Review

Proteomic Approaches for the Global Analysis of Proteins

BioTechniques 33:1308-1316 (December 2002)

Gregory A. Michaud and Michael Snyder¹

Protometrix, Guilford, CT, and
¹Yale University, New Haven, CT, USA

INTRODUCTION

Basic science research in biology for the past 50 years has been determining the functions of gene products using classical approaches that typically involve studying one or a few genes at a time. However, the complete sequencing of genomes revolutionized the manner in which biological research is conducted. In the field of genomics, it is now possible to prepare microarrays that probe the expression of thousands of genes and even the entire gene repertoire of a single organism. Many expression experiments have described the developmental profile of several organisms and alterations associated with different environmental conditions (16). Mutant banks that contain either a deletion or inactivating construct for every gene in the genomes of many different organisms have recently been described or are underway (5,8,10,27).

More recently, attention has turned to proteomics, the study of proteins and their activities expressed by a given genome. It is important to study proteins rather than just genes because (i) protein levels and activities can differ significantly from RNA levels, (ii) many functions of genes can only be readily studied using biochemical techniques, and (iii) much regulation of gene activity occurs at the protein level, including protein stability, modification, and localization. Below we discuss the current high-throughput methods for the global identification of protein abundance and protein localization. Approaches for revealing protein function through protein-protein interactions and large-scale biochemical as-

says are presented. Finally, we will discuss how bioinformatic analyses of proteomic data are critical to understanding biological problems.

Many of the projects described below use the budding yeast, *Saccharomyces cerevisiae*. Yeast, the first fully sequenced eukaryote, has a relatively small genome size and number of predicted genes (approximately 6300 genes have been annotated) (9). Approximately two-thirds of yeast proteins have similarity to those of vertebrates, indicating that what is learned from yeast will be relevant to humans. As such, many of the methods and information gained from working with yeast can be transferred to other organisms.

PROTEIN ABUNDANCE

Gene expression is a highly regulated process that is influenced by genotype, environment, and cell type (22). Global analysis of protein levels can provide clues as to cellular processes that occur during development and how cells respond to environmental conditions and differ between normal and disease tissues.

Traditionally, protein abundance has been examined using 2-D gel electrophoresis (2D-GE) in which complex protein mixtures are separated into several thousands spots (24). In the first dimension, proteins are isoelectrically focused in a pH gradient, allowing for separation based on pI. In the second dimension, proteins are separated in a mass resolving gel. By staining the gel with protein-binding dyes, semi-accurate determinations for the numbers

ABSTRACT

Improvements in technology that allow miniaturization and high-throughput analyses of thousand of genes and gene products have changed the focus and scope of research and development in both academia and industry. It is now possible to study entire proteomes with the goals of elucidating protein expression, subcellular localization, biochemical activities, and their regulation. Alterations in different cell types and conditions and in normal and disease states can be revealed. This wealth of information not only has facilitated our basic understanding of many biological processes but also has enormous potential for drug discovery and development.

and masses of expressed proteins can be determined. However, there are two problems with this approach. The first is that the resolution of gel electrophoresis is limited. Second and more importantly, the identity of the expressed protein remains unknown.

The coupling of 2D-GE to mass spectrometry (MS) provides for both precise mass determination and protein identification (Figure 1). Accurate mass determination is best gathered when mass mapping is coupled to matrix-assisted laser desorption ionization (MALDI) or electrospray ionization (ESI) time-of-flight (TOF) MS. Typically, a protein sample is cut from a gel, digested with a specific protease (e.g., trypsin), and the resulting peptides are accurately sized by MS and compared to predicted sizes of all the proteins present in the databases. The high-resolution mass determination of the peptides allows accurate identification if several peptides closely match the expected peptide profile of proteins in a database. More recently, the use of tandem mass spectrometry (MS/MS) has become a more popular method of protein identification (2,11) (Figure 1). Tandem mass spectrometry involves the separation of a peptide mixture in the first MS; individual peptides are isolated in the mass spectrometer and fragmented at their peptide bonds, and the fragments are sized in the second MS. The fragmentation pattern allows determination of the peptide sequence,

which, when compared to those predicted in the database, allows for the direct identification of a protein.

