# Quantification in Proteomics through Stable Isotope Coding: A Review

**ARTICLE** *in* JOURNAL OF PROTEOME RESEARCH · JUNE 2004

Impact Factor: 4.25 · DOI: 10.1021/pr0340734 · Source: PubMed



**2 AUTHORS:**

Samir Julka
Dow Chemical Company
**14** PUBLICATIONS **534** CITATIONS
SEE PROFILE

Fred E Regnier
Purdue University
**354** PUBLICATIONS **17,499** CITATIONS
SEE PROFILE

Journal of proteome research



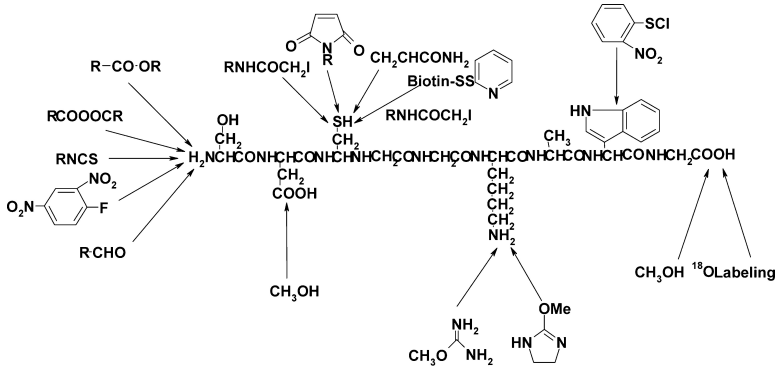
Reviews

# Quantification in Proteomics through Stable Isotope Coding: A Review

Samir Julka, and Fred Regnier

*Journal of Proteome Research*, **2004**, 3 (3), 350-363• DOI: 10.1021/pr0340734 • Publication Date (Web): 13 January 2004






## More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Links to the 17 articles that cite this article, as of the time of this article download
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

View the Full Text HTML



Journal of Proteome Research is published by the American Chemical Society. 1155 Sixteenth Street N.W., Washington, DC 20036

Journal of proteome research reviews

# Quantification in Proteomics through Stable Isotope Coding: A Review

**Samir Julka and Fred Regnier***

*Department of Chemistry, Purdue University, West Lafayette, Indiana 47907*

**Received September 10, 2003**

This review focuses on techniques for quantification and identification in proteomics by stable isotope coding. Methods are examined for analyzing expression, post-translational modifications, protein:protein interactions, single amino acid polymorphism, and absolute quantification. The bulk of the quantification literature in proteomics focuses on expression analysis, where a wide variety of methods targeting different features of proteins are described. Methods for the analysis of post-translational modification (PTM) focus primarily on phosphorylation and glycosylation, where quantification is achieved in two ways, either by substitution or tagging of the PTM with an isotopically coded derivatizing agent in a single process or by coding and selecting PTM modified peptides in separate operations. Absolute quantification has been achieved by age-old internal standard methods, in which an isotopically labeled isoform of an analyte is synthesized and added to a mixture at a known concentration. One of the surprises is that isotope coding can be a valuable aid in the examination of intermolecular association of proteins through stimulus:response studies. Preliminary efforts to recognize single amino acid polymorphism are also described. The review ends with the conclusion that (1) isotope ratio analysis of protein concentration between samples does not necessarily relate directly to protein expression and rate of PTM and (2) that multiple new methods must be developed and applied simultaneously to make existing stable isotope quantification methods more meaningful. Although stable isotope coding is a powerful, wonderful new technique, multiple analytical issues must be solved for the technique to reach its full potential as a tool to study biological systems.

**Keywords:** quantification • expression • post-translational modifications • protein:protein interactions • single amino acid polymorphism • absolute quantification

## 1. Introduction

**Background.** Living systems are dynamic, each developing, surviving, and proliferating in a different way. But how does this occur and how are biological processes regulated? A key to solving this puzzle is the realization that in responding to change, organisms must themselves change. Signaling, gene activation or suppression, transcription, translation, post-translational modifications, intracellular transport, metabolic processing, and feedback control at both the gene and protein level all involve material changes in cells. A major portion of life science oriented research today is directed toward recognizing these stimulus-related changes and associating them with a response mechanism.

The concept that physical change is associated with stimuli has been used widely in medicine and the life sciences for centuries, first based on the physical appearance of a patient or organism and now with the addition of chemical measurements that target a broad variety of biological processes. A critical component of using deviations from "normal" to assess health and the impact of stimuli on biological systems is the need for quantification. The degree of divergence from the norm often indicates the magnitude of a stimulus, the extent to which a biological system is being stressed, or the state of disease progression. Clearly, analytical tools that quantify change play an important role in life sciences.

But the problem of understanding system dynamics is much deeper than just quantifying change. An even more important question is what to quantify. In the case of many diseases, one has no idea which of the 10−20 thousand proteins in blood or a cell is associated with the disease. A very important element in finding what to quantify is recognizing that a small number of proteins are associated with the onset and progression of either a regulatory event or disease and that they change in concentration, and sometimes structure, during the event.

But can a disease or regulatory event really be defined by a "needle-in-a-haystack"? Having recognized a protein that is up- or down-regulated in concert with an event, the question is whether the change in that protein is associated exclusively with that single, specific event or other stimuli as well. The

* To whom correspondence should be addressed. Fred Regnier, Department of Chemistry, Purdue University, 1393 Brown Bldg., West Lafayette, IN, 47907-1393, USA. Phone: (765) 494-1648. Fax: (765) 494-0359. E-mail: fregnier@purdue.edu.


10.1021/pr0340734 CCC: $27.50 
Published on Web 01/13/2004

possibility that a disease or regulatory stimulus can be defined by a single marker is remote. Even a glance at a metabolic chart shows a high degree of connectivity in biological systems and that multiple proteins are likely to be impacted by a stimulus. This is the concept behind "Systems Biology". It is now understood that the biological response to a stimulus is likely to involve many proteins and that change is likely to occur in a temporal pattern producing changes in signaling, transcription, translation, post-translational modification, and metabolic alterations, not necessarily in that order. But still, the issue is how to find and quantify these hidden patterns of change.

An additional question is what changes to expect. Although quantification of expression at the mRNA and protein levels has been the focus of most studies on cellular regulation, it is important to note that a wide variety of enzymatic and nonenzymatic post-translational modifications play a decisive role in regulation as well. In fact, aberrations in post-translational modification are a dominant feature of many diseases, along with variations in mRNA processing and single amino acid polymorphism. There is surprisingly little literature on recognizing and quantifying changes in post-translational modifications.

**The Birth of Global Coding as a Way to Look for Change.** Although global coding with stable isotopes has recently become popular in proteomics, the technique evolved from other sources. One of the earliest applications of stable isotope coding was in the synthesis of internal standards that could only be differentiated from analytes with a mass spectrometer in the final step of the analysis. [1,2] The concept of attaching coding agents in a derivatization reaction followed.[3] Isotope coding was even used to determine the relative concentration of analyte between samples.[4] But the major conceptual advance that has sparked so much interest in stable isotope coding in proteomics came from genomics. A procedure was described by Brown in which cDNA species were tagged with a fluorescent dye during the course of converting mRNA species to their cDNA analogues with reverse transcriptase.[5] Using a different fluorescent dye for control and experimental samples, it was possible to achieve sample specific coding. When these differentially dyed or coded samples were mixed, and the individual cDNA species selected by hybridization with a complimentary immobilized oligonucleotide, expression of a particular species could be compared by the difference in fluorescence. Through the use of DNA chip technology, it is now possible to examine an entire transcriptome in a single analysis.[6,7]

Brown's observations were important to proteomics because they showed a way to code and recognize change in large numbers of species simultaneously. The major difference between the Brown approach and stable isotope based quantification in proteomics is that isotopically coded tags and mass spectral detection was substituted for Brown's fluorescent tags and fluorescence spectroscopy (Figure 1). Coded peptide isoforms appear in mass spectra as doublet clusters of ions separated by the mass difference in the coding agents. Although it is more difficult to select and identify coded peptides and proteins than polynucleotides, the stable isotope coding methods emerging for proteomics are in general more powerful than those used in genomics. One of the major advantages of the new proteomic methods is that changes in polypeptides of unknown structure can be recognized and quantified. A number of coding methods have emerged that differ in the manner of labeling, the method of selecting coded peptides, and the application, but not in the basic principle described by Brown. This review will focus on these differences and the probable utility of various quantification methods in a broad range of applications.

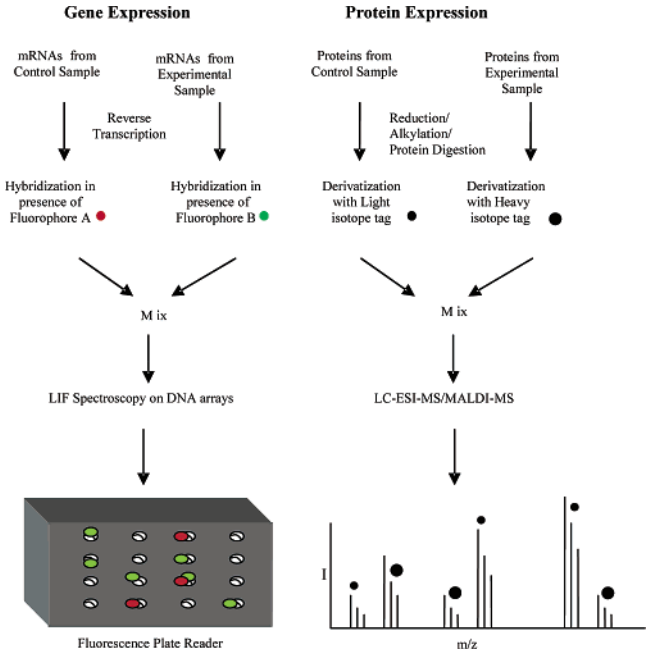



**Figure 1.** Comparison of gene expression and protein expression analysis. The schematic demonstrates the substantial similarity between these two methods.

**The Relationship between Change and Expression.** Differential coding is a superb way to quantify relative differences in concentration between samples. When the turnover rate of the species being measured is high relative to the time frame of the experiment, the relative difference (*R*) in amount of an analyte between samples is directly proportional to the relative rate of analyte synthesis. In this situation, all of the analyte measured in both the control and experimental samples would have been synthesized during the course of the experiment. This is generally the case in transcriptome analysis, in which all of the mRNA being measured was formed during the time course of the experiment.

Stable isotope ratio *(R)* measurements of coded peptides in proteomics probably have a very different meaning. The objective of the discussion below is to shed light on the meaning of these measurements where the amount ($A_{\text{experimental}}$) of a protein in an experimental sample is compared to the amount ($A_{\text{control}}$) in the control and (*R)* is the relative amount of protein in the two samples. There is no intent to actually describe cellular dynamics in the analysis of (*R*) below.

$$R = A_{\text{experimental}}/A_{\text{control}}$$

The turnover rate of proteins is generally much slower than that of mRNA species. The average rate of degradation in 50 proteins from yeast was 2.2%/h based on metabolic labeling experiments with stable isotopes.[8] Moreover, proteins vary dramatically in pool size. The amount (*A*) of a protein in a compartment after time (*t*) is given by the equation

$$A = A_{\text{o}} + t{\cdot}r - t{\cdot}d$$

where ($A_{\text{o}}$) is the initial amount of the protein in the compartment, (*r*) is the rate of synthesis, and (*d*) is the rate of degradation or migration of the protein from the compartment

being analyzed. Actually, there can be multiple pathways for removal of a protein from a compartment, but a single mechanism has been assumed here for simplicity. At homeostasis, an organism is in a steady state where $t \cdot r = t \cdot d$ and $A = A_o$. When a cell or organism undergoes a regulatory change that causes protein concentration to shift, the protein concentration ($A_t$), a short time ($t$) after the perturbation will be

$$A_t = A_o + t \cdot r_t - t \cdot d_t$$

where ($r_t$) and ($d_t$) are the new rates of protein synthesis and degradation, respectively. The ratio ($R$) of change in concentration between an experimental and control at time ($t$) will be

$$R = 1 + \{[t(r_t - d_t)]/A_o\}$$

This equation has a series of interesting implications. Clearly, it shows that we have to be more careful about claiming that ($R$) is a measure of protein expression. The value of ($R$) for different proteins is strongly dependent on ($A_o$), even when they all have the same rate of expression ($r_t$). Changes in ($R$) can also be due as much to the changes in the rate of degradation ($d_t$) as synthesis. This will almost certainly be found to be the case of signaling with phosphoproteins. ($R$) also has a temporal element. Fixed time measurements of ($R$) only identify system components that are changing unless ($t$) is large and the system has come to a new steady state.

