

# Proteomics methods for probing molecular mechanisms in signal transduction<sup>1</sup>

L. G. Sheffield\* and J. J. Gavinski†

\*Department of Dairy Science, University of Wisconsin, Madison 53706, and

†Endocrinology-Reproductive Physiology Program, University of Wisconsin, Madison 53706

**ABSTRACT:** mRNA splicing and various posttranslational modifications to proteins result in a larger number of proteins than genes. Assessing the dynamic nature of this proteome is the challenge of modern proteomics. Recent advances in high throughput methods greatly facilitate the analysis of proteins involved in signal transduction, their production, posttranslational modifications and interactions. Highly reproducible two dimensional polyacrylamide gel electrophoresis (2D-PAGE) methods, coupled with matrix assisted laser desorption-time of flight-mass spectrometry (MALDI-TOF-MS) allow rapid separation and identification of

proteins. These methods, alone or in conjunction with other techniques such as immunoprecipitation, allow identification of various critical posttranslational modifications, such as phosphorylation. High throughput identification of important protein-protein interactions is accomplished by yeast two hybrid approaches. In vitro and in vivo pulldown assays, coupled with MALDI-TOF-MS, provide an important alternative to two hybrid approaches. Emerging advances in production of protein-based arrays promise to further increase throughput of proteomics-based approaches to signal transduction.

Key Words: Proteomics, Signal Transduction, Protein, Mammary Gland

©2003 American Society of Animal Science. All rights reserved.

J. Anim. Sci. 2003. 81(Suppl. 3):48–57

## Introduction

In the broadest sense, signal transduction can be considered the mechanisms by which a cell responds to its environment. Most often, signal transduction reflects responses to hormones, growth factors, neurotransmitters, and other such molecules. However, similar pathways also mediate responses to extracellular matrix, inorganic ions and various other factors. Often, distinctions are made among endocrine, neuronal, paracrine, autocrine, and juxtacrine signaling. Such distinctions are useful in the context of organ or animal physiology, but the signaling pathways activated are often very similar.

Activation of a signal transduction pathway requires a sensing molecule, such as a hormone receptor. Activation of this receptor elicits one or more of a wide variety of pathways that lead to alteration of the transcription/translation process, alteration of the structure of ex-

isting proteins or both. These pathways are activated in the context of a specific genome. Obviously, transcription of a particular set of genes requires the presence of appropriate genes and regulatory elements. Additionally, genotype can affect nontranscriptional pathways by its impact on the specific signaling molecules present.

Although genomics has proven to be a powerful field, and the sequencing of several important organisms, as well as the amount of information available in others is impressive and useful, genome organization alone is insufficient to explain physiological processes. Critical questions emerging in the postgenome era involve how genetic information is expressed as a phenotype and how that expression is altered in response to environment. Emerging methods place considerable emphasis on high throughput techniques. DNA arrays, capable of simultaneous analysis of the expression of thousands of genes, are the current archetype of high throughput analysis in biology (Cheung et al., 1999). A frequent use of these arrays is to assess the cell or tissue content of several thousand mRNA simultaneously (Duggan et al., 1999). As useful as this information is, it still provides an incomplete picture of cell biology (as described below). More complete understanding of how genotype and its interactions with the environment result in the expressed phenotype requires additional information

Received July 8, 2002.

Accepted April 7, 2003.

Corresponding author: L. G. Sheffield; e-mail: lgsheffi@facstaff.wisc.edu.

<sup>1</sup>Supported by University of Wisconsin College of Agriculture and Life Sciences and USDA Project WIS 3108 and WIS 4665.

on protein production and posttranslational modifications. Unfortunately, protein-based methods cannot yet achieve the throughput of DNA-based techniques. However, rapid progress in proteomics promises to greatly improve the throughput of protein-based analytical procedures.

The objective of the present review is to describe some of the issues and techniques involved in improving our understanding of the broad field of signal transduction, with an emphasis on protein-based approaches. Toward this end, this review will focus on proteomics issues, including identification of specific proteins, posttranslational modification and protein-protein interaction, including examples from mammary gland biology.

### Proteome Concept

If the genome is defined as the collection of all genes in an organism, one could define the collection of all expressed genes (at the mRNA level) and their translated proteins as the transcriptome. Furthermore, the proteome can be defined as all proteins present in a cell or tissue at any given time. This includes all posttranslational modifications that occur. Thus, unlike the genome that is stable, the transcriptome and proteome are highly dynamic. The genome of mammals is generally considered to consist of about 30,000 genes. Surprisingly, this number is only about twofold greater than that found in what are generally considered less complex organisms, including insects and the nematode *Caenorhabditis elegans* (Rubin et al., 2000). However, mammals generally produce a considerably more complex array of proteins from their genome. This additional complexity in the mammalian proteome arises from a number of sources, including alternative splicing of mRNA, multiple pathways of proteolytic processing of proteins and various other posttranslational modifications, such as glycosylation, perenylation, and phosphorylation. Furthermore, subcellular localization of a protein can affect its function (Petsko, 2001). Although these mechanisms exist in all eucaryotes, they are employed to a greater extent in some phyla, such as mammals, giving rise to more complex proteomes.

#### *mRNA Splice Variants*

The presence of multiple mRNA species corresponding to a single transcribed region of DNA is well known and exist for 40 to 60% of expressed genes (Modrek and Lee, 2002). In some cases, this represents the presence of multiple states of processing (mature mRNA vs. unspliced pre-mRNA). However, in many cases, multiple splice variants exist. In the case of prolactin receptor, all species examined to date have multiple isoforms of the prolactin receptor (Bole-Feysot et al., 1998; Horseman, 2002). These prolactin receptor isoforms are often termed long and short isoforms, although the so-called short isoform is a variety of specific isoforms. In addition, an intermediate isoform has been identified in rat