Recent improvements in separation methods have alleviated the need for 2-D gels, a time-consuming, low-resolution, and labor-intensive process. Liquid chromatographic (LC) methods in which separation columns are directly linked to the MS allow protein mixtures or a mixture of tryptic digests to be directly separated in the columns and fractions injected into the MS for separation (24). The use of tandem MS (LC/MS/MS) further allows the direct identification of each peptide. This process, which takes several hours to one day, allows the relatively rapid analysis of protein mixtures. One problem with these approaches is that not all proteins present in a mixture are quantitatively separated in peaks; thus, although it is useful for determining which proteins are present, their relative abundance is more suspect.

To circumvent this problem, several groups have developed MS methods for accurately determining relative amounts of proteins in two samples (e.g., proteins isolated from cells grown in two growth conditions or in normal versus disease states). This involves labeling proteins from each source with a different isotope. One group does this by incubating cells in medium containing either a N^{14} or N^{15} nitrogen source (26). Peptide mixtures from two samples are

mixed and subjected to LC/MS/MS. Proteins that differ in the two mixtures can be rapidly identified and sequenced. The other group carries out a similar approach but labels the proteins *in vitro* after isolation with either an isotope-coded affinity tag or a solid-phase isotope tag (12,32) (Figure 2). Briefly, cysteinyl-containing proteins/peptides are specifically labeled with a mass tag that contains seven hydrogen ($d0$) or seven deuterium ($d7$) atoms. By linking the mass tag to either biotin or a solid-phase support, the labeled proteins can be purified, thus simplifying the mixture. Proteins/peptides are then subjected to LC/MS/MS and analyzed as described above. Both of these approaches show considerable promise for analyzing relative protein expression.

PROTEIN MODIFICATION

To date, more than 200 posttranslational modifications have been described (19). One approach for identifying both the type of modification and position on the protein is mass spectroscopy. MS is ideally suited for this approach because it is extremely sensitive. Proteins expressed at low levels having few modifications can be detected.

So far, this approach has been used primarily to measure protein phosphorylation. Typically, a protein of interest is isolated using an affinity tag or an anti-

Figure 1. MS methods for protein identification. A protein mixture is usually treated with a proteolytic enzyme (e.g., trypsin) to fragment proteins into peptides. The resulting peptides can be subjected to either MS or MS/MS for protein identification. In MS, the samples are ionized in the ionization chamber, analyzed by mass/charge (m/z) ratio in the mass analyzer, and detected by the ion detector. In MS/MS, peptides are fragmented in the collision cell at their peptide bonds before entering the second MS. Better sample resolution can be achieved by processing proteins/peptides before running MS. This is achieved by 2D-GE and gel purification of protein bands before tryptic digestion. Also, passing peptide mixtures through LC columns or capillary electrophoresis (CE) before MS decreases the complexity of the sample and increases confidence of protein identification.

Proteomic Technologies

body. The protein is digested with a protease such as trypsin and the resulting peptides analyzed using MALDI or ESI MS. A phosphorylated peptide will have a molecular weight that is 80 Da greater than the nonphosphorylated form. To determine the location of the modification, the peptide mixture can also be subjected to MS/MS. Sequencing of the peptide reveals the modified residue.

Recently, several approaches have been described that perform large-scale analysis of protein phosphorylation; these take advantage of the fact that many phosphorylated peptides exhibit tight affinity to specialized immobilized metal affinity columns (4,25). These columns contain iminodiacetic or nitrilotriacetic acid resins activated with either Fe^{3+} or Ga^{3+} . In one recent study, peptides from a yeast extract were bound to an immobilized metal affinity column, washed, and eluted, and the phosphopeptides were subjected to MS. These researchers were able to identify 383 sites of phosphorylation on 216 peptide sequences (4). Most of these modifications were novel, reflecting the fact that little is known about protein modification on a large scale.