The equation below shows that when it is the objective to measure the actual expression rate ($r_t$) of a protein, it is necessary to know the initial protein concentration in the compartment ($A_o$) along with ($R$) and the rate of degradation ($d_t$) after the stimulus

$$r_t = \{[(R - 1)]/t\} + d_t$$

When it is assumed that the stimulus does not impact the rate of degradation, i.e., $d = d_t$, this equation reduces to

$$r_t = [A_o(R - 1)]/t$$

Because it is necessary to know $A_o$, quantification of the absolute amount of a protein initially in a compartment is an important issue in proteomics. But so is the initial rate of expression ($r$) and degradation ($d$). Another way to solve for expression rate is in terms of the rate of degradation ($d$). When $r = r_t$, then

$$d_t = -[A_o(R - 1)]/t$$

These equations describe what might happen to the rate of expression shortly after a perturbation, as might be the case with a drug or some other transitory stimulus. Given enough time at a new rate of expression or degradation, the system will come to a new steady state in which the concentration of a protein is either higher or lower. These longer-term effects are what we associate with many diseases.

Although terms such as "quantification of expression" are widely associated in the literature with measurements of ($R$), and will be used in this review as well, it is important to understand that ($R$) does not measure the rate of expression ($r$).

**Need Assessment And Classification Of Quantification Strategies.** The value of analytical methods is in what they can do, not their novelty. For this reason, the analytical methods reviewed here will be examined in terms of their potential to address specific needs. The discussion that follows is based on the reviewers' perception of prominent issues and needs in proteomics.

It has been shown above that protein expression, post-translational modification, protein:protein interactions, absolute quantification, single amino acid polymorphism, and temporal changes in concentration are major issues of interest in proteomics. Coding and quantification strategies for the study of these phenomena will be most useful when (1) all the peptides in a sample are coded, (2) a predictable mass window is created between coded isoforms of peptides that precludes spectral overlap, (3) peptides are quantitatively labeled, (4) isotopic isoforms are highly enriched, (5) sample treatment after coding is minimal, (6) spectral interpretation is not impacted by the presence of coding agents, (7) ionization efficiency is not diminished by coding, (8) coding does not cause chromatographic resolution of peptide isoforms, (9) coding allows multiplexing of more than two samples, and (10) the method can be used with all types of organisms, including humans.

## 2. Separation Component

**The Sample Complexity Problem.** The proteome of many types of cells can contain more than 10 000 proteins. Assuming a tryptic digest of an average protein will have 30−50 peptides, it is easily seen that a tryptic digest of a proteome can have $3 \times 10^5$ to $5 \times 10^5$ or more peptides. Considering that a reversed-phase chromatography column has a peak capacity of 300−500 at best and no more than 20−40 peptides can be analyzed simultaneously in most electrospray ionization mass spectrometers, the analytical capacity of an LC−MS will be $2 \times 10^4$ peptides in the best case. This is at least an order of magnitude lower than required to analyze all the peptides in a proteome. Obviously, additional fractionation steps must be added to deal with samples of this complexity.

A second major problem is the enormous difference in protein concentration in most proteomes. There is a $10^6$ difference in protein abundance in *E. coli* and $10^9$ in man. But the dynamic range of most mass spectrometers is in the range of $10^4$. This problem is compounded by the fact that high and low abundance peptides can coelute. Ionization of low abundance peptides can be suppressed and their spectra masked by high abundance species in this case. It is for this reason that high abundance proteins are often removed from samples and multidimensional separations are applied to reduce the number of proteins that are coeluting. Affinity selection methods are the most popular for sample simplification, but ion exchange and a second dimension of reversed-phase chromatography are also used. These methods will be discussed below in association with the various quantification procedures.

A major portion of this review will be spent on labeling strategies for stable isotope quantification, (Figure 2) based on coding experimental samples in one way and control samples another. In many cases, the labeling procedures target specific amino acids or post-translational modifications. Coded and uncoded peptides must be resolved for these procedures to be of greatest utility. With extreme changes in protein concentration, peptides will appear in mass spectra as single clusters of ions instead of the normal doublet clusters seen when both isoforms of a peptide are present. It is difficult without sequencing both sets of peptides to differentiate between un-coded singlet clusters and those that are coded but appear as singlets because of large regulatory changes. This problem is aggravated by the

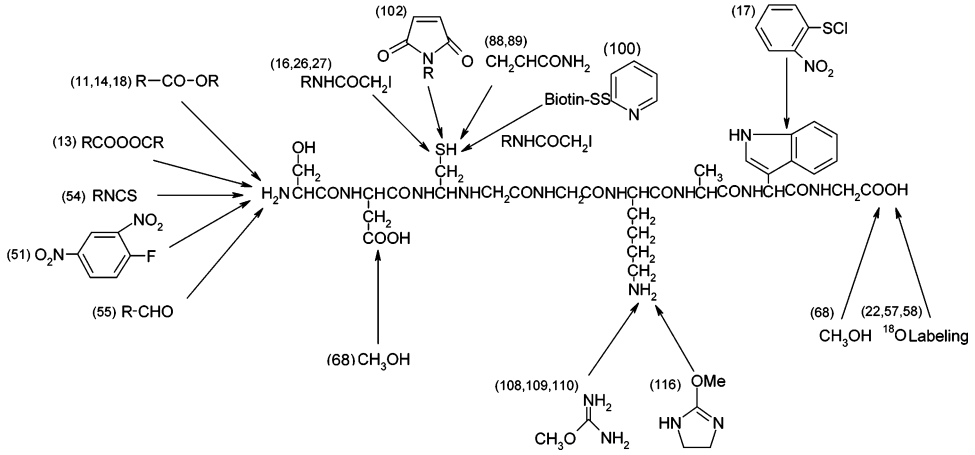



**Figure 2.** Generic summary of coding strategies currently employed to exploit various reactive centers in peptides. All reactions shown to occur on the amino terminus also apply to the $\epsilon$-amino group of lysine residue. Numbers in the parentheses indicate the respective references.

fact that 80% or more of the peptides may be un-coded. Using a separation system that resolves coded from un-coded peptides solves this problem.

**Chromatographic Isotope Effects.** Stable isotopes vary in their utility as labeling agents. Although deuterium is inexpensive and easily incorporated into organic compounds, it is well documented that deuterium labeling causes the retention time of low molecular weight organic compounds to decrease during reversed-phase chromatography (RPC).[9] The same is true with peptides.[10] This means that peptide isoforms often fail to coelute, and their isotope ratio will vary during the course of elution from RPC columns. Quantification can only be achieved by comparing their chromatographic peak areas after the analysis is complete. Although it is relatively easy to compute chromatographic peak areas using MS data, real-time isotope ratio analysis is precluded. There is the additional problem of recognizing low abundance peptide isoforms in complex mixtures when their retention times differ by 30−60 s. They appear as "singlets" that are grossly up- or down-regulated.

The degree to which the resolution of deuterated peptides occurs depends on several factors including (1) the number of deuterium atoms in the peptide, (2) coding agent structure, (3) peptide sequence, and (4) peptide size. In most cases, resolution tends to increase as the number of deuterium atoms increase relative to peptide molecular weight. But it has been shown with 4-trimethylammonium butyrate that when 9 deuterium atoms were placed in the methyl groups around the quaternary amine, there was almost no isotope effect.[11] Grouping deuterium atoms around a polar functional group that does not interact with the RPC stationary phase dramatically reduces the chromatographic isotope effect. This means that deuterium-labeled reagents can be used for coding but the structure of the reagent and the position of deuterium atoms in the reagent must minimize their interaction with the RPC stationary phase. Labeling with $^{15}N$ has also been reported to cause small chromatographic resolution.[12]

Coding samples with reagent labeled with $^{13}C$ and un-labeled reagent (ulr), in contrast to deuterium labeling, causes no observable difference in elution time of the coded isoforms.[13] Reagents labeled with $^{13}C$ that have been examined for in vitro coding are succinic anhydride ($4^{13}C$/ulr),[13] succinimidyl propionate ($3^{13}C$/ulr),[14] iodoacetanilide ($6^{13}C$/ulr),[15] cleavable ICAT reagent ($9^{13}C$/ulr),[16] 2-nitrobenzenesulfenyl chloride ($6^{13}C$/ulr),[17] and sulfo-succinimidyl benzoate ($6^{13}C$/ulr).[18] In vivo labeling with $6^{13}C$ labeled arginine[19] and $6^{13}C$ lysine[20,21] has also been reported. The much higher cost of $^{13}C$ labeling is a potential limitation. $^{18}O$ has also been reported to cause no chromatographic isotope effects.[22]

Chromatographic isotope effects can also be suppressed by internal compensation within the coding reagents, such is in the tandem mass tags (TMTs) strategy.[23] The TMTs work in a pair. A TMT reagent consists of (i) a sensitization group (guanidino moiety) attached to heavy coding reagent, (ii) a cleavage enhancer group (proline) and (iii) a light mass normalization group (N-terminus specific activated acid). Its counterpart has the positions of heavy coding reagent and light normalization group swapped. Peptides from control and experimental samples can be differentially labeled at their N-terminus with the two forms of tandem mass tags. TMT coded peptides are of identical overall mass and similar chemical structure and thus coelute during reversed-phase chromatography. The TMT approach differs from other isotope coding techniques in that quantification is carried out in the 2nd dimension of mass spectrometry. The TMT fragment (sensitization group and heavy/light coding group) is tethered to its respective mass normalization groups by a facile bond, which is easily cleaved during MS/MS. Quantitation of protein expression is carried out by comparing the ratio of the cleaved TMT fragments. The TMT strategy also offers an increase in peak capacity, improvement in sensitivity, and a high signal-to-noise ratio. A crucial limitation is the need for performing a product ion scan on every coded peptide. This is because a change in protein expression cannot be recognized until an MS/MS is performed. Other possible concerns are the possibility of overlap of fragment ions from nonrelated peptides and the difficulty in synthesizing TMTs.

Silent isotope labeling is yet another way of overcoming chromatographic isotope effects.[24] The technique involves differential incorporation of an amino acid, having identical overall mass, but differing in its isotopic enrichment. As an example, phenylalanine enriched with one $^{13}C$ (at the carboxylic carbon) and phenylalanine enriched with one $^{15}N$ were differentially incorporated in vivo into control and experimental HeLa cells, respectively. Phenylalanine containing proteolysis-product peptides from the two samples are of identical overall mass. As with the TMT approach, protein expression is quanti-

fied in the 2nd dimension of mass spectrometry. The intensity of the differentially coded immonium ion of phenylalanine in peptides is compared in MS/MS. Immonium ions in this case are separated by +1 mass due to absence of $^{13}C$ carbon atom in the phenylalanine immonium ion. Silent isotope labeling shares the same advantages and disadvantages with the TMT approach. An additional concern with silent isotope labeling is that there is only a +1 amu mass difference of immonium ions, necessitating the need for deconvolution of the enriched and natural abundance peaks.

## 3. Expression

Two very different stable isotope coding strategies are being used to analyze protein expression. One is to isotopically code all peptides in a proteolytic digest of a proteome, either before or after proteolysis, and then chromatographically select classes of peptides for analysis in a second step that has no relationship to the first. The advantage of global coding is that it allows any peptide, from any place in a protein to be used for some type of quantitative measurement, even post-translational modifications. Multiple structure specific selection strategies can then be applied to the same control and experimental sample set. This is necessary when the specific structural feature being targeted is absent in a portion of the proteome. For example, neither histidine nor cysteine is present in all proteins. But analysis of both cysteine- and histidine-containing peptides covers the entire proteome in most cases.[25]

The other approach to expression analysis is to introduce isotopic coding and an affinity tag into peptides in a single step with one reagent.[26,27] Labeling and chromatographic selection are closely connected in this approach. Complex digests are easily simplified and database searches can be restricted to peptide candidates that contain a single amino acid with this method. The great advantage of this approach is simplicity. The relative merits of both the global and targeted labeling strategies will be analyzed below.

