

Whole-proteome analyses have considerable appeal in systems biology, but the previous sections highlight that the digestion–fractionation–LC–MS/MS paradigm might be approaching a practical limit. However, alternative paradigms related to targeted spectral acquisitions are emerging, which may provide greater dynamic range, simplicity in sample processing, and higher confidence in identifications. When large-scale studies are surveyed, only ~5% of all MS/MS scans collected lead to the identification of unique peptides [55]. This may seem surprisingly low, but it reflects the many areas where redundancy and error can arise in such experiments. The notion of targeted proteomics has emerged, were pre-existing data are mined to identify only unique peptides, with the attributes required for serving as "biomarkers" for their respective proteins of origin [78–80]. Targeting the mass spectrometer may therefore deliver a more effective use of available scan time and increase the depth of proteome coverage.

4.1. SRM methods

The targeting approach implements triple quadrupole (QQQ) mass spectrometers for monitoring unique peptides in a selective and sensitive fashion. Driven by the long history of applications pharmaceutical analysis, these instruments permit the isolation of a peptide ion and a highly-discriminating peptide fragment ion, and apply this dual isolation capability as a very selective sample "filter" referred to as a transition [81,82]. This process is referred to as selective reaction monitoring (SRM, the favored term) or multiple reaction monitoring (MRM). Transition monitoring can offer extremely high sensitivity and dynamic range to the process of peptide monitoring, relative to the data-driven approach described earlier. The stochastic nature of ion selection is removed entirely, in favor of monitoring known peptides with known elution times. Modern QQQ instruments can monitor well over 1000 peptides/hour and the methods are very portable [80,83]. As a result, inter-lab reproducibility can greatly exceed that of the conventional method. The main challenge in establishing appropriate filters involves the selection of the peptides and transitions, and then establishing assays for these selections. Here, the repositories of data from prior large-scale sequencing efforts are highly useful, combined with bioinformatic approaches [84]. When considering yeast, on average there is a high expectation for multiple proteotypic peptides per protein [85]. This should offer sufficient flexibility in assay development.

As protein standards are not available in most cases, one approach to SRM assay development involves cost-effective synthesis of the unique peptides. Large-scale initiatives are attempting to define proteome-wide collections of such peptides with corresponding validations of transitions. Monitoring all transitions for peptides would not be very effective, so the best approach involves a scheduling of transition sets, according to peptide elution time [86]. This provides the capacity required (>1000 peptides), which would otherwise be restricted to 50 or less, based on considerations of instrument duty cycle.

However, SRMs do not necessarily offer the perfect filter. In highly complex mixtures redundancies will occur, and caution has been advised that assay design consider selectivity not just sensitivity [87,88]. A partial response has involved the selection of three or more transitions for any given peptide, allowing for inclusion of an intensity pattern to help generate specificity [89]. While this is promising, it involves a larger number of transitions and therefore a reduction in multiplexing capacity. An interesting alternative seeks to implement an extra dimension to the transition, in the hope of increasing selectivity [90]. A further alternative involves raising antibodies to proteotypic peptides, and using these in immunoaffinity-style cartridges for both increased selectivity and sensitivity in multiplexed SRM assays [91]. While this does increase the cost associated with implementation, the format offers excellent opportunities for targeting low-abundance proteins within a simple workflow (Fig. 8) [92]. Nevertheless, the potential for

Please cite this article in press as: M.F. Khan, et al., Proteomics by mass spectrometry—Go big or go home? J. Pharm. Biomed. Anal. (2011), doi:10.1016/j.jpba.2011.02.012

redundancies will ensure that high resolution reversed-phase chromatography remains a key part of any monitoring solution.

4.2. SRM applications

The most enlightening studies to date involve targeted analyses of the yeast proteome. These offer an opportunity for comparative analysis with "conventional" data-driven high throughput undertakings, of the sort described earlier. One study tracks the dynamical effect of glucose repression on a targeted segment of the proteome. This involved monitoring 45 different proteins selected from the networks engaged in carbon metabolism, and tracking their flux as a function of glucose consumption, on the way to a metabolic shift towards fermentative growth in the resulting ethanol-rich media [79]. The analytical procedures need only concern us here. Generally, the SRM multiplexed assay required two peptides and three transitions per peptide, for every protein in the list. This assay was built with the aid of synthetic peptides, as discussed, and spanned proteins expressed at over three orders of magnitude in concentration. Once developed, the assay required less than 1 h per sample without any proteome fractionation. Because some of these proteins were of low abundance, using a data-driven approach would require extensive fractionation, where each sample "data point" would need over 1 month of analysis time.

