Chips for Proteomics: A New Tool or Just Hype?

Peter James

Lund University, Lund, Sweden

BioTechniques 33:S4-S13 (December 2002)

INTRODUCTION

Recently, biological vernacular has been expanded with a series of “omes”: the genome, the DNA sequence of an organism; the transcriptome, the mRNA being expressed at a given time in a cell; and the proteome, the protein equivalent. The latest in the family has been dubbed the metabolome and is a catchall term for all small molecules that are a product of enzymatic and chemical activity within the cell. In contrast to the genome, which is fairly inert, the latter three molecular groups are highly dynamic and vary greatly according to the endo- and exogenous conditions and throughout the life cycle of an organism. The revolution that led to the “omics” explosion has been one based on technology development. The sequencing of whole genomes began with the development of a simple method for determining the sequence of DNA some 30 years ago (27). There were several key steps in the jump from being able to sequence a few short DNA strands to the determination of whole chromosomes. Each of these critical steps was essentially just a refinement of this basic sequencing method: the use of fluorescent dyes, capillary electrophoresis, etc. The critical breakthrough was the realization and subsequent demonstration that, by using the already available techniques and automating them, one could employ shotgun sequencing on a genomic level (10). Once this happened, the next challenge was to develop software capable of managing all the data. Now, large-scale genomics is firmly established.

The next step was the development of new techniques to allow the global analysis of gene expression, transcriptomics, and function. This can be carried out in a quantitative and qualitative manner using binding assays based on DNA arrays (28) or, alternatively, sequencing methods based on either differential display PCRs (19), or a tagged DNA approach termed serial analysis of gene expression (SAGE) (33) can be used. These tools now allow global experiments, such as systematic analyses, of total gene expression following the perturbation of a system to be carried out (34). Proteomics, at least that subbranch generally called expression proteomics (the analysis and quantitation of all proteins and their modifications in a cell at a specified time), is at a similar watershed. The techniques of protein separation 2-dimensional polyacrylamide gel electrophoresis (2D-PAGE) and peptide separation by high-performance liquid chromatography (HPLC) have been established for more than a quarter of a century now. In parallel, the development of new ionization techniques, electrospray ionization (ESI), and matrix-assisted laser

desorption and ionization (MALDI) has allowed peptides and proteins to become accessible to analysis by mass spectrometry (8,18). The major breakthrough in the last 10 years, which has paved the way to allow a proteomics approach to be truly global, has been the development of software algorithms to identify proteins on the basis of their digestion patterns or peptides from their fragmentation patterns. The next step was the development of completely automated systems that allow high-throughput analysis of protein expression in a systematic manner. In this review, I address the various types of proteomics and how protein array technology can be applied, but I also highlight the pitfalls of “blind” chips and the need to interface these arrays with mass spectrometers for protein identification.

WHAT IS PROTEOMICS?

Before going into detail as to how protein expression analysis can be approached by array methods, one should define the term proteomics. Originally, it was defined as the protein complement of the genome; however, since the whole genome is never expressed in a cell, a more restricted definition must be used. The proteome is the set of gene products and their covalent modifications that occur within a given type of cell at a specific stage and time in its development. Proteome analysis can be divided into (i) expression proteomics, which analyzes protein expression and modification; (ii) cell-map proteomics, which attempts to define all protein–protein interactions occurring in a cell under given conditions (3); and (iii) functional proteomics. Expression proteomics relies heavily on quantitative 2D-PAGE to map protein expression in defined cells and is used to follow how protein expression changes in response to perturbation, be it genetic modification or environmental. Cell-map proteomics can be carried out either by high-throughput genetic screening using two-hybrid systems (11) or by isolation and characterization of protein complexes.

These new methods for proteome and gene expression analysis are quantitative and will allow new systematic approaches to the investigation of function and regulation of unknown genes. N.L. Anderson has defined three major areas for the analysis of gene function and regulation: (i) molecular anatomy (protein composition of cells and tissues); (ii) molecular pathology (analysis of disease in terms of changes in protein expression and modification); and (iii) molecular pharmacology/toxicology

(the effects of drugs and xenobiotics on protein expression and modification). A fourth area, molecular physiology, can be added, the change in protein expression in response to changes in the cells micro- or macroenvironment.