Another issue is whether some peptides in a protein should be excluded in quantification of expression. Proteins are often expressed and then post-translationally modified into multiple isoforms. This means that un-modified peptides will be derived from all the isoforms. In contrast, peptides derived from sites that are subsequently post-translationally modified will be found in only the un-modified protein. When newly expressed protein is added to a pool of preexisting isoforms, tryptic peptides associated with modification sites will differ in concentration from those that are common to all isoforms. Only peptides common to all isoforms of a protein can be used in expression analysis. But how can this be known unless all possible post-translational modifications of a protein have been described. One way is through redundant peptide analysis. A peptide that appears to differ in relative amount from sister peptides derived from the same protein is likely to be associated with a post-translational modification.

**3.1.1 Metabolic Coding in Isotopically Enriched/Depleted Media.** Metabolic incorporation of stable isotope ($^{15}N/^{13}C$) labeled nutrients in growth media into cultured cells is a popular method for global coding of proteomes. *S. cerevisiae*,[28,29] *E. coli*,[30] *D. radiodurans*,[31] and mammalian cell lines[12] have all been used in the analysis of protein expression.[32] Using $^{15}N$ metabolic labeling in *S. cerevisiae* and a multidimensional cation exchange and reversed-phase chromatography method for fractionating tryptic peptides, mass spectral analysis has identified more than 800 proteins and found up to 10-fold differences in expression.[28]

Recently, metabolic labeling has been extended from coding at the cell culture level to multicellullar organisms and mammals. In separate experiments, coding was achieved in the fruit fly *D. melanogaster* and nematode *C. elegans* by feeding them *S. cerevisiae* and *E. coli* respectively that had been grown in $^{15}N$ enriched media.[33] Control fruit flies and nematodes were grown on nonenriched *S. cerevisiae* or *E. coli*. Up to 98% enrichment in $^{15}N$ was demonstrated. In a further validation of the method, protein expression was compared in two strains of *C. elegans*, with and without a germ line. The anticipated absence of sperm protein was observed. Metabolic labeling has even been used with small mammals. Rat littermates were differentially grown on $^{15}N$ enriched and depleted media for 44 days, respectively.[34] One was treated with cycloheximide, while the other was used as a control. Using multidimensional liquid chromatography to fractionate a tryptic digest of the combined control and experimental proteomes, 369 proteins from liver were detected and relative changes in protein expression quantified.

Growing cells on media devoid of heavy isotopes has also been used to produce coded proteins. Since isotopic enrichment/depletion is achieved throughout the protein sequence, the technique can be used to globally address various protein quantitation and modification issues. The effect of Cd(II) stress on protein expression in *E. coli* was studied by growing stressed cells in $^{13}C$, $^{15}N$, and $^{2}H$ depleted media while unstressed cells were grown in normal media.[30] The two samples were mixed, digested, fractionated, and the isotope ratios of coded peptides used to determine the degree to which Cd(II) stress impacted expression. A similar approach was used to study phosphorylation.[29]

Metabolic coding meets all of the criteria required of an ideal coding system with a few exceptions. One is that studies are precluded in human subjects. The other is that the mass window between coded isoforms of peptides is not uniform. The number of heavy isotopes incorporated into a peptide will vary with amino acid composition and molecular weight. This makes it more difficult to recognize the coded isoforms of a peptide. This problem is exacerbated by that fact that the mass increment with an average peptide can be 12−14 amu. The fact that some cell types will not grow in isotopically enriched or depleted media is a problem.[35] Still another concern is differential transport (migration) of the coding agent between cellular compartments and organs.

**3.1.2 Derivatization of Primary Amines.** Amines are easily derivatized, providing an in vitro means for coding all peptides derived from a proteome except C-terminal arginine-containing peptides that are N-terminally blocked. A host of amine derivatizing agents have now been described for stable isotope coding in proteomics, including tandem mass tags (TMTs),[23] activated esters of 4-trimethylammoinium butyrate (TMAB),[11] and an activated benzoate ester.[18] The advantage of benzoate derivatization is that sequence coverage is increased by retention of low molecular weight, hydrophilic peptides. Most of the reagents noted above are available in $^{13}C$ isoforms that show no chromatographic isotope effect. An attractive feature of amine coding reagents is that excess reagent remaining after derivatization does not interfere with subsequent chromatographic steps.

**Acylation.** Coding by acylation is a global coding strategy often referred to as a global internal standard technique

(GIST).[36] Peptide acylation in water at neutral pH is generally achieved with an activated form of an acid, such as an *N*-hydroxysuccinimide (NHS) ester, or catalysis with a water-soluble carbodiimide. Adding a sulfonic acid group to succinimide increases the solubility of NHS esters. N-Terminal amines and the $\epsilon$-amino group of lysine are derivatized under these conditions. This means that tryptic peptides with a C-terminal lysine are coded differently than those with a C-terminal arginine and are recognizable in mass spectra as having double the mass shift between coded isoforms. One of the problems with acylation is the partial esterification at tyrosine. But these esters are easily hydrolyzed by the addition of a few drops of hydroxylamine at the end of the acylation reaction.[37] Using parallel enzyme assays,[38] quantification by acylation has been shown to have a linear dynamic range greater than 3 orders of magnitude.[39] A detection range of $6 \times 10^4$ has been seen with MALDI−MS.

There is also the option of labeling only the N-terminus of peptides by derivatizing proteins with succinic anhydride before proteolysis.[40] After proteolysis, the only primary amines available for acylation or alkylation will be at the N-terminus of peptides. Because trypsin does not cleave proteins at succinylated lysine residues, some of the peptides obtained are relatively large. Guanidination of the $\epsilon$-amino group on lysine is another means of differentiating between these two types of amino groups. Because the $\alpha$-amino group at the N-terminus of a peptide has a different p*K* than the $\epsilon$-amino group on lysine, it is possible to selectively derivatize the $\epsilon$-amino group on lysine. This leaves the $\alpha$-amino group free for coding. The attractive feature of this procedure over acylation with acetate or propionate alone is that the guanidino group enhances ionization and detection sensitivity of lysine-containing peptides.

Although acylation allows global coding of proteome digests, there is still the issue of sample complexity. This issue can be dealt with in either of two ways. One is to use a much higher resolution instrument, such as one with an ion mobility separator (IMS) between the reversed-phase chromatography column and mass spectrometer.[41] Peptides leaving the RPC column are electrosprayed into either an IMS directly[42] or stored in an ion trap and then loaded into the IMS.[43] Although the IMS has some similarity to a time-of-flight (TOF) mass spectrometer, it operates at much higher pressure and separates peptides largely on the basis of their conformation. Peak capacity of the IMS is roughly 50[41] with drift times of less than 20 ms.[44] When coupled to a TOF−MS, it is possible to take mass spectra of each of the peaks from the IMS. An IMS-TOF−MS alone has an analytical capacity of approximately 1000 peptides. When coupled to a high-resolution RPC column the analytical capacity of this system will increase 300−500-fold. Peptides differentially coded through acylation are readily quantified with an IMS-TOF−MS instrument.[45]

A second solution to the sample complexity issue is to select a portion of the proteome for expression analysis. When immobilized metal affinity chromatography (IMAC) columns are loaded with Cu(II) they have been shown to selectively bind histidine-containing peptides and provide significant reduction in sample complexity.[46] Selectivity is impacted by both the column support matrix and derivatization. The binding of nonhistidine containing peptides is almost eliminated by the use of very polar support matrixes such as polysaccharides and silica. Reduction of charge at the N-terminus through acylation has also been shown to impact selectivity in Cu(II)IMAC. Acylation diminished the selection of peptides with a single histidine residue, skewing the selection toward larger peptides and those with multiple histidine residues. Acylation with 4-trimethylammonium butyrate overcomes this problem by replacing the charge at the N-terminus of peptides. Quantification of expression through a double selection strategy has also been described.[25] Cu(II)IMAC was used to select histidine-containing peptides from a tryptic digest of an *E. coli* lysate in the first dimension, whereas covalent chromatography of cysteine-containing peptides was used in the second dimension of selection.

Acylation has been used in studies of peptide expression,[47] post-translational modifications,[48,49] single amino acid polymorphism,[50] and protein cross-linking.[51,52] The problem of quantifying isobaric peptides has been dealt with by performing comparative proteomics in the second dimension of mass spectrometry.[36]

A disadvantage of acylation is that it can compromise MALDI−MS detection sensitivity with C-terminal lysine-containing peptides that lack either an additional arginine or histidine residue to give them a positive charge. Derivatization with succinate provides the most serious problem by both neutralizing a positive charge and making the peptide more electronegative. This can easily be circumvented by (i) using a quarternary amine coding agent, such as TMAB (ii) or by guanidination of lysine before acylation.[53] Acylation seems to have less impact on sensitivity in ESI−MS.

**Other N-Termini Labeling Reagents.** There are a variety of other reactions that can be used for coding primary amines. One is the reaction with 2,4 dinitrofluorobenzene.[51] Another is the reaction with phenylisothiocynate (PIC) to form a phenyl carbamylate.[54] Yet another is a Schiff base formation with a coded aldehyde pair followed by reduction to a secondary amine.[55] None of these have been widely used for coding.

**3.1.3 Oxygen-18 Labeling.** A hydroxyl group from water is introduced into the carboxyl group formed during amide bond hydrolysis. When proteolysis is carried out in $H_2{}^{18}O$, all peptides will be labeled except the peptide originating from the carboxy-terminus of the protein. When hydrolysis of control and experimental samples is carried out in $H_2{}^{16}O$ and $H_2{}^{18}O$, respectively, the peptides are differentially coded according to sample origin. Opoid peptides were among the first to be compared in this way.[56] This approach has since been used to code peptides in studies of C-terminal characterization,[57] expression,[58] post-translational modifications,[22] de-novo sequencing,[59] protein:protein interactions,[60] and to identify disulfide linkages.[61]

Trypsin,[62−63] chymotrypsin,[64] lys-C,[58,62] and glu-C[22] can be used to code the C-termini of cleaved peptides. These proteases are reported to introduce two $^{18}O$, i.e., +4 amu into peptides in a two-step process during the course of proteolysis in pure $H_2{}^{18}O$. Because these enzymes form an ester intermediate with peptides during peptide bond hydrolysis, the first mole of $H^{18}O$ is incorporated during hydrolysis of the peptide from the enzyme. After proteolysis is complete, these enzymes continue to reversibly form esters with the amino acid at the C-terminus of the peptide. During the course of many cycles of esterification and hydrolysis in $H_2{}^{18}O$, all $^{16}O$ is eventually displaced from the C-terminal carboxyl. Unfortunately, the rate of exchange differs with peptide size,[62,65] type of amino acid,[64] between enzymes[58] and with peptide sequence.[62,65] The rate of exchange is also impacted by the presence of high concentrations of urea during proteolysis.[38] It is even possible to $^{18}O$ label peptides

after proteolysis by a back exchange reaction in which $H_2{}^{16}O$ is removed and replaced with $H_2{}^{18}O$.[64] An issue with $^{18}O$-based quantification of expression is the need to evaporate samples to dryness before the introduction of $H_2{}^{18}O$. Redissolution of peptides is sequence dependent and prone to losses, requiring special sample preparation methods.[65] There is also the problem that peptides from the C-terminus of proteins appear as singlets. These singlets can be interpreted as having arisen from large changes in expression.

**3.1.4 TACT (Tagging Amino and Carboxyl groups at peptide Termini).** Because $^{18}O$ coding does not label peptides arising from the C-terminus of proteins and none of the primary amine coding procedures label N-terminally blocked tryptic peptides with an arginine residue at their C-terminus, none of the in vitro coding procedures discussed to this point is truly universal. This has been addressed by combining the $^{18}O$ and amino acylation coding strategies in a procedure referred to as tagging amino and carboxyl termini (TACT).[66] Although TACT can be used for comparative expression analysis and aids in de-novo sequencing, the technique is of greatest utility in recognizing single amino acid polymorphism (SAAP). SAAPs appear in the spectra of TACT coded mixtures as singlets while all other peptides appear as doublet clusters of ions.

**3.1.5 Esterification.** Esterification of carboxyl groups in peptides has been used to enhance de-novo sequencing and in retention of hydrophilic peptides in liquid chromatography based methods.[67] This technique was recently re-visited as a method for stable isotope coding. Differentially coding control and experimental samples with $CH_3OH$ and $CD_3OH$ respectively has been used to study the expression of lipid raft associated proteins.[68] Similar studies have been done with ethanol coding as a means to improve the identification of proteins.[69] The technique is global in approach and shares the advantages linked to other global coding methods. Peptides devoid of a carboxyl group will obviously not be coded. This would be the case in C-terminally blocked peptides lacking glutamate or aspartate residues. The concern with this method is that peptides with 3−4 carboxyl groups will be difficult to esterify[70] and esters could be partially hydrolyzed by acidic mobile phases during reversed-phase chromatography.

