Available online at www.sciencedirect.com
SCIENCE DIRECT®
Methods 29 (2003) 124–130
METHODS
www.elsevier.com/locate/ymeth

Mass spectrometric-based approaches in quantitative proteomics

Shao-En Ong, Leonard J. Foster, and Matthias Mann*

Center for Experimental Bioinformatics, Department of Biochemistry and Molecular Biology, University of Southern Denmark, Campusvej 55, M 5230 Odense, Denmark

Accepted 15 October 2002

Abstract

Classically, experiments aimed at studying changes in protein expression have always followed a small set of proteins. This focused approach was necessary since tools to efficiently analyze large numbers of proteins were simply not available. Large-scale quantitative proteomics promises to produce reams of data that previously would have taken decades to measure with classical methods. Mass spectrometry is already a well-established protein identification tool and recent methodological developments indicate that it can also be successfully applied to extract quantitative data of protein abundance. From the first reports 4 years ago, numerous schemes to take advantage of stable isotope nuclei incorporation in proteins and peptides have been developed. Here we review the benefits and pitfalls of some of the most commonly used protocols, focusing on a procedure now being used extensively in our laboratory, stable isotope labeling with amino acids in cell culture (SILAC). The basic theory, application, and data analysis of a SILAC experiment are discussed. The emerging nature of these techniques and the rapid pace of technological development make forecasting the directions of the field difficult but we speculate that SILAC will soon be a key tool of quantitative proteomics.

Keywords: Quantitative proteomics; Stable isotope; Amino acid labeling; SILAC; Isotope-coded affinity tag; Isotope labeling; Quantitation

1. Introduction

Proteomics as a field has grown over the years to mean many different things. For our purposes, it is most useful to think of proteomics as the study of the protein complement of a cell. Largely dependent on the methods employed, the subset of proteins of interest could range from a small complex of interacting proteins obtained in pulldown experiments, to subcellular fractions, to even whole cell lysates from tissue samples. Prior to the development of these tools in quantitative proteomics, biologists studying cellular changes were limited to the investigation of either smaller numbers of proteins or gross morphological changes. Quantitative studies were largely dependent on the use of antibodies, and this was semiquantitative at best due to the variation in antibody binding affinities of antibodies. Armed with the methods that we discuss below, biologists now have the ability to monitor global protein expression and obtain much needed quantitative data on the molecular basis of cellular change.

2. Methods

2.1. Quantitation by isotope ratio

At present, the use of stable isotopic nuclei (^{13}C, ^{15}N, ^{2}H) in concert with mass spectrometry (MS) for quantitative analyses is commonplace in the small-molecule field. Known amounts of isotopically labeled versions of drugs or metabolites are spiked into a sample and compared with the unlabeled variant, in what is widely regarded as the best way to apply an internal standard for the purpose of quantitation. Mass spectrometric methods for quantitative analyses of proteins/peptides seek to distinguish two samples by the comparison of isotopically "light" or "heavy" forms. The ratio of intensities of the peptide peaks in a given mass spectrum give a relative ratio of abundance of the two species (see

* Corresponding author. Fax: +45-65933018.
E-mail address: mann@bmb.sdu.dk (M. Mann).

1046-2023/03/$ - see front matter © 2003 Elsevier Science (USA). All rights reserved.
PII: S1046-2023(02)00303-1

