Quantitative proteomics?

Matthias Mann

cDNAs or short oligonucleotides arrayed on chips have become powerful tools to probe cellular mRNA populations. They permit the comprehensive determination of the expression of individual genes in a given cell type or cell state, the so-called "transcriptome"[1]. Although such analyses provide important information, it would be equally useful to measure individual protein levels directly by proteomic analysis because steady-state levels of mature gene products are subject to additional levels of control and quite often, mRNA and protein levels are not correlated. However, "differential display proteomics" has been slow to take off because of various technical limitations, in particular the difficulty in accurately measuring the differences in individual protein levels between two samples. In this issue, Aebersold and colleagues[2] demonstrate an ingenious method to obtain accurate quantification in a proteomics experiment.

In "classical proteomics," the total protein complement of a cell or tissue is applied to a two-dimensional gel where proteins are electrophoresed to separate them on the basis of their charge and mass. Several hundred proteins can typically be visualized by radioactive or fluorescent labeling or silver staining. Differences between the reference and altered states are measured by quantifying the ratios of spot intensities between independent two-dimensional gels. Unfortunately, spot recognition and quantification is time-consuming and not very accurate, even if computer-assisted.

In array experiments, gene transcripts are identified by their position on the chip. However, proteins on two-dimensional gels do not have fixed, precalculable positions. Thus, for many years spot identification with high sensitivity and throughput was impossible. This bottleneck has now been removed by powerful new methods of mass spectrometry (MS)[3]. Now, protein spots can be excised and incubated with trypsin, which reduces the protein to peptides that are more easily analyzed by MS. The masses of the peptides are measured and the protein is identified by comparison of the resulting "mass fingerprint" to corresponding mass predictions contained in a database. In a second step, if the mass fingerprinting was not sufficient for identification, peptides are sequenced by MS and the fragmentation pattern used to identify the protein in a database, even on the basis of one or a few peptide sequences. Unfortunately, despite the revolutionary impact of the mass spectrometric identification methods, proteins cannot be quantified by mass spectrometric analysis because peptide signals in the mass spectrometer are extremely variable.

To tackle this problem, several groups have borrowed a method from small-molecule MS–isotopic labeling (see Figure 1). In this procedure a stable isotope, such as deuterium, is introduced into the protein sample of one of the states to be analyzed. Although most of the physical characteristics of the peptides remain the same, the masses of the isotopically labeled molecules are shifted in the spectrum. The ratio between the labeled and the natural mass peaks permits accurate quantification of differences. This principle has been used before to compare yeast grown on ^{15}N medium to control yeast grown on normal medium[4]. Protein samples from both states are then mixed and analyzed together, such that each peptide mass peak will now have a "companion" shifted in mass by its number of nitrogen atoms. Such a strategy is limited to microorganisms, however, and Aebersold and coworkers decided to introduce the isotope after the protein extract had been prepared. Standard protocols call for the protein sample to be reduced and the cysteines blocked, to prevent the latter from reacting nonspecifically with gel components and each other. Instead of the normal reagent for this purpose—usually iodoacetamide—these workers used a reagent that is both isotope labeled (contains either eight or no deuterium atoms) and linked to a biotin affinity tag. Proteins from cells grown under a first set of conditions were blocked by the reagent containing eight deuteriums, and those from cells grown under a second condition were blocked by the normal reagent.

Two separate factors contribute to the beauty of the approach. First, once the tag has been synthesized, no additional chemical work-up is required (which would reduce the sensitivity of the analysis). Second, the affinity tag allows selective retrieval of the cysteine-containing pep-

