

# Proteomics and Mass Spectrometry in Nutrition Research

Helen Kim, PhD, Grier P. Page, PhD, and Stephen Barnes, PhD

From the Department of Pharmacology and Toxicology, Section on Statistical Genetics, Department of Biostatistics, the 2D-Proteomics Laboratory, and the Mass Spectrometry Shared Facility, Comprehensive Cancer Center, University of Alabama at Birmingham, Birmingham, Alabama, USA

## INTRODUCTION

The announcement of the first versions of the human genome sequence in 2000<sup>1,2</sup> ushered in a qualitatively different approach to scientific investigation in this new millennium. Whereas investigators previously examined the effects of treatments on individual genes or proteins in well-controlled, hypothesis-driven experiments, newly engineered technologies such as DNA microarrays generated from the genomic information currently allow for the measurement of the effect of a treatment on the expression of thousands, if not all, of the genes in a cell or tissue. As a result, nutrition journals in 2003 abound with data on multiple genes whose transcription to mRNA is affected by individual nutritional agents. The wealth of information is staggering, and it creates the need for substantial improvements in the methods of statistical analysis of the genomic data to assess the most "significant" changes and in experimental design to reduce the complexity of the experiment and remove sources of error.<sup>3</sup> Investigators are confronted with the challenge of determining the importance of hundreds of genes that they never suspected were involved in the system they are studying. Further, a full understanding of genomic data requires a complementary analysis of the changes in the proteome of the system, in other words, what changes occur in protein expression and/or modification downstream of the changes in gene expression. Without such an understanding, genomic and mRNA results are largely meaningless to the investigator but provide an abundance of "data."

Fortunately, developments in the ability to study gene expression at the genome level have coincided with the development of high-dimensional methods for the analysis of the *proteome*, which

is defined as the complement of expressed proteins in a biological system, whether it be a cell, its organelles, or the entire organism. In nutrition research, the organism can be the eater or the eaten; in other words, one can study what happens to an organism after exposure to a nutrient or the source of the nutrient. *Proteomics*, the sum of the methods that discover and quantitate proteins and their biochemical changes, has been enhanced by the development of two critical technologies, electrospray ionization (ESI) and matrix-assisted laser desorption ionization (MALDI), for the evaporation of peptides and proteins and their analysis by mass spectrometry (MS). It is noteworthy that the 2002 Nobel Prize in Chemistry was awarded jointly to John Fenn of Virginia Commonwealth University and Koichi Tanaka of the Shimadzu Corporation in recognition of their contributions to the development of these techniques and to the science of proteomics.

These MS-based technologies have removed many of the roadblocks to the identification and analysis of proteins and peptides. Attention has now shifted to the front end of proteomics analysis—the issue of how to resolve and detect changes in the multiple components of a proteome so that they can be analyzed by MS. Although there is no clear consensus,<sup>4,5</sup> the total number of genes in the human genome is 24 000 to 30 000. In any given cell, the expressed genes may total 10 000. However, the number of components in the proteome of that cell may exceed 100 000 due to post-translational modifications, alternative splicing, or proteolytic processing, thus creating a formidable analytical challenge. The post-translational modifications can involve phosphorylations that modulate the function of a protein, glycosylation that may guide the protein through organelles or may be critical for the physical or immunologic properties of the protein, or modifications of a protein resulting from chemical reactions with a reactive species generated by ingestion of certain xenobiotics and/or by oxidative stress. Unlike DNA microarray analysis, proteomics does not (currently) have the equivalent of the polymerase chain reaction to enhance the signal. Thus, careful consideration must be made of the expected copy number of a protein in the cell type being studied. Current proteomics technology requires at least 100 fM of a particular polypeptide; in other words, the sample must have  $6.023 \times 10^{10}$  molecules of the protein. This figure can be used to calculate the amount of cell or tissue sample needed for the experiment (Table I).

To address proteomics effectively, a full range of analytical biochemical methodologies has to be employed, many of which are already familiar to investigators in nutrition research, namely differential centrifugation, immunoprecipitation, one-dimensional and two-dimensional (2D) gel electrophoresis, and various types of chromatography. These must be combined with careful consideration of experimental design and implementation of the study and of the statistical techniques used to analyze the proteomic data to generate the most rigorous and meaningful data.

Unlike microarray analysis in which a complete genome can be "spotted" onto a glass slide, there is (currently) no working tech-

S.B. and H.K. are members of the Purdue and University of Alabama at Birmingham (UAB) Botanicals Center for Age-Related Disease, which is funded by the National Center for Complementary and Alternative Medicine, the National Institutes of Health Office of Dietary Supplements (P50 AT-00477), and the UAB Specialized Cooperative Center for Nutrient-Gene Interaction in Cancer Prevention (U54 CA-100949). G.P.P. and S.B. are supported by a National Science Foundation Plant Biology grant (0217651). The mass spectrometers used to generate the data shown in this article were purchased with funds provided by Shared Instrumentation Grants (S10 RR11329 and S10 RR13795) from the National Center for Research Resources (NCRR) and an award from the UAB Health Services Foundation General Endowment Fund to S.B. Instrumentation for 2D-gel processing was purchased with an NCRR Instrumentation Grant (S10 RR16849) and two awards from the UAB Health Services Foundation General Endowment Fund to H.K.

Correspondence: Helen Kim, PhD, Department of Pharmacology and Toxicology, Room 460, McCallum Building, University of Alabama at Birmingham, 1530 3rd Avenue South, Birmingham, AL 35294, USA. E-mail: helenkim@uab.edu

TABLE I.

BASICS OF PROTEOMICS—PROTEIN COPY NUMBER AND TISSUE/CELLS*		
Copy number/cell	Cells/100 fmol	Tissue (approximate)
1	$6.023 \times 10^{10}$	600 g
10	$6.023 \times 10^9$	60 g
100	$6.023 \times 10^8$	6 g
1 000	$6.023 \times 10^7$	0.6 g
10 000	$6.023 \times 10^6$	0.06 g
100 000	$6.023 \times 10^5$	0.006 g

\* The amount of protein needed for an assay is based on the minimum detectable (10 fM) for analysis on a standard nano-liquid chromatography, electrospray ionization, quadropole orthogonal, time-of-flight instrument. This is multiplied by 10 (to 100 fM) to account for the volume in which the sample is reconstituted at the final stage of the workup procedure.

nology that can readily display complete proteomes qualitatively or quantitatively. Nonetheless, the methods in use, although each has limitations, generate many, many more proteins for follow-up analysis and characterization than was possible a decade ago. Moreover, a proteomics approach can reveal proteins that are involved in a response to a stimulus or nutrient of which one may have been totally unaware. Thus, the value of a good proteomics analysis can far outweigh the drudgery of having to produce enough sample for analysis. This article discusses the current proteomics methodologies for nutrition research and the experimental design strategies that will allow for rigorous statistical analysis of quantitative changes at the proteome level. In addition, we discuss why metabolomics (also referred to as *metabolomics*) may emerge as the most important methodology in nutrition research.

