

A Functional Proteomics Approach to Signal Transduction

PAUL R. GRAVES* AND TIMOTHY A.J. HAYSTEAD*[†]

**Department of Pharmacology and Cancer Biology, Duke University,
Durham, North Carolina 27710-3686; [†]Serenex Inc., 323 Foster Street,
Durham, North Carolina 27710*

ABSTRACT

The purpose of this review is to highlight how proteomics techniques can be used to answer specific questions related to signal transduction in a wide variety of systems. In our laboratory, we utilize proteomic technologies to elucidate signal transduction pathways involved in smooth muscle contraction and relaxation, cell growth and tumorigenesis, and the pathogenesis of malaria. We see the real application of this technology as a tool to enhance the power of existing approaches such as classical yeast and mouse genetics, tissue culture, protein expression systems, and site-directed mutagenesis. Our basic approach is to examine only those proteins that differ by some variable from the control sample. In this way, the number of proteins to be processed by electrophoresis, Edman degradation, or mass spectrometry is greatly reduced. In addition, since only those proteins that change in response to a given biological treatment are analyzed, the experimental outcome provides information about specific signaling pathways. Examples of typical experiments in our laboratory are measurement of changes in protein phosphorylation in response to treatment of cells with growth factors or specific drugs, characterization of proteins associated with a bait protein in a “pull-down” experiment, or measurement of changes in protein expression. Frequently, in these experiments, it is necessary to define complex protein mixtures. To achieve this goal, we utilize a variety of techniques to isolate specific types of proteins or “subproteomes” for further analysis. In this review, we discuss strategies used in our laboratory for studying signaling pathways, including subproteome isolation, proteome mining, and analysis of the phosphoproteome.

I. Introduction

Functional proteomics can be defined as a method to identify specific proteins in a cell, tissue, or organism that undergo changes in abundance, localization, or modification in response to a specific biological condition. In functional proteomics, the aim is not to identify or characterize every protein in the cell but rather to provide information about a small number of proteins that are directly relevant to the biological question being studied. To achieve this goal, functional proteomics often is combined with complementary techniques such as protein biochemistry, molecular biology, and cell physiology. In our laboratory, we use functional proteomics to answer questions about signal transduction in a variety of systems. In this review, we describe how proteomics

techniques can be applied to study changes in global protein expression or modification in response to different conditions. In particular, we discuss how the phosphoproteome can be analyzed and provide examples of our own work to illustrate the utility of these methods.

II. The Need for Protein Studies in Signal Transduction

The completion of genome-sequencing projects will facilitate the analysis of signaling pathways in various organisms. As a first step towards this goal, changes in the expression level of mRNA have been measured using techniques such as serial analysis of gene expression (SAGE) (Velculescu *et al.*, 1995) or DNA chip microarrays (Schena *et al.*, 1995; Shalon *et al.*, 1996). These methods will provide valuable information that can be used in conjunction with proteomic studies. However, these methods alone cannot provide accurate information about protein levels in the cell because mRNA levels do not necessarily reflect protein abundance. Indeed, several studies have shown a poor correlation between mRNA levels and protein expression (Anderson and Seilhamer, 1997; Gygi *et al.*, 1999b; Ideker *et al.*, 2001).

Another reason protein analysis is important is because proteins can undergo post-translational modifications and these modifications cannot be predicted from genomic data. Proteins can be modified by the attachment of groups such as phosphates, sulfates, carbohydrates, and lipids. These modifications can change the function of proteins in cells. This is especially important in signal transduction, where signals frequently are transmitted by protein modifications. One type of protein modification that is particularly abundant is protein phosphorylation. It is estimated that 30% of all cellular proteins contain covalently bound phosphate (Hubbard and Cohen, 1993). Thus, to obtain a full understanding of signal transduction pathways, it is essential to study protein modifications at the molecular level.

Finally, protein studies are necessary in signal transduction because signaling pathways are composed of proteins. Many proteins exist in protein complexes whose function may be elucidated only through a study of the intact complex. Currently, the composition and regulation of many protein complexes in signaling remain unknown. In a recent study of protein interactions in *Saccharomyces cerevisiae*, hundreds of multiprotein complexes were detected, of which more than 90% contained at least one previously uncharacterized protein (Gavin *et al.*, 2002; Ho *et al.*, 2002). To fully understand signaling pathways, the individual proteins in a protein complex will have to be identified and characterized during signaling events. These types of experiments will require a proteomics approach.

III. A Proteomics Approach to Signal Transduction

In signal transduction, experiments often are conducted to activate or inhibit specific pathways in order to elucidate the components and regulation of the pathway. For example, signaling pathways can be modulated by treatment of cells with growth factors, hormones, or small molecules. Interference of specific pathways sometimes can be achieved with drugs, gene knockouts, or mRNA interference. Whatever the condition analyzed, in a proteomics approach, the entire cell, tissue, or animal is homogenized and proteins are extracted for study. Proteins that differ by some variable then are selected for further study. The initial results can raise a host of questions. For example, what is the protein that changes in response to the condition? Is the protein modified? What form of modification is it? What are the enzymes responsible for the modification? Finally, what is the biological significance of the protein change observed?

Because all these questions must be answered to understand signal transduction, proteomics represents only the first step in this endeavor. It is no longer sufficient to produce long lists of proteins that undergo changes in a given system. It is also necessary to validate the changes by other methods. Validation of a change in a specific protein may involve cloning and expression of the protein in a heterologous system, protein biochemistry to confirm modifications, and cell biology to determine biological significance. Validation of a *single* protein or its modification often requires 6–12 months of work. Therefore, a proteomics experiment should be designed carefully so that only proteins relevant to the biological problem are studied. Otherwise, significant time can be spent investigating irrelevant proteins.

One of the major hurdles in proteomics occurs at the beginning of the experiment. That is, how can the complex mixture of proteins produced by homogenization of whole cells, tissues, or organisms be analyzed simultaneously? One of the difficulties inherent in a proteomics approach is that a cell extract contains thousands of distinct proteins. These proteins have to be resolved from each other so that changes in individual proteins can be observed. Another difficulty is the detection of low-copy proteins in a cell extract. Finally, because a large number of proteins are involved in proteomic analysis, strategies must be devised to focus the investigation on certain types of proteins.

A. ANALYSIS OF PROTEIN MIXTURES BY TWO-DIMENSIONAL GEL ELECTROPHORESIS

An example of how two-dimensional gel electrophoresis (2-DGE) can be used in proteomics is shown in Figure 1A. In this approach, a cell extract is prepared from two different samples and resolved by 2-DGE for comparison of protein expression or changes in protein modification. This method can provide valuable information and has been used to investigate specific signaling path-