PROTEIN LOCALIZATION

Knowing that a protein is expressed does not provide information about activity. The localization of a protein within the cell is, however, a step closer to revealing function (Table 1). For example, a protein that localizes to the mitochondria or spindle pole likely has a role in respiration or microtubule function, respectively. The fact that a protein localizes to a specific subcellular compartment is also informative in that it tells us what roles it most likely does not have. This is very powerful when integrating diverse datasets (see below). Ascribing function based on localization is crude but certainly better than just knowing that a protein is expressed.

Perhaps the best example of a large-scale protein localization study was undertaken in *S. cerevisiae*. A majority of the yeast proteome was epitope-tagged by several methods, and the subcellular localization of 2744 yeast proteins was determined by high-throughput immunolocalization and fluorescence micro-

scopy (20). It is estimated that approximately 47% of the yeast proteome is cytoplasmic, 20% transmembrane, 27% nuclear, 13% mitochondrial, and 13% exocytic (endoplasmic reticulum and secretory vesicles). Almost 1000 of the localized proteins are of unknown function. The localization of these proteins should help in designing phenotypic screens and assays aimed at further elucidating their function.

Ziauddin and Sabatini (35) have recently described a "reverse transfection" protocol for transfecting mammalian cells on microarrays for analyzing a variety of activities including subcellular localization. Protein expression constructs are printed on slides with a standard arrayer. Subsequently, the mi-

croarray is coated with transfection reagent and then immersed in a plate of actively growing cells for approximately two days. On the slide, cells attach to the surface and take up the expression plasmid printed in that area. After a period of approximately two days, the cells are fixed, the proteins are detected by indirect immunofluorescence, and their locations are determined with a fluorescence scanner. Nuclear localization of the known transcription factors (ETR101, MEF2C, and SP1-like transcription factor) was observed. This approach is very attractive because in principle approximately 10 000 different localizations could be performed on one slide.

Determining sites of protein localization is most reliable when protein lo-

Figure 2. Protein quantification by solid-phase isotope tagging. Protein abundance is measurable using an isotope tagging methodology coupled to MS. (A) Solid-phase isotope tag structure. The tag contains a SH-reactive group that binds to cysteinyl-containing proteins covalently attached to a leucine molecule either containing seven hydrogens (d0) or seven deuterium atoms (d7). This SH-reactive mass tag is linked to beads through a photocleavable linker. (B) Quantification of protein abundance between two samples is achieved by first trypsinizing the samples, reducing the proteins to increase reactivity with the mass tag, and then capturing peptides with either d0 or d7 mass tag beads. After an incubation period, the samples are mixed, washed, and then exposed to light to cleave the linker. The eluted samples are passed through an LC column before MS/MS protein identification and quantitation.

calization patterns are observed under normal levels of gene expression. However, for many proteins, sites of protein localization can only be determined when genes are overexpressed. Caution should be taken when interpreting such data because spurious localizations may arise from gene overexpression.

PROTEIN-PROTEIN INTERACTIONS

One very popular method for elucidating protein function and deciphering the biochemical and regulatory pathways in which proteins function is to determine interacting partners. By identifying interacting partners of known function with those whose role is uncharacterized, it is often possible to infer the biological process and/or activity of the uncharacterized protein.

The most popular method for detecting protein-protein interactions is the two-hybrid approach (Table 1). In these assays, a protein is fused to a DNA-binding domain and tested for interaction in yeast against a panel of con-

structs that have coding sequences fused to a transcriptional activation domain. A positive protein-protein interaction is recorded through the activation of a reporter gene(s) (Figure 3A). This approach has now been applied to several organisms, including two large-scale efforts for the budding yeast *S. cerevisiae* (18,30). These studies have revealed more than 5100 protein-protein interactions and established a large network of interactions. This network has many clusters in which many proteins of similar function group together. Proteins that have not been characterized that fall into these groups are inferred to have the same function as other members of the group (Figure 3B). Using this approach, functional predictions could be made for 364 proteins of unknown function (28).