DO WE NEED PROTEOMICS WHEN WE HAVE DNA CHIPS?

To fully understand the workings of such a complex system as a cell, global analyses (both spatial and temporal) of transcription, translation, posttranslational modifications (PTMs), and metabolites must be carried out. There is an obvious need to complement the well-established genome-wide mRNA expression methods with global analyses of protein expression and PTM (35,38). There have been very few comprehensive analyses of the correlation between mRNA profiles and protein expression in any biological system (1,13,23). The initial evaluations seem to indicate that there is only a significant correlation between mRNA and protein levels for half of the genes being expressed. The reason(s) for this discrepancy is entirely unknown at the moment. There are several key objections to the reduction of biological studies to following changes in mRNA: (i) the level of mRNA does not allow one to predict the level of protein expression; (ii) protein function is controlled by many PTMs; and (iii) protein maturation and degradation are very dynamic processes, which dramatically alter the final amount of active protein independent of mRNA level. A large-scale protein expression study would be an invaluable aid to understanding this phenomenon, as well as for identifying markers missed by mRNA studies. These studies can also indicate defects in cell signaling mechanisms by showing the changes occurring, for example, in phosphorylation patterns. This would be an important tool in understanding the mechanism underlying the development and progression of a disease. Finally, in a similar vein to analyzing to the relationship of mRNA to protein, the level of metabolites in a tissue is only partially related to the protein expression profile (9). The analysis of metabolite profiles may provide a very useful tool for diagnostics and prognostics (6).

LIMITATIONS FOR ALL PROTEOMICS METHODS

The development of methods for the global analysis of the protein complement of the cell is still in its infancy. Since proteins are vastly more physicochemically diverse than nucleic acids, a universal separation method is unlikely to be found, and this is further compounded by the lack of an amplifying method analogous to PCR. There are only two partially satisfactory methods for analyzing the state of expression of the majority of proteins in a cell. One method is based on 2D-PAGE (24), and the other, new non-gel-based method will be discussed later. In the former technique, proteins are separated in the first dimension according to their isoelectric point, i.e., by migrating to a point in the gel where the pH causes the net charge on the protein to become neutral. In the second dimension, they are separated according to their mobility in a porous gel, which is proportional to the amount of detergent, sodium dodecyl sulfate (SDS)-bound, which is approximately mass-dependent. However,

2D-gel technology suffers many drawbacks. The separation on a single gel can show up to 10 000 species, although many of these are due to PTMs and the number of gene products being visualized is probably only on the order of 1000. The increase in reproducibility that has been brought about by the introduction of commercial immobilized pH gradient (IPG) first dimension gels (2) allows very accurate and quantitative comparative 2D gel mapping. Detailed "proteome maps" can be created with advanced computer imaging programs and then analyzed by subtractive or cluster methods to find relationships between the protein spots. The weakness of 2D-PAGE lies in its inability to deal with certain classes of proteins, mostly the highly hydrophobic ones (membrane and cytoskeletal especially) and those with isoelectric points at either extreme of the pH scale (such as acidic hyperphosphorylated and alkaline DNA binding proteins). There are also problems with quantitation because of the low dynamic range of stains.

There is now a challenge to replace gels with alternative methods of expression analysis. The most promising method is 2D chromatography of whole cell protein digests coupled to mass spectrometry (multidimensional protein identification technology [MudPIT]) (35). The actual level of complexity is greatly reduced. Since a single human gene will exist, on average, in 10 different protein forms, due to splicing or posttranslational modification, the digestion of the proteins reduces 10 proteins to 50 peptides plus 10–20 modified ones, instead of 10×50 after protein separation. The complexity of the problem can be further reduced by selective affinity capture of, for example, cysteine-containing peptides using an isotopic label to enable quantification (12). The isotope-coded affinity tags (ICAT) approach in combination with MudPIT is rapidly gaining popularity.