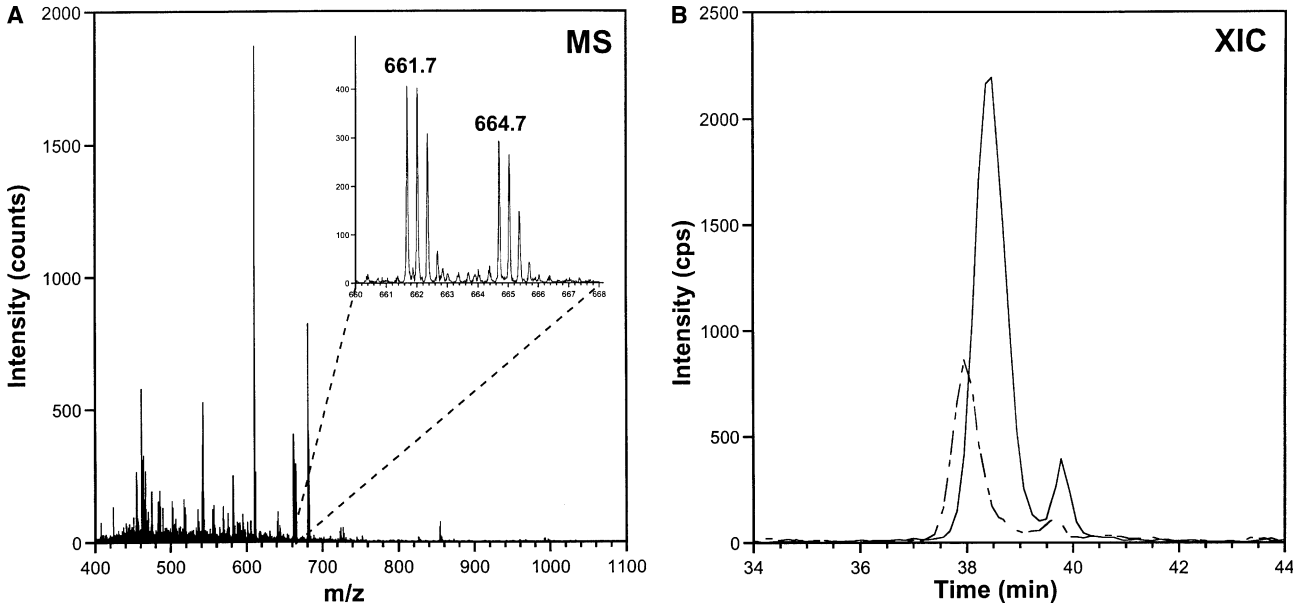

Fig. 1. (A) Typical MS survey scan from a QSTAR Pulsar electrospray ionization quadrupole time-of-flight hybrid mass spectrometer. (Inset) Magnified view of mass region containing heavy (labeled with L-leucine-5,5,5-*d*) and light forms of LLESSLSSSEGEEPVEYK from the integrin α5 protein. Note that the peptide contains three leucines but is also triply charged, resulting in a m/z difference of only 3. (B) Extracted ion currents of the first isotope peaks of the above peptides (661.7 and 664.7). The heavier peptide eluted from the column approximately 0.4 min earlier due to the nine deuterium atoms.

Fig. 1). The labeling protocols can be broadly classified into two groups based on the method of incorporation of the quantitative tag: biological incorporation, where labeling of the peptide/protein is achieved by growing cells in media enriched in stable isotope-containing anabolites, and chemical incorporation, which relies on the use of a derivatization reagent for chemical modification of proteins in a site-specific manner after harvest of the proteins. A summary of some of the methods available is listed in Table 1.

In some of the earliest reports, Langen and co-workers [1] made use of ^{15}N-labeling in bacteria and yeast for quantitation; Oda and co-workers [2] grew yeast on ^{15}N-enriched media for the quantitation of phosphopeptides, whereas Gygi et al. [3] made use of a chemical reagent containing a linker region bearing either none or eight deuterium atoms in the isotope-coded affinity tag (ICAT) approach. After this first report by Gygi et al. [3], several variations on the chemical reagent theme emerged. These include derivatization schemes that target other reactive sites on peptides such as the carboxy termini of peptides, the amino groups in the N-terminal portion, and even the phosphate moieties of phosphopeptides in the PhiAT [4] approach. Furthermore, the use of the serine protease-catalyzed attack of ^{18}O water in the digestion step originally used for the de novo sequencing of peptides [5,6] has also been applied to quantitative proteomics [7,8]. More recently, our laboratory has developed a uniquely powerful method using stable isotope-containing amino acids in mammalian cell culture (SILAC) and applied it to the quantitation of protein expression changes in a muscle differentiation model [9]. We have also introduced the use of ^{13}C substituted amino acids like ^{13}C-arginine with SILAC, which affords some additional advantages for quantitative analyses with liquid chromatography-mass spectrometry.