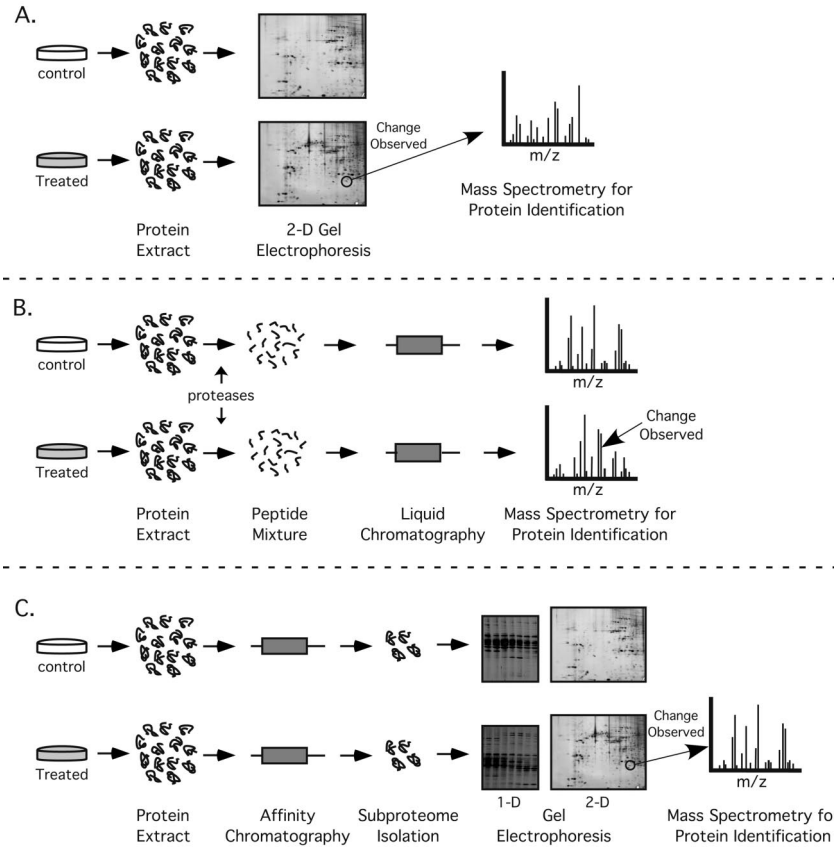


FIG. 1. Strategies for proteome analysis. (A) Analysis of whole proteomes by two-dimensional gel electrophoresis (2-DGE). In this approach, a protein extract is prepared from two different samples and the proteins are resolved and visualized by 2-DGE. Proteins that differ by some variable then are selected for identification by mass spectrometry (MS). Although this method allows for the selection of relevant proteins, few low-copy proteins can be identified. (B) Analysis of whole proteomes by MS. In this approach, all cellular proteins are converted to peptides. The peptides are resolved by liquid chromatography and analyzed by MS. This method allows for the identification of low-abundance proteins but, since there is no selection for relevant proteins, all proteins must be analyzed by MS. (C) Analysis of subproteomes. In this approach, a protein extract is separated into individual subproteomes by fractionation or specialized affinity chromatography and proteins are resolved by 1-DGE or 2-DGE. This allows for the enrichment of low-copy proteins and their selection for further analysis.

ways (Lewis *et al.*, 2000). However, this approach also suffers from many limitations, including poor protein representation and an inability to identify low-copy proteins (Gygi *et al.*, 1999b; Graves and Haystead, 2002). In addition,

it is difficult to screen large numbers of samples for changes because 2-DGE is very labor intensive and time consuming. Therefore, alternative strategies are needed.

B. ALTERNATIVES TO ELECTROPHORESIS

The limitations of 2-DGE have inspired the development of methods to circumvent protein gels entirely. In one such method, the entire protein mixture isolated from a cell is converted to peptides, then the peptides are resolved by liquid chromatography. Following separation, the peptides are injected directly into a mass spectrometer in an “on-line” configuration for mass analysis and protein identification (Figure 1B). A variety of techniques have emerged using this general strategy, including multidimensional liquid chromatography (Link *et al.*, 1999), cation-exchange and reverse-phase chromatography (Opiteck *et al.*, 1997), and liquid chromatography combined with ion mobility spectrometry and tandem mass spectrometry (LC-IMS-TOF-MS/MS) (Kindy *et al.*, 2002). To detect changes in protein expression levels, a method known as isotope coded affinity tags (ICAT) was introduced (Gygi *et al.*, 1999a). In this approach, peptides from two different samples are labeled covalently with a reagent that allows for their subsequent purification and discrimination in the mass spectrometer (Gygi *et al.*, 1999a).

The advantage of these methods is that, theoretically, more proteins can be identified, especially low-copy proteins. In addition, because protein gels are avoided, time is saved from loading and extracting proteins from gels, a process that is not easily automated. However, protein gels perform an important function. They allow for the visual selection of specific proteins from a complex mixture. In the absence of such visual selection, every protein (or peptide) in a cell extract has to be analyzed to find those that undergo changes in response to the biological condition. Frequently, it is important to activate specific signaling pathways and determine early events to understand how signals are transmitted. In these experiments, activation of a specific signaling pathway will not result in changes in the majority of cellular proteins. Therefore, if all proteins in a cell are analyzed, most of the analysis time will be spent on irrelevant proteins. In addition, the analysis of large numbers of proteins requires specialized equipment and computer systems to handle massive amounts of data.

C. TARGETED ANALYSIS OF SUBPROTEOMES

In our laboratory, we utilize an approach that enriches for low-copy proteins and allows for their visual selection. In this method, affinity chromatography is used to isolate subproteomes or specific types of proteins related to the biological question. These proteins are then visualized using one-dimensional gel electrophoresis (1-DGE) or 2-DGE (Figure 1C). The isolation of subproteomes of the

cell serves two important functions. First, it reduces the complexity of the analysis because fewer proteins are analyzed. This saves data analysis time and focuses attention on only those proteins relevant to the biological question. Second, it enriches for low-abundance proteins that otherwise would go undetected in cell extracts. Because of the reduction in protein complexity achieved with this approach, the protein samples can be analyzed for differences using conventional 1-DGE. As a result, many more proteins are represented in the gels and large numbers of samples can be screened quickly.

An example of subproteome enrichment is demonstrated by the use of the specific affinity resin, γ -phosphate-linked adenosine triphosphate (ATP)-Sepharose, originally developed in our laboratory for the purification of protein kinases (Haystead *et al.*, 1993). Unlike previous nucleotide resins, we linked ATP in an orientation favorable for binding protein kinases based upon the crystal structure of cyclic AMP (cAMP)-dependent protein kinase (Knighton *et al.*, 1991). From this structure, it was shown that the γ -phosphate group of ATP extended to the solvent. Therefore, this part of ATP was used for linkage to the matrix. We have shown that, in addition to binding protein kinases, γ -ATP-Sepharose has the ability to bind a large variety of purine-utilizing enzymes, including dehydrogenases, heat shock proteins, nonprotein kinases, DNA ligases, and miscellaneous purine-utilizing enzymes (Graves *et al.*, 2001). This subproteome can be termed the “purine-binding proteome.”

Because many components of signal transduction pathways (e.g., protein kinases) are present in the purine-binding proteome, γ -ATP-Sepharose can be used as a tool to dissect protein-signaling pathways. In this approach, the entire purine-binding proteome is captured from control cells or cells that have undergone some form of treatment. Proteins then are eluted from the matrix with free ATP and either resolved by 2-DGE or further fractionated by ion-exchange chromatography (Figure 2). This approach can provide information about changes in protein expression, in the activation state of protein kinases (as reflected in their ability to bind γ -ATP-Sepharose), or in the post-translational modification of proteins.

IV. Protein Identification

Protein identification is an essential part of almost every proteomics experiment. Here, we outline how mass spectrometry (MS) can be used to identify proteins and describe our approach to protein identification. (For a more comprehensive review of MS, see Burlingame *et al.*, 1998; Yates, 1998.)