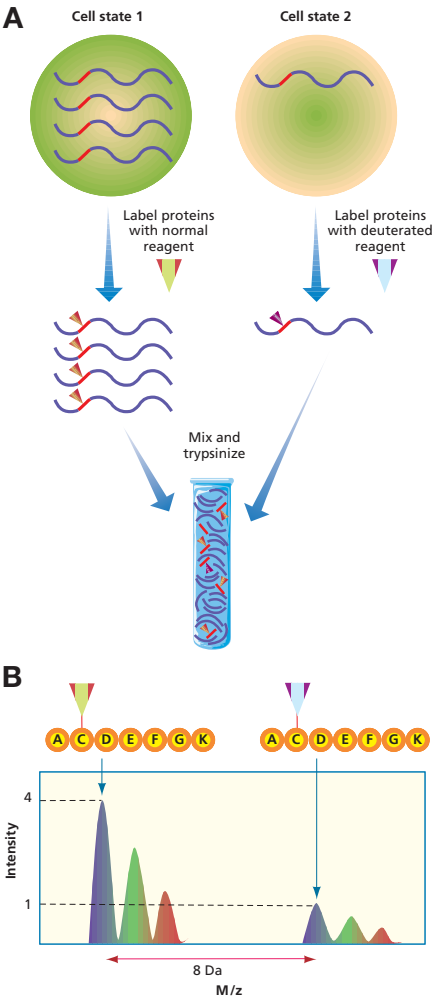

Figure 1. Principle of quantitation by incorporating stable isotope-labeled tags. (A) A protein occurs in higher copy number in cell state 1 than in cell state 2. In the example the ratio is 4:1. Cysteine-containing proteins from a cell extract are labeled with the normal (cell state 1) and the deuterated form of a blocking reagent (cell state 2). After digestion with trypsin, the protein extracts are mixed and cysteine-containing peptides selectively retrieved via a biotin tag on the blocking reagent. (B) Peptides are measured by mass spectrometry, and the otherwise identical peptides will be separated in a mass spectrum by the mass difference between the labeled and unlabeled reagent (8 Da in the example). The ratio of abundances of the gene in cell state 1 compared to cell state 2 is preserved in the two peak heights or peak areas. There is more than one peak for each peptide because of the natural ^{13}C contents of proteins. Sequencing of either of the two identical peptides in the same experiment identifies the gene product being quantitated. In the same experiment several cysteine-containing peptides from a single gene are often analyzed, and as many peptides as possible from all proteins in the extract are measured.

Prof. Matthias Mann is head of the Protein Interaction Laboratory, University of Southern Denmark (mann@cebi.sdu.dk), and CSO at Protana A/S (www.protana.com), Odense, Denmark.

tides, reducing the complexity of the peptide mixture. This latter feature comes in handy when analyzing very crude peptide mixtures, such as those produced by the simultaneous digestion of complex protein mixtures[5] by chromatographic separation coupled to MS. The cysteine-containing peptides of the mixture from the first (blocked by normal reagent) and second (blocked by deuterated reagent) set of growth conditions are simultaneously retrieved using the biotin tag; their ratios to the mixed-in control now provide an accurate relative measurement of protein amount (see Figure 1). Ideally a few cysteine-containing tryptic peptides should be present in the protein to give redundancy in protein identification and quantification, and these peptides should be in a size range to permit easy analysis of mass spectrometric sequencing (5–25 residues). This is generally the case for all but some of the smaller proteins. Although the method is validated here in yeast, in principle it is just as applicable to human tissue. The only difference is that a greater number of proteins will be expressed in human tissue than in yeast, and the human genome contains many more genes making the protein identification task more challenging. However, modern mass spectrometers now acheive extremely high mass spectral quality, which allows confident identification even in very large databases.

Will this method mount a credible challenge to DNA chips for global gene expression analysis? It is not likely that this will be the area of widest application for this technique, because problems with dynamic range are still present. Crude protein mixtures are always dominated by a few species that make up the bulk of the material. When analyzing trypsinized crude peptide mixtures, most of the chromatogram is dominated by peptides from the most abundant proteins. Furthermore, the one or few cysteine-modified peptides must be measured at sufficient signal strength to allow accurate peak quantitation. This is a potential limitation when compared to standard protein identification, where almost any peptide from the protein will suffice and a very weak signal can still permit identification. To quantify most genes expressed in yeast, one would have to sequence thousands of peptides and detect some peptides in amounts thousands of times smaller than others, a formidable challenge to MS. However, the utility of the method to accurately measure ratios of abundant proteins has clearly been demonstrated. Indeed, it will probably be most useful in more targeted "functional proteomics" experiments, for the analysis of members of protein complexes[6,7], proteins binding to an affinity bait, or other situations where protein populations have been preselected by functional purification. Functional groups other than cysteine could also be derivatized with isotopically and affinity-labeled reagents. This could be very useful in the mass spectrometric identification and quantitation of some posttranslational modifications, a largely unsolved problem in high-throughput proteomics.

1. DeRisis, J., Iyer, V.R. & Brown, O.P. *Science* **278**, 680–686 (1997).
2. Aebersold, R. *Nat. Biotechnol.* **17**, 994–999 (1999).
3. Shevchenko, A. *et al. Proc. Natl. Acad. Sci. USA* **93**, 14440–14445 (1996).
4. Oda, Y., Huang, K., Cross, F., Cowburn, D. & Chait, B. *Proc. Natl. Acad. Sci. USA* **96**, 6591–6596 (1999).
5. Yates, J., McCormack, A., Schieltz, D., Carmack, E. & Link, A. *J. Protein Chem.* **16**, 495–497 (1997).
6. Neubauer, G. et al. *Proc. Natl. Acad. Sci USA* **94**, 385–390 (1997).
7. Neubauer, G. et al. *Nat. Genet.* **20**, 46–50 (1998).