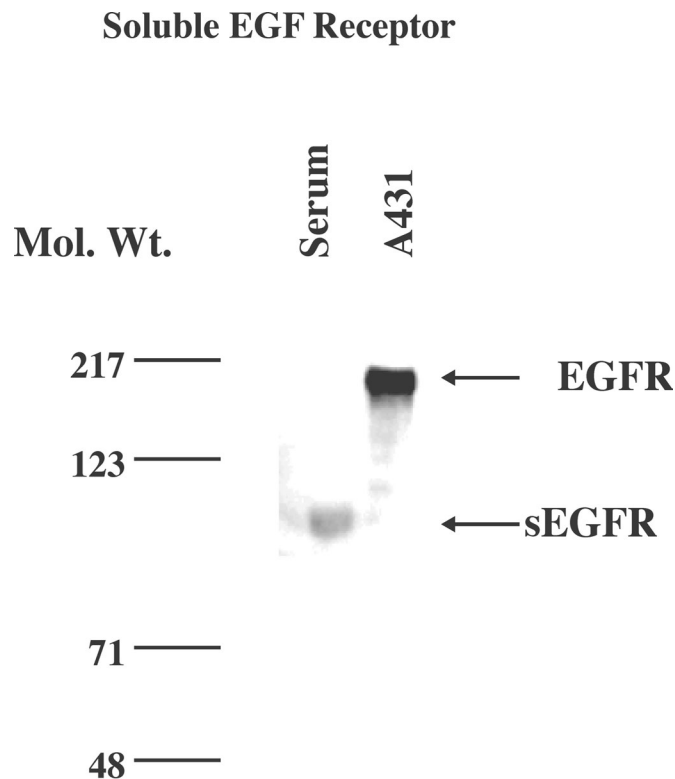
and human. The long isoform of the receptor is capable of both mitogenic and lactogenic signaling, whereas the short isoforms are generally considered incapable of supporting these functions (Chang and Clevenger et al., 1996; Chang et al., 1998). The rat intermediate isoform of prolactin receptor has been reported to be mitogenic and activate JAK2, while a human intermediate isoform has minimal mitogenic activity, although it activates JAK2 (Kline et al., 1999). Short isoforms of the prolactin receptor have been shown to inhibit lactogenic signaling by long isoforms, probably because of the formation of inactive heterodimers of long and short isoform receptors (Berlanga et al., 1997). However, in the rat corpus luteum (in which prolactin is luteotropic), short isoforms of prolactin receptor appear to be involved in activating specific pathways, including binding of a specific prolactin receptor associated protein (Duan et al., 1997).

In addition to transmembrane isoforms of prolactin, a splice variant that gives rise to a soluble form of the prolactin receptor lacking transmembrane and intracellular domains has recently been identified (Trott et al., 2003). Similar findings have also been observed for growth hormone receptor in rodents (Postel-Vinay and Finidori, 1995). Interestingly, while the soluble forms of growth hormone receptor in rodents arises from alternative splicing, a similar protein in humans appears to arise from proteolytic processing of the growth hormone receptor (Postel-Vinay and Finidori, 1995).

Several splice variants of epidermal growth factor (EGF) receptor have also been reported (Reiter and Maihle, 1996; Reiter et al., 2001). A soluble 110-kDa EGF receptor, consisting of the extracellular domain, is present in human serum, appears to be regulated by reproductive hormones (Baron et al., 2001) and is elevated in certain cancers, most notably ovarian (Baron et al., 1999). Recently, we have also detected a similar protein in bovine serum (Figure 1).

At least three splice variants of transforming growth factor alpha (TGF $\alpha$ ), an EGF receptor agonist, are known to exist. Although they interact with members of the ErbB family, they do so with varying specificities. The so-called wild-type variant coimmunoprecipitated with ErbB4, whereas variants I and II coimmunoprecipitated with ErbB2. All of these variants activated ErbB2, but the biological consequences differed, with Variants I and II, but not wild type, producing autonomous growth (Xu et al., 2000).

As a final example of splice variants, Yu et al. (1992) observed a splice variant of acidic fibroblast growth factor. This variant produces a frameshift mutation, resulting in a different C-terminal sequence. Unlike acidic fibroblast growth factor, which is mitogenic to a number of cells, including fibroblasts, this variant appears to be an acidic fibroblast growth factor antagonist. Unfortunately, its physiological significance is unclear.



**Figure 1.** Soluble epidermal growth factor (EGF) receptor (sEGFR) in bovine serum. Bovine serum or A431 cell membranes (a human epidermoid carcinoma that expresses high levels of EGF receptor and is used as a positive control) were separated by SDS-PAGE (10% gel), transferred to polyvinylidene fluoride (PVDF) membranes and probed with anti-EGF receptor as previously described (Hendrix et al., 1996). Note the presence of a 110-kDa protein immunologically reactive with anti-EGF receptor in bovine serum.

### Protease Processing

Signal peptides are generally removed from proteins during processing. The identity of consensus signal peptides is known, and can be readily predicted from DNA sequence information (Baldi et al., 2000). Furthermore, protein sequence is also capable of predicting a number of other processing features, including likely subcellular localization (Emanuelsson and von Heijne, 2001). However, other processing events are more difficult to predict, and many proteins are processed by multiple pathways, giving rise to multiple processing variants.

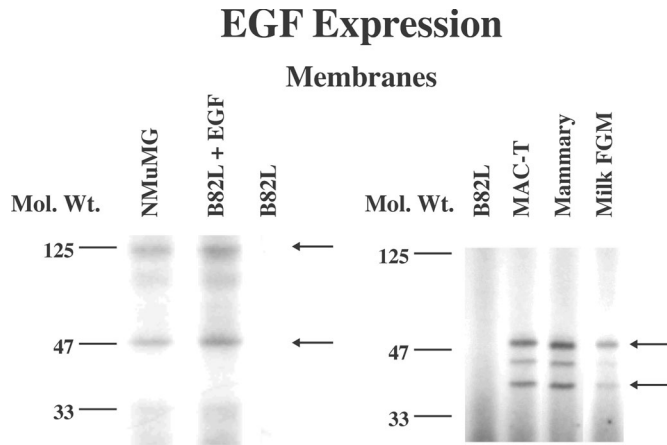
Transforming growth factor alpha is one important regulator of mammary development that is processed by multiple pathways. Transforming growth factor alpha is produced as a 160-amino acid transmembrane precursor. In most cases, this precursor is processed to a 50-amino acid mature peptide that is biologically active. However, transmembrane forms of TGF $\alpha$  have been identified (Luettiki and Lee, 1990). Juxtacrine actions of these forms of TGF $\alpha$  have been suggested (Bush et al., 1998; Xiao and Majumder, 2001). In addition,

dimerization of transmembrane TGF $\alpha$  molecules with anti-TGF $\alpha$  induces protein phosphorylation reminiscent of receptor activation (Shum et al., 1994). Because the intracellular domain of the transmembrane form of TGF $\alpha$  does not appear to contain kinase activity, this is most likely due to association of transmembrane TGF $\alpha$  with other signaling molecules. This appears to require palmitoylation of Cys 153 and 154 in the intracellular domain of transmembrane TGF $\alpha$  (Shum et al., 1996). Recently, Shi et al. (2000) determined that CD9 physically associates with transmembrane forms of TGF $\alpha$ , while Kuo et al. (Kuo et al., 2000) found that the Golgi associated protein GRASP55 was also associated with transmembrane TGF $\alpha$ . These results indicate that transmembrane forms of growth factors may have functions beyond serving as juxtacrine ligands, although this possibility has received little attention.

Epidermal growth factor is also produced as a large (>1200 amino acid) glycoprotein precursor that contains a transmembrane domain and a short intracellular domain (Mroczkowski and Carpenter, 1988). In some tissues, preproEGF is almost completely processed to mature peptides, including the 53-amino acid mature EGF. However, in others, such as the kidney, very little is processed in this way, and most remains as a transmembrane protein. In the mammary gland, some EGF does appear to be processed and secreted into milk (Grosvenor et al., 1993), although a substantial amount of EGF remains as partly processed preEGF (Mroczkowski and Reich, 1993). Recently, we have observed that bovine mammary epithelial cells also produce pre-EGF like molecules, although they do not appear to produce mature EGF in detectable amounts (Figure 2). The biological role of these forms of EGF remains unclear. Mroczkowski et al. (1989) suggested that they might give rise to juxtacrine signaling, by activating EGF receptors in adjacent cells. Maheshwari et al. (2001) has also observed that expression of EGF in transmembrane forms results in directional migration of cells, whereas soluble EGF stimulates random migration.