The variety of chemistries available for modifying reactive groups in a typical peptide combined with the numerous structures possible for a quantitative tag creates a large number of possibilities for chemically incorporating an isotope label. Conversely there are fewer methods for biological incorporation that are viable within the limits imposed by nature—there are only 20 common amino acids made from five elements—since any unusual amino acid analogs may not be as well tolerated by cells. ^{15}N-labeling is viable only for simpler organisms such as yeast [2,10] and bacteria, mainly due to cost—media bearing isotope-enriched reagents are very expensive. Additionally, ^{15}N-labeling adds another layer of complexity to data analysis. Because both the backbone nitrogen atoms and all the side-chain nitrogens are labeled, the varying mass differentials in unlabeled and labeled peptide species cannot be predicted without a priori knowledge of the peptide sequence.

3. Description of procedures

3.1. Quantitative proteomics in cell culture systems: SILAC

Since the bulk of current research in proteomics has a medical focus, quantitative proteomic methods should at least be applicable to mammalian systems since these

Table 1
Labeling approaches used in mass spectrometric-based quantitative proteomics

Method	Biological incorporation		Chemical incorporation		
	SILAC [9]	^{15}N labeling	ICAT (residue specific) [3]	Acetylation	Esterification
Δ Mass	Fixed, depending on choice of label	+1 Da for each nitrogen, variable depending on peptide sequence	8 Da/Cys	3 Da/(deuterated *N*-acetoxysuccinimide) [13]	3 Da for each methoxy [15]
Type of sample	Mammalian cell culture, lower organisms (yeast, bacteria, worm)	Yeast, bacteria [1,2], mammals possible [19] but cost is prohibitive on larger scale	All protein samples, including tissue		
Postlabel fractionation	Optional protein separation methods are applicable, such as SDS–PAGE			Labeling only at the peptide level, postlabeling fractionation limited to peptide fractionations	
Labeling target	Proteins, selected amino acid	Proteins, N termini of peptide, side chains of Lys, Arg, His, Asn, and Gln	Proteins, Cys	Peptides, N termini of peptides, primary amines of Lys	Peptides, C terminus, Asp, Glu
Cross-reactivity (specificity of labeling)	None	None, but knowledge of sequence required to know mass differential	Labeling efficiency in a complex mixture needs optimization. Overlabeling may occur (2 ICAT reagents per peptide)	Ser, Thr, His, Cys (pretreatment with *N*-hydroxylamine)	Only carboxylic groups modified

are often used as model systems for pathophysiological disorders in humans. In SILAC, two groups of cells are grown in culture media identical in all respects save one: the first medium contains the "light" and the second medium contains the "heavy" form of a particular amino acid (e.g., L-leucine or deuterated L-leucine). Through the use of essential amino acids (those not synthesizable by the cell type) the cells are forced to use the particular labeled or unlabeled form. Thus, with each cell doubling the cell population replaces at least half of the original form of the amino acid, eventually incorporating 100% of a given "light" or "heavy" form of the amino acid. We have tested many different attached cell lines (NIH 3T3, mouse fibroblasts; HeLa, human adenocarcinoma-derived; HepG2, human hepatocellular carcinoma), suspension cell lines (HeLa S3, PC-12 rat pheochromocytoma), and primary cultures. In all cells observed thus far we find that SILAC labeling has no detrimental effect on cells in terms of growth rates, morphology, or biological activity.

The SILAC approach neither requires additional purifications to remove excess labeling reagent nor involves multistep labeling protocols. Hence the labeling process is straightforward and highly efficient—100% of the sample is available for analysis. Cells are cultivated in labeling media under typical cell culture conditions. Unlabeled and labeled samples can be combined prior to lysis of the cells and treated as a single sample in all subsequent steps. This allows the experimenter to use any method of protein or even peptide purification (after enzymatic digestion) without introducing error into the final quantitative analysis. The amount of labeled protein in SILAC required for quantitative analyses is far less than that in the case of chemical incorporation where a large excess of labeling reagent and sample is required to ensure an unbiased labeling reaction.