Despite the debates raging about their relative merits, all share the same limitations. The dynamic range of protein expression in a cell is around 10^7 , and in fluids such as blood serum, it is probably around 10^{12} . The two main detection methods used, mass spectrometry and fluorescence, can only deal with four orders of magnitude. Thus there is no tool, chip, gel, or HPLC that can cover the range of protein expression found in nature. Moreover, all of the techniques are dependent on protein digestion to get fragments for identification or analysis of PTMs, and all proteases cease to be effective at substrate concentrations at the nanomolar level. Even antibody chips must be analyzed to see if the fluorescence is caused by the antigen protein or to cross-reactivity or even the binding of the protein antigen plus a series of other proteins in a complex, since the binding is done under nondenaturing conditions. Thus, whichever approach one takes, one will only see a subset of the proteins that are present in the cell or serum. The protein array as a competitor is lagging behind at the moment because of several technical and conceptual problems that I will outline below.

STANDARD PROTEOMICS OR CHIPS?

Before launching into a discussion of protein arrays, advantages, and limitations, a general outline of the types of environments in which the various proteomics techniques could be valuable will be given. Because of the cost of equipment involved in automation, there will probably be only a limited

number of truly high-throughput proteomics laboratories similar to the DNA sequencing “factories” used for the human genome project. Currently, there are two types of proteomics centers—those with a heavy clinical leaning, aimed at processing huge numbers of clinical specimens to find protein expression–modification correlations for diagnosis and monitoring treatments, and the more problem-orientated center aimed at understanding basic biological phenomena. For both types, the basic questions to be answered before asking how to automate the procedures are ones of experimental design. Should I use a protein-based (2D-PAGE) or a peptide-based (ICAT and MudPIT) approach, or should I be using protein chips? The first step is to define the experimental design factors, which are the parameters within which the experiment will operate. The main factors are: (i) How much material is available? (ii) What coverage of protein expression and modification is required? (iii) How many samples are in the experiment? (iv) How long will the experiments last? and (v) What are you looking for?

For a clinician, material is usually limiting. Tissue biopsies and material from biological specimen banks are usually extremely hard to come by and considered valuable material. For the clinician who may be looking for a diagnostic or prognostic tool, the need for total protein expression coverage is not critical; it is the pattern of expression that is the best indicator, not just one or two proteins. For a biologist, however, all proteins are critical when trying to define functional networks, and coverage is critical. Usually the biologist has access to more material and can afford to do exhaustive prefractionation or multiple zoom gels to increase coverage. The clinicians will, on the other hand, probably be dealing with large patient cohorts to extract meaningful data over a population, so repetition of a single separation type is the most likely course and is the approach most open to automation. Since the proteins of interest (the markers) will not be known until all the patient samples have been analyzed, the time of the experiment is usually long, on the order of months if not years. The ability to store gels and to go back to them for cutting and protein analysis is a great advantage here and reduces the demands on the later stages of the proteome analysis pipeline, in digestion and mass spectrometry protein identification. For the biologist, a MudPIT-type approach, in which an entire cell is digested with a protease and all of the resulting peptides are analyzed by multidimensional peptide HPLC coupled to mass spectrometry for on-line identification, is probably preferable. However, one must note that these exhaustive analyses are extremely slow; the analysis of a single yeast experiment takes around 28 hours of chromatography. The 2D-gel approach allows parallelization, and a small laboratory can easily run 50 gels (corresponding to 100–150 samples using differential in-gel electrophoresis [DIGE] technology) (31) in a week. Both of these techniques are completely dwarfed by the possibilities of protein array technology.