### Phosphorylation

A major modification important in signal transduction is protein phosphorylation. This can arise from receptor kinases, receptor-associated kinases or kinases downstream of the initial signal. In addition, cells contain a diverse array of phosphatases that dephosphorylate proteins. Although one often thinks of hormone action as stimulating protein phosphorylation and phosphatase activity reversing those actions, this is not always the case. For example, the kinase src is actually inhibited by a specific tyrosine phosphorylation (Courtneidge et al., 1993). Whereas the details of kinase regulation are beyond the scope of this review, the ability to detect this modification is critical to studying signal transduction.



**Figure 2.** Forms of epidermal growth factor (EGF) in mammary tissue. Left: Cell membranes from mouse mammary epithelial cell line (NMuMG), mouse fibroblasts lacking EGF expression (B82L), or B82L cells transfected with murine EGF were separated by SDS-PAGE on a 10% gel, transferred to polyvinylidene fluoride (PVDF) membranes and probed with anti-EGF essentially as previously described (Hendrix et al., 1996), except that anti-EGF was used instead of other antibodies. Right: Cell membranes were isolated from B82L cells (negative control), a cultured bovine mammary epithelial cell line (MAC-T), lactating bovine mammary tissue (Mammary) or milk fat globule membranes. Proteins were separated by SDS-PAGE on a 10% gel, transferred to PVDF membranes and probed with anti-EGF essentially as previously described (Hendrix et al., 1996), except that anti-EGF was used instead of other antibodies. Note the presence of two major partly processed forms of EGF (arrows) and other less intense bands. B82L cells are a negative control, lacking EGF.

## Studying the Proteome

### *Protein Separation and Identification*

A major advance in the study of protein expression has been in two-dimensional PAGE. Initial methods (O'Farrell, 1975) involved initial separation by isoelectric focusing (usually in a small tube gel) followed by transferring this gel to an SDS-PAGE gel for the second dimension separation. The initial isoelectric focusing gel was usually generated in situ by the use of carrier ampholytes added to samples, gels and/or running buffers. Such gels provided high resolution, but were difficult to reproduce, in large part because of inconsistencies in reproducing the pH gradient of the first dimension gel. More recently, the availability of immobilized pH gradients has greatly facilitated two-dimensional PAGE (Righetti et al., 1983; Fichmann, 1999; Gianazza, 1999; Matsui et al., 1999; Gorg et al., 2000). Immobilized pH gradient gels of various lengths and pH ranges are available from a number of commercial suppliers. Further advances in computer assisted image analysis

methods has improved the ability to accurately compare specific spot intensities.

For two-dimensional PAGE to be particularly useful, methods of identifying specific proteins are necessary. Some clues may be obtained from molecular weight and isoelectric point estimations, but identification by these methods is not particularly reliable. If antibodies to a suspected protein are available, immunological detection (Western blot) is useful. However, this method is not particularly useful for the identification of a large number of proteins or for novel proteins.

Historically, protein sequencing, either of the N-terminus or of tryptic fragments, is a useful method of protein identification. Even without a full sequence, fragments can often be matched to a specific DNA sequence for identification. More recently, developments in mass spectrometry provide a rapid method of protein identification. Entire proteins are often not particularly useful for mass spectrometry, but protease digestion products (such as tryptic fragments) are readily separated by modern methods. In particular, matrix assisted laser desorption time of flight mass spectrometry (MALDI-TOF-MS) techniques are emerging as the method of choice for identifying tryptic fragments (Hamdan et al., 2001; Leushner, 2001; Mann et al., 2001; Zaluzeć et al., 1995).

Modern mass spectrometry instruments have a resolution well below 1 atomic mass unit. This allows ready matching of a particular tryptic fragment to a specific amino acid composition, since only a single combination of amino acids will provide an appropriate match in almost all cases. However, this does not generally provide an unambiguous determination of the specific sequence of amino acids, since any sequence combination of the same amino acids match the molecular weight estimates. Therefore, automated search routines that compare the tryptic map with that of other proteins are necessary. Importantly, these maps can be theoretical maps, based on DNA sequence alone.

In addition to appropriate algorithms for analyzing mass spectra, the quality of the underlying database is also critical to the success of proteomics efforts (Stupka, 2002). In the case of researchers using human, mouse, or other widely used models, very complete and high-quality databases are available and readily accessible. In other cases, such as most livestock and many less studied species, database completeness can be a major limitation. In the case of genomics or gene expression profiling, identifying a sequence with 80 or 90% identity to a human sequence is a reasonable prospect. In the case of tryptic mapping by mass spectrometry, this is likely to prove difficult or impossible.

Statistical approaches allow calculation of the probability that a particular match of a peptide map with a protein in a database is due to chance (Berndt et al., 1999; Eriksson et al., 2000; Eriksson and Fenyo, 2002). However, complete match is rarely obtained. In some cases, predicted fragments are absent from the mass spectra, whereas in others, some mass spectra peaks

are not predicted for the particular protein. Often, these extraneous peaks are caused by known contaminants, such as trypsin, or to acrylamide contaminants. Regardless, verification by other methods, such as microsequencing or immunological approaches is desirable for proteins identified for further study.

Previously, Beaton et al. (1997) used a proteomics approach to identify proteins, including glucose-regulated protein 78 and protein disulphide isomerase, associated with lactation and prolactin action in mammary glands. In other studies, proteomics approaches have been used to map the proteome of the endoplasmic reticulum and milk fat globule membrane, and to examine the changes in the golgi associated with transition from a basal state to maximum secretion (Wu et al., 2000a, 2000b).

As an additional example of the applications of proteomics methods to mammary gland biology, we have investigated the possibility that the previously described soluble variants of EGF receptor induce signaling by associating with transmembrane forms of EGF. Such reverse signaling has been demonstrated for transmembrane forms of TGF $\alpha$  (using antibodies, rather than soluble receptors, to induce signaling by the transmembrane growth factor) (Shum et al., 1994). For these studies, the extracellular domain of mouse EGF receptor was cloned into pIZ/V5-His in frame with a polyhistidine tag. Cabbage looper embryo cells (High-Five, Invitrogen, Carlsbad, CA) were transfected, and stable lines were generated that were resistant to Zeocin. Soluble EGF receptor was found to be secreted into culture media in these cells. The soluble EGF receptor was purified from culture media by wheat germ agglutinin chromatography (Quijano and Sheffield, 1998) followed by Ni-affinity chromatography (Lilius et al., 1991). Mammary epithelial cells previously shown to express transmembrane forms of EGF (NMuMG line, American Type Culture Collection, Rockville, MD) were treated with soluble EGF receptor for various times, proteins were extracted with 8 M urea, 4% (3-[(3-cholamidopropyl)dimethylamino]-1-propane sulfonate, 40 mM Tris base, 1 mM phenylmethanesulfonyl fluoride, 100  $\mu$ g of aprotinin/ml and 100  $\mu$ g of leupeptin/ml. Proteins were separated by two-dimensional PAGE using immobilized pH gradient strips for the first dimension and a 10% PAGE gel for the second (Amersham Pharmacia Biotech, Piscataway, NJ). After 24 h of treatment with EGF receptors, substantial differences were apparent in proteins resolved by two-dimensional PAGE (Figure 3). Protein spots were chosen for subsequent identification by MALDI-TOF-MS based on their change in spot intensity and resolution in the area of the spot. Among the proteins increased are cytokeratin 8, phospholipase C alpha, and tubulin  $\alpha$ 6. Cytokeratin 8 resolved as three independent spots, probably because of different phosphorylation states of the protein.