It is important to note that this procedure has a considerable number of both false positive and false negatives. Indeed, from the two yeast studies that were performed, of the 691 and 1533 core binary interactions found by Uetz et al. (30) and Ito et al. (18), respectively, only 141 interactions are shared. This is most like-

Figure 3. Yeast two-hybrid. (A) A yeast cell expressing two proteins: a target protein fused to a transcriptional activation domain (AD) and a bait protein fused to a DNA binding domain (DBD). Interaction of the two proteins in the nucleus of yeast results in the expression of reporter gene(s) that have upstream DBD enhancer elements. (B) A two-hybrid cluster of proteins involved in spindle pole body function (18). Three uncharacterized proteins (shown in yellow), Ydr016Cp, Ykr083Cp, and Ylr423Cp, interact with known spindle pole body proteins.

Proteomic Technologies

ly because the screens were not comprehensive. Moreover, of the interactions revealed, only approximately 50% are expected to be true interactions (29).

An alternative high-throughput method for revealing protein-protein interactions for large numbers of proteins has recently been described. This method involves affinity purification of proteins coupled with their identification using mass spectroscopy (Figure 4 and Table 1). Using this approach Gavin et al. (6) and Ho et al. (15) identified a number of protein complexes from *S. cerevisiae*. Independent purification of approximately 600 complexes from each study showed that approximately 80% of the purified baits contain copurifying proteins. Gavin et al. (6) report complexes ranging from 2 to 83 components. This reflects that this approach allows for the purification and identification of proteins in both small and large complexes. In addition, the approach is very sensitive in that proteins expressed at only 15 copies per cell are observed in protein complexes.

As with the two-hybrid approach, analyses of the datasets allowed for the identification of numerous protein complexes that could be linked through common copurifying proteins. Proteins of similar function were observed in complexes, and proteins of unknown function that purified with proteins of known function could be assigned potential functional roles.

These datasets provide crucial information regarding multiprotein assemblies within the cell that have not been obtained by two-hybrid analyses alone. A drawback of this approach is that only stable protein-protein interactions are likely to be observed. Weak or transient interactions for binary protein complexes most likely will not survive the purification process. However, multiple weak interactions may be stabilized within larger protein complexes. Another consideration is that similar purification conditions are used for the isolation of protein complexes. Because stringent washing conditions are not used, it is likely that these datasets, like two-hybrid, also contain a high number of false positives.

In summary, both two-hybrid and mass spectrometric identification of copurifying proteins in yeast have gen-

erated extensive proteome protein interaction maps. These maps provide a global view of the networks that link biological pathways in the budding yeast *S. cerevisiae*.

HIGH-THROUGHPUT SOLUTION SCREENING FOR BIOCHEMICAL ACTIVITIES

Innumerable proteins and their corresponding genes have been identified through biochemical fractionation coupled with a specific assay for protein activity. A high-throughput version of this approach was used to reveal biochemical activities for proteins from *S. cerevisiae* (23). Briefly, Martzen et al. (23) generated a library of GST-fusions, transformed the library into yeast, prepared pools of strains expressing different yeast genes fused to GST, purified proteins from the pools, and assayed the

pools for desired activities. Using this approach, they identified activities (a cyclic phosphodiesterase, an Appr-1'-p-processing activity, and a cytochrome c methyltransferase) for three previously uncharacterized genes. This approach has also been used to identify genes that code for sequence-specific DNA-binding transcription factors (14).

GLOBAL ANALYSIS OF BIOCHEMICAL ACTIVITIES USING PROTEIN MICROARRAYS

More recently, it has been possible to analyze the activities of thousands of proteins using protein microarrays (21,33,34). Protein microarrays contain defined set of proteins spotted and analyzed at high density (Figure 5). These have enormous potential for protein profiling and analyzing biochemical activities on a very large scale (Table 1).