This remarkable improvement comes at the cost of targeting—only those proteins hypothesized to be involved would be monitored. However, this sort of capability can stimulate many interesting experiments. With a moderate degree of fractionation, for example peptide separation using off-gel electrophoresis, one can achieve significant improvements in dynamic range and still be in a position to monitor multiple samples in a realistic lab setting [79]. This sort of capability has been applied to the selective monitoring of all kinases and phosphatases in yeast [80], and a related version to phosphotyrosine profiling in human mammary epithelial cells [93].

Applications to plasma analysis are emerging, but this biofluid remains the most challenging of samples. Depletion of high abundance proteins may still be required, in order to access the lower abundance components of the proteome with high confidence. At this stage of development, expectations in terms of precision and inter-lab reproducibility are being considered, and assays are being developed for higher abundance proteins [94].

5. Conclusions and perspective

Proteomics remains a field driven by developments in peptide mass spectrometry. Bottom-up methods still offer the most powerful means of achieving deep-proteome coverage. We have attempted to show that approaches built upon a continuous "discovery" cycle, where peptides are selected for mass analysis based upon data-driven approaches, are very labor intensive. It is unlikely that they will be applied to larger-scale initiatives, where proteomes require replicative analysis over multiple timepoints. However, these initiatives will remain essential. At a minimum, they will populate databases of proteotypic peptides upon which targeted, multiplexed SRM assays can be built. This type of assay will require considerable investment in peptide selection and transition selection, an activity that will always require a research component, although certain assays may approach an "off the shelf" quality. For example, targeted assays of kinases may find early adoption in the pharmaceutical and drug discovery. Requests for assay development and implementation will work their way into the labs of analysts more frequently engaged in small molecule SRM development. This is the natural home for such activities, and therefore we encourage this community to engage in discussions with proteomics researchers around the optimization of such assays.