LIMITATIONS OF CHIP-BASED ANALYSIS

Protein, peptide, and ligand arrays promise to be of enormous importance in clinical, pharmaceutical, and basic biology settings. The advantages are clear: large whole proteome array chips, specific diagnostic chips, and chips for ligand and/or

small molecule screening. The area is enormous and clearly holds much promise both for technology developers and users. In terms of the applications in proteomics, chips can be used in all areas: expression proteomics, cell-map, and functional proteomics. The technology for printing the proteins–peptides or small molecules in a high-density fashion has been lifted more-or-less intact from the technologies used in the DNA chip counterparts. Much more work needs to be done on the surface chemistry, in terms of attachment chemistries and providing an environment in which the proteins are stable and do not denature by unfolding onto the surface. These considerations are not part of this review and are not the major problems involved in generating specific antibodies or whole genomes worth of proteins to make these arrays. I will limit the discussion of the analysis and interpretation of arrays to the affinity binding-based arrays, such as protein capture by antibody, affibody, or aptamer, protein:protein interaction probing, and ligand binding arrays (either ligand attached to chip or protein).

Given the considerable number of problems involved in immobilizing the bait molecule on the chip in a fully functional form, the next stage, binding of molecules to the immobilized bait or capture molecule will also be difficult. There are three main areas that have to be dealt with: (i) getting the target to the capture molecule; (ii) preventing nonspecific binding; and (iii) identifying what has bound to the capture molecule. The first problem encountered is diffusion-limited binding. If a protein is present at low concentration, it will take a long time to reach the correct binder immobilized on the chip. Microfluidic channeling and pumping is required to keep the protein solution applied to the chip in a very concentrated form, which can be recirculated past the all the binders in the array. The second problem is closely linked to the first, namely the prevention of nonspecific binding of proteins to the capture molecules. There are several reasons for this. First, this slows down the equilibration of the proteins with the chip and drops the sensitivity drastically. The large dynamic range of protein expression (between 10^7 in cells and 10^{12} in body fluids) compared to mRNA (approximately 10^3) and the number of isoforms of proteins and their modified forms exacerbates the problem. There is no analogue to increasing the salt concentration or temperature to ensure correct hybridization of proteins as can be done for DNA chips. However, the main drawback to the current chip technologies being used is the uncertainty of what is actually binding to the capture molecule.

WHAT AM I BINDING?

The main problem of protein array technology is identifying what is really being bound. The common problems are outlined in Figure 1. The most difficult problem is clearly to obtain antibodies to all proteins that show no cross-reactivity with any other. Anyone who has ever done a Western blot analysis will know how difficult, if not virtually impossible, this is. Once these antibodies are available, more antibodies against all exon products must be made to allow for isoform analysis and to help correct for nonspecific binding. Even if the specificity is absolute, false positives can still be generated since the binding must be done under nondenaturing conditions. This could cause problems,

since most proteins exist in the cell as complexes with other proteins, and these will lead to an inaccurate determination of how much protein is binding a specific point in the array. The opposite problem can also occur, if a protein is recruited to a complex

or is posttranslationally modified in such a way as to destroy the binding epitope, false negatives can be generated. In order to deal with these problems, all chips should be capable of being directly probed to allow the determination of what is being bound.

Figure 1. Problems with sample capture by arrays. This shows the binding of a protein from two theoretical experiments (A, expected result and B, alternative explanations), each with a different fluorescent label. (B1) This indicates that the interpretation is correct after mass spectrometry verification. (B2) This panel would be incorrectly interpreted as a threefold increase in protein expression, whereas with verification, it would be shown that the protein has been incorporated into a complex. (B3) This panel would be interpreted falsely as a decrease in expression, whereas the expression is constant, but after a modification, the epitope is no longer available to the capture molecule.

INTERFACING ARRAYS WITH MASS SPECTROMETRY

The simplest conceptual method for analyzing what is bound is to place the array in a mass spectrometer. The easiest method is to coat the array either completely or in a spot-wise fashion using a dispenser with a UV-absorbing compound, a matrix. The sample can then be desorbed and ionized directly using a UV laser and analyzed by time-of-flight mass spectrometry (TOF-MS) (Figure 2). This method of protein immobilization on either biological arrays (antibodies, protein binder, etc.) or on standard

REVIEW

chromatographic supports immobilized on a MALDI target plate has been described (16a). Termed SELDI (surface-enhanced laser desorption–ionization mass spectrometry), this method has recently proven to be successful in profiling complex protein mixtures such as blood plasma, nipple aspirates, and cerebrospinal fluid to find biomarkers (16b). The drawback here is that only the molecular weight of the intact binding protein can be determined, and this will not be enough to allow a positive determination of the identity (however, see the final section on future developments in mass spectrometry).