**3.2 Targeting Specific Amino Acids.** With complex proteomes, targeting structural features is an attractive strategy for simplifying mixtures and bringing them into conformity with the analytical capacity of LC−MS systems. Moreover, knowledge of amino acid composition reduces the size of the database that must be searched. A number of amino acids viz. cysteine, lysine, N-terminal serine/threonine, tryptophan and methionine have been targeted in various in vitro methods discussed below.

**3.2.1 Metabolic Incorporation of Stable Amino Acids.** Like the use of enriched/depleted $^{15}N$ and $^{13}C$ media, in vivo incorporation of isotopically coded amino acids has been a popular method to perform comparative proteomics. Selected amino acids are incorporated into proteins by growing cell lines that are auxotrophic in the chosen amino acids. One cell line is grown in normal media and another in media rich in an isotopically coded amino acid. After a period of time, the two cell lines are harvested and analyzed in the same manner as other quantification studies based on metabolic labeling.

Both essential and nonessential amino acids have been quantitatively incorporated into cell lines.[71] These methods have come to be known by the acronyms SILAC[19,72] (Stable Isotope Labeling by Amino acids in Cell culture) and AACM[71] (Amino Acid Coded Mass tagging). The inherent advantage of this type of metabolic labeling is that it provides predictable mass shifts between peptide pairs. Moreover, the mass difference of differentially coded peptide pairs provides information on the number of amino acids in a peptide, thereby constraining database searches. Choice of the amino acid is crucial. Amino acid abundance will determine the number of tryptic peptides in mixtures of control and experimental samples that are differentially labeled. Lysine is a popular choice[73] due to its relatively high abundance, the fact that proteases such as lys-C and trypsin produce peptides with a single lysine residue, and the lack of metabolic scrambling of the label during protein synthesis. Leucine is another abundant amino acid that gives broad labeling of a proteome.[35,74,75]

The method also has several limitations**.** It is impossible to globally code all of the proteins in a proteome, not even with mixtures of labeled amino acids.[76,77] The use of mixtures also leads to sequence dependent mass shift, thereby complicating spectral interpretation. Metabolic scrambling and the concomitant uncertainty introduced in quantification is another issue. The amino acid being incorporated should be at the end of a metabolic pathway to avoid conversion to other amino acids. For example, catabolism of $^{13}C$-arginine into $^{13}C$-proline via the arginase pathway in HeLa cells led to the labeling of both proline and arginine residues in proteins.[19] Glycine is also reported to scramble to other amino acids.[78] This phenomenon can result in an unpredictable isotope dilution, partial loss of labeling, and a difference in mass shift. A possible way to overcome this problem is by use of amino acids: lysine, leucine, methionine, tyrosine, and valine which have reported to experience no isotope scrambling.[78] Another possible way is by in vitro incorporation of amino acids in a cell-free protein synthesis system.[79] Yet another problem is the inability to use the technique on human subjects. Finally, for this technique to be of utility in PTM studies, there must be a global correlation between an amino acid and a particular post-translational modification (PTM), such as aspartic acid and N-glycosylation.

**3.2.2 Cysteine.** The low abundance of cysteine and high nucleophilicity of its sulfhydryl group make it an attractive choice for residue specific coding. According to *in silico* analysis, 91.6% of the proteins in the yeast proteome contain cysteine while only 9.3% of the tryptic peptides derived from yeast proteins will have a cysteine residue.[25] Because proteins are most easily related to DNA databases through their peptides, it is generally necessary that they be reduced and alkylated before proteolysis. During alkylation, moieties such as biotin have been introduced that later facilitate the selection of alkylated cysteine containing peptides from proteolytic digests of a proteome with avidin.[80−82] Isotopic coding of cysteine in proteins[83] began with the use of $d_0/d_3$ isoforms of acrylamide to improve the identification of cysteine peptides during mass spectrometry. Combining both a biotin affinity tag and isotope coding in a single alkylating agent led to what is now known as the ICAT (Isotope Coded Affinity Tag) reagent.[26,27,84,85] Through this reagent, it is possible to both isotopically code and affinity select cysteine-containing peptides. Although a variety of other reagents have been described for targeting cysteine-containing peptides,[86−90] the ICAT reagent is by far the most popular.

The initial ICAT reagent had an iodoacetyl group tethered to a biotin tag via a deuterated/non deuterated oxyethylene linker.[26] Control and perturbed samples were separately alky-

lated using light and heavy isoforms of the reagent, respectively. The differentially coded samples were then mixed and subjected to trypsin digestion. After a cation exchange chromatography step to remove un-reacted biotinylating reagent, biotin tagged peptides were selected with an avidin (monomer) column. An avidin monomer column is generally used to reduce the binding affinity of biotinylated peptides and increase their recovery. Peptides recovered from the avidin column are then fractionated further on a reversed-phase chromatography column before mass spectral analysis. This process allows simultaneous simplification and quantification. A few examples of successful application of the technique are in the study of low abundance proteins,[91] membrane bound proteins,[92] and a comparative expression analysis of 145 proteins from *Pseudomonas aeruginosa*, a pathogen colonizing the airways of patients with cystic fibrosis.[93]

Because the ICAT reagent is commercially available and the method has been so widely used, it has probably been more critically evaluated than other methods. A series of limitations noted in the literature will be discussed below. One is that some proteins do not contain cysteine and will thus be missed.[94] While true, this problem is not unique to ICAT. All methods that target low abundance amino acids have this limitation. This is the reason that more than one amino acid should be targeted to cover the entire proteome of an organism. Another is that it cannot be used to detect post-translational modifications unless they contain cysteine. Again this is true of all methods that target specific amino acids. Targeted amino acid methods are better for looking at expression. A third issue is that alkylation with iodoacetamide based alkylating reagents may not completely alkylate all cysteine residues in a protein.[95, 96, 97] It is also probable that the reaction is not completely specific for cysteine and that partial reaction with other amino acids might occur. Another complaint is that not all cysteine-containing peptides are captured.[98] Although these are important, serious issues, the degree to which they impact, or compromise quantification is not clear. One of the great advantages of internal standard based methods is that they will compensate for massive problems as long as both control and experimental samples experience the problem the same way. When dealing with mixtures that contain half a million peptides, it will be easy to find peptides that do not "fit the rule".

The initial commercial ICAT reagents were a pair of $d_0$ and $d_8$ labeled isoforms. A problem with these reagents was that peptide isoforms created by differential derivatization with this reagent pair could be partially resolved during reversed-phase chromatography. That problem has now been addressed by going to a $9^{13}C$ coded reagent in place of the initial $d_8$ material. The chromatographic isotope effect is eliminated with the $9^{13}C$ reagent.[99] Another problem with the initial reagent was that the tethered biotin fragmented from peptides and complicated the interpretation of MS/MS spectra. That problem has now been solved by introducing an acid cleavable site in the tether between biotin and iodoacetamide that allows the removal of biotin before MS/MS analysis. The number of proteins detected with this second-generation cleavable reagent was larger than with the first generation reagent.[16] The new cleavable isotope tag shows promise.

**ICROC (Isotope Coded Reduction Off of a Chromatographic support).** This is a technique in which cysteine residues in proteins are tagged with biotin by disulfide exchange.[100,101] After removal of excess biotinylating reagent, tagged proteins are digested and biotinylated peptides are captured on avidin beads. The disulfide bonds linking peptides to the beads are then reduced and alkylated with an isotopically coded alkylating agent such as *N*-ethyliodoacetamide. The cysteine-containing peptides thus purified and coded are then quantified with a differential coding scheme using control and experimental samples. Insufficient data is currently available to fully evaluate this method.

**Solid-Phase Based Coding of Cysteine.** This approach exploits cleavable covalent chemistry for removal of isotopic tags tethered to a suitable solid support. Unlike solution based cysteine derivatization, isotope labeling in solid-phase methods is achieved after proteolysis to circumvent steric constrains. One such reagent, termed ALICE (Acid Labile Isotope Coded Extractants) consists of a maleimide group attached to an isotopically coded linker.[102] Isoforms of the linker are tethered to a polymeric resin via an acid labile amide. Post labeling, experimental, and control samples are mixed, and the maleimide group (attached to the isotopic linker) is cleaved under mild acidic conditions. A modest 90% recovery of cysteine-containing peptides was reported. In another approach, an iodoacetamide based cysteine specific reagent was used.[103] Isoforms of leucine ($d_0/d_7$) were derivatized with an iodoacetyl group at their C-terminus and a photocleavable *o*-nitrobenzyl linker at their N-terminus and then attached to aminopropyl coated glass beads. Control and experimental samples were separately digested, selected and labeled, and then mixed. The nitrobenzyl linker was then cleaved by exposure to UV light for 2 h. A comparison of this reagent with the original solution based ICAT reagent revealed an increased coverage of yeast proteins with the solid-phase technique. Both of these strategies allow simultaneous coding and selection of cysteine peptides. An additional advantage is that rigorous wash conditions can be applied to peptides captured on the solid-phase sorbent, eliminating nonspecific binding.

**3.2.3 Lysine.** One of the problems with MALDI−MS of tryptic peptides is that arginine-containing peptides ionize much better than those that contain lysine alone.[104] It is postulated that the higher ionization efficiency of arginine-containing peptides is due to very high basicity of guanidino functionality in side chain of arginine residue. Conversion of lysine to homoarginine by reaction of the $\epsilon$-amino group with *O*-methylisourea[105] enhances the ionization efficiency of lysine-containing peptides[106−108] and leads to an increase in the number of peptides identified in samples. Thirty percent more proteins from the *Caulobacter* proteome were identified after guanidination.[109] Versions of this approach are known as either MCAT[110] (Mass-Coded Abundance Tags) or QUEST[111] (Quantitation Using Enhanced Sequence Tags). Beyond increasing the representation of lysine containing peptides in LC−MS analyses of tryptic digests, tagging at lysine residues aids de-novo sequencing of peptides by facilitating the identification of "y" ions. Confusion over the assignment of "b" and "y" ions is precluded because mass coded doublets reveal "y" ions. Guanidination also makes it easier to differentiate between glutamine and lysine residues that are very close in mass.

Guanidination of lysine has also been coupled with N-terminal derivatization to facilitate de-novo sequencing, such as in the case of acylation with chlorosulfonylacetyl chloride.[112] Guanidination increases the ionization efficiency while sulfonation has been shown to enhance charge-site initiated cleavage of amide bonds, thereby aiding "y" ion detection. Derivatization with phenylthiocarbamyl (PTC)[113] is another way to impact fragmentation at the N-terminal amide bond.

Most of the internal standard quantification methods discussed to this point involved the use of stable isotope coding. In contrast, there are cases where coding has been achieved by derivatization of one sample and not the other, as in mass-coded abundance tagging (MCAT). Derivatization of one sample with *O*-methylisourea and not the other causes a 42 amu difference between coded peptides. Guanidination of lysine residues with amidine moieties differing by a methyl group has also been used.[111] Addition of an extra methyl group does not alter the ionization efficiency of tagged peptides in this case. Concerns relating to the guanidination approach are that the liquid chromatography and mass spectral behavior of isoforms will differ, excess methylisourea interferes with peptide ionization and has to be removed,[114] derivatization may be aberrant with low and high molecular weight peptides,[110,108] and N-terminal glycine residues are modified.[115]

**An Imidazole Based Reagent.** In another approach, derivatization of lysine with a methoxy-imidazole moiety (2-methoxy-4,5-dihydro-1H-imidazole) has been used to increase the ionization efficiency of lysine-containing peptides.[116] Like guanidination, imidazole derivatization improved sequence coverage. But, unlike guanidination, no side reaction was reported at the N-terminus of peptides. When proteome extracts are first digested with lys-C and then derivatized, all peptides in the digest will be labeled except those arising from the C-termini of proteins. Arginine residues will be included in lys-C peptides. It is also important to note that some of the lys-C peptides will be larger than tryptic peptides. Whether this approach will be of value in protein and peptide quantification remains to be determined.