An important consideration in interpreting these results, as in other analyses of protein content by two-dimensional PAGE is the mechanism by which a de-

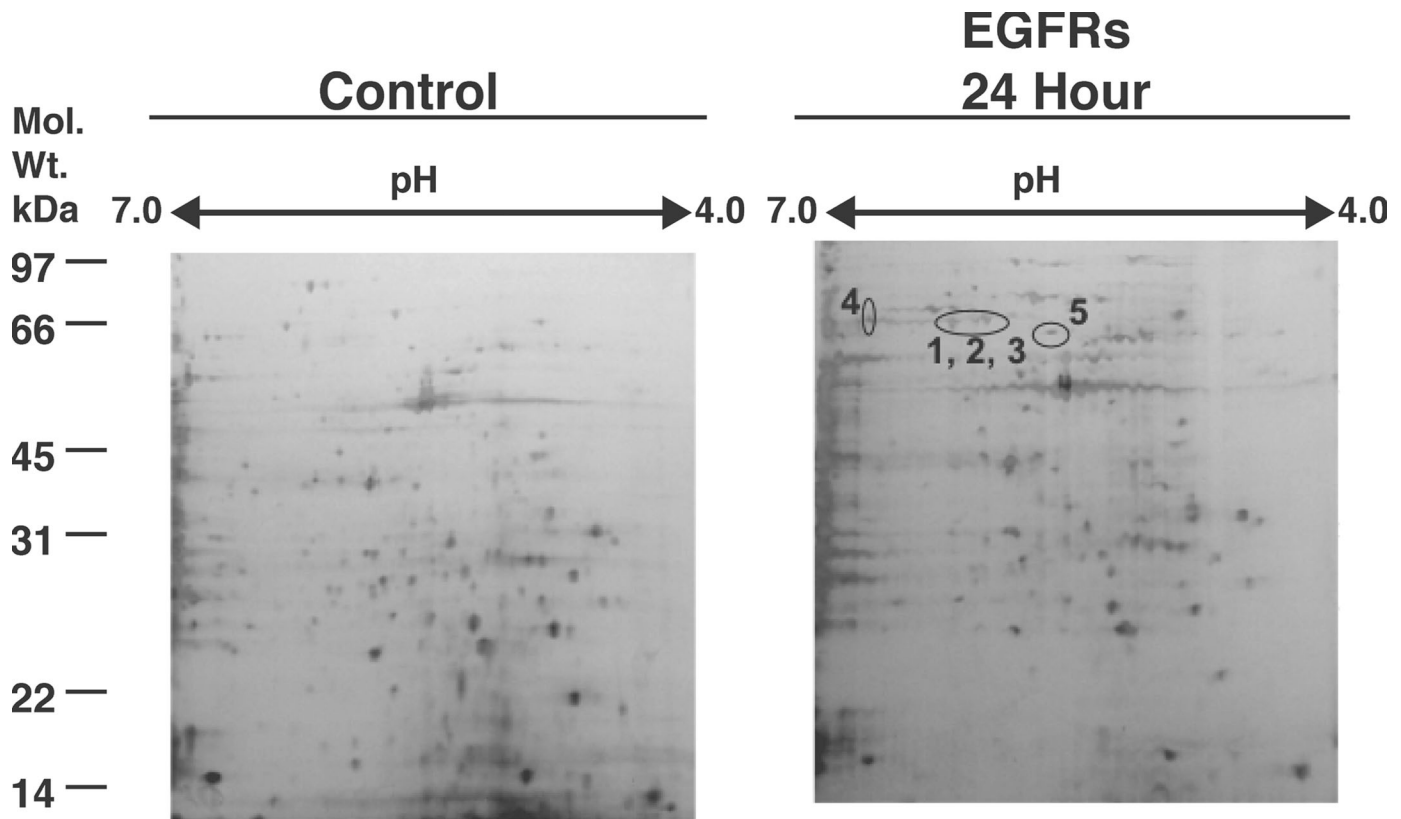
crease in spot intensity is achieved. Several possibilities exist in addition to the most obvious alteration in protein production. Protein stability could be altered. Alternatively, the solubility of the protein in the lysis buffer used could be altered. This could be due to either posttranslational modifications to the protein itself or alterations in its association with other cell components, such as caveoli or the cytoskeleton (Prusheik et al., 1997; Schlegel and Lisanti, 2001; Schroeder et al., 2001). These alternatives are currently under investigation for the impacts of EGF receptors on cellular protein content.

### *Assessing Phosphorylation Status*

Classically, protein phosphorylation was detected by incubating cells with  $^{32}$ P and assessing incorporation into proteins. In addition, immunological approaches are also available, particularly for assessing tyrosine phosphorylation. Immunological approaches to determining phosphorylation of a specific protein formerly involved immunoprecipitation of the protein and probing with anti-phosphotyrosine or immunoprecipitation from  $^{32}$ P-labeled cells and autoradiography. More recently, antibodies that recognize specific phosphopeptides have been developed and used to differentiate phosphorylated and nonphosphorylated states of proteins. These have been widely used to assess tyrosine phosphorylation of a number of proteins, including mitogen-activated protein kinase, c-src, focal adhesion kinase, various members of the STAT (signal transducers and activators of transcription) family, and others (Kaufmann et al., 2001; Nagata et al., 2001). However, one should note that these antibodies recognize only a short peptide sequence. If that sequence is present in other proteins, this approach cannot distinguish between them. For example, the tyrosine phosphorylation of STAT5a and 5b are identical, and so STAT5a tyrosine phosphorylation cannot be resolved from STAT5b tyrosine phosphorylation by using phosphorylation state-specific antibodies alone (although prior immunoprecipitation with specific antibodies allows separation of STAT5a and STAT5b).

Another approach to protein phosphorylation is the use of mass spectra. Because the addition of phosphate to a peptide results in a distinctive shift in mass spectra, this can be used as a signal for the presence of phosphate in a protein (Wilkins et al., 1999). Often, the specific phosphorylation site can be mapped using this technique. However, the method is technically demanding, in that very high sensitivity or enrichment of samples is usually required (Sickmann and Meyer, 2001).