Figure 4. Tandem affinity purification. (A) The tandem affinity purification tag contains Protein A and a calmodulin binding peptide linked to the C-terminus of an open reading frame (ORF). Protein A and calmodulin binding peptide are separated by a TEV protease cleavable site. (B) Purification of tandem affinity purification-tagged ORFs from yeast lysates involves first a high-affinity purification using immunoglobulin-coated beads followed by treating the immobilized tandem affinity purification-tagged proteins with TEV protease to elute ORF-CBP tagged proteins. The second chromatographic step involves capturing with calmodulin-coated beads. The proteins are washed, eluted, and electrophoresed by SDS-PAGE. Proteins are purified by cutting bands from the gel, and the identity of any copurifying proteins is identified by MALDI-TOF MS.

Method	Application	Features
2D-GE/MS and LC/CE/MS	Protein abundance/abundance/identification/posttranslational modifications	Very sensitive
Protein Localization	Subcellular localization of proteins	Defines subcellular compartment Data helps increase confidence of protein interaction data Proposes function solely from localization
Two-Hybrid	Protein-protein interactions	Mostly binary interactions Able to detect transient interactions Very automatable High false-positives/false-negatives
MS Identification of Copurifying Proteins	Protein-protein interactions	Able to detect components in small-to-large protein complexes Proteins are in their native environment Most interactions not binary High false-positives
Solution Screening Biochemical Activities	Demonstrate biochemical activities for proteins of known/unknown function	Ability to vary assay conditions Bias for well expressed proteins
Protein Microarrays	protein-protein interactions protein-RNA interactions protein-DNA interactions protein-sugar interactions protein-lipid interactions protein-small-molecule interactions interaction profiling posttranslational modifications enzymatic assays	Very versatile for detecting molecular interactions, including, but not limited to, protein-protein interactions Able to control experimental parameters Very high-throughput Ability to prove binary interactions Proper folding/function for large numbers of immobilized proteins unknown

Figure 5. Protein microarrays. Protein microarrays have typically been prepared by arraying proteins in high density on glass slides. In this example, the slide contains 48 subarrays with 144 spots per subarray. Therefore, the slide would contain $48 \times 144 = 6912$ samples. Probing of microarrays generally involves adding a probe, incubating for a period time at a desired temperature, washing away any unbound probe, and subsequently performing detection (fluorescence, radioactive, etc.) of bound probe. In the bottom subarray, hits are shown in red. Hits are defined as interactions between the probe and samples immobilized on the array.

Two types of protein arrays have been generated: antibody arrays and functional protein arrays. Antibody arrays are used for protein profiling in which protein abundance and/or alterations are measured (13). In this approach, protein mixtures are exposed to an antibody array that reacts with a number of components of the mixture. Each antibody binds its protein antigen, which is usually detected using a second antibody to the protein. In this manner, the amount of protein can, in principle, be quantified simultaneously for large number of samples and among different cell and tissue types. It can also be used to detect specific protein modifications. The biggest hurdle with this approach is the preparation of high-quality antibodies. Most antibodies are not specific enough for these assays. For this reason, antibody arrays for thousands of proteins have yet to be generated.

Functional protein arrays contain proteins arrayed on a solid surface and can be used to screen many proteins for function. The first proteome-wide analyses of biochemical activities were

Proteomic Technologies

performed using the proteome of *S. cerevisiae* (33). Zhu et al. (33) constructed a library of approximately 5800 yeast protein GST-his6-yeast open reading frame fusions. The proteins were overexpressed in yeast, purified, and printed at high-density on nickel or aldehyde slides. This yeast proteome microarray was exposed to a variety of molecules. Proof-of-principle experiments demonstrated that the array could be used to screen protein-protein and protein-lipid interactions. By adding biotinylated calmodulin to the chip, 39 calmodulin-interacting proteins were identified. Within the set, several expected calmodulin-binding proteins (i.e., Cmk1p, Cmk2p, Cmp2p, Dst1p, and Arc35p) were revealed. In addition, 33 novel calmodulin-interacting proteins were revealed, suggesting that these proteins are regulated by binding to calmodulin. A feature worth noting regarding protein microarrays for the

high-throughput mapping of protein interactions is that, of the currently available high-throughput approaches, only protein arrays can allow for confirmation of a binary protein interaction. If a reconstituted interaction between two purified proteins in vitro is not due to the presence of a third protein (which can be determined by MS), then the interaction must be direct. The yeast proteome chip was also probed with liposomes containing different phosphoinositides. One hundred and fifty targets were identified, showing varying specificity dependent on the lipid head group.