**3.2.4 N-Terminal Threonine and Serine.** The presence of a primary amine and hydroxyl groups on adjacent carbon atoms in peptides with an N-terminal serine or threonine residue allows them to be cleaved by periodate oxidation. This fact has been exploited to select peptides with this structural feature.[117] Ninety-eight percent of the proteins from *C. elegans* have at least one peptide containing an N-terminal serine or threonine residue. Periodate oxidation of N-terminal serine and threonine produced an aldehyde on what was originally the α-carbon that in turn reacts with biocystin hydrazide. Biotin coded peptides were subsequently selected with streptavidin coated beads. The method, however, will cause simultaneous oxidation of methionine residues and cis-diols in glycopeptides. With appropriate stable isotope labeling, this method has the potential to be a useful quantitative methodology but the problem of glycopeptide oxidation must be overcome.

**3.2.5 Tryptophan.** The reaction of tryptophan and cysteine with 2,4-dinitrobenzenesulfenyl chloride is well-known.[118−120] This reaction has recently been employed to code tryptophan-containing peptides through the use of a $6^{13}C$ isoform of 2-nitrobenzenesulfenyl chloride.[17] After derivatization of tryptophan and cysteine residues, polypeptide mixtures were reduced and alkylated, removing the coding agent from cysteine residues and leaving only tryptophan residues with the specific tag. Although not described, it is likely that a nitrobenzene targeting antibody could be used to select peptides that contain tryptophan residues.

**3.2.6 Methionine.** Methods have been developed to select methionine-containing peptides but they have not been applied in isotope based quantification. In one approach, methionine peptides were selected using bromoacetyl groups tethered to glass beads. A sulfonium ion was quantitatively formed under acidic conditions. Captured peptides were released by nucleophillic substitution with mercaptoethanol under alkaline conditions. Methionine selection was coupled with N-terminal labeling to quantify expression and metalloprotease-mediated shedding of mouse cell surface proteins.[100]

## 4. Quantification of Post-translational Modifications

Proteins are often modified in some way after leaving the ribosome, either (1) through derivatization with new functional groups as in the cases of phosphorylation, sulfonation, acetylation, and methylation; (2) addition of larger structural units such as ubiquitin, a prenyl moiety, monosaccharides, glycans, or a lipid; (3) by proteolytic processing or degradation; (4) deglycosylation; or (5) some form of nonenzymatic modification such as glycation, functional group oxidation, or oxidative cleavage. These are but a few of the roughly one hundred types of post-translational modifications that have been described.

Quantification of post-translational modifications is accompanied by a series of complicating issues. One is that proteins may exist in modified and un-modified forms. It can be difficult to differentiate between them. A second is that there can be multiple, unrelated modifications in the same protein, either at the same site or at multiple sites. For example, a protein may occur in a large number of glycoforms. This makes the differentiation problem even more difficult. A third issue is that further steps of post-translational modifications at the same site, as in the case of glycan processing, may or may not be related. Finally, some forms of post-translational modification at a site may be disease related while others are not.

Proteomics is heavily dependent on DNA databases for protein identification. Through in silico transcription and translation of DNA sequences from databases, protein sequence can be quickly predicted and protein parents identified based on matching either their predicted tryptic peptide mass fingerprint or the sequence of a small number of signature peptides to experimental data. A major problem with this approach in the case of post-translationally modified proteins is that one must know that a tryptic peptide is carrying a PTM and understand the nature of the modification before the mass of these modified peptides can be predicted from DNA databases. This is almost never the case. It is in this context that PTM targeted selection of modified tryptic peptides is of so much interest.

**4.1 Glycosylation.** Among the many types of PTM, glycosylation occurs most frequently. The glycoproteome plays a key role in cellular regulation and function[121] in addition to being prominently associated with disease. Alzheimer's disease, diabetes, stress, some autoimmune diseases, cystic fibrosis, arthritis, cancer, certain types of heart disease, respiratory illnesses, renal function diseases, and diseases related to cellular adhesion are all associated with aberrations in glycoproteins.[122−124] Proteomics, when coupled with glycobiology, provides a powerful new method to study glycosylation and glycopathologies.

Glycoproteins occur in two forms. O-linked glycosylation occurs through a post-translational attachment of either mono- or oligosaccharides to the hydroxyl group on serine, threonine, or occasionally hydroxyproline. O-glycosylation occurs independent of protein expression. N-linked glycosylation differs in that a glycan is attached to the amide group of asparagine at -Asn-X-Ser- or -Asn-X-Thr- sequences while the protein is still on the ribosome, i.e., during protein synthesis. Absence of N-glycosylation during protein synthesis occurs in only very rare cases associated with genetic diseases. This means that

the initial N-glycosylation is directly proportional to protein synthesis and is not really a post-translational modification. Both O- and N-linked oligosaccharides can undergo extensive post-translational processing involving the deletion or addition of individual sugars.[125,126] Multiple genes regulate this glycan processing through a series of enzymes located at different sites in cells.

Changes in glycoprotein concentration have been used to recognize and quantify changes in glycosylation. One way is to differentially code tryptic peptides with succinic anhydride[38] or succinimidyl acetate[127] using the GIST protocol combined with lectin based glycopeptide selection. Another is through periodate oxidation and covalent capture along with succinic anhydride coding.[128]

Lectin affinity chromatography is by far the most widely used and specific method for selecting glycoproteins and glycopeptides. Lectins are available that target either O-linked or N-linked oligosaccharides, broad structural features of a glycan, a single sugar residue, or even specific glycopathologies at both the glycoprotein and glycopeptide levels. [129,130] The long history of exploiting lectin specificity as a histological staining agent and in oligosaccharide structure analysis is of great value in proteomics. On the basis of lectin staining, it is well-known that aberrations in fucosylation and other forms of glycosylation occur in cancer and that which lectins recognize these cancer specific changes. But few of these aberrations have been connected to specific proteins. One way to address this problem is through the use of a lectin that targets a disease specific aberration in glycosylation, such as fucosylation (Figure 3). Immobilized *Lotus tetragonolobus* agglutinin (LTA) has been used to select fucose containing tryptic peptides from tryptic digests of blood and show that fucosylation of a large number of proteins was suppressed in lymphosarcoma patients during chemotherapy.[127]

One of the complications in glycoproteomics is that glycan heterogeneity in glycopeptides complicates direct MS/MS sequencing and identification of peptides. Deglycosylation by either enzymatic or chemical means is necessary before peptide identification is possible, as noted above. The strong point of this approach is that a large number of well-characterized lectins are available. The weak point is that there will be types of glycosylation for which there is no specific lectin.

Another recently described route to select glycoproteins is by periodate oxidation and covalent capture of the resulting dialdehydes.[128] Periodate cleaves cis-diols in the oligosaccharide portion of glycoproteins to form aldehydes that were captured by a hydrazide resin. Following the capture of oxidized proteins on the resin, bound proteins were trypsin digested and non-glycosylated peptides eluted. Following derivatization with stable isotope coded succinic anhydride that was used later in quantification; the peptide portion of the glycopeptide was removed from the resin by hydrolysis with PNGase F, an enzyme specific for N-glycosylation. Released peptides were further fractionated by reversed-phase chromatography and identified by mass spectral analysis. All glycoforms that contain one or more diols will be oxidized by periodate. The strong point of this method is that a broad range of N-glycosylated isoforms of glycoproteins will be selected. The weak point will be the difficulty in differentiating between them.

Phenylboronates form a cyclic ester with vicinal diols, the rate of which depends on diol stereochemistry. This has led to extensive use of *m*-aminophenylboronate chromatography columns to select carbohydrates and release them under acidic

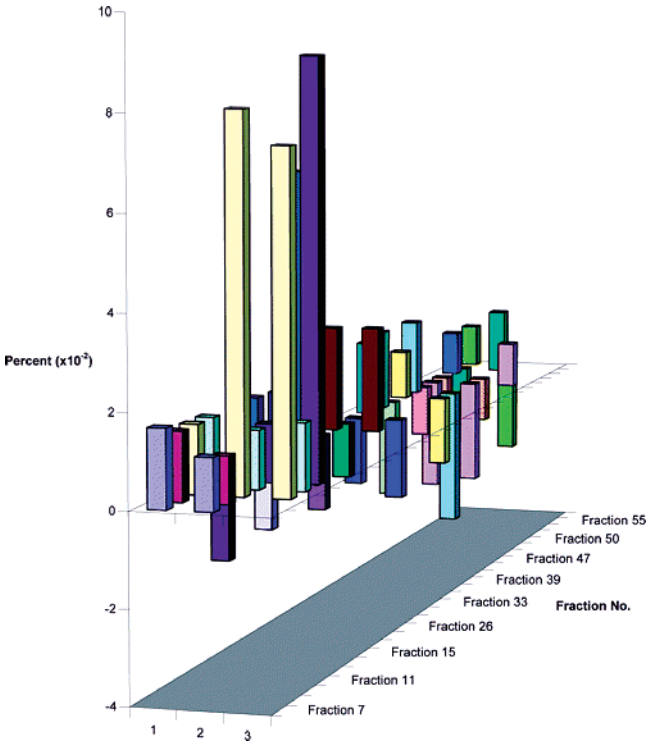



**Figure 3.** Differential display of chemotherapy-induced changes in the relative concentration of fucose-containing glycopeptides selected from canine serum. A total of 57 peptide pairs were observed to change in concentration. Among these, one-third changed more than 2-fold in concentration after chemotherapy. Concentration changes are represented as percentage change when concentrations before and after chemotherapy are compared. Peptides with a positive change decreased in concentration during therapy where those with negative values increased.[127] (Xiong, L.; Andrews, D.; Regnier, F. E. Comparative Proteomics of Glycoproteins Based on Lectin Selection and Isotope Coding *J. Proteome Res.,* ASAP).

conditions.[131,132] These columns have been used in the capture of ribonucleotides and should be of equal utility in glycoprotein proteomics.

Although there can be broad oligosaccharide heterogeneity at any one site in a glycoprotein, only a few of these isoforms may be associated with a disease. It is likely that in the future the quantification methods described above will be used to recognize glycoforms associated with regulatory events or disease and study their role in biological processes.

**4.2 Phosphorylation.** Rapid modulation of signaling is a key component of intracellular regulation and homeostasis. Signaling anomalies can lead to uncontrolled growth or death. In either case, the ability of cells to carry out their genetically determined role is compromised by faulty signaling. It is not surprising that almost 10% of all references in the phosphoprotein literature relate to disease. It is in this context that both phosphorylation and dephosphorylation of proteins in signaling pathways is of such great interest in biology, medicine, and drug discovery. In some respects, monitoring phosphoprotein dynamics is equivalent to "wire tapping" a cell.

There are two major classes of phosphoproteins; serine or threonine phosphorylated proteins and those that are phosphorylated on tyrosine. Other types of phosphorylation are seen infrequently or are transitory, as in the case of histidine phophorylation.[133] Tyrosine phosphorylated proteins constitute

roughly about ~0.5% of the phosphoproteome[134] and are thought to be involved in most crucial signaling. Phosphorylation at multiple sites is very important in regulation. This presents an analytical problem in that phosphopeptides with 3−4 modified serine/threonine or tyrosine residues may not be retained by reversed-phase columns or may not ionize well. Marked differences in the chemical properties of serine/threonine and tyrosine phosphate esters have been exploited in their determination. Phosphate esters of serine and threonine readily $\beta$-eliminate in base, whereas the tyrosine esters do not. There are also enzymes that differentially hydrolyze these two classes of phosphate esters.

Antibodies specific for either tyrosine phosphorylation or serine/tyrosine phosphorylation have been used extensively to select and recognize phosphoproteins, either through immunoprecipitation, immunosorbent chromatography, or in western blotting.[135,136] A recent paper describes a protocol for global selection of phosphotyrosine modified proteins using antiphosphotyrosine immunoprecipitation followed by trypsin digestion of the captured proteins, differential coding of control and experimental samples with $CH_3OH$ and $CD_3OH$ respectively, mixing the coded samples, IMAC selection of the phosphopeptides, and reversed-phase mass spectrometry of the selected phosphopeptides.[137] In general, antibodies targeting phosphorylation seem to be more specific for phosphoproteins than phosphopeptides.