Recently, considerable interest has arisen on the use of protein array-based approaches to study protein phosphorylation. Conceptually, antibody arrays would be very useful in identifying proteins present in a cell extract as well as their post-translational modifications (Borrebaeck, 2000; Srinivas et al., 2001; Yarmush and



**Figure 3.** Two-dimensional PAGE analysis of soluble epidermal growth factor (EGF) receptor effects on mammary epithelial cells. NMuMG mouse mammary epithelial cells were treated as controls or for 24 h with 500 ng/ml of recombinant soluble EGF receptor (EGFRs). Proteins were extracted and separated by two dimensional PAGE on a 4.0–7.0 pH isoelectric focusing gel, followed by a 12% SDS-PAGE gel. A number of proteins spots were seen to increase and decrease upon treatment. 1, 2, 3 = cytokeratin 8, 4 = phospholipase C $\alpha$ , 5 = tubulin  $\alpha$ 6.

Jayaraman, 2002). Antibodies attached to a solid surface could be designed to capture a wide variety of proteins. Anti-phosphotyrosine, serine, or threonine could then be used to probe the array to determine phosphorylation status of specific proteins. Alternatively, a large number of phosphorylation state-specific antibodies could be arrayed. A major disadvantage of this approach is that production of antibodies is considerably more time consuming than production of DNA fragments. In addition, choosing DNA fragments that require similar hybridization conditions is relatively straightforward, particularly when compared to developing a large number of antibodies that have similar optimal conditions. As a result, few such arrays are currently available.

To investigate the possible signaling by soluble forms of EGF receptor, we also used standard phosphotyrosine western blots to determine whether EGF receptors induced changes in cellular phosphotyrosine. These studies determined that EGF receptors increased protein tyrosine phosphorylation in Triton X-100-insoluble proteins, but not in soluble proteins (Figure 4), suggesting an increased tyrosine phosphorylation of cytoskeletally associated proteins (Prusheik et al., 1997). Immunoprecipitation followed by SDS-PAGE and MALDI-MS has identified some of the tyrosine phos-

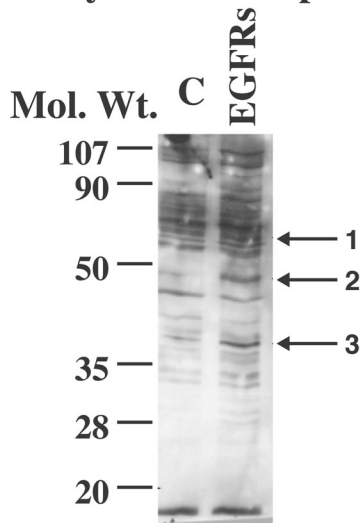
phorylated proteins. Interestingly, these appear to be members of the cytokeratin family.

Cytokeratins are tyrosine phosphorylated in responses to various cell signals, including EGF and the phosphotyrosine phosphatase inhibitor pervanadate (Aoyagi et al., 1985; Feng et al., 1999). Because phosphorylation changes are often associated with changes in molecular structure and function, and because cytokeratins are major components of intermediate filaments, these results suggest possible roles for growth factor signaling in modifying cytoskeleton and intermediate filament organization. Interestingly, such organizational changes have been associated with normal mammary development, differentiation, and tumorigenesis (Ben-Ze'ev, 1987; Taylor-Papadimitriou et al., 1992; Hendrix et al., 1996; Lelievre et al., 1996; Rudland et al., 1998).

#### *Intermolecular Interactions*

In addition to identifying proteins directly modified by a signaling pathway, identifying interactions among proteins is critical to understanding signal transduction (Legrain et al., 2001; Tucker et al., 2001). Currently, major efforts are under way to map all possible protein-

## Cytoskeleton Tyrosine Phosphorylation



**Figure 4.** Tyrosine phosphorylation in response to soluble epidermal growth factor (EGF) receptor. NMuMG mouse mammary epithelial cells were treated as controls or for 10 min with 500 ng/ml of recombinant soluble EGF receptor (EGFRs). Triton X-100 soluble proteins (cytosol) and insoluble (cytoskeleton) proteins were separated on a 10% SDS-PAGE gel, transferred to polyvinylidene fluoride (PVDF) membranes and probed with anti-phosphotyrosine as previously described (Hendrix et al., 1996). Tyrosine phosphorylation of several proteins was modified by treatment. 1 = cytokeratin 8, 2 = cytokeratin 18, 3 = cytokeratin 19.

protein interactions in simple models, such as *Saccharomyces cerevisiae*. In more complex mammalian systems, this is currently an intractable problem, but some specific subsets of the problem are being investigated. These efforts contain a number of major challenges, including not only the complexity of the mammalian proteome itself, but also the fact that intermolecular interactions are often highly dynamic and respond rapidly to changes in cellular environment (Auerbach et al., 2002; Figeys, 2002).

Classically, ligand binding methods, such as radioreceptor assays, were standard methods of determining protein interactions. These same methods could be used to assess interactions among nonreceptor proteins. Additionally, coimmunoprecipitation studies are commonly used to assess protein-protein interactions (for example, Johnson et al., 1996). Pull-down assays (in vitro and in vivo) are extensions of ligand binding methods. In these methods, a target is expressed in a cell (in vitro) or added to a cell lysate (in vitro), usually fused with a tag, such as glutathione S transferase (Smirnova et al., 2001) or polyhistidine (Lu et al., 1993). The glutathione S transferase tag is then immunoprecipitated, and associating proteins are identified by immunological methods, sequencing, or mass spectrometry.

A major advance in high-throughput analysis of protein-protein interactions is the Yeast two hybrid screen and its various relatives (Gietz and Woods, 2002). The basic principal of this assay is that a number of transcription factor DNA-binding domains are capable of activating transcription only when associated with a transcription activating domain. This activation domain do not need to be the same protein provided it is in appropriate proximity to the DNA binding domain. If the DNA binding domain and transcription activation domain are expressed separately but fused with proteins that interact, transcription can be activated via the interaction of the fusion proteins. In practice, one usually expresses a known protein (often called "bait") fused with a DNA binding domain. A cDNA library is then fused with the transcription activation domain, such that transcription of a reporter gene (often  $\beta$ -galactosidase or green fluorescent protein driven by an appropriate promoter) is activated if the known protein and the protein coded by a cDNA fragment interact. In this way, thousands of possible interacting proteins can be screened simultaneously, and only positive clones need be sequenced and identified.

Although useful, two hybrid approaches are subject to a number of artifacts, including relatively large numbers of false positives and false negatives. As a result, additional methods to verify a specific interaction are generally considered necessary. A large number of methods are available (Lakey and Raggett, 1998). In addition to coimmunoprecipitation, ligand binding and affinity purification methods discussed above, such methods as microcalorimetry or plasmon resonance spectrometry can provide information on binding kinetics and thermodynamics, although their use requires specialized equipment (Phipps and Mackin, 2000; Cooper, 2002).