3.2. Extracting quantitative data from mass spectra

A typical LC/MS-MS experiment involves a series of cycles comprising a single MS survey scan followed by three or four product ion scans where peptides are sequenced. The survey scans (see Fig. 1A for an example) represent the total ion currents for each peptide eluting from the HPLC column at that given time. The extracted ion current (XIC) monitors the intensity in the survey scan of a particular mass over time (see Fig. 1B). The area under this peak then represents the "quantity" of that peptide in the sample and may be validly compared with the corresponding peak from an isotopically distinct peak measured in the same experiment to obtain an abundance ratio for that peptide.

This quantitation procedure is simple, both conceptually and in practice. To assist the mass spectrometrist, the acquisition software (such as Analyst which drives ABI/MDS-Sciex mass spectrometers used in our laboratory) usually has tools for extracting single XICs. However, due to the enormous volume of data contained in a single LS/MS experiment (>4000 peptides to extract from 1500 to 3000 survey scans) these extractions are time consuming, even with the assistance of the computer. To complicate matters, the addition of several deuterium atoms to a peptide has a small but significant effect on the interaction of that peptide with reversed-phase materials [11] (note the XIC peak shift in Fig. 1B): the more hydrophillic deuterated peptides elute from reversed-phase material in lower organic solvent content than do natural abundance peptides due to the increased polarizability of the ^{2}H–C bond versus the ^{1}H–C bond. So, in addition to extracting ion currents, the spectrometrist must also account for shifts in elution times when assigning ratios. Currently all this analysis must be done manually as software is not yet available to handle proteome-sized datasets. To circumvent this limitation in our laboratory we have developed "in-house" software to handle these large datasets and mass spectrometer manufacturers plan to have similar software available soon.

In quantitative datasets generated by the SILAC approach, we routinely observe relative standard deviations well within 20% and hence quantitative changes of 1.5× are of value. MS-based quantitative approaches generally give good quantitative data in comparison to microarray-based approaches.

3.3. Additional information derived from SILAC labeling

The resulting increased mass of peptides labeled in SILAC provides structural information by defining additional constraints useful for increasing specificity and, correspondingly, providing higher confidence in protein identifications. In MS/MS spectra, fragmentation patterns generated by unlabeled and labeled peptide pairs are essentially identical except for the expected shift in the masses of ion fragments specific to the location of labeled amino acid [12] (see Fig. 2). With chemical incorporation, the contribution of fragment ions in MS/MS spectra arising from fragmentation of the labeling reagent (especially reagents with larger linker regions, biotin moieties, etc.) may complicate the process of protein identification. However, this pitfall is sometimes balanced by the fact that the reagent fragmentation results in a diagnostic fragment ion of a labeled peptide (product specifications; Applied Biosystems).

Optimal quantitation results are obtained if more peptides are used for quantitating a given protein. In SILAC, the option to select specific amino acids for quantitative labeling improves the odds of increasing the "quantitation coverage" of a given protein; i.e., in the case of leucine, almost 70% of unique tryptic peptides in the human genome contain at least a single leucine residue (unique tryptic peptides of more than five amino