## DESIGN AND ANALYTICAL CONSIDERATIONS OF PROTEOMIC EXPERIMENTS

As with any experiment, there are three phases to a proteomics experiment: design, implementation, and interpretation of the data. Because these experiments usually include many confounding and biasing factors, the success of an experiment is greatly aided by the experimentalist collaborating with a statistician or at least keeping in mind the issues that will be raised (see below) during all stages of the proteomics experiment.

The goal of any proteomic experiment is to generate data with as high a quality as possible so as to be able to derive valid inferences about the true state of any sample. The first step in any experiment is a scientifically based hypothesis with a clearly stated objective. Although the hypothesis in proteomics-type experiments may be broader than in traditional hypothesis-driven science, there should be a clear understanding of the system being studied and of the objective at the end of the experiment, because it is from this perspective that many of the aspects of an experimental design will be derived.<sup>6</sup>

It is well known that high-dimensional biological experiments, such as microarrays,<sup>7,8</sup> are liable to large non-biological sources of error. These issues are just as important in proteomics.<sup>9</sup> Known sources of non-biological variability can be the reagent lots, day, gel lot, technicians, temperature, and dye/stain incorporation. Even the use of differential gel electrophoresis (DIGE) technologies<sup>10</sup> may result in parallel dye biases similar to those that have been observed in two-color gene microarray analyses. In addition to

non-biological variability, there undoubtedly will be biological differences that are not directly related to the biological question such as circadian rhythms, sex, age, smoking status, or other factors, many of which may be unmeasured. Although one can statistically adjust for a portion of the impact of the factors on an experiment, it is better to reduce the impact by reducing the variability in these factors as much as possible or by spreading (orthogonalizing) the variability equally across all groups.

Biological replication and sample size must be considered in proteomic experiments.<sup>9</sup> There are extensive biological variability between samples and a large technical variability; thus, it is critical to analyze more than one sample per group. Only by comparing multiple true biological replicates will truly generalizable biological differences be observed as opposed to individual idiosyncratic differences. There are no good methods for calculating power or sample size for proteomic experiments, but some of the methods of microarray analysis, in which power is extrapolated based on a set cutoff<sup>11</sup> or from pilot experiments, may be appropriate.<sup>12,13</sup>

In microarray experiments there are several levels of statistical analysis after image processing before the actual statistical comparison of groups. In microarray analysis and 2D gel analysis, spot intensity is what is measured. However, gene arrays and 2D gel protein arrays are qualitatively different, so that, by definition, gene arrays are 2D uniform arrays of spots of cDNAs or oligonucleotides, and the analytical challenge is to generate a uniform background and a reproducible binding of the fluorescent probe to each spot. For 2D gel arrays, the position of the spot itself, whether it has changed as a result of the protein being modified, and its intensity are parameters that are major issues. In other words, the *x* and *y* coordinates of the protein spots in the gel comprise the data in an experiment as much as changes in spot intensity. In microarrays, one knows what gene a particular spot represents; whereas the identities of most of the spots on a 2D gel are not known, at least initially, in an experiment. The single greatest analytical challenge in 2D gels arises from the fact that, even when two identical gels are run under identical conditions, the same protein spot may end up at a slightly different *x* coordinate from the spot in the "control" gel. Ideally, spot movement should be due to a real chemical modification leading to an altered isoelectric point (causing horizontal movement) in a 2D gel. However, more often, a protein can be found at slightly different *x* coordinates between two gels due to small gel-to-gel differences. A protein also may undergo vertical movement in the gel due to molecular weight changes. Thus, the statistical challenge is in determining when a positional shift in a gel spot is "real." Further, the ability to obtain quantitative proteomic data by using instruments and methods such as MALDI time of flight (TOF) and multidimensional protein information technology (MUDPIT; see below) are still in development, and statistical analysis may not yet be appropriate. In addition, techniques such as immunoprecipitation will enrich for some proteins but not for others, thereby biasing the quantification methods. All of this adds great complexity to the analysis of proteomic data; currently, many of the pertinent issues are not fully addressed by available software or programs, but it is an active area of research.

Once a list of quantitative proteins levels can be derived, then traditional statistical analyses may be conducted. There are at least three classes of statistical tests that can be used to analyze proteomic data depending on the specific question being asked: 1) if the purpose is to identify the proteins that are different between two or more experimental conditions, then class descriptive methods or supervised analytical methods are used<sup>14,15</sup>; 2) if the goal is to predict which group new samples belong to, then the key is to use class predictive statistics<sup>16,17</sup>; 3) if the goal is to divide a non-homogeneous set of samples into one or more homogeneous groups, then the use of class discovery or unsupervised analytical methods<sup>18,19</sup> is called for. Experiments should be planned in such as way that the statistical analyses to be performed will be known

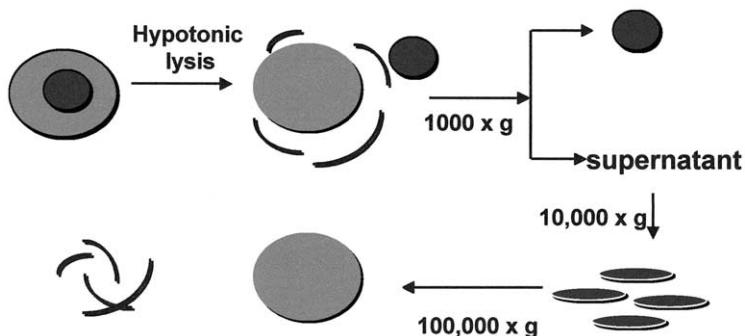


FIG. 1. Enrichment for cellular sub-proteomes by traditional differential centrifugation. Current two-dimensional gel methodology and follow-up mass spectrometry represent "modern" methods, so the best two-dimensional gels and, therefore, the most meaningful mass spectrometry data are obtained by initial enrichment of proteomes of interest before using the two-dimensional gels with traditional methods of cellular subfractionation, such as differential centrifugation. Typically, cells or tissues are lysed to break open cells without solubilizing intracellular compartments. This preparation is subjected to a series of centrifugations, at higher and higher speeds, each step yielding a pellet and supernatant that represents a cellular compartment. Each of these fractions can then be analyzed on a separate two-dimensional gel, or selected fractions can be analyzed, depending on what is known about the system, and the proteins suspected to undergo change in response to the nutrient or stimulus. In this scheme, the round compartment that sediments at 1000g is the nucleus, the rounded broken lines represent the microsomal membrane fraction that is recoverable as a pelleted fraction by centrifugation at 100 000g, and the ovoid organelles that sediment at 10 000 to 20 000g are mitochondria.

before the experiment starts and that appropriate analytical tools are available.