To identify the protein or proteins, the protein must be digested in smaller peptides. This allows the protein to be identi-

fied by either protein mass fingerprinting or by peptide mass fingerprinting. The idea that a set of (poly)peptide masses obtained by specific enzymatic or chemical cleavage can be used as a unique fingerprint, allowing a protein to be identified in a database, was first put forward in 1977 (4). The set of masses of polypeptides generated by in gel proteolysis (estimated from SDS-PAGE gels) was used to identify viral coat proteins. The mass accuracy obtained by gel electrophoresis was sufficient to allow the identification, since the size of the database was so small. The concept appeared to lay dormant until 1993, when five independent laboratories published methods and algorithms for database searching using mass spectral data

Figure 2. Time-of-flight mass spectrometry. The laser produces intense pulses of UV light. These are absorbed by the matrix, and the energy causes the ionization and desorption. The pulse starts a timer, which measures the time taken for the ions to fly to the detector, and hence, the mass can be deduced if the instrument is calibrated with compounds of known mass.

Figure 3. Protein identification by protein fingerprinting.

(15,17,21,26,37). The same general approach was common to all the papers. The general strategy of the method is outlined in Figure 3. First, a protein is digested by a chemical or a protease with high sequence specificity (such as cyanogen bromide [CNBr] or trypsin) to produce a set of peptides. The molecular masses are determined by MALDI- or ESI-MS and form the mass fingerprint. This set of experimentally determined set of masses are subsequently used to search a set of mass profiles generated by theoretical fragmentation of a protein database in order to find the protein generating the most similar pattern.

If there are not enough peptides present in the digest, because of poor digestion or low amounts, the peptide mixture can be introduced into a mass spectrometer capable of carrying out MS/MS. This involves the selective isolation (by the mass spectrometer) of a peptide that is then fragmented using a collision gas to generate a fragmentation spectrum that is characteristic of that peptide (Figure 4). This can be used as a peptide fingerprint to identify the protein in the database (Figure 5). The John Yates group developed the SEQUEST program, which fully automates the database searching (7). An alternative approach was put forward by Matthias Mann's group and developed as the PeptideSearch program (22). The algorithm developed for the database searching, PeptideSearch, requires the intact peptide mass, a small manually determined tag sequence, and the tag sequence start and end masses.

FUTURE TRENDS IN MASS SPECTROMETRY COUPLING TO CHIP ARRAYS

Currently, the major limiting factor in analyzing proteins captured on chips by mass spectrometry is sample handling. Enzymatic digestion of proteins at concentrations lower than 500

fmol/ μL is very inefficient. This is simply due to the fact that proteases show 50% maximal activity in the 5–50 pmol/ μL range. The alternative is either to develop chemical methods (but these still suffer from sample handling losses) or to carry out the fragmentation directly in the mass spectrometer. MS/MS of intact proteins electrosprayed into various mass spectrometry instruments has been achieved with a wide range of techniques such as collisionally induced dissociation (29), infrared multiphoton dissociation (IRMPD) (20), UV laser-induced and surface-induced fragmentation (36), and finally, activated electron capture dissociation (ECD; 16). All of these techniques can be carried out by Fourier-transform ion cyclotron resonance (FT-ICR), where the high mass accuracy, resolution, and ability to carry out MS^n experiments allow one to make sense of the complex fragmentation patterns.