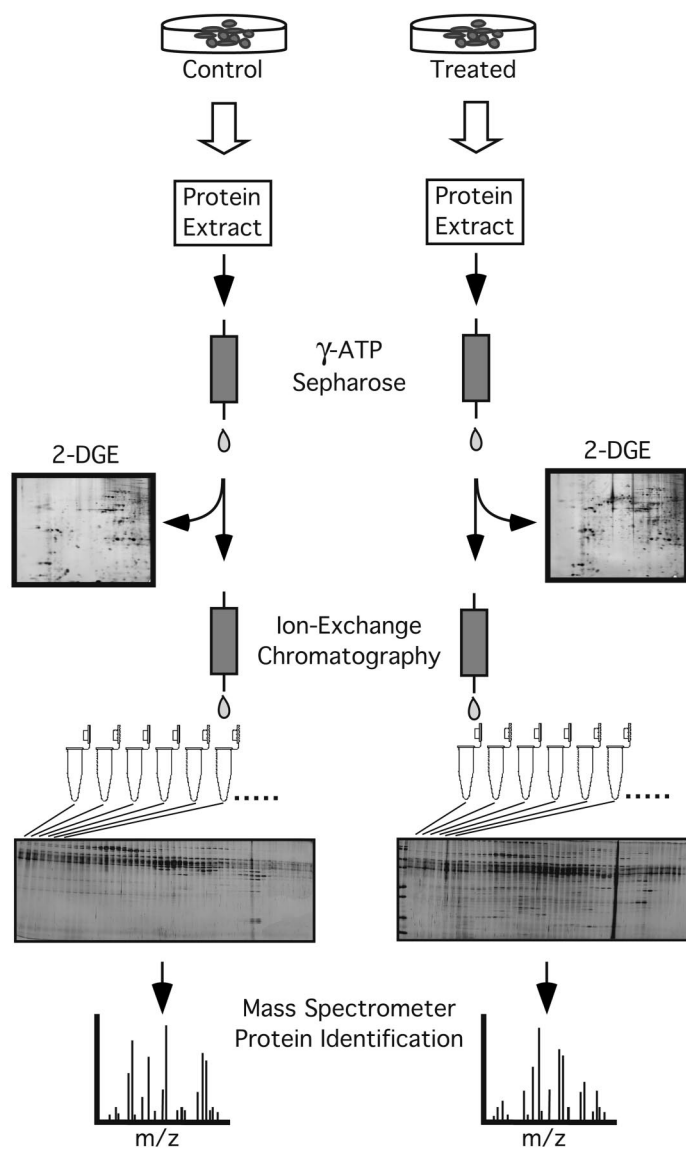


FIG. 2. Analysis of changes in the purine-binding proteome of cells. In this approach, the entire purine-binding proteome from control or treated cells is captured on γ -ATP-Sepharose and the matrix is washed to remove nonspecific proteins. Proteins are eluted with free ATP and analyzed by 2-DGE. As an alternative, the protein eluate can be further resolved by ion-exchange chromatography and 1-DGE. Proteins that differ by some variable are identified by MS. This method can allow for changes in protein expression or modification to be detected in the purine-binding proteome.

A. MASS SPECTROMETRY

MS is now the dominant technology for protein identification (Pandey and Mann, 2000). Using MS, most proteins can be identified at the femtomolar level. However, the success of protein identification can depend upon the sample preparation and on the type of mass spectrometer used. In functional proteomics, considerable effort frequently is expended to obtain a relatively small number of proteins for analysis. In this case, the most important feature of a mass spectrometer is high sensitivity and ability to identify a protein with the highest level of confidence. To this end, we have found that electrospray ionization and tandem mass spectrometry (MS/MS) are the most useful. For MS analysis, samples must be converted to gas-phase ions. Two of the most common methods to accomplish this are matrix-assisted laser desorption ionization (MALDI) and electrospray ionization (ESI).

1. MALDI-MS

In MALDI, a laser is used to promote the ionization of analyte molecules embedded in a crystalline matrix (Karas and Hillenkamp, 1988). The single biggest advantage to MALDI analysis is speed. In MALDI, the entire process of sample preparation and analysis in the mass spectrometer can be automated. This is not yet the case for ESI. As a result, MALDI is the preferred method for high-throughput proteomics operations. However, for many functional proteomics applications, MALDI may not be the best choice. This is because the type of data obtained may not be specific enough for protein identification. In MALDI, the masses of individual peptides derived from a protein are measured in what is known as a peptide mass “fingerprint.” In general, peptide mass fingerprint analysis is not as reliable for protein identification as peptide amino acid sequence obtained by MS/MS. Because of this deficiency, a new breed of MALDI mass spectrometers have emerged that can, in addition to peptide mass fingerprinting, also obtain peptide amino acid sequence. Examples of these machines include MALDI-Q-TOF (Loboda *et al.*, 2000; Shevchenko *et al.*, 2000) and MALDI-TOF-TOF (Medzihradszky *et al.*, 2000). It remains to be seen which machine will be most successful at protein identification.

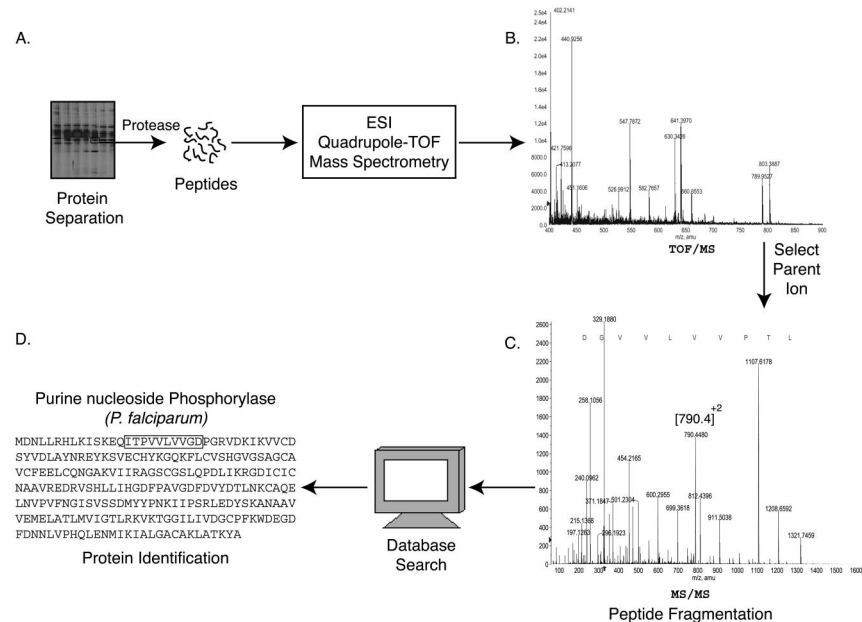
2. ESI-MS

In ESI, a sample is directly introduced into the mass spectrometer as a fine mist of charged droplets (Fenn *et al.*, 1989; Hunt *et al.*, 1992). In an adaptation of electrospray known as nanospray, samples are introduced through micro-capillary tubes at very low flow rates (Wilm and Mann, 1996; Wilm *et al.*, 1996). ESI has a number of advantages over MALDI in functional proteomics applications and, as a result, is the method of choice for sample analysis in our

laboratory. First, nano-electrospray MS can be used to obtain peptide amino acid sequence in a procedure known as MS/MS (Wilm *et al.*, 1996). This is the most specific method for protein identification and greater confidence can be achieved in protein assignment. Second, from our experience, nanospray provides a higher sensitivity than MALDI, allowing for the identification of low-copy proteins. Finally, and perhaps most importantly, MS/MS is less affected by protein mixtures than peptide mass fingerprinting. This is critically important because, frequently, protein “bands” isolated from polyacrylamide gels contain more than one protein.

B. OUR APPROACH TO PROTEIN IDENTIFICATION

The general approach for protein identification in our laboratory is outlined in Figure 3. In this method, a protein band is excised from a gel, digested to peptides by trypsin, and the peptides purified by the method of Wilm and colleagues (Wilm *et al.*, 1996). The resultant peptides are ionized by nano-



electrospray and mass analyzed with a quadrupole-time-of-flight (TOF) mass spectrometer (Q-STAR, PE-SCIEX) to obtain a spectrum of the peptides. From this spectrum, a parent ion is selected for fragmentation and amino acid sequencing. We are capable of identifying faint, silver-stained bands that equal approximately 5–10 ng of protein. However, because MS is not quantitative, frequently several different proteins can be identified from a single band isolated from a polyacrylamide gel. Thus, additional experiments often are required to confirm that the protein identified in the mass spectrometer correlates with a given protein observed on a gel.

C. DATABASE SEARCHING

Using data produced by mass spectrometers, proteins can be identified by searching DNA and protein-sequence databases. The success of protein identification depends upon the type of data utilized, the type of search conducted, and the databases searched. In cases where databases are incomplete, different search strategies are necessary.

1. Peptide Mass Fingerprint Database Searching

In this approach, peptide masses obtained from MALDI-MS are compared against theoretical spectra obtained from primary-sequence databases (Eng *et al.*, 1994). Although this type of search method is rapid, it is not error tolerant and requires a high degree of agreement for success. Factors that alter peptide masses from those in the database (e.g., post-translational modifications) can invalidate this search method. Since many proteins in signaling pathways are modified, this search method may not result in correct protein identification. Another factor that can contribute to inaccurate protein identification is mass degeneracy. For example, a peptide with the sequence VSLNP will have the same mass as the peptide LSNVP. In addition, the masses of certain amino acids, when combined, can equal the mass of other amino acids. For example, the mass of glycine plus glutamate or aspartate plus alanine equals the mass of tryptophane. Because of these limitations, peptide mass fingerprinting is not well suited for protein identification from organisms with large genomes (e.g., human) or towards genomes that are not completely annotated (Qin *et al.*, 1997).