Direct selection of all phosphopeptides from trypsin digests with Al(III), Fe(III), or Ga(III) loaded IMAC columns is another approach.[138] The problem with this method is that peptides with multiple aspartate or glutamate residues are also selected. Half or more of the selected peptides may not be phosphorylated. Several solutions to this problem have been explored. One is to methylate the carboxyl groups in phosphopeptides.[139] But the tendency of aspartate esters to internally cyclize under acidic conditions through aspartamide formation with the loss of methanol, makes them labile during reversed-phase chromatography with acidic mobile phases. Another approach is to digest with endoprotease glu-C.[140] This enzyme cleaves proteins at acidic residues and greatly reduces the number of acidic peptides. Ga(III)IMAC selection of endoprotease glu-C digests indicate that roughly 80% of the captured peptides are phosphorylated. Quantification in this method has been achieved with GIST and isotope coding of methyl esters.

Specific selection of serine/threonine peptides has been achieved by replacing the phosphate group with a moiety, such a biotin, which can be affinity selected.[141] Peptides that are O-phosphorylated on serine or threonine undergo $\beta$-elimination in base with the formation of a conjugated diene −(CH═CHC═O)−. Addition of 1,2-ethanedithiol to this alkene under basic conditions forms a derivative with a free sulfhydryl group. Alkylation of these thiol derivatized peptides with ICAT reagent places an affinity tag on peptides. Quantification can be achieved with the ICAT method, as has been shown in yeast. But O-glycosylated peptides also undergo $\beta$-elimination. Prior to application of this method the mixture must first be deglycosylated.

Another method for selecting all phosphopeptides through attachment of an affinity tag to phosphate residues has been described.[142] In this process, primary amines in peptides are first protected using the *tert*-butyl-dicarbonate (*t*-BOC) chemistry. An aqueous solution of the protected peptides is then treated with ethanolamine and a water-soluble carbodiimide to convert carboxyl groups to amides and phosphate groups to phosphoramidates. Phosphate groups in peptides were subsequently regenerated by treatment with weak acid and again converted to phosphoramidates with carbodiimide catalysis, but with cystamine in the second phosphoramidation. Reduction of cystamine generates a free sulfhydryl group on phosphate groups in peptides. Following removal of nonpeptide reactants through reversed-phase chromatography these sulfhydryl-containing peptides are covalently captured by allowing them to react with iodoacetyl groups attached to glass beads. Phosphoramidate bonds in the captured peptides are again cleaved with trifluoroacetic acid, but at a concentration that cleaves the *t*-BOC protecting group as well. Phosphopeptide yield was roughly 20%.

**4.3 Nitration.** Nitrotyrosine plays a direct role in cellular signaling[143] and is frequently found in proteins. Fluctuation in the concentration of nitrotyrosine has been associated with oxidative stress related diseases,[144] diet,[145] and even aging.[146] Because nitrophenyl groups are immunogenic, it is possible to prepare antibodies that target nitrotyrosine and use them in the same manner as with phosphotyrosine.

## 5. Protein:Protein Interactions

Protein:protein interactions have been associated with various metabolic and signaling pathways.[147] N-Terminally directed acylating agents have been used to isotopically code proteins in the polyubiquitin[51] complex and between microtubule destabilizing protein (Op 18) and tubulin[52] that were covalently linked in vitro. After proteolysis, characteristic isotopic-patterns unambiguously detected cross-linked peptides in both cases.

At a more complex cellular level, tandem affinity purification (TAP) has been applied to purify protein complexes.[148,149] A common limitation of TAP has been the inability to distinguish nonspecific coprecipitation of proteins from specifically interacting proteins of interest. As high as 30% spurious interactions were estimated.[149] Isotope coding was recently applied to the study of expression in specifically interacting protein complexes in which the nonspecific background had been eliminated.[150] Protein−protein interactions in the epidermal growth factor receptor pathway were studied using the SILAC approach. Isotopically coded arginine ($^{13}C$) was differentially incorporated into epidermal growth factor (EGF) by stimulation of HeLa cells. It was shown that of the 228 proteins coprecipitated with TAP, 28 were overexpressed. On the basis of the specific stimulation of the EGF pathway, it could be concluded that these proteins had functional specificity in the EGFR complex.

In another study, specific protein components in macromolecular RNA Polymerase II preinitiation complex (PIC) were determined.[151] The TATA binding protein (TBP) was used to examine proteins specifically linked to the PIC complex. Only experimental samples were stimulated with TBP. This was followed by coding control and experimental samples with light and heavy forms of the ICAT reagent, respectively. Among the 206 proteins in the isolated pool, 49 were identified as specifically interacting with the macromolecular RNA Polymerase II preinitiation complex (PIC). Greater than 90% of the proteins associated with the PIC complex were quantified. This was due to the absence of cysteine containing tryptic peptide in the mass range used.

Both of these studies show how differential stimulation of a biological pathway can be used to identify proteins associated a pathway. This is a powerful way to overcome problems of nonspecific binding and false positives in protein:protein interaction studies. In view of the very large amount of

nonspecific binding in interaction systems, isotope coding strategies are a valuable asset.

## 6. Single Amino Acid Polymorphism

Single nucleotide polymorphism (SNP) mining is currently being pursued extensively at the genome level to address issues of genetic variation,[152] in evolution studies,[153] and in pursuit of "personalized medicine".[154] In the human genome, less than 1% of $2.1 \times 10^6$ SNPs are expected to result in variation at the protein level.[51] This means that an average of 7000 proteins are expected to vary in sequence between two individuals. Very few studies have been reported to address the single amino acid polymorphism (SAAP) issue at the proteome level.[155,156] The TACT strategy was recently applied to this problem.[51] Peptides that are identical appear as doublet pairs whereas those with a SAAP appear as singlets varying no more than ± 139 amu, i.e., the difference in mass between glycine and tryptophan.

## 7. Absolute Quantification

It was noted above that protein expression analysis depends on knowledge of the absolute concentration of a protein in the cellular pool. Absolute quantification of neuropeptides was reported using $^{18}O$ labeling[56] and isotopically coded synthetic peptide.[157] This idea has been extended to quantification of known proteins using a synthetic, isotopically coded peptide from the protein.[158−160] In a similar manner, membrane proteins were quantified by comparing their concentration to that of a chemically synthesized peptide present in the soluble region of the protein.[159] Using the same approach, low abundance yeast proteins and known post-translational modifications were quantified by synthesizing isotopically coded peptides containing the supposed PTM.[160] For complex mixtures, information on retention time and ionization efficiency of synthesized reference peptides eliminates any possible errors originating from peptide mass degeneracy. High digestion efficiency is critical to ascertain absolute quantification of a protein. It is desirable to subject the reference peptides to proteolysis, by incorporating an amino acid specific to the chosen protease within the chemically synthesized peptide, thereby minimizing errors resulting from alterations in digestion efficiency.

## Conclusions

There has been an explosion of activity in stable isotope quantification of proteins during the past four years. On the basis of the literature sited in this review, it can be concluded that there is both good news and bad news for life scientists wishing to use these methods to study biological systems. The good news is that there are now a wide variety of methods available for both relative and absolute quantification of proteins, these methods appear to be very powerful, they are designed to quantify large numbers of proteins simultaneously, and the chemistry involved in using them seems relatively simple. There is nothing in the literature of the past that comes close to these methods in potential utility. The bad news is that the bulk of the effort in quantification has been directed toward expression analysis alone, few of the requisite reagents are commercially available, the methods do not actually measure changes in the rate of expression, and few have been validated with complex mixtures to produce the same answer multiple times. But the worst news of all is that a single isotope ratio measurement gives a very incomplete picture of the regulatory status of a biological system. The rate of protein degradation and inter-compartmental transport, post-translational modifications, relative pool sizes of protein isoforms, absolute concentration of proteins, and the rate of isoform interconversion are but a few of the additional measurements required to understand how a given stimulus impacts a biological system. It is concluded that a series of new methods have to be developed and implemented to make existing stable isotope quantification methods more meaningful. As we are learning from systems biology, many natural processes are interrelated and single measurements are of limited value in interpreting biological phenomena.

Finally, we would conclude with the obvious. There is a pressing need to deal with all the issues above that are bad news. It would be nice if more reagents were commercially available. New methods that quantify the interconversion of protein isoforms would also be high on the list of needs. Most proteins undergo some form of post-translational modification in their lifetime, including degradation. The fact that there are more than a hundred possible post-translational modifications of proteins means there is a lot to be done. Quantification of proteins in intermolecular complexes is prominent on the list of needs, especially in view of the fact that it provides a way to differentiate between specific and nonspecific association. Improving the dynamic range and quantification of low copy number proteins is a further dream of the proteomics community. But probably the most critical need of all is to validate that the methods we already have are reproducible and transferable.

**Acknowledgment.** The authors gratefully acknowledge financial support from Grant No. 5R01 GM 59996-04.

## References

(1) Cho, A. K.; Lindeke, B.; Hodshon, B. J.; Jenden, D. J. *Anal. Chem*. **1973**, *45*, 570−574.
(2) Horning, M. G.; Stillwell, W. G.; Nowlin, J.; Leltratanangkoon, K.; Carroll, D.; Dzidic, I.; Stillwell, R. N.; Horning, E. C.; Hill, R. M. *J. Chromatogr.* **1974**, *91*, 413−423.
(3) Cowen, A. E.; Hofmann, A. F.; Hachey, D. L.; Thomas, P. J.; Belobaba, D. T.; Klein, P. D.; Tokes, L. *J. Lipid Res.* **1976**, *17*, 231−238.
(4) Heck, H.; Buttrill, S. E., Jr.; Flynn, N. W.; Dyer, R. L.; Anbar, M.; Cairns, T.; Dighe, S.; Cabana, B. E. *J. Pharmacokinet. Biopharm*. **1979**, *7*, 233−248.
(5) Schena, M.; Shalon, D.; Davis, R. W.; Brown, P. O. *Science* **1995**, *270*, 467−470.
(6) Lashkari, D. A.; DeRisi, J. L.; McCusker, J. H.; Namath, A. F.; Gentile, C.; Hwang, S. Y.; Brown, P. O.; Davis, R. W. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 13057−13062.
(7) DeRisi, J. L.; Iyer, V. R.; Brown, P. O. *Science* **1997**, *278*, 680−686.
(8) Pratt, J. M.; Petty, J.; Garcia-Riba. I.; Robertson, D. H. L.; Gaskell, S. J.; Oliver, S. G.; Beynon, R. J. *Mol. Cell Proteomics* **2002**, *1*, 579−591.
(9) Filer, C. N. *J. Labeled Compd. Radiopharm*. **1999**, *42*, 169−197.
(10) Zhang, R.; Sioma, C.; Wang, S.; Regnier, F. E. *Anal. Chem*. **2001**, *73*, 5142−5149.
(11) Zhang, R.; Sioma, C.; Thompson, R. A.; Xiong, L.; Regnier, F. E. *Anal. Chem*. **2002**, *74*, 3662−3669.
(12) Conrads, T. P.; Alving, K.; Veenstra, T. D.; Belov, M. E.; Anderson, G. A.; Anderson, D. J.; Lipton, M. S.; Pasa-Tolic, L.; Udseth, H. R.; Chrisler, W. B.; Thrall, B. D.; Smith, R. D. *Anal. Chem*. **2001**, *73*, 2132−2139.
(13) Zhang, R.; Regnier, F. E. *J. Proteome. Res.* **2002**, *1*, 139−147.
(14) Xiong, L.; Zhang, R.; Regnier, F. E., in preparation.
(15) Niwayama, S.; Kurono, S.; Matsumoto, H. *Bioorg. Med. Chem Lett*. **2003**, *13*, 2913−2916
(16) Oda, Y.; Owa, T.; Sato, T.; Boucher, B.; Daniels, S.; Yamanaka, H.; Shinohara, Y.; Yokoi, A.; Kuromitsu, J.; Nagasu, T. *Anal. Chem*. **2003**, *75*, 2159−2165.