One final issue to be considered is that proteomic technologies can generate truly huge amounts of data, orders of magnitude more than microarray experiments.<sup>20</sup> For example, a hybrid quadrupole-orthogonal TOF mass spectrometer or TOF-TOF mass spectrometer will generate approximately 2 GB of data *per day*. Thus, careful consideration must be made with regard to the organization and storage of such volumes of data so that the investigator neither drowns in data nor archives the information only to "lose" it forever in disorganized cyberspace.

## SEPARATING THE PROTEOME

The proteome is much more complex than the transcriptome. It also exhibits the problem of a very wide dynamic range: whereas the albumin concentration in blood is 3 to 5 g/100 mL (0.5 to 0.8 mM), circulating cytokines are only in the range of 10 to 100 pg/mL (1 to 10 pM), nine orders of magnitude less.<sup>20</sup> Thus, unless effective pre-analysis fractionation methods are employed, the investigator can observe only the most abundant and not always interesting proteins.

So where does the investigator start? In the case of complex multicellular organisms such as mammals, birds, and fish, the investigation is usually focused on a specific organ (e.g., brain, heart, liver, or muscle). These organs can be selectively removed (but one should remember to flush the organ with ice-cold isotonic saline to eliminate blood protein contamination) for proteomics analysis. Investigators using transgenic mouse models also may be able to label proteins of interest with visible tags such as the green fluorescent protein.<sup>21</sup> This allows separation by flow cytometry of only those cells that express the protein above a certain level. This approach can be very valuable for those investigating protein-protein interactions by ensuring that such interactions are limited to proteins that are present in the same cell *in vivo*.<sup>21</sup>

Proteins have a property that largely distinguishes them from DNA and mRNA; they are localized to specific regions of a cell as opposed to just the nucleus (DNA) or just the ribosomal machinery (RNA). Thus, exactly where a protein is matters a great deal to the biochemical outcome. For example, brain-derived neurotrophic factor normally is found in secretory granules or synapses; however, a methionine-to-valine polymorphism causes a failure of brain-derived neurotrophic factor to localize to brain secretory

granules or synapses. The Met allele is associated with poorer episodic memory in humans; thus, for brain-derived neurotrophic factor, a single mutation has dramatic functional consequences due to a change in its intracellular location.<sup>22</sup> Fractionation of subcellular parts of cells can importantly decrease the complexity of the proteome to be analyzed and provide information about disturbances in cellular organization (Figure 1). The classic methods of ultracentrifugation to isolate nuclear, plasma membrane, mitochondrial, lysosomal, peroxisomal, microsomal, and cytosol fractions were developed 50 y ago, but are still relevant today. Sucrose, CsCl<sub>2</sub>, and Percoll gradients are used for more refined resolution of intracellular organelles, as is free-flow electrophoresis.<sup>23</sup>

Even with subcellular fractionation, further multidimensional separations are needed to examine the proteome protein by protein. If the interest is in a particular protein and its partners in a pathway, then the protein can be immunoprecipitated or, better, immunoabsorbed (Figure 2). This usually requires a specific antibody to that protein. Covalent coupling of the antibody to an agarose matrix is the preferred method because it can be used to pull the "needles out of the haystack" and then elute the members of the complex(es) without contamination by the antibody. The resulting set of proteins can usually be separated by one-dimensional sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). In this scenario, the proteins will be in similar molar amounts dictated by the nature of the complexes they form. Another way of identifying protein complexes is the use of blue-native gels.<sup>24</sup> In these polyacrylamide gels, the SDS is replaced by mild detergents and Coomassie blue; the latter binds to the proteins and serves as the carrier of charge for the complex. Combining this approach with an orthogonal second dimension using conventional SDS-PAGE analysis leads to ladders of proteins that define each complex.<sup>25</sup>

At the present time, most of the published protein separations make use of 2D gel electrophoresis, involving isoelectric focusing (IEF) in the first dimension and SDS-PAGE in the second. Developed by O'Farrell in 1975,<sup>26</sup> these 2D gels separate proteins according to differences in their isoelectric points in the first dimension and in their molecular mass in the second dimension (Figure 3). The IEF separation often enables post-translationally modified forms of the same protein to be separated from each other. The addition of phosphate groups to serine, threonine, and tyrosine residues typically decreases their isoelectric points by 0.1

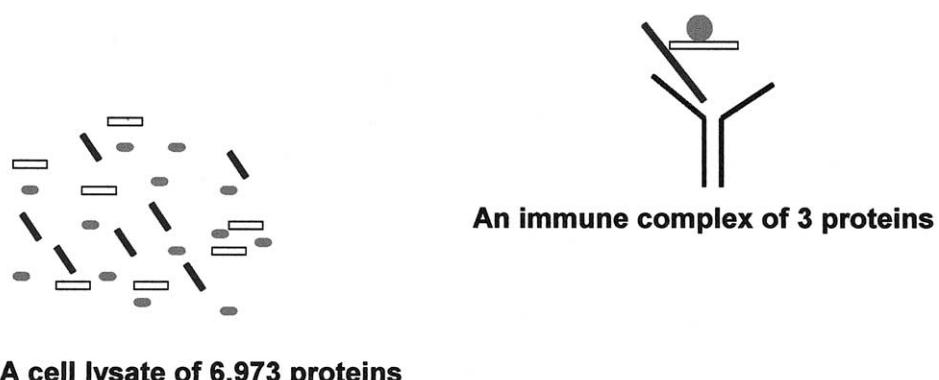


FIG. 2. Reduced two-dimensional gel complexity with immunoaffinity purification of proteins of interest. If a protein of interest is known from the outset, this can greatly simplify the proteomic analysis of the fate of that protein in an experiment; compare the complexity of proteins in the cell lysate versus that of the immune complex. Immunoaffinity purification can also reveal proteins that associate with the antigen and, hence, change in this miniproteome. The number of proteins here is usually low enough that a one-dimensional gel can resolve most of the bands; the advantage of the two-dimensional gel is the resolving power of the first-dimension isoelectric focusing step. Up to 20 times more protein can be excised from a two-dimensional gel spot than from the same size gel spot excised from a "band" of the same amount of protein run on a one-dimensional gel.

pH unit. However, nitration of tyrosine residues or alkylation of cysteine would not be expected to alter the isoelectric point of a protein.