2. Peptide Mass Tag Database Searching

Peptide mass tag database searching can be conducted using peptide sequence obtained from MS/MS. In this approach, a partial amino acid sequence, known as the sequence tag, is combined with the mass of the peptide to search relevant databases (Eng *et al.*, 1994; Mann and Wilm, 1994). The advantage to this method is that it is fast and, if enough amino acid sequence is obtained, can

be a very specific method for protein identification. However, this search method also depends upon a complete and well-annotated sequence database.

3. FASTS Database Searching

The FASTS program uses amino acid sequence obtained from MS/MS to search databases from organisms whose genomes have not been sequenced completely or whose databases are not fully annotated (Mackey *et al.*, 2002). In this approach, *de novo* peptide sequence is obtained from MS, entered into the FASTS program, and all peptide sequences are searched against a database simultaneously. Because the program utilizes amino acid sequence, it is error tolerant (Mackey *et al.*, 2002). This is important because sequencing errors, polymorphisms, and conservative amino acid substitutions often can prevent protein identification with other search methods. A schematic of how the FASTS program functions is shown in Figure 4. The FASTS program is available to the public at <http://fasta.bioch.virginia.edu/>.

V. Phosphoproteome Analysis

Much of proteomics has centered around the quantitation of differences in protein expression between two samples. However, a comparison of protein expression levels provides only one aspect of protein regulation. Proteins also undergo a large number of post-translational modifications that can affect

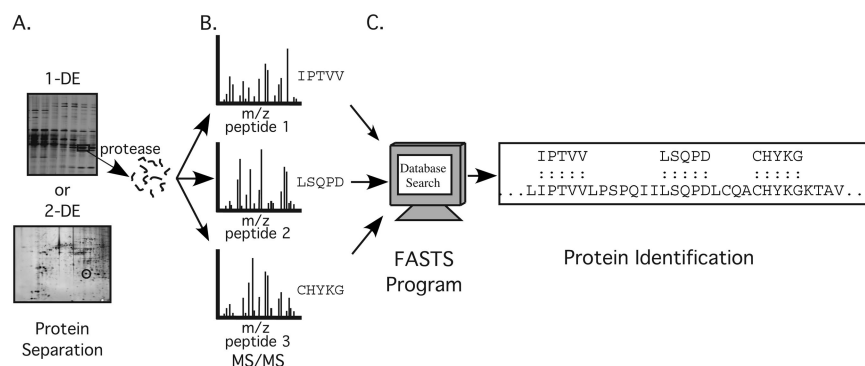


FIG. 4. Protein identification by the FASTS program. The FASTS program allows database searching with a combination of short peptide sequences obtained from *de novo* sequencing by MS. (A) Proteins of interest are separated by 1-DGE or 2-DGE, excised from the gel, and digested to peptides. (B) MS/MS is performed and *de novo* amino acid sequence is obtained for multiple peptides. (C) All the peptides are entered into the FASTS program and a search is conducted with the peptide sequences simultaneously against the relevant database.

function. Reversible protein phosphorylation is one of the most widespread protein modifications. Thus, methods must be devised to study the phosphoproteome, defined as all phosphorylated proteins in the cell. Here, we summarize recent advances and outline some of our own strategies for analysis of the phosphoproteome.

A. PHOSPHOPROTEOME ANALYSIS BY PROTEIN RADIOLABELING

Protein labeling with either inorganic phosphate ($^{32}\text{P}_i$) or γ -(^{32}P)-ATP is still one of the most practical ways to study protein phosphorylation. In proteomics, a common approach is to label cells *in vivo* with $^{32}\text{P}_i$, then resolve the ^{32}P -labeled proteins by 2-DGE. The major drawback to this approach is that although many phosphoproteins can be visualized by autoradiography, they are not present in sufficient amounts to be identified. To overcome this problem, we utilize methods to enrich for specific subproteomes prior to analysis by electrophoresis (Figure 5).

1. Analysis of Protein Phosphorylation *in Vivo*

In this approach, cells are labeled *in vivo* with ^{32}P -orthophosphate, treated in some way, then cell extracts are prepared. The cell extracts are applied to the affinity matrix ATP-Sepharose for capture of the purine-binding proteome. Following washing to remove nonspecific proteins, ATP-binding proteins (including protein kinases) are recovered by elution with ATP and then further resolved by ion-exchange chromatography. Protein phosphorylation is visualized by electrophoresis of fractions on 1-DGE followed by autoradiography. Proteins that undergo changes in phosphorylation in response to the treatment are sequenced and identified by MS (Figure 5A). In an alternative approach, radiolabeled lysates are fractionated by ion-exchange chromatography, then directly analyzed by 1-DGE or 2-DGE.

2. Protein Kinase Profiling *in Vitro*

Protein kinase profiling is a way to identify protein kinases and their substrates under different biological conditions. In this strategy, a cell extract is fractionated by ATP-Sepharose or ion-exchange chromatography and the individual fractions are incubated with kinase buffer containing γ - ^{32}P -ATP to allow labeling of phosphorylated proteins (Figure 5B). Individual fractions are resolved by 1-DGE and proteins that undergo changes in phosphorylation are selected by autoradiography. If specific proteins undergo changes in phosphorylation, they are excised from the gel and identified by MS. If individual proteins cannot be identified in the fraction, then the entire fraction can be reapplied to another ion-exchange column under different conditions. Alternatively, the fraction can

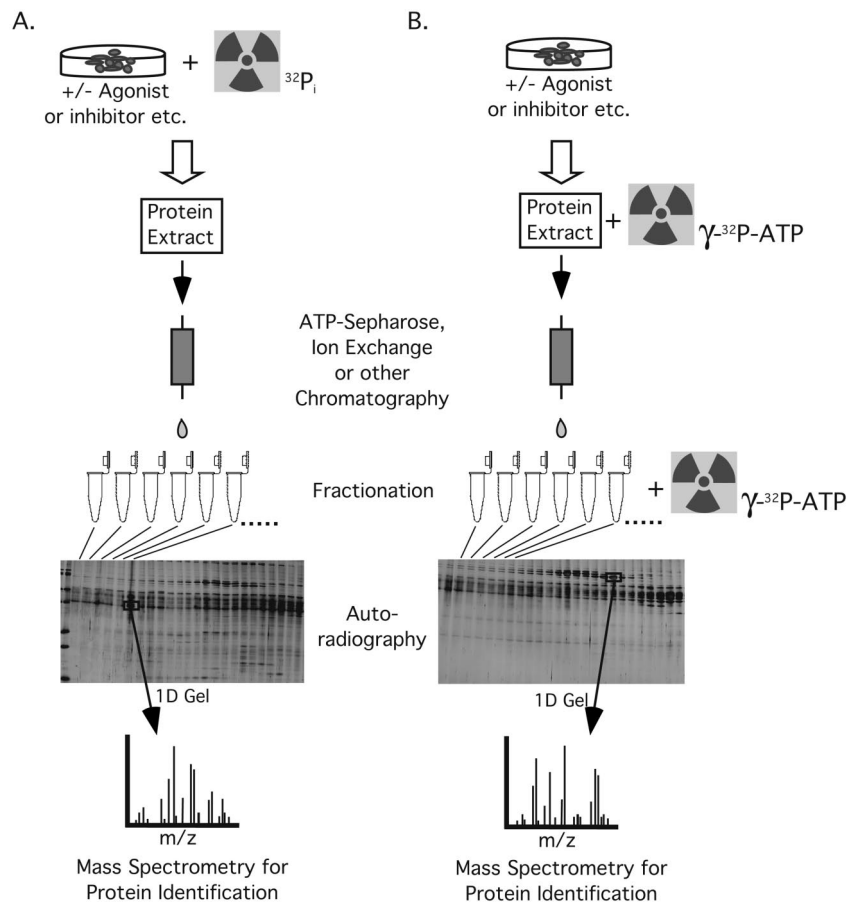


FIG. 5. Strategies for phosphoproteome analysis. (A) Analysis of *in vivo* protein phosphorylation. In this approach, cells are grown in the presence of ^{32}P -inorganic phosphate to allow labeling of all phosphoproteins. Following cell treatment, cell extracts are prepared and the proteins fractionated by chromatography to enrich for specific types of proteins. Proteins are resolved further by electrophoresis and changes in phosphorylation are visualized by autoradiography. Proteins that undergo changes in phosphorylation are excised and sequenced by MS. (B) Analysis of *in vitro* protein phosphorylation. In this approach, a protein extract is incubated with $\gamma\text{-}^{32}\text{P}\text{-ATP}$ or, alternatively, the proteins are fractionated and the individual fractions incubated with $\gamma\text{-}^{32}\text{P}\text{-ATP}$. This approach can allow the identification of protein kinases and their substrates under different conditions.