References

[1] J.R. Yates, C.I. Ruse, A. Nakorchevsky, Proteomics by mass spectrometry: approaches, advances, and applications, Annu. Rev. Biomed. Eng. 11 (2009) 49–79.
[2] B. Domon, R. Aebersold, Mass spectrometry and protein analysis, Science 312 (2006) 212–217.
[3] R. Aebersold, M. Mann, Mass spectrometry-based proteomics, Nature 422 (2003) 198–207.
[4] A. Motoyama, J.D. Venable, C.I. Ruse, J.R. Yates 3rd, Automated ultra-high-pressure multidimensional protein identification technology (UHP-MudPIT) for improved peptide identification of proteomic samples, Anal. Chem. 78 (2006) 5109–5118.
[5] M.P. Washburn, D. Wolters, J.R. Yates 3rd, Large-scale analysis of the yeast proteome by multidimensional protein identification technology, Nat. Biotechnol. 19 (2001) 242–247.
[6] S. Ghaemmaghami, W.K. Huh, K. Bower, R.W. Howson, A. Belle, N. Dephoure, E.K. O'Shea, J.S. Weissman, Global analysis of protein expression in yeast, Nature 425 (2003) 737–741.
[7] L.M. de Godoy, J.V. Olsen, J. Cox, M.L. Nielsen, N.C. Hubner, F. Frohlich, T.C. Walther, M. Mann, Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast, Nature 455 (2008) 1251–1254.
[8] S.P. Schrimpf, M. Weiss, L. Reiter, C.H. Ahrens, M. Jovanovic, J. Malmstrom, E. Brunner, S. Mohanty, M.J. Lercher, P.E. Hunziker, R. Aebersold, C. von Mering, M.O. Hengartner, Comparative functional analysis of the *Caenorhabditis elegans* and *Drosophila melanogaster* proteomes, PLoS Biol. 7 (2009) e48.
[9] G.E. Merrihew, C. Davis, B. Ewing, G. Williams, L. Kall, B.E. Frewen, W.S. Noble, P. Green, J.H. Thomas, M.J. MacCoss, Use of shotgun proteomics for the identification, confirmation, and correction of *C. elegans* gene annotations, Genome Res. 18 (2008) 1660–1669, Epub 2008 July 1624.
[10] N.E. Castellana, S.H. Payne, Z. Shen, M. Stanke, V. Bafna, S.P. Briggs, Discovery and revision of Arabidopsis genes by proteogenomics, Proc. Natl. Acad. Sci. U.S.A. 105 (2008) 21034–21038, Epub 2008 December 21019.
[11] K. Baerenfaller, J. Grossmann, M.A. Grobei, R. Hull, M. Hirsch-Hoffmann, S. Yalovsky, P. Zimmermann, U. Grossniklaus, W. Gruissem, S. Baginsky, Genome-scale proteomics reveals *Arabidopsis thaliana* gene models and proteome dynamics, Science 320 (2008) 938–941, Epub 2008 April 2024.
[12] S.P. Gygi, Y. Rochon, B.R. Franza, R. Aebersold, Correlation between protein and mRNA abundance in yeast, Mol. Cell. Biol. 19 (1999) 1720–1730.
[13] Y. Ishihama, T. Schmidt, J. Rappsilber, M. Mann, F.U. Hartl, M.J. Kerner, D. Frishman, Protein abundance profiling of the *Escherichia coli* cytosol, BMC Genomics 9 (2008) 102.
[14] V. Sharov, K.Y. Kwong, B. Frank, E. Chen, J. Hasseman, R. Gaspard, Y. Yu, I. Yang, J. Quackenbush, The limits of log-ratios, BMC Biotechnol. 4 (2004) 3.