The development of flat plastic IEF strips in which the pH gradient is immobilized in the acrylamide bonded to the plastic strip has enormously improved the reproducibility and user-friendliness of this method. This can be seen in the 2D gels shown in Figure 4. The gels represent the total brain proteome

from mice treated with a diet containing isolated soy protein as the protein source. Two sources of isolated soy protein were used, one extracted with alcohol (Figure 4A) and the other not extracted (Figure 4B). The pattern of spots between gels was very repro-

### 1st dimension: isoelectric focusing in IPG strips

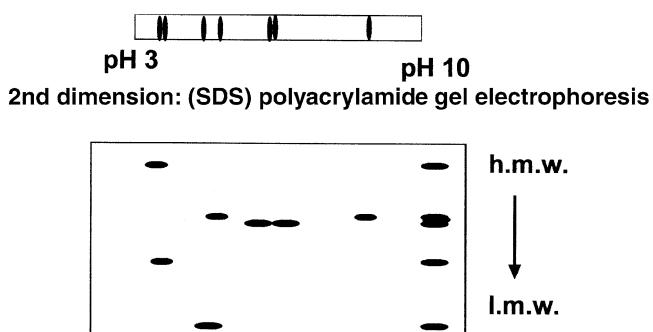


FIG. 3. Schematic of two-dimensional electrophoresis. This is the “standard” two-dimensional gel electrophoresis, whereby polypeptides are dissociated in high urea and detergent and become focused under an electric current at their isoelectric points in a low percentage (usually 4% to 5%) acrylamide strip containing an IPG. The second dimension is the familiar SDS-PAGE gel, which resolves the polypeptides according to differences in molecular mass. This method results in a qualitatively higher resolution of the polypeptides in a biological sample than in either dimension alone, because the polypeptides are being separated according to two different parameters. In addition, a polypeptide “spot” on a two-dimensional gel often contains significantly higher amounts of protein than does a spot in the corresponding band for that protein on a one-dimensional SDS gel, because the first dimension has literally focused it into a tight spot. This has ramifications for mass spectrometric analysis, because the less acrylamide gel in the sample, the better. Reprinted from Kim H, Chaves L, Hall P, DeSilva T, Coward L, Kirk M, Barnes S. The use of proteomics technology to study brain proteins affected by soy isoflavones. In Descheemaeker K, Debruyne I, eds. *Clinical Evidence—Dietetic Applications*. Antwerp-Apel-dorm: Garant, 2002:155. IPG, immobilized pH gradient; h.m.w., high molecular weight; l.m.w., low molecular weight; SDS-PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis.

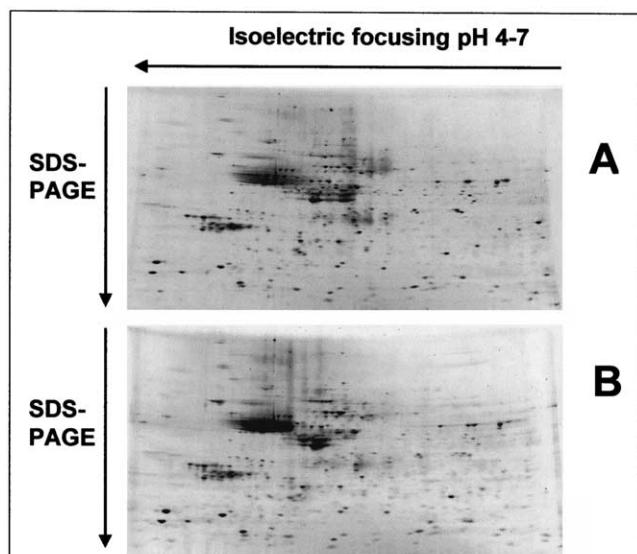


FIG. 4. Two-dimensional gels from a nutrition experiment. (A, B) This pair of two-dimensional gels represents an experiment testing the hypothesis that specific proteins in rodent brain would be affected if soy isoflavones were added to the animal's diet. Mice were segregated into two dietary groups from weaning and maintained in these groups for 12 mo. At death, whole brains were dissected out, frozen in liquid nitrogen, and archived at  $-80^{\circ}\text{C}$  until analyzed. Two-dimensional gel electrophoresis of the brain homogenates was carried out using 11-cm immobilized pH gradient strips (see Figure 3) incorporating a pH gradient of 4 to 7. The second-dimension gel incorporated a 10% to 20% acrylamide gradient. The gels were stained with colloidal Coomassie Brilliant Blue, and the images were acquired with a Bio-Rad Calibrated Imaging Densitometer. Preliminary image analysis using PDQuest determined that the indicated gel spots representing a variety of isoelectric points and molecular weights were different in the brains from the animals that ingested soy isoflavones from those in the brains from the animals that ingested a diet depleted of the isoflavones. Courtesy of Jessy Deshane and Helen Kim (unpublished data). SDS-PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis.