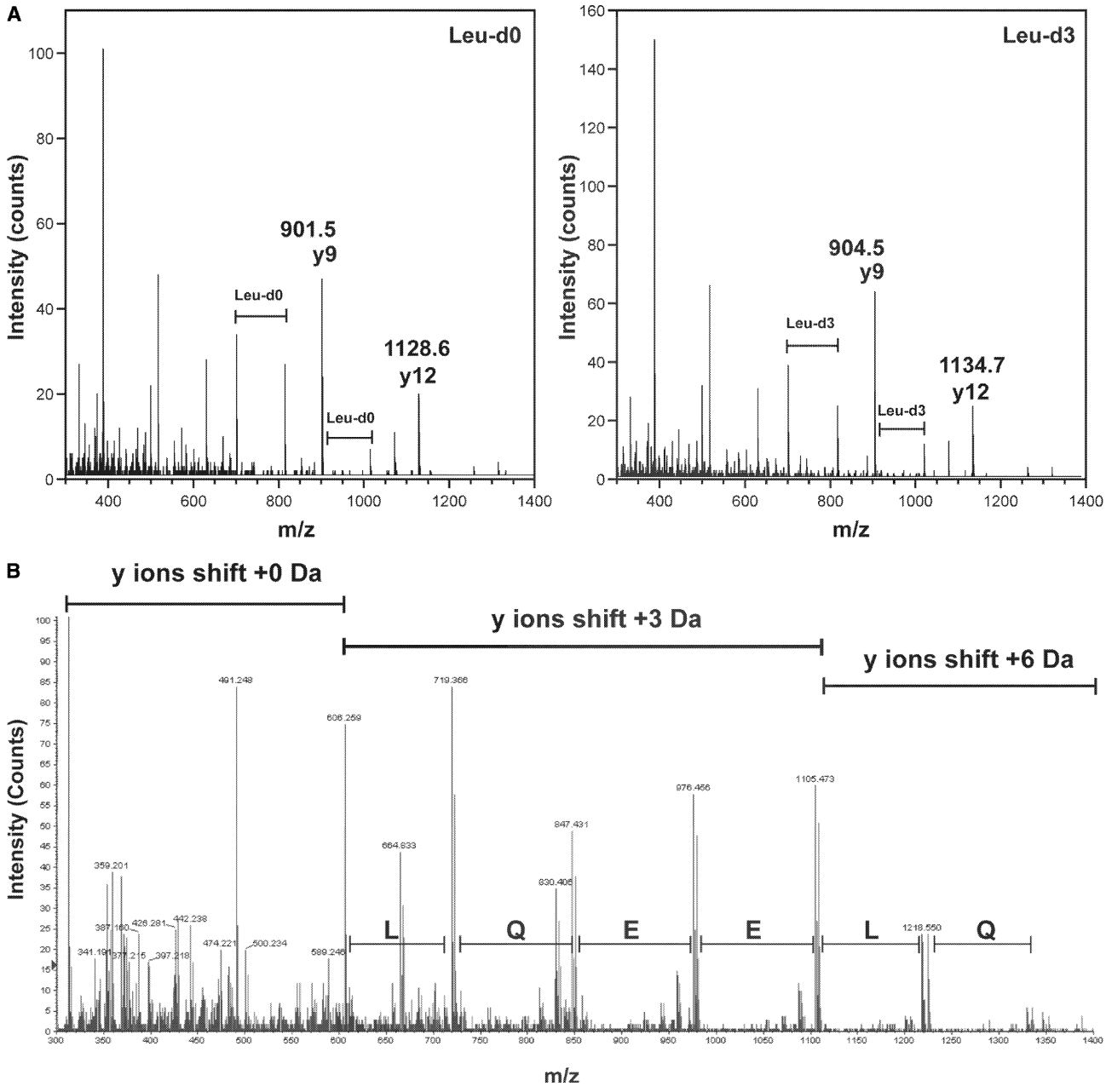

Fig. 2. (A) Product ion spectra (MS/MS) from SILAC peptides. The left panel is the MS/MS spectra of the unlabeled peptide DAGE-GGLSLAIEGPSK, the right panel shows product ion spectra of the same peptide labeled with leucine-5,5,5-*d*. The y9 ions in the two spectra are separated by 3 Da, and y12 ions are separated by 6 Da. (B) Overlaid product ion mass spectra from SILAC peptides. MS/MS spectra of the peptide IAQLEEQLDNETK labeled with normal leucine and leucine-5,5,5-*d* are overlaid to highlight the characteristic shift in fragment masses determined by the position of leucine in the peptide sequence.

acids based on calculations made with the IPI database) and there are almost no proteins without a single leucine. In contrast, chemical incorporation methods that target the sulfhydryl groups of cysteines dictate that only peptides containing cysteines are quantitatable, theoretically limiting the quantitation coverage possible with these approaches to approximately 25% of peptides and roughly 90% of all proteins. Further, if the goal is to study posttranslational modifications, the option to use other labeled amino acids in the SILAC approach could be exploited to increase quantitation coverage or directly target a modified residue.