(17) Kuyama, H.; Watanabe, M.; Toda, C.; Ando, E.; Tanaka, K.; Nishimura, O. *Rapid Commun. Mass Spectrom*. **2003**, 17, 1642−1650.
(18) Julka, S.; Regnier, F. E., in preparation.
(19) Ong, S. E.; Kratchmarova, I.; Mann, M. *J. Proteome. Res*. **2003**, *2*, 173−181.
(20) Wierenga, S. K.; Zocher, M. J.; Mirus, M. M.; Conrads, T. P.; Goshe, M. B.; Veenstra, T. D. *Rapid Commun. Mass Spectrom*. **2002**, *16*, 1404−1408.
(21) Berger, S. J.; Lee, S. W.; Anderson, G. A.; Tolic, L. P.; Tolic, N.; Shen, Y.; Zhao, R.; Smith, R.D. *Anal Chem*. **2002**, *74*, 4994−5000.
(22) Reynolds, K. J.; Yao, X.; Fenselau, C. *J. Proteome. Res.* **2002**, *1*, 27−33.
(23) Thompson, A.; Schaefer, J.; Kuhn, K.; Kiene, S.; Schwarz, J.; Schmidt, G.; Johnstone, R.; Neumann, T.; Hamon, C. *Anal. Chem*. **2003**, *75*, 1895−1904.
(24) Olsen, J. V.; Andersen, J. S.; Ong, S.-E.; Mann, M. In Abstracts of the 51st Conference in Mass Spectrometry and Allied Topics, **2003**.
(25) Wang, S.; Zhang, X.; Regnier, F. E. *J. Chromatogr. A* **2002**, *949*, 153−162.
(26) Gygi, S. P.; Rist, B.; Gerber, S, A.; Turecek, F.; Gelb, M, B.; Aebersold, R. *Nat. Biotechnol*. **1999**, *17*, 994−999.
(27) Aeberosld, R.; Gygi, S. P.; Griffin, T. J.; Han, D, K, M.; Yelle, M. *J. Am. Genom./Proteom. Tech*. **2001**, *1*, 26−27.
(28) Washburn, M. P.; Ulaszek, R.; Deciu, C.; Schieltz, D. M.; Yates, J. R., III. *Anal. Chem*. **2002**, *74*, 1650−1657.
(29) Oda, Y.; Huang, K.; Cross, F. R.; Cowburn, B.; Chait, B. T. *Proc. Acad. Nat. Acad. U.S.A.* **1999**, *96*, 6591−6596.
(30) Pasa-Tolic, L.; Jensen, P. K.; Anderson, G. A.; Lipton, M. S.; Peden, K. K.; Martinovic, S.; Tolic, N.; Bruce, J. E.; Smith, R. D. *J. Am. Chem. Soc*. **1999**, *121*, 7949−7950.
(31) Smith, R. D.; Anderson, G. A.; Lipton, M. S.; Tolic, L.P.; Shen, Y.; Conrads, T.P.; Veenstra, T. D.; Udseth, H. R. *Proteomics* **2002**, *2*, 513−523.
(32) Lahm, H. W.; Langen, H. *Electrophoresis* **2000**, *21*, 2105−2114.
(33) Krijgsveld, J.; Ketting, R. E.; Mahmoudi, T.; Johansen, J.; Sanz, M. A.; Verrijzer, C. P.; Plasterk, R. H. A.; Heck, A. J. R. *Nat. Biotechnol*. **2003**, *21*, 927−931.
(34) Wu, C. C.; MacCoss, M. J.; Howell, K. E.; Matthews, D. E.; Yates, J. R. In Abstracts of the 51st Conference in Mass Spectrometry and Allied Topics, **2003**.
(35) Jiang, H.; English, A. M. *J. Proteome Res*. **2002**, *1*, 345−350.
(36) Chakroborty, A.; Regnier, F. E. *J. Chromatogr. A* **2002**, *949*, 173−184.
(37) Wang, S.; Regnier, F. E. *J Chromatogr. A* **2001**, *924*, 345−357.
(38) Regnier, F. E.; Riggs, L.; Zhang, R.; Li, X.; Liu, P.; Chakraborty, A.; Seeley, E.; Sioma, C.; Thompson, R. A. *J. Mass Spectrom*. **2002**, *37*, 133−145.
(39) Ji, J.; Chakraborty, A.; Geng, M.; Zhang, X.; Amini, A.; Bina, M.; Regnier, F. E. *J. Chromatogr. B* **2000**, *745*, 197−210.
(40) Munchbach, M.; Quadroni, M.; Miotto, G.; James, P. *Anal. Chem*. **2000**, *72*, 4047−4057.
(41) Valentine, S. J.; Kulchania, M.; Srebalus Barnes, C. A.; Clemmer, D. E. *Int. J. Mass Spectrom*. **2001**, *212*, 97−109.
(42) Srebalus Barnes, C. A.; Hilderbrand, A. E.; Valentine, S. J.; Clemmer, D. E. *Anal. Chem*. **2002**, *74*, 26−36.
(43) Lee, Y. J.; Hoaglund-Hyzer, C. S.; Srebalus Barnes, C. A.; Hilderbrand, A. E.; Valentine, S. J.; Clemmer, D. E. *J. Chromatogr. B* **2002**, *782*, 343−351.
(44) Hoaglund-Hyzer, C. S.; Lee, Y. J.; Counterman, A. E.; Clemmer, D. E. *Anal. Chem*. **2002**, *74*, 992−1006.
(45) Kindy, J. M.; Taraszka, J. A.; Regnier, F. E.; Clemmer, D. E. *Anal. Chem*. **2002**, *74*, 950−958.
(46) Ren, D.; Penner, N. A.; Slentz, B. E.; Mirzaei, H.; Regnier, F. E. *J. Proteome. Res*. **2003**, *2*, 321−329.
(47) Che, F.-Y.; Fricker, L. D. *Anal. Chem*. **2002**, 74, 3190−3198.
(48) Riggs, L.; Seeley, E. H.; Regnier, F. E., in preparation.
(49) Xiong, L.; Regnier, F. E. *J. Chromatogr. B* **2002**, *782*, 405−418.
(50) Liu, P.; Regnier, F, E. *Anal. Chem*. **2003**, *75*, 4956−4963
(51) Chen, X. H.; Chen, Y. H.; Anderson, V. E. *Anal. Biochem*. **1999**, *273*, 192−203.
(52) Muller, D.R.; Schindler, P.; Towbin, H.; Wirth, U.; Voshol, H.; Hoving, S.; Steinmetz, M. O. *Anal. Chem*. **2001**, *73*, 1927−1934.
(53) Zhang, X.; Jin, Q.K.; Carr, S.A. Annan, R. S. *Rapid. Commun. Mass Spectrom*. **2002**, *16*, 2325−2332.
(54) Mason, D. E.; Liebler, D. C. *J. Proteome. Res.* **2003**, *2*, 265−272.
(55) Zhang, H.; Bart, B. M.; Eng, J.; Aebersold, R. In Mass Spectrometry and Allied Topics, **2003**.
(56) Desiderio, D. M.; Kai, M. *Biomed. Mass Spectrom*. **1983**, *10*, 471−479.
(57) Kosaka, T.; Takazawa, T.; Nakamura, T. *Anal. Chem*. **2000**, *72*, 1179−1185
(58) Yao; X.; Freas, A.; Ramirez, J.; Demirev, P. A.; Fenselau, C. *Anal. Chem*. **2001**, *73*, 2836−2842.
(59) Qin, J.; Herring, C. J.; Zhang, X. *Rapid Commun Mass Spectrom*. **1998**, *12*, 209−216.
(60) Back, J. W.; Notenboom, V.; de Koning, L. J.; Muijsers, A. O.; Sixma, T. K.; de Koster, C. G.; de Jong, L. *Anal. Chem*. **2002**, *74*, 4417−4422.
(61) Wallis, T. P.; Pitt, J. J.; Gorman, J. J. *Protein Sci*. **2001**, *10*, 2251−2271.
(62) Schnolzer, M; Jedrzejewski. P; Lehmann, D. *Electrophoresis*, **1996**, *17*, 945−953.
(63) Schevchenko, A.; Schevchenko, A. *Anal. Biochem*. **2001**, *296*, 279−283.
(64) Yao, X.; Afonso, C.; Fenselau, C. *J. Proteome Res*. **2003**, *2*, 147−152.
(65) Stewart, I.; Thomson, T.; Figeys, D. *Rapid Commun. Mass Spectrom*. **2001**, *15*, 2456−2465.
(66) Liu, P.; Regnier, F. E. *J. Proteome. Res.* **2002**, *1*, 443−450.
(67) Falick, A. M.; Matlby, D. A. *Anal. Biochem*. **1989**, *182*, 165−169.
(68) Goodlett, D. R.; Keller, A.; Watts, J. D.; Newitt, R.; Yi, E. C.; Purvine, S.; Eng, J. K Haller, P.-V.; Aebersold, R.; Kolker, E. *Rapid Commun. Mass Spectrom*. **2001**, *15*, 1214−1221.
(69) Kim, T.-Y.; Reilly, J. P. In Abstracts of the 51st Conference in Mass Spectrometry and Allied Topics, **2003**.
(70) Shevchenko, A.; Chernushevich, I.; Ens, W; Standing, K. G.; Thomson, B.; Wilm, M.; Mann, M. *Rapid. Commun. Mass Spectrom*. **1997**, *11*, 1015−1024.
(71) Pan, S.; Gu, S.; Bradbury, E. M.; Chen, X. *Anal.Chem*. **2003**, *75*, 1316−1324.
(72) Ong, S. E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D. B.; Steen, H.; Pandey, A.; Mann, M. *Mol. Cell Proteomics* **2002**, *1*, 376−386.
(73) Gu, S.; Pan, S.; Bradbury, E. M.; Chen, X. *Anal. Chem*. **2002**, *74*, 5774−5785.
(74) Pratt, J. M.; Robertson, D. H. L.; Gaskell, S. J.; Garcia, I. R.; Hubbard, S. J.; Siddhu, K.; Oliver, S. G.; Butler, P.; Hayes, A.; Petty, J.; Beynon, R. J. *Proteomics* **2002**, *2*, 157−163.
(75) Martinovic, S.; Veenstra, T. D.; Anderson, G. A.; Tolic, L. P.; Smith, R. D. *J. Mass Spectrom*. **2002**, *37*, 99−107.
(76) Zhu, H.; Pan, S, Gu, S.; Bradbury, E. M.; Chen, X. *Rapid Commun. Mass Spectrom*. **2002**, *16*, 2115−2123.
(77) Chen, X.; Smith, L. M.; Bradbury, E. M. *Anal. Chem*. **2000**, *72*, 1134−1143.
(78) Engen, J. R.; Bradbury, E. M.; Chen, X. *Anal. Chem*. **2002**, *74*, 1680−1686.
(79) Yabuki, T.; Kigawa, T.; Dohmae, N.; Takio, K.; Terada, T.; Ito, Y.; Laue, E. D.; Cooper, J. A.; Kainosho, M.; Yokoyama, S. *J. Biomol. NMR* **1998**, *11*, 295−306.
(80) Spahr, C. S., Susin, S. A.; Bures, E. J.; Robinson, J, H.; Davis, M. T.; McGinley, M. D.; Kroemer, G.; Patterson, S. D. *Electerphoresis* **2000**, *21*, 1635−1650.
(81) Deziel, M. R.; Mau, M. M. *Anal. Biochem*. **1990**, *190*, 297−303.
(82) Chollet, A.; Bonnefoy, J. Y.; Oderlatt, N. *J. Immunol. Methods* **1990**, *127*, 179−185.
(83) Sechi, S.; Chait, B. *Anal. Chem*. **1998**, *70*, 5150−5158.
(84) Smolka, M. B.; Zhou, H.; Purkayastha, S.; Aebersold, R. *Anal. Biochem*. **2001**, *297*, 25−31.
(85) Von Haller, P, D.; Yi, E.; Donohoe, S.; Vaughn, K.; Keller, A.; Nesvizhskii, A.I.; Eng, J.; Li, X.-J.; Goodlett, D. R.; Aebersold, R.; Watts, J. D *Mol. Cell Proteomics* **2003**, *2*, 428−442.
(86) Goodlett, D. R.; Bruce, J. E. Anderson, G. A. Rist, B.; Pasa-Tolic, L.; Fiehn, O.; Smith, R. D.; Aebersold, R. *Anal. Chem*. **2000**, *72*, 1112−1118.
(87) Adamczyk, N.; Gebler, J. C.; Wu, J. *Rapid. Commun. Mass Spectrom*. **1999**, *13*, 1813−1817.
(88) Sechi, S. *Rapid Commun. Mass. Spectrom*. **2002**, *16*, 1416−1424.
(89) Gehanne, S.; Cecconi, D.; Carboni, L.; Righetti, P. G.; Domenici. E.; Hamdan, M. *Rapid Commun. Mass Spectrom*. **2002**, *16*, 1692−1698.
(90) Niwayama, S.; Kurono, S.; Matsumoto, H. *Bioorg. Med. Chem Lett*. **2001**,*11*, 2257−2261.
(91) Gygi, S. P.; Rist, B.; Griffin, T. J.; Eng, J.; Aebersold, R. *J. Proteome Res.* **2002**, *1*, 47−54.
(92) Han, D. K.; Eng, J.; Zhou, H.; Aebersold, R. *Nat. Biotechnol*. **2001**, *19*, 946−951.
(93) Gulna, T.; Purvine, S. O.; Yi, E. C.; Eng, J.; Goodlett, D. R.; Aebersold, R.; Miller, S. I. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 2771−2776.
(94) Rabilloud, T. *Proteomics* **2002**, *2*, 3−10.