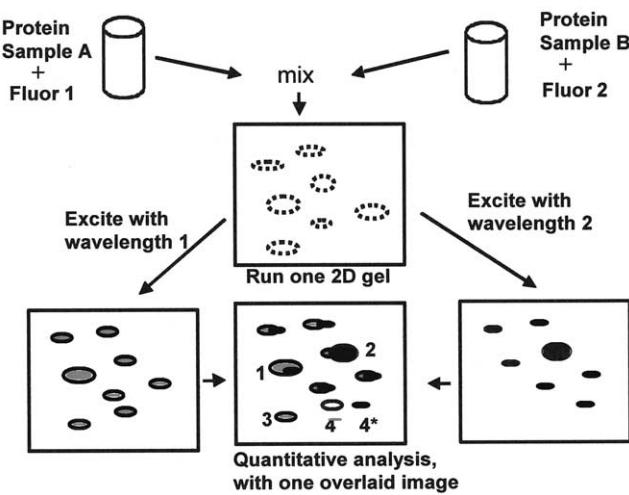


FIG. 5. Schematic of difference gel electrophoresis. In difference gel electrophoresis, similar but not identical biological samples, such as the brain homogenates shown in Figure 2, are each reacted with a different cyanine fluorescent dye, as developed by Unlu and coworkers.<sup>10</sup> Sample 1 is reacted with cyanine-3 and sample 2 is reacted with cyanine-5 (Amersham Biosciences). A two-dimensional-gel is then run of the separately reacted samples loaded on the same gel, so that like proteins migrate in the same position. The cyanine-3 spots are visualized as one color, typically green, by excitation and emission with one set of wavelengths, and the cyanine-5 spots are visualized as another color, typically red, with a different set of wavelengths. With this method, differences in expression of a protein spot are indicated by one color predominating over the other. This method is especially useful for detecting changes in the horizontal migration of a protein. A protein that shifts to a new location will appear in one gel as a "new" spot in the second color and disappear from its original position in the first color. Because like proteins by definition co-migrate, even slight shifts in position can be detected with this method.

ducible, allowing identification and quantification of small but significant changes in expression and in protein post-translation modification.<sup>27</sup>

A recent development in proteomics technology has been in fluorescent dyes<sup>10</sup> that covalently label the proteins in two similar but different samples representing, e.g., the effect of a treatment versus a control procedure. In this methodology, instead of running the samples on two different gels, the proteins labeled separately with the different dyes are mixed together and run on the same gel, thus the nomenclature, difference gel electrophoresis, or DIGE (Figure 5). Because the same protein in two different samples *migrates with itself*, differences in 2D location on the resultant gel due to variations between gels are eliminated. DIGE thus has the capability of substantially reducing gel-to-gel errors in position and in quantitative assessment of each protein spot. It should be noted, however, that the current dyes available for DIGE label proteins minimally on lysines, to minimize alterations in molecular weight; this introduces the caveat of most of a protein actually migrating at a slightly different position on a 2D gel from the dual-labeled spot, requiring post-staining for all proteins after the initial detection of the dual-labeled spots. The identification of the protein spots that change in expression or that have undergone significant shifts in isoelectric point requires sophisticated image data analysis. Even though some investigators have used software such as MELANIE or PDQuest (Bio-Rad Laboratories, Hercules, CA), the need to simultaneously analyze all the gels and all the replicates from an experiment at one time in an operator-neutral and, hence, unbiased way demands more robust software with greater capacity such as Progenesis Discovery (Nonlinear Dynamics, Ltd., Newcastle on Tyne, UK), although DeCyder software (Amersham Biosciences, Piscataway, NJ) is specific for the dual

dye-labeling technology. It should be noted, however, that none of the current image analysis packages provide adequate statistical analytical tools, and the astute investigator will export the image analysis data to a statistician to complete the analysis. The reader is encouraged to review a recent evaluation of the commercially available image analysis software.<sup>28</sup>

An oft-quoted limitation of 2D electrophoresis is the limited dynamic range for the detection of individual proteins in the second-dimension gel.<sup>29</sup> Typically, approximately 1000 spots are observed on a 2D gel. This is a very small proportion of the proteome and relative to a whole cell or tissue is unlikely to probe very deep into the underlying biochemistry of the system being studied. However, incorporating any of the fractionation techniques mentioned earlier can substantially enrich for the proteins expressed at lower levels (Figures 1 and 2). Also, IEF strips with a narrow pH range, e.g., 4.5 to 5.5, can be used to better resolve the proteins with isoelectric points in that range.

2D-IEF/SDS-PAGE analysis has one distinct advantage over proteomics separation methods based on proteolysis of the peptides and their subsequent 2D liquid chromatography (LC): each subform of the protein is individually separated and then analyzed. Another multidimensional method, in which the proteins are separated intact, has been pioneered by Chong et al.<sup>30</sup> In this method, proteins are first separated in the liquid phase by chromatofocusing. Individual fractions are then separated by reverse-phase liquid chromatography by using a pellicular, non-porous phase. This method can be automated and is now available commercially. The protein fractions isolated by this 2D procedure can then be subjected to proteolysis and analyzed by using peptide mass fingerprinting and/or nano-LC ESI tandem MS.

## PROTEIN PROTEOLYSIS

An advantage of hydrolyzing proteins in gel bands (one-dimensional analysis) or spots (2D analysis) is that the proteins are denatured, making them easier to digest. The in-gel digestion method was developed by the members of the University of California-San Francisco Mass Spectrometry Facility, and a detailed description of it is available on-line.<sup>31</sup> The peptides recovered from the digest are usually purified by absorbing them onto a ZipTip reverse-phase column. They are mixed with an equal volume of a saturated 50% aqueous acetonitrile solution of  $\alpha$ -cyano-4-hydroxycinnamic acid. This mixture is spotted (1  $\mu$ L) onto the MALDI target plate. Investigators have reported that using different solvents for the peptides can alter the quality of the MALDI-TOF mass spectra.<sup>32</sup> Because this is very dependent on the protein being studied, it is recommended that changes in the solvent conditions be restricted to when the protein cannot be clearly identified. A summary of the overall process of separating and analyzing proteins from 2D electrophoresis is shown schematically in Figure 6.

## MASS SPECTROMETRY AND PROTEOMICS

The ability to "evaporate" peptides and proteins from the liquid (ESI) or the solid (MALDI) phase has been a major advance in the study of proteomics. Peptide mass fingerprinting using MALDI-TOF-MS has substantially helped investigators without much experience in MS.<sup>33</sup> Because MALDI-TOF mass spectrometers are moderately priced, many universities have core facilities offering protein identification services to their faculty. The principle of peptide mass fingerprinting is that enzymes such as trypsin, Glu-C, and chymotrypsin cut proteins reproducibly at specific residues, thereby generating a predictable group of peptides with known sequence and, hence, precise molecular weights. Because a modern MALDI-TOF instrument can measure the molecular weight of each observable peptide with an accuracy better than 100 ppm (0.1