be applied to $\gamma\text{-ATP}$ -Sepharose to isolate protein kinases. The advantage to this approach is that a large number of samples can be screened easily and quickly for changes in protein phosphorylation and both the substrates and the protein

kinases can be identified. One caveat to this approach is that because the phosphorylation is conducted *in vitro*, the results must be confirmed by *in vivo* studies.

B. METHODS FOR PHOSPHOPROTEOME ISOLATION

A major obstacle in the study of phosphorylated proteins is that phosphoproteins comprise only a small fraction of the total protein in a cellular lysate. As a result, many phosphoproteins cannot be identified in a cell extract. To address this problem, methods have emerged to enrich for phosphorylated proteins from a complex mixture of proteins. These techniques may facilitate the isolation and study of the entire phosphoproteome under different biological conditions.

1. Immobilized Metal Affinity Chromatography

Immobilized metal affinity chromatography (IMAC) has been used to purify phosphopeptides (Andersson and Porath, 1986) and, more recently, has been combined with MS to identify phosphoproteins from cell lysates (Posewitz and Tempst, 1999; Cao and Stults, 2000; Stensballe *et al.*, 2001; Ficarro *et al.*, 2002). However, in addition to binding phosphopeptides, IMAC is known to bind nonphosphorylated, acidic peptides through their carboxylate groups. To overcome this problem, Ficarro and colleagues converted the carboxylate groups in peptides to their corresponding methyl esters. This step was reported to reduce the binding of nonphosphorylated peptides to the IMAC matrix (Ficarro *et al.*, 2002).

In the approach of Ficarro and colleagues, a protein extract is prepared from the sample and digested to peptides by the action of a protease. The resultant peptide mixture is treated with methanolic hydrochloric acid (HCl) to convert carboxylate groups to methyl esters and the mixture is applied to an IMAC column. The recovered peptides are separated further by high-performance liquid chromatography (HPLC) and analyzed online by an ion-trap mass spectrometer (Ficarro *et al.*, 2002). However, if one assumes that approximately 30% of all proteins are phosphorylated (Hubbard and Cohen, 1993), then the total number of phosphoproteins in a cell lysate easily could exceed 1000. This approach requires that all phosphoproteins in a cell lysate (i.e., thousands of phosphoproteins) be analyzed to detect those that undergo changes.

2. Chemical Derivatization of Phosphorylated Amino Acids

One approach to isolate phosphorylated proteins or peptides is to take advantage of the unique chemistry of phosphoamino acids in peptides. Previously, Meyer and colleagues developed a method for the identification of phosphoserine in peptides by its conversion to S-ethylcysteine and Edman

sequencing (Meyer *et al.*, 1991). Under alkaline conditions, phosphoserine will undergo β -elimination to dehydroalanine, which can be converted to S-ethylcysteine by Michael addition with ethanethiol. In our laboratory, we optimized conditions for the derivitization of phosphoserine with 1,2-ethanedithiol (EDT) for the purpose of fluorophore labeling of proteins (Fadden and Haystead, 1995). This same strategy can be used to attach biotinylated moieties to purify phosphoproteins or peptides. An example of this approach recently was provided by Oda and colleagues who, using the same chemistry described herein, converted phosphoserine residues to biotinylated residues in proteins. The biotinylated proteins were purified by avidin chromatography, digested to peptides, and identified by MS (Oda *et al.*, 2001). In an alternate method, Zhou and colleagues devised a technique to isolate peptides phosphorylated on serine, threonine, and tyrosine residues (Zhou *et al.*, 2001).

The main disadvantage of these two methods is that the current chemistries require significant amounts of protein or peptide for identification by MS to be successful. As a result, we have not found these methods to be a practical approach for isolation of phosphoproteins. In addition, the selectivity of these methods has not been confirmed. Several proteins were isolated using these methods that are not known phosphoproteins (e.g., glyceraldehyde 3-phosphate dehydrogenase). Therefore, further work needs to be done to confirm these results and to determine the specificity of this approach (Oda *et al.*, 2001; Zhou *et al.*, 2001).

C. IDENTIFICATION OF PROTEIN PHOSPHORYLATION SITES

In addition to identifying phosphoproteins, it is important to identify the phosphorylation sites in proteins, so that information can be obtained about the function of the phosphorylation event. The two predominant techniques for phosphorylation site analysis are Edman degradation and MS.

1. Phosphorylation Site Identification by Edman Degradation

Edman degradation is still one of the most practical methods to determine phosphorylation sites in peptides. This is because the technique is relatively simple, is very sensitive, and can be applied to a large variety of peptides (Aebersold *et al.*, 1991; Boyle *et al.*, 1991; Wettenhall *et al.*, 1991). If enough radioactivity can be incorporated into the phosphoprotein of interest, sites can be determined at the sub-fmol level. In our hands, this can be as little as 1000 cpm (not ideal). In this approach, a ^{32}P -labeled protein is digested with a protease and the resulting phosphopeptides are purified by reverse-phase HPLC or thin-layer chromatography (TLC) (Figure 6). The isolated peptides then are cross-linked via their C-termini to an inert membrane and the radioactive membrane is subjected to several rounds of Edman cycles. The radioactivity is collected after each