Another very exciting application of protein arrays is their potential for screening small-molecule interactions with proteins. For many drugs, the mode-of-action is unknown. By screening a protein array with drugs, candidate targets can be determined. This is certainly possible because MacBeath and Schreiber (21) have demonstrated on

glass slides that the FK506 binding protein of yeast, FKBP12, interacts with small molecules with varying affinity. The disassociation constants of FKBP12 with the tested compounds are in the micromolar to nanomolar range. By controlling the concentration of the immobilized protein or the probe, protein microarrays afford the ability to study interactions with varying affinities. In addition, these experiments can be done under a variety of conditions. The ability to control experimental parameters makes protein microarrays unique.

Because many drug targets are membrane proteins, it was anticipated that protein microarrays would not be suitable for studying interactions with this class of proteins. An elegant study by Fang et al. (3) demonstrates that functional G-protein coupled receptors in a lipid bilayer can be reconstituted on glass slides. The results from photobleaching and inhibition studies suggest

that the membranes are fluidic and that the proteins retain ligand specificity.

Protein chips are also useful for screening enzymatic activities. One of the most studied classes of proteins is protein kinases. Protein kinases are desirable drug targets because of their critical roles in regulating biological signaling pathways. Zhu et al. (34) purified 119 of the 122 kinases from the budding yeast *S. cerevisiae* and screened the kinases with 17 substrates on polydimethylsiloxane (PDMS) chips in microwells. These assays show that many of the kinases are active in this format and exhibit substrate specificity. One of the kinases, Ste20, shows activity towards Axl2p. Both Ste20p and Axl2p have a similar polarized localization during the cell cycle of yeast, indicating that Ste20p may regulate the activity of Axl2p by phosphorylation during bud morphogenesis (34).

Another promising chip technology

involves the use of gold-modified slides to present peptides (17). After arraying a *src* peptide substrate and peptide controls and incubating the slide with *c-Src* kinase, it was observed by phosphorimaging, scintillation counting, surface plasmon resonance, and immunofluorescence that only the *src* substrate is phosphorylated. In addition, Houseman et al. (17) show that kinase inhibition experiments can be performed directly on the array. The compatibility of several detection methods with PDMS chips or gold-modified glass slides makes these supports attractive alternatives to standard commercially available glass slides. In addition, the microwells of PDMS chips offer the ability to do solution/mixing-type experiments in a high-throughput format.

In addition to performing binding and enzymatic assays on slides, protein chips could also be used to detect post-translational modifications. For exam-

ple, phosphoproteins or glycosylated proteins could be detected using phosphospecific antibodies or lectins respectively. Antigen-antibody interactions are readily detectable on protein arrays (13). The drawbacks of proteome chips are the time and cost of producing expression clones for an entire proteome. Also, the long-term stability of proteins on slides remains to be determined. It will be challenging to determine the extent to which proteins are properly folded/functional when immobilized.

DATA INTEGRATION

The various projects described above have generated a wealth of information, particularly for yeast. There is currently considerable interest in integrating the data from the various studies for two reasons, accuracy and coverage. Most of the studies described above

Proteomic Technologies

have high false-positive and even higher false-negative rates. For example, it is estimated that approximately 50% of two-hybrid interactions do not occur in vivo (although this is difficult to prove). An even greater problem is false negatives, since most studies are not comprehensive. By integrating the information from a variety of studies, both better coverage and increased confidence in functional assignments can be achieved. The yeast interaction studies provide an example of this issue. Of approximately 80 000 interactions found from different high-throughput methods, only approximately 2400 interactions are found by more than one study (31). However, when interactions are found by multiple approaches, particularly different approaches, their accuracy is increased significantly. For example, by adding yeast protein localization information to the analysis of yeast interaction datasets by removing interactions of proteins that are localized in different subcellular compartments, the accuracy of the data increases 3-fold, with very little loss in coverage (7).