3.4. Potential pitfalls in quantitation

3.4.1. Labeling issues

Accurate quantitation with biological incorporation methods depends on the level of incorporation (i.e., any

natural abundance isotope nuclei in the heavy sample will add to the signal from the light sample) so the labeling process must be sufficiently long to allow full incorporation of the heavy label. In our experience we find that growing the cells for five population doublings results in full incorporation within the detection limits of the mass spectrometer. In this situation there is theoretically no more than 3.13% ($1/2^5$) of the natural abundance label left in the sample even without consideration of the increased label incorporation from cellular protein turnover.

3.4.2. Sample complexity

The large number of proteins analyzed in whole-proteome quantitation adds an extra dimension of complexity to the measurements. To illustrate, mass spectrometers operating in LC-MS and LC-MS/MS modes (as described above) would only be able to sequence approximately one-twentieth of the total number of tryptic peptides available in an average run (assuming a complex sample composed of 2000 proteins). The increased number of peptides in a quantitative experiment arising from isotopic peptide pairs implies that prefractionation is even more critical prior to MS analysis. With biological incorporation methods, sample complexity can be reduced at both the protein (i.e., SDS–PAGE) and the peptide (i.e., multidimensional ion-exchange chromatography–reversed-phased (IEX-RP) liquid chromatographicseparation) levels. In some chemical incorporation methods, the quantitative tag is introduced only after enzymatic cleavage (e.g., acetylation [13,14], esterification [15], and solid-phase ICAT [16]), thus making peptide subfractionation mandatory (or requiring multiple labeling reactions and increased cost).

As an added boon, some modification-based strategies employ an enrichment step (as with solid-phase capture of peptides on beads [16] or biotinylated ICAT-labeled peptides captured over an avidin column) that helps to reduce the complexity of the sample. Despite this enrichment step, many standard ICAT protocols involving very complex mixtures still require peptide fractionation (involving multidimensional IEX-RP chromatography or the like) before MS analyses to yield useful results.

3.5. To fractionate or not to fractionate

Intracellular trafficking of proteins and protein complexes form an important mechanism for cellular regulation. So, simply quantitating protein changes in whole-cell lysates under differential conditions may not provide an accurate picture of those conditions due to dilution of those changes through the entire cellular protein complement. For example, to accurately compare the relative abundance of nuclear proteins in two different samples with a chemical incorporation method would require two separate nuclei preparations, protein isolation, and labeling reactions before finally mixing the samples for MS analysis. Given the expected losses in each purification step, the amount of starting material required for such an experiment suggests that biological incorporation would be a more practical choice. However, biological incorporation is not feasible in many situations; in systems where tissue explantation and labeling of primary cells in culture are not workable, chemical incorporation methods are the most obvious choice.

3.6. Relative vs absolute changes

Precision in quantitation is difficult to determine, especially on a global scale. The labeling methods discussed here provide relative quantitation. Absolute quantitation is far more difficult but clearly not impossible since it was recently demonstrated for a relatively small dataset. Here, Gygi and co-workers [17] synthesized phosphopeptides and "spiked" known amounts in a protein digest as an internal standard for the quantitation of phosphorylation. This approach requires prior estimation of the absolute amount and the identity of the peptide species to be quantitated. Although this approach has been suggested on a larger scale [18], its application to global proteomics would require a huge investment of resources and would present a formidable barrier for smaller research groups.

3.7. What range of differences can be measured?

The dynamic range of quantitative measurements is an important issue to bear in mind. Relative quantitation hinges on the fact that the mass spectrometric readout is able to tolerate large swings in peptide abundance. Space–charge limitations of the ion-trap (Paul trap) type instruments make them less accurate for quantitative purposes in comparison to quadrupole-time-of-flight type instruments. Further complicating this dynamic range issue is the importance of signal-to-noise and sufficient ion statistics used in peak recognition and quantitation. Since all these factors depend on the instrument and type of analysis employed, it also stands to reason that improved technological developments in future MS instruments will further improve their quantitative capabilities.