FT-ICR mass spectrometry is potentially one of the most sensitive and widely applicable of all mass spectrometry instruments and has recently undergone a revival in interest, although it was first described in 1974 by Comisarow and Marshall (5). The instrument has impressive performance capabilities. A resolving power of over 100 000 has been demonstrated for proteins with a mass of 100 kDa and an accuracy of ± 3 Da. A resolution of almost 1 000 000 was achieved with a peptide mass/ion content ratio (m/z) of 650, allowing not only the separation of the isotopes ^{12}C , ^{13}C , etc., but also the complete resolution of the ^{34}S and ^{13}C isotopes, which are separated by only 0.011 Da (30). The mass range is also enormous—up to 1.1×10^8 . Single ions can be measured, although in general, a few hundred ions should be trapped to give a clear current. Subattomolar sensitivity (<600 000 molecules) for mass spectrometry and MS/MS of intact peptides and proteins up to 30 kDa has been demonstrated (32).

IRMPD allows a high degree of control over the energy added (or removed), is highly efficient, and requires no gas load.

Figure 4. MS/MS peptide fragmentation scheme.

Figure 5. Protein identification by peptide fingerprinting (MS/MS).

Thus, fragmentation can be controlled in a precise manner, which is important for MSⁿ experiments. Combinations of fragmentation methods could be very useful for analyzing PTMs. Most single methods produce confusing hard-to-interpret spectra, but recently, it has been shown that glycopeptides can be sequenced both on the sugar and peptide chains by a combination of ECD and IRMPD (14). The ECD provided sequence ions derived from the peptide backbone, with no observed loss of sugars, whereas IRMPD provided abundant fragment ions, primarily through dissociation at glycosidic linkages. The main drawback is the expense of the instruments and the level of expertise required to run them (as well as their physical size and powerful magnets). A small, easy-to-use benchtop instrument would revolutionize the field of mass spectrometry if it were not too expensive.

CONCLUSIONS

I have outlined a few of the technical disadvantages and pointed out some of the developments occurring in mass spectrometry that may help overcome these. Basically, in exactly the same manner that a cDNA chip can be used to sift out a few hundred candidate genes from tens of thousands, chip arrays for proteins will be able to do the same thing. However, even with cDNA chips, the results need to be confirmed by reverse-transcription PCR (RT-PCR) to retain a large enough confidence level. The same thing must be done for protein chips: the antibodies must be validated for specificity, cross-reactivity, and quantitation. The proteins of interest discovered by these techniques must also be validated to avoid any of the complications outlined above for false positives.

Clearly, the simplicity of the chip format and fluorescent read-out makes it highly attractive, though the price will initial-

ly be high. It is not inconceivable given the recent trends in mass spectrometry miniaturization that a bench-top mass spectrometer could be developed to analyze the binders in a more academic and/or drug discovery environment. Mass spectrometry validated chips will need to be produced for the clinical environment, since reasonably, only a fluorescent read-out can be expected if wide-scale use and acceptance is the target. However, since the methods used for following disease progression and diagnosis are usually based on pattern interpretation techniques, the exact identity of the proteins binding is less critical, since a large number of parameters are used in the diagnosis (this is the conceptual opposite of the needs of the academic biologist who is interested in the specifics of the mechanism). For the foreseeable future at least, all the proteomics techniques (i.e., gel, chromatographic, and array-based) will continue to be used and further developed and find their individual niches.

Clearly, chip arrays for protein analysis will have a large part to play in the development of proteomics. The initial euphoria led to a situation where the arrays have not been delivering what had been promised. However, this is just a question of time and the availability of the appropriate reagents. Protein arrays of the whole human genome are becoming available; the UNIPROT set being constructed by Dolores Cahill from Berlin and the FLEX consortium set from Joshua Labaer at Harvard are two examples. These protein arrays can, in turn, be used to validate the specificities of the antibodies being generated by new automated systems under construction in several laboratories.