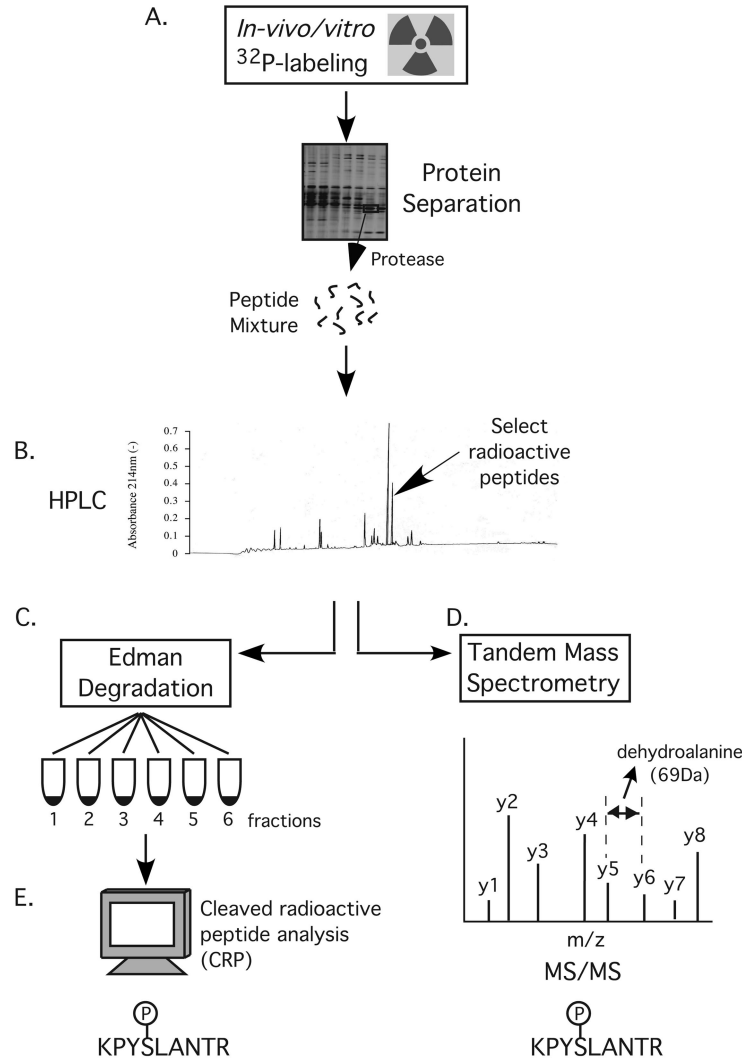


FIG. 6. Strategies for phosphorylation site identification. (A) Proteins that are labeled with ^{32}P are resolved by electrophoresis and converted to peptides by the action of proteases. (B) Peptides are resolved by high-performance liquid chromatography (HPLC) and phosphorylated peptides are selected by measurement of radioactivity. (C) Phosphorylated peptides are subject to Edman degradation and the resultant fractions are monitored for release of ^{32}P . (D) Analysis of phosphorylated peptides by MS/MS. The location of phosphoserine in a peptide can be determined by obtaining amino acid sequence and characterizing the mass difference between peptide fragments. (E) Cleaved radioactive peptide analysis is a bioinformatics program that assists the assignment of phosphorylation sites in peptides.

cleavage step and the released ^{32}P is measured in a scintillation counter. This method positionally places the phosphoamino acid within the sequenced phosphopeptide (Figure 6). Of course, this is meaningful only if the sequence of the phosphopeptide is known. The analysis ceases to become quantitative beyond 30 Edman cycles (even with efficient, modern Edman machines) due to well-understood issues with repetitive yield associated with Edman chemistry.

Recently, our laboratory has extended the usefulness of phosphorylation site characterization by Edman chemistry through the development of the cleaved radioactive peptide (CRP) program (MacDonald *et al.*, 2002). In CRP analysis, one requires only that the sequence of the protein be known. Purification and sequencing of individual peptides is not required. Radiolabeled proteins (isolated following immunoprecipitation from ^{32}P -labeled cells, for example) are cleaved at predetermined residues by the action of a protease. The phosphopeptides are then separated by HPLC or TLC (or, if only one site is present, no peptide separation is required), cross-linked to the inert membrane, and carried through 25–30 Edman cycles. The sequence of the target protein is entered into the CRP program. This program predicts how many Edman cycles are required to cover 100% of the serines, threonines, and tyrosines from the site of cleavage. Generally, one round of CRP analysis narrows the number of possible sites to 5–10 for most proteins. Phosphoamino acid analysis can be used to reduce the number of possibilities still further. The CRP analysis is repeated following cleavage with a second protease (usually one cutting at R but M and F are alternatives). The second round of CRP usually unambiguously localizes the phosphoamino acid to one possible site. The technique does not work if sites are greater than 30 amino acids away from all possible cleavage sites. The finding that CRP analysis is not applicable may, in itself, confine a phosphorylation site to a segment of the protein that is likely to produce very large proteolytic fragments. The CRP program is accessible at <http://fasta.bioch.virginia.edu/crp/> and was written in collaboration with Aaron Mackey and Bill Pearson of the University of Virginia (MacDonald *et al.*, 2002).

2. Phosphorylation Site Identification by MS

Phosphorylation sites in peptides can also be analyzed by MS/MS. In this approach, the phosphopeptide is sequenced in the mass spectrometer and the site of phosphorylation is determined unambiguously (Figure 6). Phosphopeptides can be identified from a mixture of peptides by a method known as precursor ion scanning (Neubauer and Mann, 1999). Serine and threonine phosphorylated peptides undergo loss of 98 Da (H_3PO_3^-) and 80 Da (HPO_3^-) from the molecular ion. These ions can be used to identify phosphorylated peptides. Having identified a phosphopeptide, the peptide mixture is sprayed under acidic conditions and the phosphopeptide sequenced by conventional MS/MS. Upon fragmentation of

the phosphopeptide, phosphoserine can be identified by the formation of dehydroalanine (69 Da), the β -elimination product of phosphoserine. Similarly, phosphothreonine can be identified by the formation of its β -elimination product, dehydroamino-2-butyric acid at 83 Da (Neubauer and Mann, 1999).

VI. Proteomics Applications

A. YEAST GENETICS AND PROTEOMICS TO ANALYZE SIGNALING PATHWAYS

One application of proteomics that holds great potential is the analysis of signal transduction pathways in yeast. Proteomics techniques, when combined with yeast genetics, can allow dissection of signaling events at the molecular level. In our laboratory, we have combined proteomics and yeast genetics to identify physiological targets of protein phosphatases. In *S. cerevisiae*, protein phosphatase-1 (Glc7p) and its binding protein, Reg1p, are essential for regulation of glucose-repression pathways. However, little was known about the physiological substrates of this phosphatase. Therefore, in a collaborative effort with Dr. Marion Carlson, we undertook a proteomics approach to identify the physiological substrates of the phosphatase (Alms *et al.*, 1999). To accomplish this goal, we examined the effects of deletion of the REG1 gene on the yeast phosphoproteome. Analysis by two-dimensional phosphoprotein mapping identified two distinct proteins that were greatly increased in phosphate content in the *reg1* deletion mutants. Peptide microsequencing identified these proteins as hexokinase II (Hxk2p) and the E1 α subunit of pyruvate dehydrogenase.

We went on to validate these findings in a comprehensive biochemical study. Consistent with increased phosphorylation of Hxk2p in response to *REG1* deletion, fractionation of yeast extracts by anion-exchange chromatography identified a Hxk2p phosphatase activity in wild-type strains that was selectively lost in the *reg1* Δ mutant. Having carried out these studies, we attempted to rescue the *reg1* Δ phosphoprotein phenotype by overexpressing both wild-type and mutant Reg1p in the deletion strains. Here, both the phosphorylation state of Hxk2p and Hxk2p phosphatase activity were restored to wild-type levels in the *reg1* Δ mutant by expression of a LexA-Reg1p fusion protein. In contrast, expression of a LexA-Reg1p protein containing mutations at phenylalanine in a putative PP-1C (the catalytic subunit) binding site motif (K/R)(X)(I/V)XF was unable to rescue Hxk2p dephosphorylation in intact yeast or restore Hxk2p phosphatase activity. These results demonstrate that Reg1p targets PP-1C to dephosphorylate Hxk2p *in vivo* and that the peptide motif (K/R)(X)(I/V)XF is necessary for its PP-1 targeting function. These studies, therefore, demonstrate how a proteomics approach can be used to first identify enzyme targets in cells, then direct further analysis to verify the findings. Clearly, a combined proteomics

and genetics approach greatly enhances one's ability to directly answer key biological questions. We believe that a similar strategy could be adopted with transgenic or knockout mouse work, particularly in cases where there is no obvious phenotype.

B. SMOOTH MUSCLE CELL PHYSIOLOGY

Proteomics, when combined with cell physiology, can be a valuable tool to understand basic cellular processes. In our laboratory, we apply proteomics techniques to study the regulation of smooth muscle contraction (SMC). SMC is regulated principally by free intracellular calcium levels and the level of myosin light-chain phosphorylation (Sellers and Adelstein, 1985). The steady-state level of myosin light-chain phosphorylation is, in turn, regulated by the relative activities of myosin light-chain kinase and myosin light-chain phosphatase (SMPP-1M). SMPP-1M itself is regulated by phosphorylation and its activity can be inhibited by phosphorylation of its targeting subunit. However, until recently, the protein kinase responsible for this phosphorylation was unknown. Therefore, we utilized a proteomics strategy to identify the relevant endogenous kinase responsible for phosphorylation and inhibition of SMPP-1M.

Because γ -ATP-Sepharose is capable of binding protein kinases (Haystead *et al.*, 1993), we applied the entire SMC complement of proteins to γ -ATP-Sepharose. Using this affinity purification, we successfully obtained amino acid sequence for the previously unidentified MYPT1-kinase (MacDonald *et al.*, 2001a). Further experiments, including addition of recombinant enzyme to permeabilized SMC, confirmed a role for the MYPT1 kinase as a regulator of SMC contractile state (MacDonald *et al.*, 2001a,b; Borman *et al.*, 2002).