Array approaches similar to DNA arrays have recently begun to emerge for identifying protein-protein interactions (Emili and Cagney, 2000; Figeys and Pinto, 2001; Li, 2000; Yarmush and Jayaraman, 2002). A major limitation is the difficulty of producing large protein fragments in substantial numbers. As a result, most applications to date have focused on arraying relatively small peptide fragments that are readily synthesized by chemical methods. For example, these can be useful for mapping phosphotyrosine sites interacting with a particular src homology 2 (SH2) domain (Yeh et al., 2001). However, a number of protein-protein interactions may require structures larger than those represented by short peptides, or may be modified by the global protein structure, such that binding assays based on peptide fragments may be misleading. Nonetheless, the potential throughput for these methods is considerable.

## Conclusions

Developments in protein chemistry, including reproducible two-dimensional PAGE and MALDI-TOF-MS,

coupled with database and statistical developments, allow rapid and high throughput identification of proteins and posttranslational modifications. Yeast two hybrid approaches and other high throughput methods allow analysis of a large number of possible protein-protein interactions of potential importance in signal transduction. Recent advances in antibody and protein arrays, analogous to DNA arrays, hold considerable promise for further increases in the throughput of proteomics approaches to signal transduction.

### Acknowledgments

Authors would like to acknowledge the technical assistance of Kimberly Smuga-Otto. Research was supported by University of Wisconsin College of Agricultural and Life Sciences and USDA project WIS 3108 and WIS 4665.

### Literature Cited

- Aoyagi, T., H Suya, K. Umeda, N. Kato, O. Nemoto, H. Kobayashi, and Y. Miura. 1985. Epidermal growth factor stimulates tyrosine phosphorylation of pig epidermal fibrous keratin. *J. Invest. Dermatol.* 84:118–121.
- Auerbach, D., S. Thaminy, M. O. Hottiger, and I. Stagljar. 2002. The post-genomic era of interactive proteomics: Facts and perspectives. *Proteomics* 2:611–623.
- Baldi, P., S. Brunak, Y. Chauvin, C. A. Andersen, and H. Nielsen. 2000. Assessing the accuracy of prediction algorithms for classification: An overview. *Bioinformatics* 16:412–424.
- Baron, A. T., J. M. Lafky, C. H. Boardman, S. Balasubramaniam, V. J. Suman, K. C. Podratz, and N. J. Miahle. 1999. Serum sErbB1 and epidermal growth factor levels as tumor biomarkers in women with stage III or IV epithelial ovarian cancer. *Cancer Epidemiol. Biomarkers Prev.* 8:129–137.
- Baron, A. T., J. M. Lafky, V. J. Suman, D. W. Hillman, M. C. Buenafe, C. H. Boardman, K. C. Podratz, E. A. Perez, and N. J. Miahle. 2001. A preliminary study of serum concentrations of soluble epidermal growth factor receptor (sErbB1), gonadotropins, and steroid hormones in healthy men and women. *Cancer Epidemiol. Biomarkers Prev.* 10:1175–1185.
- Beaton, A., R. J. Wilkins, and T. T. Wheeler. 1997. Lactation-associated and prolactin-responsive changes in protein synthesis in mouse mammary cells. *Tissue Cell* 29:509–516.
- Ben-Ze'ev, A. 1987. The role of changes in cell shape and contacts in the regulation of cytoskeleton expression during differentiation. *J. Cell Sci. Suppl.* 8:293–312.
- Berlanga, J. J., J. P. Garcia-Ruiz, M. Perrot-Appianat, P. A. Kelly, and M. Edery. 1997. The short form of the prolactin (PRL) receptor silences PRL induction of the beta-casein gene promoter. *Mol. Endocrinol.* 11:1449–1457.
- Berndt, P., U. Hobohm, and H. Langen. 1999. Reliable automatic protein identification from matrix-assisted laser desorption/ionization mass spectrometric peptide fingerprints. *Electrophoresis* 20:3521–3526.
- Bole-Feysoot, C., V. Goffin, M. Edery, N. Binart, and P. A. Kelly. 1998. Prolactin (PRL) and its receptor: Actions, signal transduction pathways and phenotypes observed in PRL receptor knockout mice. *Endocrinol. Rev.* 19:225–268.
- Borrebaeck, C. A. 2000. Antibodies in diagnostics—from immunoassays to protein chips. *Immunol. Today* 21:379–382.
- Bush, M. R., J. M. Mele, G. M. Couchman, and D. K. Walmer. 1998. Evidence of juxtacrine signaling for transforming growth factor alpha in human endometrium. *Biol. Reprod.* 59:1522–1529.
- Chang, W. P., Y. Ye, and C. V. Clevenger. 1998. Stoichiometric structure-function analysis of the prolactin receptor signaling domain by receptor chimeras. *Mol. Cell. Biol.* 18:896–905.
- Chang, W. P., and C. V. Clevenger. 1996. Modulation of growth factor receptor function by isoform heterodimerization. *Proc. Natl. Acad. Sci. USA* 93:5947–5952.
- Cheung, V. G., M. Morley, F. Aguilar, A. Massimi, R. Kucherlapati, and G. Childs. 1999. Making and reading microarrays. *Nat. Genet.* 21(Suppl. 1):15–19.
- Cooper, M. A. 2002. Optical biosensors in drug discovery. *Nat. Rev. Drug Discov.* 1:515–528.
- Courtneidge, S. A., S. Fumagalli, M. Koegl, G. Superti-Furga, and G. M. Twamley-Stein. 1993. The Src family of protein tyrosine kinases: Regulation and functions. *Dev. Suppl.* 1:57–64.
- Duan, W. R., T. G. Parmer, C. T. Albarracin, L. Zhong, and G. Gibori. 1997. PRAP, a prolactin receptor associated protein: Its gene expression and regulation in the corpus luteum. *Endocrinology* 138:3216–3221.
- Duggan, D. J., M. Bittner, Y. Chen, P. Meltzer, and J. M. Trent. 1999. Expression profiling using cDNA microarrays. *Nat. Genet.* 21(Suppl. 1):10–14.
- Emanuelsson, O., and G. von Heijne. 2001. Prediction of organellar targeting signals. *Biochim. Biophys. Acta* 1541:114–149.
- Emili, A. Q., and G. Cagney. 2000. Large-scale functional analysis using peptide or protein arrays. *Nat. Biotechnol.* 18:393–397.
- Eriksson, J., B. T. Chait, and D. Fenyo. 2000. A statistical basis for testing the significance of mass spectrometric protein identification results. *Anal. Chem.* 72:999–1005.
- Eriksson, J., and D. Fenyo. 2002. A model of random mass-matching and its use for automated significance testing in mass spectrometric proteome analysis. *Proteomics* 2:262–270.
- Feng, L., X. Zhou, J. Liao, and M. B. Omary. 1999. Pervanadate-mediated tyrosine phosphorylation of keratins 8 and 19 via a p38 mitogen-activated protein kinase-dependent pathway. *J. Cell Sci.* 112:2081–2090.
- Fichmann, J. 1999. Advantages of immobilized pH gradients. *Methods Mol. Biol.* 112:173–174.
- Figeys, D., and D. Pinto. 2001. Proteomics on a chip: Promising developments. *Electrophoresis* 22:208–226.
- Figeys, D. 2002. Functional proteomics: Mapping protein-protein interactions and pathways. *Curr. Opin. Mol. Ther.* 4:210–215.
- Gianazza, E. 1999. Casting immobilized pH gradients (IPGs). *Methods Mol. Biol.* 112:175–188.
- Gietz, R. D., and R. A. Woods. 2002. Screening for protein-protein interactions in the yeast two-hybrid system. *Methods Mol. Biol.* 185:471–486.
- Gorg, A., C. Obermaier, G. Boguth, A. Harder, B. Scheibe, R. Wildgruber, and W. Weiss. 2000. The current state of two-dimensional electrophoresis with immobilized pH gradients. *Electrophoresis* 21:1037–1053.
- Grosvenor, C. E., M. F. Picciano, and C. R. Baumrucker. 1993. Hormones and growth factors in milk. *Endocrinol. Rev.* 14:710–728.
- Hamdan, M., M. Galvani, and P. G. Righetti. 2001. Monitoring 2-D gel-induced modifications of proteins by MALDI-TOF mass spectrometry. *Mass Spectr. Rev.* 20:121–141.
- Hendrix, M. J., E. A. Seftor, Y. W. Chu, K. T. Trevor, and R. E. Seftor. 1996. Role of intermediate filaments in migration, invasion and metastasis. *Cancer Metastasis Rev.* 15:507–525.
- Horseman, N. D. 2002. Prolactin receptor diversity in humans: Novel isoforms suggest general principles. *Trends Endocrinol. Metab.* 13:47–48.
- Johnson, J. L., S. Fenton, and L. G. Sheffield. 1996. Prolactin inhibits epidermal growth factor-induced Ras-MAPK signaling in mammary epithelial cells. *J. Biol. Chem.* 271:21574–21578.
- Kaufmann, H., J. E. Bailey, and M. Fussenegger. 2001. Use of antibodies for detection of phosphorylated proteins separated by two-dimensional gel electrophoresis. *Proteomics* 1:194–199.
- Kline, J. B., H. Roehrs, and C. V. Clevenger. 1999. Functional characterization of the intermediate isoform of the human prolactin receptor. *J. Biol. Chem.* 274:35461–35468.