REFERENCES

1. Anderson, L., and J. Seilhamer. 1997. A comparison of selected mRNA and protein abundances in human liver. *Electrophoresis* 18:533-537.
2. Bjellqvist, B., K. Ek, P.G. Righetti, E. Gianazza, A. Görg, R. Westermeier, and W. Postel. 1982. Isoelectric focusing in immobilized pH gradients: prin-

- ciple, methodology and some applications. *J. Biochem. Biophys. Methods* 63:17-339.
3. Blackstock, W.P. and M.P. Weir. 1999. Proteomics: quantitative and physical mapping of cellular proteins. *Trends Biotechnol.* 17:121-127.
 4. Cleveland, D.W., S.G. Fischer, M.W. Kirschner, and U.K. Laemmli. 1977. Peptide mapping by limited proteolysis in sodium dodecyl sulfate and analysis by gel electrophoresis. *J. Biol. Chem.* 252:1102-1106.
 5. Comisarow, M.B. and A.G. Marshall. 1974. Fourier transform ion cyclotron resonance mass spectroscopy. *Chem. Phys. Lett.* 25:282-283.
 6. Duez, P., A. Kumps, and Y. Mardens. 1996. GC-MS profiling of urinary organic acids evaluated as a quantitative method. *Clin. Chem.* 42:1609-1615.
 7. Eng, J.K., A.L. McCormack, and J.R. Yates. 1994. Correlating tandem mass spectral data of peptides to sequences in a protein database. *J. Am. Soc. Mass Spec.* 5:976-989.
 8. Fenn, J.B., M. Mann, C.K. Meng, S.F. Wong, and C.M. Whitehouse. 1989. Electrospray ionization for mass spectrometry of large biomolecules. *Science* 246:64-71.
 9. Fiehn, O., J. Kopka, P. Dormann, T. Altmann, R.N. Trethewey, and L. Willmitzer. 2000. Metabolite profiling for plant functional genomics. *Nat. Biotechnol.* 18:1157-1161.
 10. Fleischmann, R.D., M.D. Adams, O. White, R.A. Clayton, E.F. Kirkness, A.R. Kerlavage, C.J. Bult, J.F. Tomb, et al. 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 269:496-512.
 11. Fromont-Racine, M., J.C. Rain, and P. Legrain. 1997. Toward a functional analysis of the yeast genome through exhaustive two-hybrid screens. *Nat. Genet.* 16:277-282.
 12. Gygi, S.P., B. Rist, S.A. Gerber, F. Turecek, M.H. Gelb, and R. Aebersold. 1999. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* 17:994-999.
 13. Gygi, S.P., Y. Rochon, B.R. Franza, and R. Aebersold. 1999. Correlation between protein and mRNA abundance in yeast. *Mol. Cell Biol.* 19:1720-1730.
 14. Hakansson, K., H.J. Cooper, M.R. Emmett, C.E. Costello, A.G. Marshall, and C.L. Nilsson. 2001. Electron capture dissociation and infrared multiphoton dissociation MS/MS of an N-glycosylated tryptic peptide to yield complementary sequence information. *Anal. Chem.* 73:4530-4536.
 15. Henzel, W.J., T.M. Billeci, J.T. Stults, S.C. Wong, C. Grimley, and C. Watanabe. 1993. Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases. *Proc. Natl. Acad. Sci. USA* 90:5011-5015.
 16. Horn, D.M., Y. Ge, and F.W. McLafferty. 2000. Activated ion electron capture dissociation for mass spectral sequencing of larger (42 kDa) proteins. *Anal. Chem.* 72:4778-4784.
 - 16a. Kuwata, H., T.T. Yip, C.L. Yip, M. Tomita, and T.W. Hutchens. 1997. Bactericidal domain of lactoferrin: detection, quantitation, and characterization of lactoferrin in serum by SELDI affinity mass spectrometry. *Biochem. Biophys. Res. Commun.* 245:764-773.
 - 16b. Paweletz, C.P., B. Trock, M. Pennanen, T. Tsangaris, C. Magnant, L.A. Liotta, and E.F. Petricoin, III. 2001. Proteomic patterns of nipple aspirate fluids obtained by SELDI-TOF: potential for new biomarkers to aid in the diagnosis of breast cancer. *Disc. Markers* 17:301-307.
 17. James, P., M. Quadroni, E. Carafoli, and G. Gonnet. 1993. Protein identification by mass profile fingerprinting. *Biochem. Biophys. Res. Commun.* 195:58-64.