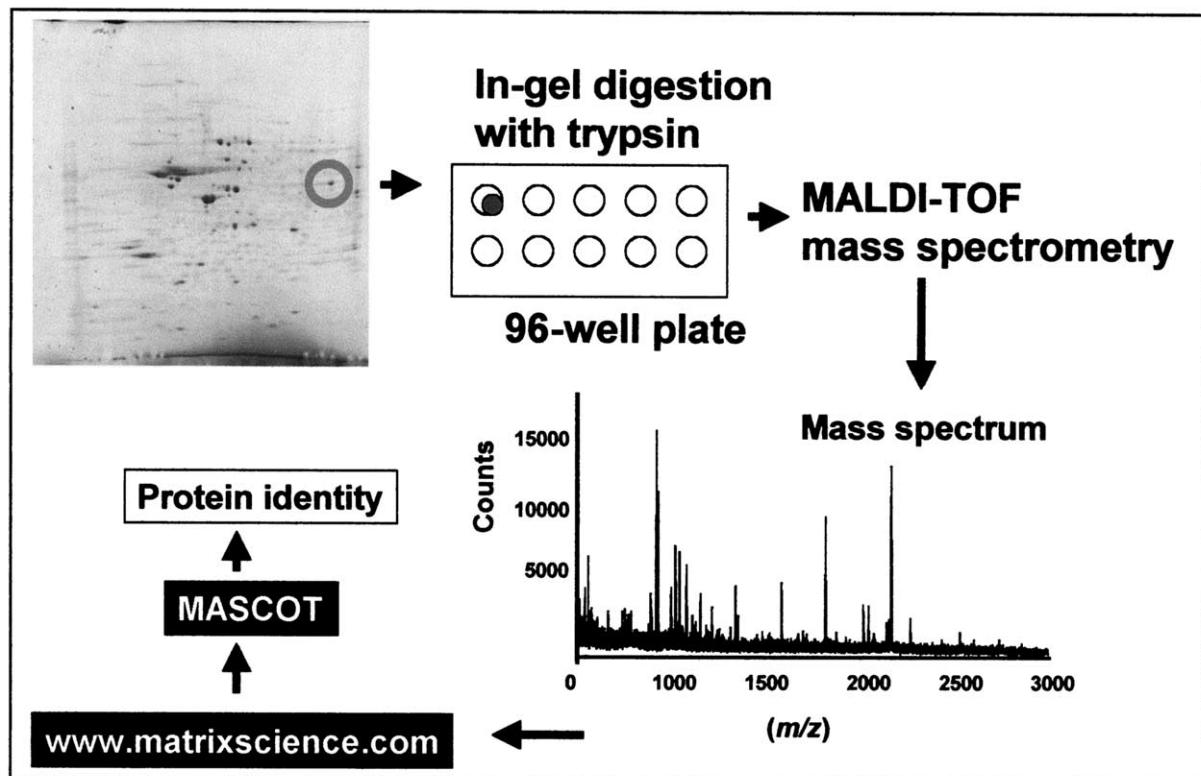


FIG. 6. Cartoon of the steps from the gel protein spot to MALDI-TOF mass fingerprinting. The gel slice is destained, dried, and reconstituted in digestion buffer with the protease (trypsin). The peptide fragments are transferred to the MALDI target plate for mass spectrometric analysis. The resulting data are transferred to a search engine (MASCOT) to identify the protein. MALDI-TOF, matrix-assisted laser desorption ionization coupled with time of flight.

Da for a 1000-Da peptide), it is possible to rapidly search *in silico*–generated databases of peptides from known proteins and those predicted by genome Open Reading Frame (ORF) search engines. The MALDI-TOF spectrum shown in Figure 7 is of the human blood protein transthyretin. In this example, the sequence coverage of this protein is high. This is not always the case.

There are several Web sites to which lists of peptides from MALDI-TOF analysis of a protein spot can be submitted (Table II). Each search engine calculates a statistical likelihood that the list of observed peptides matches that predicted for a particular protein or proteins (some records are of the same protein that was isolated by more than one investigator; Table III shows the example of calgranulin and its aliases). For each protein, the investigator is presented with a protein record accession number for the genome database used for the analysis, e.g., NCBI and SwissProt, and the predicted amino acid sequence of the protein and of each observed peptide (Table III). In 2003 the National Institutes of Health funded work to be carried out by the Swiss Protein Bioinformatics Institute to improve the curatorship of the protein databases.<sup>34</sup>

Proteolysis of proteins is usually carried out with trypsin, which is particularly useful for tandem MS (as discussed in the next section). However, depending on the amino acid sequence of the protein being studied, other proteases may be far more useful. Because it is possible to digest a protein sequence *in silico* with Web-based software such as MS-Digest,<sup>35</sup> investigators can evaluate which of the many protease options is best suited to the study of particular proteins. This is most useful when trying to locate sites of post-translational modification. Ideally the peptide containing the modification should have a molecular weight of 800 to 1500 Da.

Although peptide mass fingerprinting can produce a strong likelihood of identifying a protein, verifying the preliminary iden-

tification requires a second method, tandem MS. Because the molecular weights of the peptides have been established by MALDI-TOF-MS, we can now use ESI-MS to create a molecular ion (for tryptic peptides, this ion carries two positive charges, one on the N-terminal amino acid and the other on the C-terminal arginine or lysine). By using a quadrupole filter, each doubly charged ion can be individually isolated and then collided with argon gas. The energy of this collision causes fragmentation of the peptide at the peptide bonds. The fragmentation produces two types of singly charged ions, “y” ions that contain the intact charged C-terminus and “b” ions that contain the charge that was at the N-terminus. The intensity of each predicted ion is dependent on the ion chemistry of the peptide. Even so, a limited number of fragment ions is often sufficient to confirm the identity of the peptide. In Figure 8, the tandem mass spectrum of a nitrated peptide from the coat protein of bacteriophage P22 is shown.

## MULTIDIMENSIONAL PROTEIN INFORMATION TECHNOLOGY

The MUDPIT approach, pioneered by Washburn et al.,<sup>36</sup> takes a very different approach to proteomics from that of 2D gel electrophoresis and other methods in which the proteins are resolved first. Instead, in MUDPIT, the proteome under investigation is completely hydrolyzed by a proteolytic enzyme such as trypsin. This, of course, generates a very large number of peptides. Although some may be redundant, i.e., the same peptide may be produced from different proteins, particularly the shorter peptides, the set of peptides to be analyzed probably well exceeds 100 000. This is too large a number to be resolved even by the highest-performance LC methods. In the MUDPIT method, peptides are first fractionated

1 CPLMVKVLD**A** VRG**S**PAINV**A** VHVF**R****K**AADD TWEPFASGKT **SE**SGELHGLT  
 51 TEEEF**V**EGI**Y** KVEIDT**K**SYW KALG**I**SPF**H**E HAEVVFT**A**ND SG**P**RRY**T**IAA  
 101 LL**S**PYSY**S**TT AVVTNP**K**E