[15] R.A. Irizarry, D. Warren, F. Spencer, I.F. Kim, S. Biswal, B.C. Frank, E. Gabrielson, J.G. Garcia, J. Geoghegan, G. Germino, C. Griffin, S.C. Hilmer, E. Hoffman, A.E. Jedlicka, E. Kawasaki, F. Martinez-Murillo, L. Morsberger, H. Lee, D. Petersen, J. Quackenbush, A. Scott, M. Wilson, Y. Yang, S.Q. Ye, W. Yu, Multiple-laboratory comparison of microarray platforms, Nat. Methods 2 (2005) 345–350.
[16] J.C. Price, S. Guan, A. Burlingame, S.B. Prusiner, S. Ghaemmaghami, Analysis of proteome dynamics in the mouse brain, Proc. Natl. Acad. Sci. U.S.A. 107 (2010) 14508–14513.
[17] Q. Zhang, R. Menon, E.W. Deutsch, S.J. Pitteri, V.M. Faca, H. Wang, L.F. Newcomb, R.A. Depinho, N. Bardeesy, D. Dinulescu, K.E. Hung, R. Kucherlapati, T. Jacks, K. Politi, R. Aebersold, G.S. Omenn, D.J. States, S.M. Hanash, A mouse plasma peptide atlas as a resource for disease proteomics, Genome Biol. 9 (2008) R93, Epub 2008 June 2003.
[18] V. Kulasingam, E.P. Diamandis, Tissue culture-based breast cancer biomarker discovery platform, Int. J. Cancer 123 (2008) 2007–2012.
[19] V.M. Faca, K.S. Song, H. Wang, Q. Zhang, A.L. Krasnoselsky, L.F. Newcomb, R.R. Plentz, S. Gurumurthy, M.S. Redston, S.J. Pitteri, S.R. Pereira-Faca, R.C. Ireton, H. Katayama, V. Glukhova, D. Phanstiel, D.E. Brenner, M.A. Anderson, D. Misek, N. Scholler, N.D. Urban, M.J. Barnett, C. Edelstein, G.E. Goodman, M.D. Thornquist, M.W. McIntosh, R.A. DePinho, N. Bardeesy, S.M. Hanash, A mouse to human search for plasma proteome changes associated with pancreatic tumor development, PLoS Med. 5 (2008) e123.
[20] R. Shi, C. Kumar, A. Zougman, Y. Zhang, A. Podtelejnikov, J. Cox, J.R. Wisniewski, M. Mann, Analysis of the mouse liver proteome using advanced mass spectrometry, J. Proteome Res. 6 (2007) 2963–2972, Epub 2007 July 2963.
[21] T. Kislinger, B. Cox, A. Kannan, C. Chung, P. Hu, A. Ignatchenko, M.S. Scott, A.O. Gramolini, Q. Morris, M.T. Hallett, J. Rossant, T.R. Hughes, B. Frey, A. Emili, Global survey of organ and organelle protein expression in mouse: combined proteomic and transcriptomic profiling, Cell 125 (2006) 173–186.
[22] N.L. Anderson, N.G. Anderson, The human plasma proteome: history, character, and diagnostic prospects, Mol. Cell. Proteomics 1 (2002) 845–867.
[23] N.L. Anderson, M. Polanski, R. Pieper, T. Gatlin, R.S. Tirumalai, T.P. Conrads, T.D. Veenstra, J.N. Adkins, J.G. Pounds, R. Fagan, A. Lobley, The human plasma proteome: a nonredundant list developed by combination of four separate sources, Mol. Cell. Proteomics 3 (2004) 311–326, Epub 2004 January 2012.
[24] G. Liumbruno, A. D'Alessandro, G. Grazzini, L. Zolla, Blood-related proteomics, J. Proteomics 73 (2010) 483–507.