In summary, by combining diverse approaches and datasets, our understanding of proteomes can be increased significantly. Advances in chemical profiling and functional expression analyses with microarrays will certainly complement existing proteomic approaches (1,35). Further integration of this information will help us to understand biological pathways and basic cell processes in both single-cell and multicellular organisms.

REFERENCES

- Adam, G.C., E.J. Sorensen, and B.F. Cravatt. 2002. Proteomic profiling of mechanistically distinct enzyme classes using a common chemotype. *Nat. Biotechnol.* 20:805-809.
- Aebersold, R. and D.R. Goodlett. 2001. Mass spectrometry in proteomics. *Chem. Rev.* 101:269-295.
- Fang, Y., A.G. Frutos, and J. Lahiri. 2002. Membrane protein microarrays. *J. Am. Chem. Soc.* 124:2394-2395.
- Ficarro, S.B., M.L. McClelland, P.T. Stukenberg, D.J. Burke, M.M. Ross, J. Shanowitz, D.F. Hunt, and F.M. White. 2002. Phosphoproteome analysis by mass spectrometry and its application to *Saccharomyces cerevisiae*. *Nat. Biotechnol.* 20:301-305.
- Fraser, A.G., R.S. Kamath, P. Zipperlen, M. Martinez-Campos, M. Sohrmann, and J. Ahinger. 2000. Functional genomic analysis of *C. elegans* chromosome I by systematic RNA interference. *Nature* 408:325-330.
- Gavin, A.C., M. Bosche, R. Krause, P. Grandi, M. Marzioch, A. Bauer, J. Schultz, J.M. Rick, et al. 2002. Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* 415:141-147.
- Gerstein, M., N. Lan, and R. Jansen. 2002. Proteomics. Integrating interactomes. *Science* 295:284-287.
- Giaever, G., A.M. Chu, L. Ni, C. Connelly, L. Riles, S. Veronneau, S. Dow, A. Luca-Danila, et al. 2002. Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* 418:387-391.
- Goffeau, A., B.G. Barrell, H. Bussey, R.W. Davis, B. Dujon, H. Feldmann, F. Galibert, J.D. Hoheisel, et al. 1996. Life with 6000 genes. *Science* 274:546-563.
- Gonczy, P., C. Echeverri, K. Oegema, A. Coulson, S.J. Jones, R.R. Copley, J. Duperon, J. Oegema, et al. 2000. Functional genomic analysis of cell division in *C. elegans* using RNAi of genes on chromosome III. *Nature* 408:331-336.
- Goodlett, D.R. and E.C. Yi. 2002. Proteomics without polyacrylamide: qualitative and quantitative uses of tandem mass spectrometry in proteome analysis. *Funct. Integr. Genomics* 2:138-153.
- 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.
- Haab, B.B., M.J. Dunham, and P.O. Brown. 2001. Protein microarrays for highly parallel detection and quantitation of specific proteins and antibodies in complex solutions. *Genome Biol.* 2:1-13.
- Hazbun, T.R. and S. Fields. 2002. A genome-wide screen for site-specific DNA-binding proteins. *Mol. Cell. Proteomics* 1:538-543.
- Ho, Y., A. Gruhler, A. Heilbut, G.D. Bader, L. Moore, S.L. Adams, A. Millar, P. Taylor, et al. 2002. Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* 415:180-183.
- Horak, C.E. and M. Snyder. 2002. Global analysis of gene expression in yeast. *Funct. Integr. Genomics* 2:171-180.
- Houseman, B.T., J.H. Huh, S.J. Kron, and M. Mrksich. 2002. Peptide chips for the quantitative evaluation of protein kinase activity. *Nat. Biotechnol.* 20:270-274.