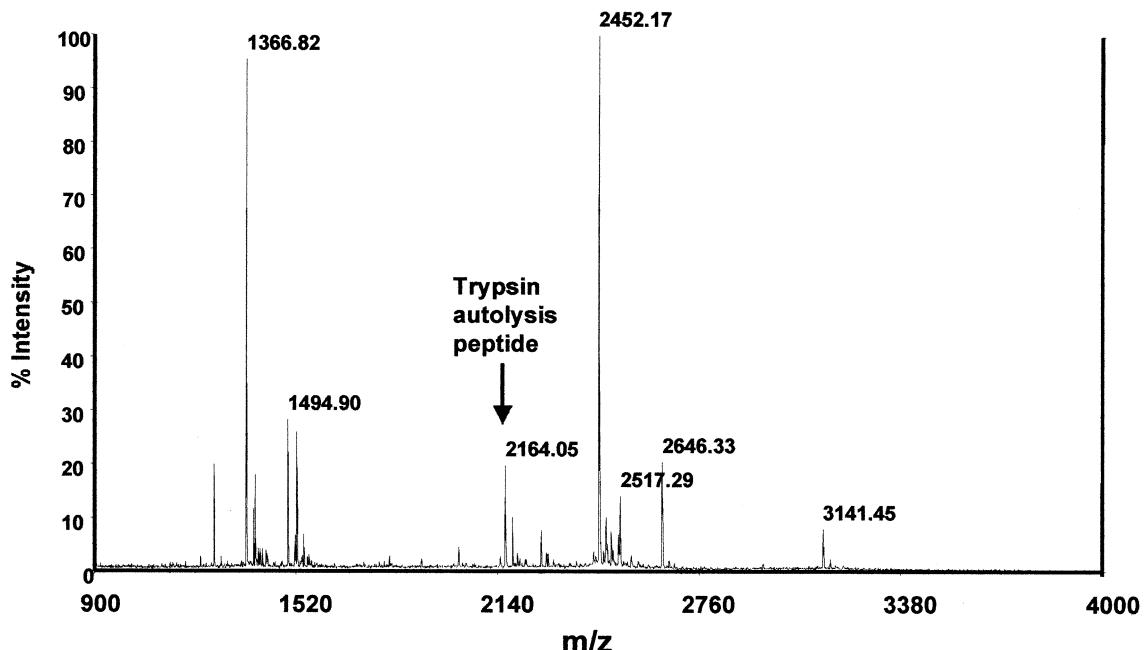


FIG. 7. MALDI-TOF mass spectrum of a trypsin-digested gel spot. This is an example of a peptide mass fingerprint obtained after trypsin digestion of a protein spot from a two-dimensional gel. The resulting peptides were mixed with a saturated solution of  $\alpha$ -cyano-4-hydroxycinnamic acid and an aliquot spotted onto a target plate. The sample was analyzed by using delay extraction on a De-Pro MALDI-TOF mass spectrometer (Applied Biosystems, Foster City, CA). The masses of all peaks were internally calibrated by using the  $m/z$  2164.05 trypsin autolysis peak. The  $^{12}\text{C}$ -ions in each peptide peak cluster were used for database searching with the MASCOT search engine. The protein was identified as human transthyretin. The sequence coverage was high; the parts of the sequence represented by the observed peptides are marked in boldface type. MALDI-TOF, matrix-assisted laser desorption ionization coupled with time of flight.

based on their charges by using a strong cation exchange column. Most peptides, except those with acylated N-terminal and with arginine and lysine residues, bind to the resin when it is in the  $[\text{H}^+]$  form. They are eluted in groups by increasing the concentration of a counter cation. Peptides that are eluted from the strong cation exchange phase are immediately captured on an analytical reverse-phase column and are subsequently eluted by a gradient of acetonitrile for ESI tandem MS analysis (Figure 9). In the version of MUDPIT of Washburn et al., the strong cation exchange and reverse-phase materials are contained in the same column. This has the advantage that dead volumes are minimized and, hence, very high sensitivities can be attained. Disadvantages include being able

to use the column only once and not being able to use strong cations such as  $\text{K}^+$  for the first step; instead, only ammonium salts can be used to elute the peptides. An arrangement with two separate columns allows for their individual regeneration and, hence, re-use.

MUDPIT has been combined with automated software (SEQUEST) for the identification of peptide sequences. Each peptide is first examined to determine its molecular weight. The virtual set of peptides for the proteome under investigation is searched for all peptides that are within 1 Da. These are then fragmented in silico and their virtual MS-MS spectra are compared with the one under examination. A statistical scoring scheme is used to identify which peptide MS-MS spectrum most closely matches that of the unknown peptide. Statistical significance has to be achieved for the software to announce a match. Because of the very large number of peptides and their tandem mass spectra that have to be analyzed, investigators make use of large computer clusters for this type of analysis.

MUDPIT methodology was initially developed to describe the members of the proteome. Recently, it has been used very effectively to identify post-translational protein modifications.<sup>37</sup> In this strategy, after the SEQUEST algorithm has identified the unmodified peptides, the data set is re-examined after first adding the expected additional mass of a modification to all the peptides in which it is believed the modification can occur. For example, if acetylation is being sought, 42 Da is added to the masses of all peptides that include the N-terminal region and to all peptides containing a lysine group. This method is remarkably efficient. However, it is limited to modifications that are currently known.

TABLE II.

WEB SITES FOR THE ANALYSIS OF PEPTIDE MASS FINGERPRINTING DATA

Program name	Web site address
MASCOT	<a href="http://www.matrixscience.com">http://www.matrixscience.com</a>
MS-Fit	<a href="http://prospector.ucsf.edu/ucsfhtml4.0/msfit.htm">http://prospector.ucsf.edu/ucsfhtml4.0/msfit.htm</a>
PeptideSearch	<a href="http://www.mann.embl-heidelberg.de/GroupPages/PageLink/peptidesearchpage.html">http://www.mann.embl-heidelberg.de/GroupPages/PageLink/peptidesearchpage.html</a>
PepMAPPER	<a href="http://wolf.bms.umist.ac.uk/mapper/">http://wolf.bms.umist.ac.uk/mapper/</a>
PeptIdent	<a href="http://us.expasy.org/tools/peptident.html">http://us.expasy.org/tools/peptident.html</a>
PepFrag	<a href="http://prowl.rockefeller.edu/">http://prowl.rockefeller.edu/</a>

TABLE III.