Please cite this article in press as: M.F. Khan, et al., Proteomics by mass spectrometry—Go big or go home? J. Pharm. Biomed. Anal. (2011), doi:10.1016/j.jpba.2011.02.012

[25] G.S. Omenn, D.J. States, M. Adamski, T.W. Blackwell, R. Menon, H. Hermjakob, R. Apweiler, B.B. Haab, R.J. Simpson, J.S. Eddes, E.A. Kapp, R.L. Moritz, D.W. Chan, A.J. Rai, A. Admon, R. Aebersold, J. Eng, W.S. Hancock, S.A. Hefta, H. Meyer, Y.K. Paik, J.S. Yoo, P. Ping, J. Pounds, J. Adkins, X. Qian, R. Wang, V. Wasinger, C.Y. Wu, X. Zhao, R. Zeng, A. Archakov, A. Tsugita, I. Beer, A. Pandey, M. Pisano, P. Andrews, H. Tammen, D.W. Speicher, S.M. Hanash, Overview of the HUPO Plasma Proteome Project: results from the pilot phase with 35 collaborating laboratories and multiple analytical groups, generating a core dataset of 3020 proteins and a publicly-available database, Proteomics 5 (2005) 3226–3245.
[26] V. Faca, S.J. Pitteri, L. Newcomb, V. Glukhova, D. Phanstiel, A. Krasnoselsky, Q. Zhang, J. Struthers, H. Wang, J. Eng, M. Fitzgibbon, M. McIntosh, S. Hanash, Contribution of protein fractionation to depth of analysis of the serum and plasma proteomes, J. Proteome Res. 6 (2007) 3558–3565, Epub 2007 August 3516.
[27] S.M. Hanash, S.J. Pitteri, V.M. Faca, Mining the plasma proteome for cancer biomarkers, Nature 452 (2008) 571–579.
[28] X. Gao, J. Zheng, L. Beretta, J. Mato, F. He, The 2009 Human Liver Proteome Project (HLPP) Workshop 26 September, Toronto, Canada, Proteomics 10 (2009) 3058–3061.
[29] L.M. de Godoy, J.V. Olsen, G.A. de Souza, G. Li, P. Mortensen, M. Mann, Status of complete proteome analysis by mass spectrometry: SILAC labeled yeast as a model system, Genome Biol. 7 (2006) R50.
[30] J. Eriksson, D. Fenyo, Improving the success rate of proteome analysis by modeling protein-abundance distributions and experimental designs, Nat. Biotechnol. 25 (2007) 651–655.
[31] G.L. Hortin, D. Sviridov, The dynamic range problem in the analysis of the plasma proteome, J Proteomics 73 (2010) 629–636.
[32] M. Vollmer, P. Horth, E. Nagele, Optimization of two-dimensional off-line LC/MS separations to improve resolution of complex proteomic samples, Anal. Chem. 76 (2004) 5180–5185.
[33] S. Eeltink, S. Dolman, G. Vivo-Truyols, P. Schoenmakers, R. Swart, M. Ursem, G. Desmet, Selection of column dimensions and gradient conditions to maximize the peak-production rate in comprehensive off-line two-dimensional liquid chromatography using monolithic columns, Anal. Chem. 82 (2010) 7015–7020.
[34] J. Rappsilber, Y. Ishihama, M. Mann, Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics, Anal. Chem. 75 (2003) 663–670.
[35] K. Gevaert, P. Van Damme, B. Ghesquiere, F. Impens, L. Martens, K. Helsens, J. Vandekerckhove, A la carte proteomics with an emphasis on gel-free techniques, Proteomics 7 (2007) 2698–2718.
[36] H. Wang, T. Chang-Wong, H.Y. Tang, D.W. Speicher, Comparison of extensive protein fractionation and repetitive LC–MS/MS analyses on depth of analysis for complex proteomes, J. Proteome Res. 9 (2010) 1032–1040.
[37] N.C. Hubner, S. Ren, M. Mann, Peptide separation with immobilized pI strips is an attractive alternative to in-gel protein digestion for proteome analysis, Proteomics 8 (2008) 4862–4872.
[38] W.H. McDonald, J.R. Yates 3rd, Shotgun proteomics: integrating technologies to answer biological questions, Curr. Opin. Mol. Ther. 5 (2003) 302–309.
[39] C.C. Wu, M.J. MacCoss, Shotgun proteomics: tools for the analysis of complex biological systems, Curr. Opin. Mol. Ther. 4 (2002) 242–250.
[40] J.V. Olsen, L.M. de Godoy, G. Li, B. Macek, P. Mortensen, R. Pesch, A. Makarov, O. Lange, S. Horning, M. Mann, Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap, Mol. Cell. Proteomics 4 (2005) 2010–2021.