- Ito, T., T. Chiba, R. Ozawa, M. Yoshida, M. Hattori, and Y. Sakaki. 2001. A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl. Acad. Sci. USA* 98:4569-4574.
- Krishna, R.G. and F. Wold. 1993. Post-translational modification of proteins. *Adv. Enzymol. Relat. Areas Mol. Biol.* 67:265-298.
- Kumar, A., S. Agarwal, J.A. Heyman, S. Matson, M. Heidtman, S. Piccirillo, L. Umansky, A. Drawid, et al. 2002. Subcellular localization of the yeast proteome. *Genes Dev.* 16:707-719.
- MacBeath, G. and S.L. Schreiber. 2000. Printing proteins as microarrays for high-throughput function determination. *Science* 289:1760-1763.
- Madden, K. and M. Snyder. 1998. Cell polarity and morphogenesis in budding yeast. *Annu. Rev. Microbiol.* 52:687-744.
- Martzen, M.R., S.M. McCraith, S.L. Spinelli, F.M. Torres, S. Fields, E.J. Grayhack, and E.M. Phizicky. 1999. A biochemical genomics approach for identifying genes by the activity of their products. *Science* 286:1153-1155.
- McDonald, W.H. and J.R. Yates, III. 2000. Proteomic tools for cell biology. *Traffic* 1:747-754.
- McLachlin, D.T. and B.T. Chait. 2001. Analysis of phosphorylated proteins and peptides by mass spectrometry. *Curr. Opin. Chem. Biol.* 5:591-602.
- Oda, Y., K. Huang, F.R. Cross, D. Cowburn, and B.T. Chait. 1999. Accurate quantitation of protein expression and site-specific phosphorylation. *Proc. Natl. Acad. Sci. USA* 96:6591-6596.
- Ross-Macdonald, P., P.S. Coelho, T. Roemer, S. Agarwal, A. Kumar, R. Jansen, K.H. Cheung, A. Sheehan, et al. 1999. Large-scale analysis of the yeast genome by transposon tagging and gene disruption. *Nature* 402:413-418.
- Schwikowski, B., P. Uetz, and S. Fields. 2000. A network of protein-protein interactions in yeast. *Nat. Biotechnol.* 18:1257-1261.
- Tucker, C.L., J.F. Gera, and P. Uetz. 2001. Towards an understanding of complex protein networks. *Trends Cell Biol.* 11:102-106.
- Uetz, P., L. Giot, G. Cagney, T.A. Mansfield, R.S. Judson, J.R. Knight, D. Lockshon, V. Narayan, et al. 2000. A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* 403:623-627.
- von Mering, C., R. Krause, B. Snel, M. Cornell, S.G. Oliver, S. Fields, and P. Bork. 2002. Comparative assessment of large-scale data sets of protein-protein interactions. *Nature* 417:399-403.
- Zhou, H., J.A. Ranish, J.D. Watts, and R. Aebersold. 2002. Quantitative proteome analysis by solid-phase isotope tagging and mass spectrometry. *Nat. Biotechnol.* 20:512-515.
- Zhu, H., M. Bilgin, R. Bangham, D. Hall, A. Casamayor, P. Bertone, N. Lan, R. Jansen, et al. 2001. Global analysis of protein activities using proteome chips. *Science* 293:2101-2105.
- Zhu, H., J.F. Klemic, S. Chang, P. Bertone, A. Casamayor, K.G. Klemic, D. Smith, M. Gerstein, et al. 2000. Analysis of yeast protein kinases using protein chips. *Nat. Genet.* 26:283-289.
- Ziauddin, J. and D.M. Sabatini. 2001. Microarrays of cells expressing defined cDNAs. *Nature* 411:107-110.

Address correspondence to:

Dr. Michael Snyder
Department of Molecular, Cellular, and Developmental Biology
Yale University
266 Whitney Avenue
New Haven, CT 06520, USA
e-mail: michael.snyder@yale.edu