We currently are using a functional proteomics approach to identify novel protein kinase substrates that are phosphorylated in response to agonist-induced smooth muscle relaxation. In our approach, permeabilized smooth muscle strips, contracted by treatment with calcium, are treated with 8-bromo-cGMP to induce relaxation. Cell extracts are prepared and proteins are resolved by 2-DGE. Proteins that undergo changes in phosphorylation in response to 8Br-cAMP are sequenced and identified by MS. Using this method, several known phosphoproteins — including HSP20, HSP27, and telokin (MacDonald *et al.*, 2000; Walker *et al.*, 2001; Borman *et al.*, 2002) — were identified. In addition to these known proteins, we have identified novel phosphoproteins. Their role in smooth muscle relaxation is being investigated (J.A. MacDonald and T.A. Haystead, unpublished results).

C. PROTEOME MINING

Proteome mining is an iterative process in which a targeted proteome is screened against compound libraries to identify highly selective drug-protein interactions (Graves and Haystead, 2002). The advantage of this approach is

the ability to identify a drug-like molecule from a compound library that is specific for a protein or group of proteins and simultaneously identify the protein targets of the drug. The process begins with the isolation of a specific subproteome from a cell, tissue, or animal source by application of saturating amounts of tissue extract to an affinity matrix (Figure 7). The affinity matrix is composed of a natural ligand immobilized in an orientation that favors interaction with its respective protein targets. For example, if the natural ligand is ATP, only those proteins that bind ATP or ATP-resembling molecules will be captured on the matrix. This can be validated by sequencing and identifying proteins that are specifically absorbed to the matrix. The captured proteome then is screened against large chemical libraries and, if a compound from the library is able to compete with the immobilized ligand for protein binding, the bound protein is displaced from the matrix and recovered. Since the drug has the potential to interact with all of the proteins bound to the matrix, information about drug specificity can be obtained by identifying the eluted proteins.

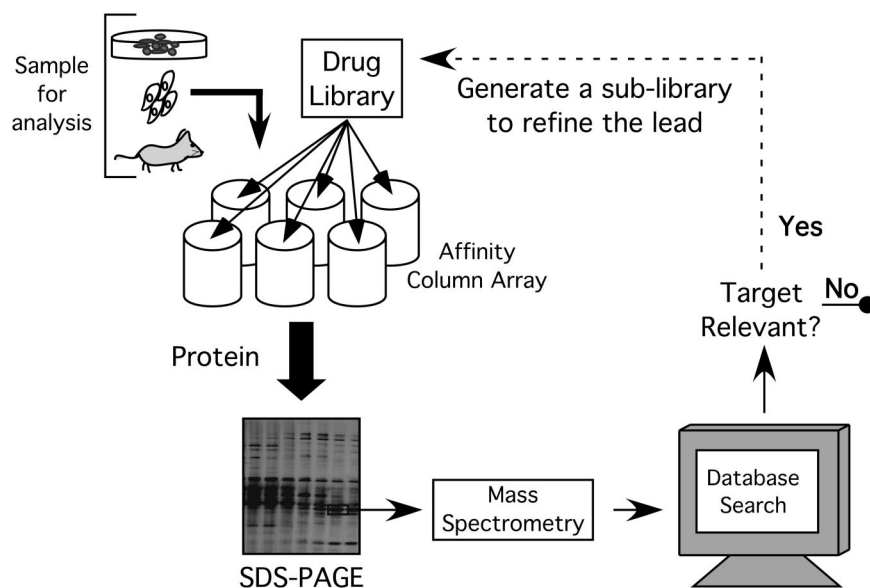


FIG. 7. Proteome mining is a method to screen combinatorial compound libraries for protein targets and gain information about compound specificity. In this strategy, proteins from a cell line, tissue, or organism are isolated on identical affinity column arrays composed of natural ligands, drugs, or inhibitors designed for the capture of specific types of proteins. After washing to remove nonspecific proteins, the arrays are exposed to a compound library and proteins that are eluted are resolved and identified by MS. If the protein is relevant, a sublibrary can be applied to the bound proteome to further define the lead molecule.

We recently applied proteome mining to investigate the protein targets of antimalarial drugs. Malaria is responsible for an estimated 1–3 million deaths annually. Because parasites have evolved resistance to many common anti-malarial drugs, there is a great demand for development of novel therapeutics (Foley and Tilley, 1998). In our approach, we captured the purine-binding proteome of human red blood cells, *Plasmodium falciparum*, on γ -ATP-Sepharose and screened the bound proteomes against several of the quinoline antimalarial drugs (e.g., chloroquine, primaquine). Interestingly, several quinoline antimalarial drugs eluted two proteins from the human red blood cell purine-binding proteome identified as aldehyde dehydrogenase and quinone reductase 2 (Graves *et al.*, 2001). We are investigating the role of these proteins in the pathology of malaria. Thus, proteome mining can allow the identification of drug targets and further investigation of those targets can reveal the molecular action of the drugs *in vivo*.

ACKNOWLEDGMENTS

We thank Elizabeth Herrick for figure making and design and Meredith Borman and Justin MacDonald for critical review of the manuscript.

REFERENCES

- Aebersold R, Watts JD, Morrison HD, Bures EJ 1991 Determination of the site of tyrosine phosphorylation at the low picomole level by automated solid-phase sequence analysis. *Anal Biochem* 199:51–60
- Alms GR, Sanz P, Carlson M, Haystead TA 1999 Reg1p targets protein phosphatase 1 to dephosphorylate hexokinase II in *Saccharomyces cerevisiae*: characterizing the effects of a phosphatase subunit on the yeast proteome. *EMBO J* 18:4157–4168
- Anderson L, Seilhamer J 1997 A comparison of selected mRNA and protein abundances in human liver. *Electrophoresis* 18:533–537
- Andersson L, Porath J 1986 Isolation of phosphoproteins by immobilized metal (Fe³⁺) affinity chromatography. *Anal Biochem* 154:250–254
- Borman MA, MacDonald JA, Muranyi A, Hartshorne DJ, Haystead TAJ 2002 Smooth muscle myosin phosphatase-associated kinase induces Ca²⁺-sensitization via myosin phosphatase inhibition. *J Biol Chem* 277:23441–23446
- Boyle WJ, van der Geer P, Hunter T 1991 Phosphopeptide mapping and phosphoamino acid analysis by two-dimensional separation on thin-layer cellulose plates. *Meth Enzymol* 201: 110–149
- Burlingame AL, Boyd RK, Gaskell SJ 1998 Mass spectrometry. *Anal Chem* 70:647R–716R
- Cao P, Stults JT 2000 Mapping the phosphorylation sites of proteins using on-line immobilized metal affinity chromatography/capillary electrophoresis/electrospray ionization multiple stage tandem mass spectrometry. *Rapid Commun Mass Spectrom* 14:1600–1606
- Eng JK, McCormack AL, Yates JR 1994 An approach to correlate tandem mass-spectral data of peptides with amino-acid-sequences in a protein database. *J Am Soc Mass Spectrom* 5:976–989