(95) Galvani, M.; Rovatti, L.; Hamdan, M.; Herbert, B.; Righetti, P. G. *Electrophoresis* **2001**, *22*, 2066−2074.
(96) Galvani, M.; Hamdan, M.; Herbert, B.; Righetti, P. G. *Electrophoresis* **2001**, *22*, 2058−2065.
(97) Mahmoud, H.; Righetti, P. G. *Mass Spectrom. Rev.* **2002**, *21*, 287−302.
(98) Patton, W. F.; Schulenberg, B.; Steinberg, T. H. *Curr. Opin. Biotech.* **2002**, *13*, 321−328.
(99) Hansen, K. C.; Schmitt-Ulms, G.; Chalkley, R.; Hirsch, J.; Baldwin, M. A.; Burlingame, A. L. *Mol. Cell Proteomics* **2003**, *2*, 299−314.
(100) Shen, M.; Guo, L.; Wallace, A.; Fitzner, J.; Eisenman, J.; Jacobson, E.; Johnson, R.S. *Mol. Cell Proteomics*, **2003**, *2*, 315−324.
(101) Johnson, R. S.; Wallace, A.; Guo, L.; Shen, M. In Proceedings of the 49th ASMS Conference in Mass Spectrometry and Allied Topics, **2001**.
(102) Qiu, Y.; Sousa, E. A.; Hewick, R. M.; Wang, J. H. *Anal. Chem.* **2002**, *74*, 4969−4979.
(103) Zhou, H.; Ranish, J. A.; Watts, J. D.; Aebersold, R. *Nat. Biotechnol.* **2002**, *20*, 512−515.
(104) Krause, E.; Wenschuh, H.; Jungblut, P. R. *Anal. Chem.* **1999**, *71*, 4160−4165.
(105) Kimmel, J. R. *Methods Enzymol.* **1967**, *11*, 584−589.
(106) Brancia, F. L.; Oliver, S. G.; Gaskell, S. J. *Rapid Commun. Mass Spectrom.* **2000**, *14*, 2070−2073.
(107) Hale, J. E.; Butler, J. P.; Knierman, M. D.; Becker, G. W. *Anal. Biochem.* **2000**, *287*, 110−117.
(108) Beardsley, R. L.; Karty, J. A.; Reilly, J. P. *Rapid Commun. Mass Spectrom.* **2000**, *14*, 2147−2153.
(109) Karty, J. A.; Ireland, M. M. E.; Brun, Y. V.; Reilly, J. P. *J. Proteome. Res.* **2002**, *1*, 325−335.
(110) Cagney, G.; Emili, A. *Nat. Biotechnol.* **2002**, *20*, 163−170.
(111) Beardsley, R. L.; Reilly, J. P. *J. Proteome Res.* **2003**, *2*, 15−21.
(112) Keough, T.; Lacey, M. P.; Youngquist, R. S. *Rapid Commun. Mass Spectrom.* **2000**, *14*, 2348−2356.
(113) Brancia, F. L.; Butt, A.; Beynon, R. J.; Hubbard, S.; Gaskell, S. J.; Oliver, S. G. *Electrophoresis* **2001**, *22*, 552−559.
(114) Brancia, F. L.; Openshaw, M. E.; Kumashiro, S. *Rapid Commun Mass Spectrom.* **2002**, *16*, 2255−2259.
(115) Beardshley, R. L.; Reilly, J. P. *Anal. Chem.* **2002**, *74*, 1884−1890.
(116) Peters, E. C.; Horn, D. M.; Tully, D. C.; Brock, A. *Rapid Commun. Mass Spectrom.* **2001**, *15*, 2387−2392.
(117) Chelius, D.; Shaler, T. A. *Bioconjugate Chem.* **2003**, *14*, 205−211.
(118) Schoffone, E.; Fontana, A.; Rocchi, R. *Biochemistry* **1968**, *7*, 971−979.
(119) Kharash, N.; Gleason, G.I.; Buess, C.M. *J. Am. Chem. Soc.* **1950**, *72*, 1796−1798.
(120) Havlik, A. J.; Kharash, N. *J. Am. Chem. Soc.* **1955**, *77*, 1150−1152.
(121) Rademacher, T. W.; Parekh, R. B.; Dwek, R. A. *Annu. Rev. Biochem.* **1988**, *57*, 785−838.
(122) Koscielak, J. *Acta Biochim. Pol.* **1995**, *42*, 1−10.
(123) Bhavanandan, V. P. *Glycobiology* **1991**, *1*, 493−503.
(124) Masters, C. L.; Beyreuther, K. *Arzneim.-Forsch.* **1995**, *45*, 410−412.
(125) Durand, G.; Seta, N. *Clin. Chem.* **2000**, *46* (6, Pt. 1), 795−805.
(126) Rudd, P. M.; Dwek, R. A. *Crit. Rev. Biochem. Mol. Biol.* **1997**, *32*, 1−100.
(127) Xiong, Li.; Andrews, D.; Regnier, F. E. *J. Proteome. Res.*, ASAP.
(128) Zhang, H.; Li, X.-J.; Martin, D. B.; Aebersold, R. *Nat. Biotechnol.* **2003**, *21*, 660−666.
(129) West, I.; Goldring, O. *Methods Mol. Biol.* **1996**, *59*, 177−185.
(130) Kobata, A.; Endo, T. *J. Chromatogr.* **1992**, *597*, 111−122.
(131) Jack, C. M. *Diabetologia* **1988**, *31*, 126−127.
(132) Gould, B. J.; Hall, P. M. *Clin. Chim. Acta* **1987**, *163*, 225−230.
(133) Napper, S.; Kindrachuk, J.; Olson, D.J. H.; Ambrose, S. J.; Dereniwsky, C.; Ross, A. R. S. *Anal. Chem.* **2003**, *75*, 1741−1747
(134) Zolodz, M.D.; Wood, K.V. *J. Mass Spectrom.* **2003**, *38*, 257−264.
(135) Ignatoski, K. M. *Methods Mol. Biol.* **2001**, *124*, 39−48.
(136) Kaufmann, H.; Bailey, J. E.; Fussenegger M. *Proteomics* **2001**, *1*, 194−199.
(137) Ficarro, S.; Chertihin, O.; Westbrook, V. E.; White, F.; Jayes, F.; Kalab, P.; Marto, J. A.; Shananowitz, J.; Herr, J. C.; Hunt, D. F.; Visconti, P. E. *J. Biol. Chem.* **2003**, *278*, 11579−11589.
(138) Posewitz, M. C.; Tempst, P. *Anal. Chem.* **1999**, *71*, 2883−2892.
(139) Ficarro, S. B.; McCleland, M. L.; Stukenberg, P. T.; Burke, D. J.; Ross, M. M.; Shabanowitz, J.; Hunt, D. F.; White, F. M. *Nat. Biotechnol.* **2002**, *20*, 301−305.
(140) Seeley, E. H.; Riggs, L.; Regnier, F. E., in preparation.
(141) Goshe, M. B.; Veenstra, T. D.; Panisko, E. A.; Conrads, T. P.; Angell, N. H.; Smith, R. D. *Anal. Chem.* **2002**, *74*, 607−616.
(142) Goshe, M. B.; Conrads, T. P.; Panisko, E. A.; Angell, N. H.; Veenstra, T. D.; Smith, R. D. *Anal. Chem.* **2001**, *73*, 2578−2586.
(143) Drew B.; Leeuwenburgh, C. *Ann. N. Y. Acad. Sci.* **2002**, *959*, 66−81.
(144) Greenacre, S. A.; Ischiropoulos H. *Free Radical Res.* **2001**, *34*, 541−581.
(145) Oldreive, C.; Rice-Evans C. *Free Radical Res.* **2001**, *35*, 215−231.
(146) Stadtman, E. R. *Ann. N. Y. Acad. Sci.* **2001**, *928*, 22−38.
(147) Figeys, D,; McBroom, L. D.; Moran, M. F. *Methods* **2001**, *24*, 230−239.
(148) Ho, Y.; Gruhler, A.; Heilbut, A.; Bader, G. D.; Moore, l.; Adams, S.-L.; Millar, A.; Taylor, P.; Bennett, K.; Boutilier, K.; Yang, L.; Wolting, C.; Donaldson, I.; Schandorff, S.; Shewnarane, J.; Vo, M.; Taggart, J.; Goudreault, M.; Muskat, B.; Alfarano, C.; Dewar, D.; Lin, Z.; Michalickova, K.; Willems, A. R.; Sassi, H.; Nielsen, P. A.; Rasmussen, K. J.; Andersen, I.R.; Johansen, L. E.; Hansen, I.H.; Jespersen, H.; Podtelejnikov, A.; Nielsen, E.; Crawford, J.; Poulsen, V.; Sørensen, B. D.; Matthiesen, J.; Hendrickson, R. C.; Gleeson, F.; Pawson, T.; Moran, M. F.; Durocher, D.; Mann, M.; Hogue, C. W. V.; Figeys, D.; Tyers, M. *Nature*, **2002**, *415*, 180−183.
(149) Gavin, A.-C., Boesche, M.; Krause, R.; Grandi, P.; Marzioch, M.; Bauer, A.; Schultz, J.; Rick, J.M.; Michon, M.-M.; Cruciat, C.-M.; Remor, M.; Höfert, C Schelder, M.; Brajenovic, M.; Ruffner, H merino, A.; Klein, K.; Hudak, M.; Dickson, D.; Rudi, T.; Gnau, V.; Bauch, A.; Bastuck, S.; Huhse, B.; Leutwein, C.; Heurtier, M.-A.; Copley, R. R.; Edelmann, A.; Querfurth, E.; Rybin, V.; Drewes, G.; Raida, M.; Bouwmeester, T.; Bork, P.; Seraphin, B.; Kuster, B.; Neubauer, G.; Superti-Furga, G. *Nature* **2002**, *415*, 141−147.
(150) Blagoev, B.; Kratchmarova, I.; Ong, S.-E.; Nielsen, M.; Foster, L. J.; Mann, M. *Nat. Biotechnol.* **2003**, *21*, 315−318.
(151) Ranish, J. A.; Yi, E. C.; Leslie, D. M.; Purvine, S. O.; Goodlett, D. R.; Eng, J.; Aebersold, R. *Nat. Genetics* **2003**, *33*, 349−355.
(152) Salisbury, B. A.; Pungliya, M.; Choi, J. Y.; Jiang, R.; Sun, X. J.; Stephens, J. C. *Mutat. Res.* **2003**, *526*, 53−61.
(153) Zhao, Z.; Fu, Y. X.; Hewett-Emmett, D.; Boerwinkle, E. *Gene* **2003**, *312*, 207−213.
(154) Shastry, B. S. *Int. J. Mol. Med.* **2003**, *11*, 379−382.
(155) Gatlin, C. L.; Eng, J. K.; Cross, S. T.; Detter, J. C.; Yates, J. R., III. *Anal. Chem.* **2000**, *72*, 757−763.
(156) Lim, A.; Prokaeva, T.; McComb, M. E.; O'Connor, P. B.; Theberge, R.; Connors, L. H.; Skinner, M.; Costello, C. E. *Anal. Chem.* **2002**, *74*, 741−751.
(157) Kusmierz, J. J.; Sumrada, R.; Desiderio, D. M. *Anal. Chem.* **1990**, *62*, 2395−2400.
(158) Barr, J. R.; Maggio, V. L.; Patterson, D. G., Jr; Cooper, G. R.; Henderson, L.O.; Turner, W. E.; Smith, S. J.; Hannon, W. H.; Needham, L. L.; Sampson, E. J. *Clin. Chem.* **1996**, *42*, 1676−1682.
(159) Barnidge, D. R.; Dratz, E. A.; Martin, T.; Bonilla, L. E.; Moran, L. B.; Lindall, A. *Anal. Chem.* **2003**, *75*, 445−451.
(160) Gerber, S. A.; Rush, J.; Stemman, O.; Kirschner, M. W.; Gygi, S. P. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 6940−6945.

PR0340734