[41] J.V. Olsen, J.C. Schwartz, J. Griep-Raming, M.L. Nielsen, E. Damoc, E. Denisov, O. Lange, P. Remes, D. Taylor, M. Splendore, E.R. Wouters, M. Senko, A. Makarov, M. Mann, S. Horning, A dual pressure linear ion trap Orbitrap instrument with very high sequencing speed, Mol. Cell. Proteomics 8 (2009) 2759–2769.
[42] A. Shevchenko, H. Tomas, J. Havlis, J.V. Olsen, M. Mann, In-gel digestion for mass spectrometric characterization of proteins and proteomes, Nat. Protoc. 1 (2006) 2856–2860.
[43] H. Wang, S.G. Clouthier, V. Galchev, D.E. Misek, U. Duffner, C.K. Min, R. Zhao, J. Tra, G.S. Omenn, J.L. Ferrara, S.M. Hanash, Intact-protein-based high-resolution three-dimensional quantitative analysis system for proteome profiling of biological fluids, Mol. Cell. Proteomics 4 (2005) 618–625, Epub 2005 February 2009.
[44] M.K. Dayarathna, W.S. Hancock, M. Hincapie, A two step fractionation approach for plasma proteomics using immunodepletion of abundant proteins and multi-lectin affinity chromatography: application to the analysis of obesity, diabetes, and hypertension diseases, J. Sep. Sci. 31 (2008) 1156–1166.
[45] N.A. Cellar, K. Kuppannan, M.L. Langhorst, W. Ni, P. Xu, S.A. Young, Cross species applicability of abundant protein depletion columns for ribulose-1,5-bisphosphate carboxylase/oxygenase, J. Chromatogr. B: Analyt. Technol. Biomed. Life Sci. 861 (2008) 29–39.
[46] P. Horth, C.A. Miller, T. Preckel, C. Wenz, Efficient fractionation and improved protein identification by peptide OFFGEL electrophoresis, Mol. Cell. Proteomics 5 (2006) 1968–1974.
[47] Y. Shen, J.M. Jacobs, D.G. Camp 2nd, R. Fang, R.J. Moore, R.D. Smith, W. Xiao, R.W. Davis, R.G. Tompkins, Ultra-high-efficiency strong cation exchange LC/RPLC/MS/MS for high dynamic range characterization of the human plasma proteome, Anal. Chem. 76 (2004) 1134–1144.
[48] H. Liu, R.G. Sadygov, J.R. Yates 3rd, A model for random sampling and estimation of relative protein abundance in shotgun proteomics, Anal. Chem. 76 (2004) 4193–4201.
[49] J. Saba, E. Bonneil, C. Pomies, K. Eng, P. Thibault, Enhanced sensitivity in proteomics experiments using FAIMS coupled with a hybrid linear ion trap/Orbitrap mass spectrometer, J. Proteome Res. 8 (2009) 3355–3366.
[50] L.F. Waanders, R. Almeida, S. Prosser, J. Cox, D. Eikel, M.H. Allen, G.A. Schultz, M. Mann, A novel chromatographic method allows on-line reanalysis of the proteome, Mol. Cell. Proteomics 7 (2008) 1452–1459.
[51] R.A. Zubarev, D.M. Horn, E.K. Fridriksson, N.L. Kelleher, N.A. Kruger, M.A. Lewis, B.K. Carpenter, F.W. McLafferty, Electron capture dissociation for structural characterization of multiply charged protein cations, Anal. Chem. 72 (2000) 563–573.
[52] R.A. Zubarev, N.L. Kelleher, F.M. McLafferty, Electron capture dissociation of multiply charged protein cations. A nonergodic process, J. Am. Chem. Soc. 120 (1998) 3265–3266.
[53] G.C. McAlister, D. Phanstiel, D.M. Good, W.T. Berggren, J.J. Coon, Implementation of electron-transfer dissociation on a hybrid linear ion trap-orbitrap mass spectrometer, Anal. Chem. 79 (2007) 3525–3534, Epub 2007 April 3519.
[54] D.L. Swaney, G.C. McAlister, J.J. Coon, Decision tree-driven tandem mass spectrometry for shotgun proteomics, Nat. Methods 5 (2008) 959–964.
[55] D.L. Swaney, C.D. Wenger, J.J. Coon, Value of using multiple proteases for large-scale mass spectrometry-based proteomics, J. Proteome Res. 9 (2010) 1323–1329.
[56] T.P. Conrads, G.A. Anderson, T.D. Veenstra, L. Pasa-Tolic, R.D. Smith, Utility of accurate mass tags for proteome-wide protein identification, Anal. Chem. 72 (2000) 3349–3354.
[57] L. Pasa-Tolic, C. Masselon, R.C. Barry, Y. Shen, R.D. Smith, Proteomic analyses using an accurate mass and time tag strategy, Biotechniques 37 (2004) 621–624, 626–633, 636 passim.
[58] A.V. Tolmachev, M.E. Monroe, S.O. Purvine, R.J. Moore, N. Jaitly, J.N. Adkins, G.A. Anderson, R.D. Smith, Characterization of strategies for obtaining confident identifications in bottom-up proteomics measurements using hybrid FTMS instruments, Anal. Chem. 80 (2008) 8514–8525.