- Fadden P, Haystead TA** 1995 Quantitative and selective fluorophore labeling of phosphoserine on peptides and proteins: characterization at the attomole level by capillary electrophoresis and laser-induced fluorescence. *Anal Biochem* 225:81–88
- Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM** 1989 Electrospray ionization for mass spectrometry of large biomolecules. *Science* 246:64–71
- Ficarro SB, McClelland ML, Stukenberg PT, Burke DJ, Ross MM, Shabanowitz J, Hunt DF, White FM** 2002 Phosphoproteome analysis by mass spectrometry and its application to *Saccharomyces cerevisiae*. *Nat Biotechnol* 20:301–305
- Foley M, Tilley L** 1998 Quinoline antimalarials: mechanisms of action and resistance and prospects for new agents. *Pharmacol Ther* 79:55–87
- Gavin AC, et al** 2002 Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* 415:141–147
- Graves PR, Haystead TA** 2002 Molecular biologist's guide to proteomics. *Microbiol Mol Biol Rev* 66:39–63; table of contents
- Graves PR, Kwiek J, Fadden P, Ray R, Hardeman K, Coley AM, Foley M, Haystead TAJ** 2001 Discovery of novel targets of quinoline drugs in the human purine binding proteome. *Mol Pharmacol* 62:39
- Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R** 1999a Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* 17:994–999
- Gygi SP, Rochon Y, Franza BR, Aebersold R** 1999b Correlation between protein and mRNA abundance in yeast. *Mol Cell Biol* 19:1720–1730
- Haystead CM, Gregory P, Sturgill TW, Haystead TA** 1993 Gamma-phosphate-linked ATP-sepharose for the affinity purification of protein kinases. Rapid purification to homogeneity of skeletal muscle mitogen-activated protein kinase kinase. *Eur J Biochem* 214:459–467
- Ho Y, et al** 2002 Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* 415:180–183
- Hubbard MJ, Cohen P** 1993 On target with a new mechanism for the regulation of protein phosphorylation. *Trends Biochem Sci* 18:172–177
- Hunt DF, Henderson RA, Shabanowitz J, Sakaguchi K, Michel H, Sevilir N, Cox AL, Appella E, Engelhard VH** 1992 Characterization of peptides bound to the class I MHC molecule HLA-A2.1 by mass spectrometry. *Science* 255:1261–1263
- Ideker T, Thorsson V, Ranish JA, Christmas R, Buhler J, Eng JK, Bumgarner R, Goodlett DR, Aebersold R, Hood L** 2001 Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* 292:929–934
- Karas M, Hillenkamp F** 1988 Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal Chem* 60:2299–2301
- Kindy JM, Taraszka JA, Regnier FE, Clemmer DE** 2002 Quantifying peptides in isotopically labeled protease digests by ion mobility/time-of-flight mass spectrometry. *Anal Chem* 74:950–958
- Knighton DR, Zheng JH, Ten Eyck LF, Xuong NH, Taylor SS, Sowadski JM** 1991 Structure of a peptide inhibitor bound to the catalytic subunit of cyclic adenosine monophosphate-dependent protein kinase. *Science* 253:414–420
- Lewis TS, Hunt JB, Aveline LD, Jonscher KR, Louie DF, Yeh JM, Nahreini TS, Resing KA, Ahn NG** 2000 Identification of novel MAP kinase pathway signaling targets by functional proteomics and mass spectrometry. *Mol Cell* 6:1343–1354
- Link AJ, Eng J, Schieltz DM, Carmack E, Mize GJ, Morris DR, Garvik BM, Yates JR III** 1999 Direct analysis of protein complexes using mass spectrometry. *Nat Biotechnol* 17:676–682
- Loboda AV, Krutchinsky AN, Bromirski M, Ens W, Standing KG** 2000 A tandem quadrupole/time-of-flight mass spectrometer with a matrix-assisted laser desorption/ionization source: design and performance. *Rapid Commun Mass Spectrom* 14:1047–1057

- MacDonald JA, Borman MA, Muranyi A, Somlyo AV, Hartshorne DJ, Haystead TA** 2001a Identification of the endogenous smooth muscle myosin phosphatase-associated kinase. *Proc Natl Acad Sci USA* 98:2419–2424
- MacDonald JA, Eto M, Borman MA, Brautigan DL, Haystead TA** 2001b Dual Ser and Thr phosphorylation of CPI-17, an inhibitor of myosin phosphatase, by MYPT-associated kinase. *FEBS Lett* 493:91–94
- MacDonald JA, Mackey AJ, Pearson WR, Haystead TAJ** 2002 A strategy for the rapid identification of phosphorylation sites in the phospho-proteome. *Mol Cell Proteomics* 1:314–322
- MacDonald JA, Walker LA, Nakamoto RK, Gorenne I, Somlyo AV, Somlyo AP, Haystead TA** 2000 Phosphorylation of telokin by cyclic nucleotide kinases and the identification of *in vivo* phosphorylation sites in smooth muscle. *FEBS Lett* 479:83–88
- Mackey AJ, Haystead TAJ, Pearson WR** 2002 Getting more from less: algorithms for rapid protein identification with multiple short peptide sequences. *Mol Cell Proteomics* 1:139–147
- Mann M, Wilm M** 1994 Error-tolerant identification of peptides in sequence databases by peptide sequence tags. *Anal Chem* 66:4390–4399
- Medzhradszky KF, Campbell JM, Baldwin MA, Falick AM, Juhasz P, Vestal ML, Burlingame AL** 2000 The characteristics of peptide collision-induced dissociation using a high-performance MALDI-TOF/TOF tandem mass spectrometer. *Anal Chem* 72:552–558
- Meyer HE, Hoffmann-Posorske E, Heilmeyer LM Jr** 1991 Determination and location of phosphoserine in proteins and peptides by conversion to S-ethylcysteine. *Meth Enzymol* 201:169–185
- Neubauer G, Mann M** 1999 Mapping of phosphorylation sites of gel-isolated proteins by nano-electrospray tandem mass spectrometry: potentials and limitations. *Anal Chem* 71:235–242
- Oda Y, Nagasu T, Chait BT** 2001 Enrichment analysis of phosphorylated proteins as a tool for probing the phosphoproteome. *Nat Biotechnol* 19:379–382
- Opiteck GJ, Lewis KC, Jorgenson JW, Anderegg RJ** 1997 Comprehensive on-line LC/LC/MS of proteins. *Anal Chem* 69:1518–1524
- Pandey A, Mann M** 2000 Proteomics to study genes and genomes. *Nature* 405:837–846
- Posewitz MC, Tempst P** 1999 Immobilized gallium(III) affinity chromatography of phosphopeptides. *Anal Chem* 71:2883–2892
- Qin J, Fenyó D, Zhao Y, Hall WW, Chao DM, Wilson CJ, Young RA, Chait BT** 1997 A strategy for rapid, high-confidence protein identification. *Anal Chem* 69:3995–4001
- Schena M, Shalon D, Davis RW, Brown PO** 1995 Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270:467–470
- Sellers JR, Adelstein RS** 1985 The mechanism of regulation of smooth muscle myosin by phosphorylation. *Curr Topic Cell Regul* 27:51–62
- Shalon D, Smith SJ, Brown PO** 1996 A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. *Genome Res* 6:639–645
- Shevchenko A, Loboda A, Ens W, Standing KG** 2000 MALDI quadrupole time-of-flight mass spectrometry: a powerful tool for proteomic research. *Anal Chem* 72:2132–2141
- Stensballe A, Andersen S, Jensen ON** 2001 Characterization of phosphoproteins from electrophoretic gels by nanoscale Fe(III) affinity chromatography with off-line mass spectrometry analysis. *Proteomics* 1:207–222
- Velculescu VE, Zhang L, Vogelstein B, Kinzler KW** 1995 Serial analysis of gene expression. *Science* 270:484–487
- Walker LA, MacDonald JA, Liu X, Nakamoto RK, Haystead TA, Somlyo AV, Somlyo AP** 2001 Site-specific phosphorylation and point mutations of telokin modulate its Ca²⁺-desensitizing effect in smooth muscle. *J Biol Chem* 276:24519–24524

- Wettenhall RE, Aebersold RH, Hood LE** 1991 Solid-phase sequencing of ^{32}P -labeled phosphopeptides at picomole and subpicomole levels. *Meth Enzymol* 201:186–199
- Wilm M, Mann M** 1996 Analytical properties of the nanoelectrospray ion source. *Anal Chem* 68:1–8
- Wilm M, Shevchenko A, Houthaeve T, Breit S, Schweigerer L, Fotsis T, Mann M** 1996 Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry. *Nature* 379:466–469
- Yates JR III** 1998 Mass spectrometry and the age of the proteome. *J Mass Spectrom* 33:1–19
- Zhou H, Watts JD, Aebersold R** 2001 A systematic approach to the analysis of protein phosphorylation. *Nat Biotechnol* 19:375–378