4. Concluding remarks

The existing methods in quantitative proteomics each have their strengths and limitations. These tools already appear quite promising for such analyses on a global scale and will deliver an enormous volume of much

needed quantitative data for the study of cellular changes. At the present pace of technological development we expect measurement of proteome-wide changes in protein expression to become routine in the near future.

Acknowledgments

Work in the Protein Interaction Laboratory (PIL) is supported by a generous grant by the Danish National Research Foundation to the Center for Experimental BioInformatics. Fruitful discussions with our colleagues at PIL are gratefully acknowledged.

References

[1] H.W. Lahm, H. Langen, Electrophoresis 21 (2000) 2105–2114.
[2] Y. Oda, K. Huang, F.R. Cross, D. Cowburn, B.T. Chait, Proc. Natl. Acad. Sci. USA 96 (1999) 6591–6596.
[3] S.P. Gygi, B. Rist, S.A. Gerber, F. Turecek, M.H. Gelb, R. Aebersold, Nat. Biotechnol. 17 (1999) 994–999.
[4] M.B. Goshe, T.P. Conrads, E.A. Panisko, N.H. Angell, T.D. Veenstra, R.D. Smith, Anal. Chem. 73 (2001) 2578–2586.
[5] M. Schnolzer, P. Jedrzejewski, W.D. Lehmann, Electrophoresis 17 (1996) 945–953.
[6] A. Shevchenko, I. Chernushevich, W. Ens, K.G. Standing, B. Thomson, M. Wilm, M. Mann, Rapid Commun. Mass Spectrom. 11 (1997) 1015–1024.
[7] O.A. Mirgorodskaya, Y.P. Kozmin, M.I. Titov, R. Korner, C.P. Sonksen, P. Roepstorff, Rapid Commun. Mass Spectrom. 14 (2000) 1226–1232.
[8] X. Yao, A. Freas, J. Ramirez, P.A. Demirev, C. Fenselau, Anal. Chem. 73 (2001) 2836–2842.
[9] S.E. Ong, B. Blagoev, I. Kratchmarova, D.B. Kristensen, H. Steen, A. Pandey, M. Mann, Mol. Cell Proteomics 1 (2002) 376–386.
[10] M.P. Washburn, R. Ulaszek, C. Deciu, D.M. Schieltz, J.R. Yates III, Anal. Chem. 74 (2002) 1650–1657.
[11] R. Zhang, C.S. Sioma, S. Wang, F.E. Regnier, Anal. Chem. 73 (2001) 5142–5149.
[12] M. Mann, S.E. Ong, J. Rappsilber, Y. Ishihama, J.S. Andersen, in: 50th ASMS Conference on Mass Spectrometry and Allied Topics, 2002.
[13] A. Chakraborty, F.E. Regnier, J. Chromatogr. A 949 (2002) 173–184.
[14] F.E. Regnier, L. Riggs, R. Zhang, L. Xiong, P. Liu, A. Chakraborty, E. Seeley, C. Sioma, R.A. Thompson, J. Mass Spectrom. 37 (2002) 133–145.
[15] D.R. Goodlett, A. Keller, J.D. Watts, R. Newitt, E.C. Yi, S. Purvine, J.K. Eng, P. von Haller, R. Aebersold, E. Kolker, Rapid Commun. Mass Spectrom. 15 (2001) 1214–1221.
[16] H. Zhou, J.A. Ranish, J.D. Watts, R. Aebersold, Nat. Biotechnol. 20 (2002) 512–515.
[17] O. Stemmann, H. Zou, S.A. Gerber, S.P. Gygi, M.W. Kirschner, Cell 107 (2001) 715–726.
[18] R. Aebersold, in: 50th ASMS Conference on Mass Spectrometry and Allied Topics, 2002.
[19] T.P. Conrads, K. Alving, T.D. Veenstra, M.E. Belov, G.A. Anderson, D.J. Anderson, M.S. Lipton, L. Pasa-Tolic, H.R. Udseth, W.B. Chrisler, B.D. Thrall, R.D. Smith, Anal. Chem. 73 (2001) 2132–2139.