[59] O.V. Krokhin, Sequence-specific retention calculator. Algorithm for peptide retention prediction in ion-pair RP-HPLC: application to 300-and 100-angstrom pore size C18 sorbents, Anal. Chem. 78 (2006) 7785–7795.
[60] O.V. Krokhin, S. Ying, J.P. Cortens, D. Ghosh, V. Spicer, W. Ens, K.G. Standing, R.C. Beavis, J.A. Wilkins, Use of peptide retention time prediction for protein identification by off-line reversed-phase HPLC-MALDI MS/MS, Anal. Chem. 78 (2006) 6265–6269.
[61] R. Craig, J.C. Cortens, D. Fenyo, R.C. Beavis, Using annotated peptide mass spectrum libraries for protein identification, J. Proteome Res. 5 (2006) 1843–1849.
[62] L. Kall, J.D. Storey, M.J. MacCoss, W.S. Noble, Posterior error probabilities and false discovery rates: two sides of the same coin, J. Proteome Res. 7 (2008) 40–44, Epub 2007 December 2004.
[63] J.E. Elias, S.P. Gygi, Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry, Nat. Methods 4 (2007) 207–214.
[64] L. Kall, J.D. Storey, M.J. MacCoss, W.S. Noble, Assigning significance to peptides identified by tandem mass spectrometry using decoy databases, J. Proteome Res. 7 (2008) 29–34, Epub 2007 December 2008.
[65] A. Keller, A.I. Nesvizhskii, E. Kolker, R. Aebersold, Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search, Anal. Chem. 74 (2002) 5383–5392.
[66] D. Fenyo, R.C. Beavis, A method for assessing the statistical significance of mass spectrometry-based protein identifications using general scoring schemes, Anal. Chem. 75 (2003) 768–774.
[67] H. Choi, D. Ghosh, A.I. Nesvizhskii, Statistical validation of peptide identifications in large-scale proteomics using the target-decoy database search strategy and flexible mixture modeling, J. Proteome Res. 7 (2008) 286–292, Epub 2007 December 2014.
[68] H. Choi, A.I. Nesvizhskii, False discovery rates and related statistical concepts in mass spectrometry-based proteomics, J. Proteome Res. 7 (2008) 47–50.
[69] N. Gupta, P.A. Pevzner, False discovery rates of protein identifications: a strike against the two-peptide rule, J. Proteome Res. 8 (2009) 4173–4181.
[70] A.I. Nesvizhskii, R. Aebersold, Interpretation of shotgun proteomic data: the protein inference problem, Mol. Cell. Proteomics 4 (2005) 1419–1440.
[71] F. Desiere, E.W. Deutsch, N.L. King, A.I. Nesvizhskii, P. Mallick, J. Eng, S. Chen, J. Eddes, S.N. Loevenich, R. Aebersold, The PeptideAtlas project, Nucleic Acids Res. 34 (2006) D655–D658.
[72] E.W. Deutsch, The PeptideAtlas project, Methods Mol. Biol. 604 (2010) 285–296.
[73] J.A. Vizcaino, J.M. Foster, L. Martens, Proteomics data repositories: providing a safe haven for your data and acting as a springboard for further research, J Proteomics 73 (2010) 2136–2146.
[74] D. Fenyo, J. Eriksson, R. Beavis, Mass spectrometric protein identification using the global proteome machine, Methods Mol. Biol. 673 (2010) 189–202.
[75] Z. Zhang, Prediction of low-energy collision-induced dissociation spectra of peptides, Anal. Chem. 76 (2004) 3908–3922.
[76] Z. Zhang, Prediction of electron-transfer/capture dissociation spectra of peptides, Anal. Chem. 82 (2010) 1990–2005.
[77] C.Y. Yen, K. Meyer-Arendt, B. Eichelberger, S. Sun, S. Houel, W.M. Old, R. Knight, N.G. Ahn, L.E. Hunter, K.A. Resing, A simulated MS/MS library for spectrum-to-spectrum searching in large scale identification of proteins, Mol. Cell. Proteomics 8 (2009) 857–869, Epub 2008 December 2022.
[78] R. Aebersold, A stress test for mass spectrometry-based proteomics, Nat. Methods 6 (2009) 411–412.
[79] P. Picotti, B. Bodenmiller, L.N. Mueller, B. Domon, R. Aebersold, Full dynamic range proteome analysis of S. cerevisiae by targeted proteomics, Cell 138 (2009) 795–806, Epub 2009 August 2006.