 18. Karas, M. and F. Hillenkamp. 1988. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal. Chem.* 60:2299-2301.
 19. Liang, P. and A.B. Pardee. 1992. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science* 257:967-971.
 20. Little, D.P., J.P. Speir, M.W. Senko, P.B. O'Connor, and F.W. McLafferty. 1994. Infrared multiphoton dissociation of large multiply charged ions for sequencing. *Anal. Chem.* 66:2809-2815.
 21. Mann, M., P. Hojrup, and P. Roepstorff. 1993. Use of mass spectrometric molecular weight information to identify proteins in sequence databases. *Biol. Mass Spectrom.* 22:338-345.
 22. Mann, M. and M. Wilm. 1994. Error-tolerant identification of peptides in sequence databases by peptide sequence tags. *Anal. Chem.* 66:4390-4399.
 23. Nelson, P.S., D. Han, Y. Rochon, G.L. Corthals, B. Lin, A. Monson, V. Nguyen, B.R. Franza, et al. 2000. Comprehensive analyses of prostate gene expression: convergence of expressed sequence tag databases, transcript profiling and proteomics. *Electrophoresis* 21:1823-1831.
 24. O'Farrell, P.H. 1975. High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.* 250:4007-4021.
 25. Oliver, S.G., M.K. Winson, D.B. Kell, and F. Baganz. 1998. Systematic functional analysis of the yeast genome. *Trends Biotechnol.* 16:373-378.
 26. Pappin, D.J.C., P. Hojrup, and A.J. Bleasby. 1993. Rapid identification of proteins by peptide-mass fingerprinting. *Curr. Biol.* 3:327-332.
 27. Sanger, F. 1981. Determination of nucleotide sequences in DNA. *Science* 214:1205-1210.
 28. Schena, M., D. Shalon, R.W. Davis, and P.O. Brown. 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270:467-470.
 29. Senko, M.W., J.P. Speir, and F.W. McLafferty. 1994. Collisional activation of large multiply charged ions using Fourier transform mass spectrometry. *Anal. Chem.* 66:2801-2808.
 30. Solouki, T., M.R. Emmett, S. Guan, and A.G. Marshall. 1997. Detection, number, and sequence location of sulfur-containing amino acids and disulfide bridges in peptides by ultrahigh-resolution MALDI FTICR mass spectrometry. *Anal. Chem.* 69:1163-1168.
 31. Unlu, M., M.E. Morgan, and J.S. Minden. 1997. Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. *Electrophoresis* 18:2071-2077.
 32. Valaskovic, G.A., N.L. Kelleher, and F.W. McLafferty. 1996. Attomole protein characterization by capillary electrophoresis-mass spectrometry. *Science* 273:1199-1202.
 33. Velculescu, V.E., L. Zhang, B. Vogelstein, and K.W. Kinzler. 1995. Serial analysis of gene expression. *Science* 270:484-487.
 34. Velculescu, V.E., L. Zhang, W. Zhou, J. Vogelstein, M.A. Basrai, D.E. Bassett, P. Hieter, B. Vogelstein, and K. Kinzler. 1997. Characterization of the yeast transcriptome. *Cell* 88:243-251.
 35. Washburn, M.P., D. Wolters, and J.R. Yates, III. 2001. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.* 19:242-247.
 36. Williams, E.R., K.D. Henry, F. McLafferty, J. Shabanowitz, and D.F. Hunt. 1990. Surface-induced dissociation of peptide ions in FT-ICR-MS. *J. Am. Soc. Mass Spec.* 1:413-416.
 37. Yates, III, J.R., S. Speicher, P.R. Griffin, and T. Hunkapiller. 1993. Peptide mass maps: a highly informative approach to protein identification. *Anal. Biochem.* 214:397-408.
 38. Zhou, H., J.D. Watts, and R. Aebersold. 2001. A systematic approach to the analysis of protein phosphorylation. *Nat. Biotechnol.* 19:375-378.

Address correspondence to:

Dr. Peter James
 Department of Protein Technology
 Lund University
 P.O. Box 7031
 22007 Lund, Sweden
 e-mail: Peter.James@elmat.lth.se