- Kuo, A., C. Zhong, W. S. Lane, and R. Derynck. 2000. Transmembrane transforming growth factor- $\alpha$  tethers to the PDZ domain-containing, Golgi membrane-associated protein p59/GRASP55. *EMBO J.* 19:6427–6439.
- Lakey, J. H., and E. M. Raggett. 1998. Measuring protein-protein interactions. *Curr. Opin. Struct. Biol.* 8:119–123.
- Legrain, P., J. Wojcik, and J. M. Gauthier. 2001. Protein–protein interaction maps: A lead towards cellular functions. *Trends Genet.* 17:346–352.
- Lelievre, S., V. M. Weaver, and M. J. Bissell. 1996. Extracellular matrix signaling from the cellular membrane skeleton to the nuclear skeleton: A model of gene regulation. *Recent Prog. Horm. Res.* 51:417–432.
- Leushner, J. 2001. MALDI TOF mass spectrometry: An emerging platform for genomics and diagnostics. *Expert Rev. Mol. Diagn.* 1:11–18.
- Li, M. 2000. Applications of display technology in protein analysis. *Nat. Biotechnol.* 18:1251–1256.
- Lilius, G., M. Persson, L. Bulow, and K. Mosbach. 1991. Metal affinity precipitation of proteins carrying genetically attached polyhistidine affinity tails. *Eur. J. Biochem.* 198:499–504.
- Lu, T., M. Van Dyke, and M. Sawadogo. 1993. Protein-protein interaction studies using immobilized oligohistidine fusion proteins. *Anal. Biochem.* 213:318–322.
- Luetke, N. C., and D. C. Lee. 1990. Transforming growth factor  $\alpha$ : Expression, regulation and biological action of its integral membrane precursor. *Semin. Cancer Biol.* 1:265–275.
- Maheshwari, G., H. S. Wiley, and D. A. Lauffenburger. 2001. Autocrine epidermal growth factor signaling stimulates directionally persistent mammary epithelial cell migration. *J. Cell Biol.* 155:1123–1128.
- Mann, M., R. C. Hendrickson, and A. Pandey. 2001. Analysis of proteins and proteomes by mass spectrometry. *Annu. Rev. Biochem.* 70:437–473.
- Matsui, N. M., D. M. Smith-Beckerman, J. Fichmann, and L. B. Epstein. 1999. Running preparative carrier ampholyte and immobilized pH gradient IEF gels for 2-D. *Methods Mol. Biol.* 112:211–219.
- Modrek, B., and C. Lee. 2002. A genomic view of alternative splicing. *Nat. Genet.* 30:13–19.
- Mroczkowski, B., and G. Carpenter. 1988. Epidermal growth factor. *Prog. Clin. Biol. Res.* 262:207–216.
- Mroczkowski, B., M. Reich, K. Chen, G. I. Bell, and S. Cohen. 1989. Recombinant human epidermal growth factor precursor is a glycosylated membrane protein with biological activity. *Mol. Cell. Biol.* 9:2771–2778.
- Mroczkowski, B., and M. Reich. 1993. Identification of biologically active epidermal growth factor precursor in human fluids and secretions. *Endocrinology* 132:417–425.
- Nagata, K., I. Izawa, and M. Inagaki. 2001. A decade of site- and phosphorylation state-specific antibodies: Recent advances in studies of spatiotemporal protein phosphorylation. *Genes Cells* 6:653–664.
- O'Farrell, P. H. 1975. High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.* 250:4007–4021.
- Petsko, G. A. 2001. Size doesn't matter. *Genome Biol.* 2: Comment 1003.1–1003.2.
- Phipps, M. A., and L. A. Mackin. 2000. Application of isothermal microcalorimetry in solid state drug development. *Pharmacol. Sci. Technol. Today* 3:9–17.
- Postel-Vinay, M. C., and J. Finidori. 1995. Growth hormone receptor: Structure and signal transduction. *Eur. J. Endocrinol.* 133:654–659.
- Prusheik, K., A. N. Ladd, and L. G. Sheffield. 1997. Effects of epidermal growth factor (EGF) and prolactin on EGF receptor cytoskeletal association in mammary epithelial cells. *Proc. Soc. Exp. Biol. Med.* 215:393–398.
- Quijano, V. J., Jr., and L. G. Sheffield. 1998. Prolactin decreases epidermal growth factor receptor kinase activity via a phosphorylation-dependent mechanism. *J. Biol. Chem.* 273:1200–1207.
- Reiter, J. L., and N. J. Mairle. 1996. A 1.8 kb alternative transcript from the human epidermal growth factor receptor gene encodes a truncated form of the receptor. *Nucleic Acids Res.* 24:4050–4056.
- Reiter, J. L., D. W. Threadgill, G. D. Eley, K. E. Strunk, A. J. Danielsen, C. S. Sinclair, R. S. Pearsall, P. J. Green, D. Yee, A. L. Lampland, S. Balasubramaniam, T. D. Crossley, T. R. Magnuson, C. D. James, and N. J. Mairle. 2001. Comparative genomic sequence analysis and isolation of human and mouse alternative EGFR transcripts encoding truncated receptor isoforms. *Genomics* 71:1–20.
- Righetti, P. G., E. Gianazza, and B. Bjellqvist. 1983. Modern aspects of isoelectric focusing: Two-dimensional maps and immobilized pH gradients. *J. Biochem. Biophys. Methods* 8:89–108.
- Rubin, G. M., M. D. Yandell, J. R. Wortman, G. L. Gabor Miklos, C. R. Nelson, I. K. Hariharan, M. E. Fortini, P. W. Li, R. Apweiler, W. Fleischmann, J. M. Cherry, S. Henikoff, M. P. Skupski, S. Misra, M. Ashburner, E. Birney, M. S. Boguski, T. Brody, P. Brokstein, S. E. Celniker, S. A. Chervitz, D. Coates, A. Cravchik, A. Gabrielian, R. F. Galle, W. M. Gelbart, R. A. George, L. S. Goldstein, F. Gong, P. Guan, N. L. Harris, B. A. Hay, R. A. Hoskins, J. Li, Z. Li, R. O. Hynes, S. J. Jones, P. M. Kuehl, B. Lemaitre, J. T. Littleton, D. K. Morrison, C. Mungall, P. H. O'Farrell, O. K. Pickeral, C. Shue, L. B. Vosshall, J. Zhang, Q. Zhao, X. H. Zheng, and S. Lewis. 2000. Comparative genomics of the eukaryotes. *Science* 287:2204–2215.
- Rudland, P. S., R. Barraclough, D. G. Fernig, and J. A. Smith. 1998. Growth and differentiation of the normal mammary gland and its tumours. *Biochem. Soc. Symp.* 63:1–20.
- Schlegel, A., and M. P. Lisanti. 2001. The caveolin triad: Caveolae biogenesis, cholesterol trafficking, and signal transduction. *Cytokine Growth Factor Rev.* 12:41–51.
- Schroeder, F., A. M. Gallegos, B. P. Atshaves, S. M. Storey, A. L. McIntosh, A. D. Petrescu, H. Huang, O. Starodub, H. Chao, H. Yang, A. Frolov, and A. B. Kier. 2001. Recent advances in membrane microdomains: Rafts, caveolae, and intracellular cholesterol trafficking. *Exp. Biol. Med.* 226:873–890.
- Shi, W., H. Fan, L. Shum, and R. Derynck. 2000. The tetraspanin CD9 associates with transmembrane TGF- $\alpha$  and regulates TGF- $\alpha$ -induced EGF receptor activation and cell proliferation. *J. Cell Biol.* 148:591–602.
- Shum, L., S. A. Reeves, A. C. Kuo, E. S. Fromer, and R. Derynck. 1994. Association of the transmembrane TGF- $\alpha$  precursor with a protein kinase complex. *J. Cell Biol.* 125:903–916.
- Shum, L., C. W. Turck, and R. Derynck. 1996. Cysteines 153 and 154 of transmembrane transforming growth factor- $\alpha$  are palmitoylated and mediate cytoplasmic protein association. *J. Biol. Chem.* 271:28502–28508.
- Sickmann, A., and H. E. Meyer. 2001. Phosphoamino acid analysis. *Proteomics* 1:200–206.
- Smirnova, E., D. L. Shurland, and A. M. van der Blik. 2001. Mapping dynamin interdomain interactions with yeast two-hybrid and glutathione S-transferase pulldown experiments. *Methods Enzymol.* 329:468–477.
- Srinivas, P. R., S. Srivastava, S. Hanash, and G. L. Wright, Jr. 2001. Proteomics in early detection of cancer. *Clin. Chem.* 47:1901–1911.
- Stupka, E. 2002. Large-scale open bioinformatics data resources. *Curr. Opin. Mol. Ther.* 4:265–274.
- Taylor-Papadimitriou, J., R. Wetzels, and F. Ramaekers. 1992. Intermediate filament protein expression in normal and malignant human mammary epithelial cells. *Cancer Treat. Res.* 61:355–378.
- Trott, J. F., R. C. Hovey, S. Koduri, and B. K. Vonderhaar. 2003. Alternative splicing to exon 11 of human prolactin receptor gene results in multiple isoforms including a secreted prolactin-binding protein. *J. Mol. Endocrinol.* 30:31–47.
- Tucker, C. L., J. F. Gera, and P. Uetz. 2001. Towards an understanding of complex protein networks. *Trends Cell Biol.* 11:102–106.
- Wilkins, M. R., E. Gasteiger, A. A. Gooley, B. R. Herbert, M. P. Molloy, P. A. Binz, K. Ou, J. C. Sanchez, A. Bairoch, K. L. Williams, and D. F. Hochstrasser. 1999. High-throughput mass spectrometric