Please cite this article in press as: M.F. Khan, et al., Proteomics by mass spectrometry—Go big or go home? J. Pharm. Biomed. Anal. (2011), doi:10.1016/j.jpba.2011.02.012

[80] P. Picotti, O. Rinner, R. Stallmach, F. Dautel, T. Farrah, B. Domon, H. Wenschuh, R. Aebersold, High-throughput generation of selected reaction-monitoring assays for proteins and proteomes, Nat. Methods 7 (2010) 43–46.
[81] G.M. Walsh, S. Lin, D.M. Evans, A. Khosrovi-Eghbal, R.C. Beavis, J. Kast, Implementation of a data repository-driven approach for targeted proteomics experiments by multiple reaction monitoring, J Proteomics 72 (2009) 838–852, Epub 2008 December 2006.
[82] A. James, C. Jorgensen, Basic design of MRM assays for peptide quantification, Methods Mol. Biol. 658 (2010) 167–185.
[83] R. Kiyonami, A. Schoen, A. Prakash, S. Peterman, V. Zabrouskov, P. Picotti, R. Aebersold, A. Huhmer, B. Domon, Increased selectivity, analytical precision, and throughput in targeted proteomics, Mol. Cell. Proteomics 10 (2011) doi:10.1074/mcp.M110.002931, in press.
[84] C.A. Sherwood, A. Eastham, L.W. Lee, A. Peterson, J.K. Eng, D. Shteynberg, L. Mendoza, E.W. Deutsch, J. Risler, N. Tasman, R. Aebersold, H. Lam, D.B. Martin, MaRiMba: a software application for spectral library-based MRM transition list assembly, J. Proteome Res. 8 (2009) 4396–4405.
[85] P. Mallick, M. Schirle, S.S. Chen, M.R. Flory, H. Lee, D. Martin, J. Ranish, B. Raught, R. Schmitt, T. Werner, B. Kuster, R. Aebersold, Computational prediction of proteotypic peptides for quantitative proteomics, Nat. Biotechnol. 25 (2007) 125–131, Epub 2006 December 2031.
[86] J. Stahl-Zeng, V. Lange, R. Ossola, K. Eckhardt, W. Krek, R. Aebersold, B. Domon, High sensitivity detection of plasma proteins by multiple reaction monitoring of N-glycosites, Mol. Cell. Proteomics 6 (2007) 1809–1817, Epub 2007 July 1820.
[87] J. Sherman, M.J. McKay, K. Ashman, M.P. Molloy, How specific is my SRM?: the issue of precursor and product ion redundancy, Proteomics 9 (2009) 1120–1123.
[88] M.W. Duncan, A.L. Yergey, S.D. Patterson, Quantifying proteins by mass spectrometry: the selectivity of SRM is only part of the problem, Proteomics 9 (2009) 1124–1127.
[89] T.A. Addona, S.E. Abbatiello, B. Schilling, S.J. Skates, D.R. Mani, D.M. Bunk, C.H. Spiegelman, L.J. Zimmerman, A.J. Ham, H. Keshishian, S.C. Hall, S. Allen, R.K. Blackman, C.H. Borchers, C. Buck, H.L. Cardasis, M.P. Cusack, N.G. Dodder, B.W. Gibson, J.M. Held, T. Hiltke, A. Jackson, E.B. Johansen, C.R. Kinsinger, J. Li, M. Mesri, T.A. Neubert, R.K. Niles, T.C. Pulsipher, D. Ransohoff, H. Rodriguez, P.A. Rudnick, D. Smith, D.L. Tabb, T.J. Tegeler, A.M. Variyath, L.J. Vega-Montoto, A. Wahlander, S. Waldemarson, M. Wang, J.R. Whiteaker, L. Zhao, N.L. Anderson, S.J. Fisher, D.C. Liebler, A.G. Paulovich, F.E. Regnier, P. Tempst, S.A. Carr, Multi-site assessment of the precision and reproducibility of multiple reaction monitoring-based measurements of proteins in plasma, Nat. Biotechnol. 27 (2009) 633–641.
[90] T. Fortin, A. Salvador, J.P. Charrier, C. Lenz, F. Bettsworth, X. Lacoux, G. Choquet-Kastylevsky, J. Lemoine, Multiple reaction monitoring cubed for protein quantification at the low nanogram/milliliter level in nondepleted human serum, Anal. Chem. 81 (2009) 9343–9352.
[91] N.L. Anderson, N.G. Anderson, L.R. Haines, D.B. Hardie, R.W. Olafson, T.W. Pearson, Mass spectrometric quantitation of peptides and proteins using stable isotope standards and capture by anti-peptide antibodies (SISCAPA), J. Proteome Res. 3 (2004) 235–244.
[92] N. Burdeniuk, A peptide-based MS immunoassay, Department of Chemistry, University of Calgary, Calgary, 2006, 157pp.
[93] A. Wolf-Yadlin, S. Hautaniemi, D.A. Lauffenburger, F.M. White, Multiple reaction monitoring for robust quantitative proteomic analysis of cellular signaling networks, Proc. Natl. Acad. Sci. U.S.A. 104 (2007) 5860–5865, Epub 2007 March 5826.
[94] M.A. Kuzyk, D. Smith, J. Yang, T.J. Cross, A.M. Jackson, D.B. Hardie, N.L. Anderson, C.H. Borchers, Multiple reaction monitoring-based, multiplexed, absolute quantitation of 45 proteins in human plasma, Mol. Cell. Proteomics 8 (2009) 1860–1877, Epub 2009 May 1861.

Please cite this article in press as: M.F. Khan, et al., Proteomics by mass spectrometry—Go big or go home? J. Pharm. Biomed. Anal. (2011), doi:10.1016/j.jpba.2011.02.012