- discovery of protein post-translational modifications. *J. Mol. Biol.* 289:645–657.
- Wu, C. C., K. E. Howell, M. C. Neville, J. R. Yates 3rd, and J. L. McManaman. 2000. Proteomics reveal a link between the endoplasmic reticulum and lipid secretory mechanisms in mammary epithelial cells. *Electrophoresis* 21:3470–3482.
- Wu, C. C., J. R. Yates, 3rd, M. C. Neville, and K. E. Howell. 2000. Proteomic analysis of two functional states of the Golgi complex in mammary epithelial cells. *Traffic* 1:769–782.
- Xiao, Z. Q., and A. P. Majumdar. 2001. Increased in vitro activation of EGFR by membrane-bound TGF- $\alpha$  from gastric and colonic mucosa of aged rats. *Am. J. Physiol. Gastrointest. Liver Physiol.* 281:G111–G116.
- Xu, X., K. F. Kelleher, J. Liao, K. E. Creek, and L. Pirisi. 2000. Unique carboxyl-terminal sequences of wild type and alternatively spliced variant forms of transforming growth factor- $\alpha$  precursors mediate specific interactions with ErbB4 and ErbB2. *Oncogene* 19:3172–3181.
- Yarmush, M. L., and A. Jayaraman. 2002. Advances in proteomic technologies. *Annu. Rev. Biomed. Eng.* 4:349–373.
- Yeh, R. H., T. R. Lee, and D. S. Lawrence. 2001. From consensus sequence peptide to high affinity ligand, a “library scan” strategy. *J. Biol. Chem.* 276:12235–12240.
- Yu, Y. L., H. Kha, J. A. Golden, A. A. Migchielsen, E. J. Goetzl, and C. W. Turck. 1992. An acidic fibroblast growth factor protein generated by alternate splicing acts like an antagonist. *J. Exp. Med.* 175:1073–1080.
- Zaluzec, E. J., D. A. Gage, and J. T. Watson. 1995. Matrix-assisted laser desorption ionization mass spectrometry: Applications in peptide and protein characterization. *Protein Expr. Purif.* 6:109–123.