ABRIDGED MASCOT RECORD FOR PEPTIDE MASS FINGERPRINTING\* OF A PROTEIN SPOT FROM TWO-DIMENSIONAL ELECTROPHORESIS ANALYSIS OF DIMETHYL SULFOXIDE-TREATED HUMAN LEUKEMIA HL-60 CELLS

Accession no.	MOWSE score	Name
gi21614544	95	S100 calcium-binding protein A8 Cystic fibrosis antigen Calgranulin-A†
Peptide masses (observed)	Peptide (calculated)	Predicted amino acid sequence
963.55	963.48	GNFHAVYR
982.46	982.43	SHEESHKE
1421.77	1421.71	LLETECPQYIR
1434.78	1434.72	GNFHAVYRDDLK

\* Data were fitted with the MASCOT-NCBInr database; all species; trypsin hydrolysis, one allowed missed cleavage; 100-ppm mass tolerance; and  $[M+H]^+$  ions. The protein was pretreated with dithiothreitol and iodoacetamide to carboxyamidate available cysteine residues.

† Complete amino acid sequence (matched peptides are underlined):

1 MLTELEKALN SIIDVYHKYS LIKGNFHAVY RDDLKKLLET ECP  
QYIRKKG  
51 ADVWFKELDI NTDGAVNFQE FLILVIKMGV AAHKKSHES HKE

Also, because it was developed to work with an ion trap procedure, it does not have the mass accuracy to distinguish between the addition to lysine of one acetyl group (42.0106 Da) or three methyl groups (42.0470 Da). This has proved to be very important when examining the regulation of the nuclear histones that undergo both of these modifications.<sup>38</sup> A higher level of mass resolution is needed in this case than can be provided by an ion trap detector.

## QUANTIFICATION IN PROTEOMICS

Nutrition scientists will, in the end, be far more interested in the quantification of proteins in response to gene- or nutrient-induced changes than just which proteins are involved. As we have noted elsewhere, the best science operates under the principles of epistemology.<sup>3</sup> Grand solutions to quantification of proteins have not been established. However, there are several methods that can be used with changing degrees of applicability. As noted earlier, the fluorescent labeling of proteins with cyanine dyes enables quantification of the protein spots observed when using 2D-IEF/SDS-PAGE.

Those using 2D-LC-ESI-MS techniques have taken advantage of the fact that 85% of the proteins in a proteome contain at least one cysteine group. The Cys-SH group is reacted with isotope-coded affinity technology (ICAT) reagent.<sup>39</sup> The ICAT reagent consists of three regions: a part that reacts with the -SH group, an alkyl bridge group that can be labeled with deuterium or <sup>14</sup>C, and a biotin group. The proteome sample from the control group in the experiment is reacted with a "light" ICAT that has natural isotope labeling, whereas the treated sample is reacted with "heavy" <sup>13</sup>C<sub>4</sub>- or D<sub>8</sub>-labeled ICAT. After these reactions, the two proteome samples are mixed and digested with trypsin or another protease, thereby ensuring equal digestion and recovery from that point on in the analysis. To substantially reduce the total number of peptides that need to be resolved, the ICAT peptides are then captured

on a StrepAvidin column (the un-derivatized peptides do not bind) followed by elution with biotin. Because difficulties were encountered in quantitatively recovering low-abundance peptides, an acid-labile group has been included in the ICAT reagent.<sup>40</sup> When analyzed by LC-ESI-MS, a peptide is observed as two sets of ions: the lower *m/z* ion is from the peptide from the control experiment and the *m/z* values are 8 amu higher in the treatment experiment. By comparing the ratios of intensities of these two sets of ions for the peptides that can be detected, investigators can determine which proteins are unchanged and which are significantly increased or decreased. The issues at the statistical level are very similar to those that have been encountered in DNA microarray analysis.

There are other methods to isotopically label proteins and, hence, produce quantitative data. In one method, the two sets of proteins are digested in H<sub>2</sub><sup>16</sup>O or H<sub>2</sub><sup>18</sup>O.<sup>41</sup> The <sup>18</sup>O is added to the peptide formed in the proteolysis. This is more generally useful for all peptides but creates a major problem in that all the peptides in the proteome have to be separated and analyzed. The isotope shift (+2 Da) results in overlapping isotope profiles, and the need for two separate proteolysis reactions introduces error into the analysis. For those studying protein synthesis and turnover, the treatment group (cells or even animals) is administered <sup>15</sup>N-labeled amino acid.<sup>42</sup> The ratio of <sup>15</sup>N to <sup>14</sup>N can be determined for peptides containing that amino acid. Of course, this method may be affected by metabolism of the labeled amino acid, although the large unlabeled sink of related compounds makes it unlikely that the label will reappear in other amino acids.

The question of peptide quantification can be approached by using a triple quadrupole multiple-reaction ion-monitoring method. It is based on the selection of the *m/z* for the molecular ion of the peptide of interest, collision with argon gas, and then selection of a specific daughter ion. MS scans are not obtained in this method; instead, several channels, each with its own parent-daughter ion combination, are monitored in turn for 50- to 100-ms periods on a revolving basis. Peptide mixtures are analyzed by acetonitrile gradient reverse-phase LC, as described elsewhere in this article. Individual ion chromatograms are produced and the areas under the single peak are determined by integration. By adding to the digest a known amount of a synthetic <sup>13</sup>C-labeled peptide corresponding to the peptide of interest, the amount of the biologically derived peptide can be calculated. Although widely used for the quantitative analysis of small molecules (xenobiotics, nutrients in foods, metabolites),<sup>43</sup> its application to proteomics has generated a new acronym, AQUA (absolute quantitative analysis).<sup>44</sup>

## TOP-DOWN PROTEOMICS

The future of proteomics in research may lie in what is termed as *top-down proteomics*.<sup>45</sup> This method would eliminate some of the steps described in this article, in particular 2D electrophoresis and 2D-LC tandem MS. However, effective methods for the primary separation of proteins are still required.<sup>46</sup> In the top-down approach, multiply charged, intact protein molecular ions are fragmented in the gas phase by using electron capture dissociation. The key to the success of this approach is to be able to select the protein molecular ions and then to identify the ion states of the resulting fragment ions. Recent advances in Fourier-transform ion cyclotron resonance MS have provided the necessary mass resolution for this technique. However, other MS instrumental methods such as the TOF-TOF<sup>47</sup> and ion traps<sup>48</sup> may contribute to this development.

## METABOLOMICS

In nutrition studies, where the fate of metabolic pathways that process food components may be a key issue, the best markers of

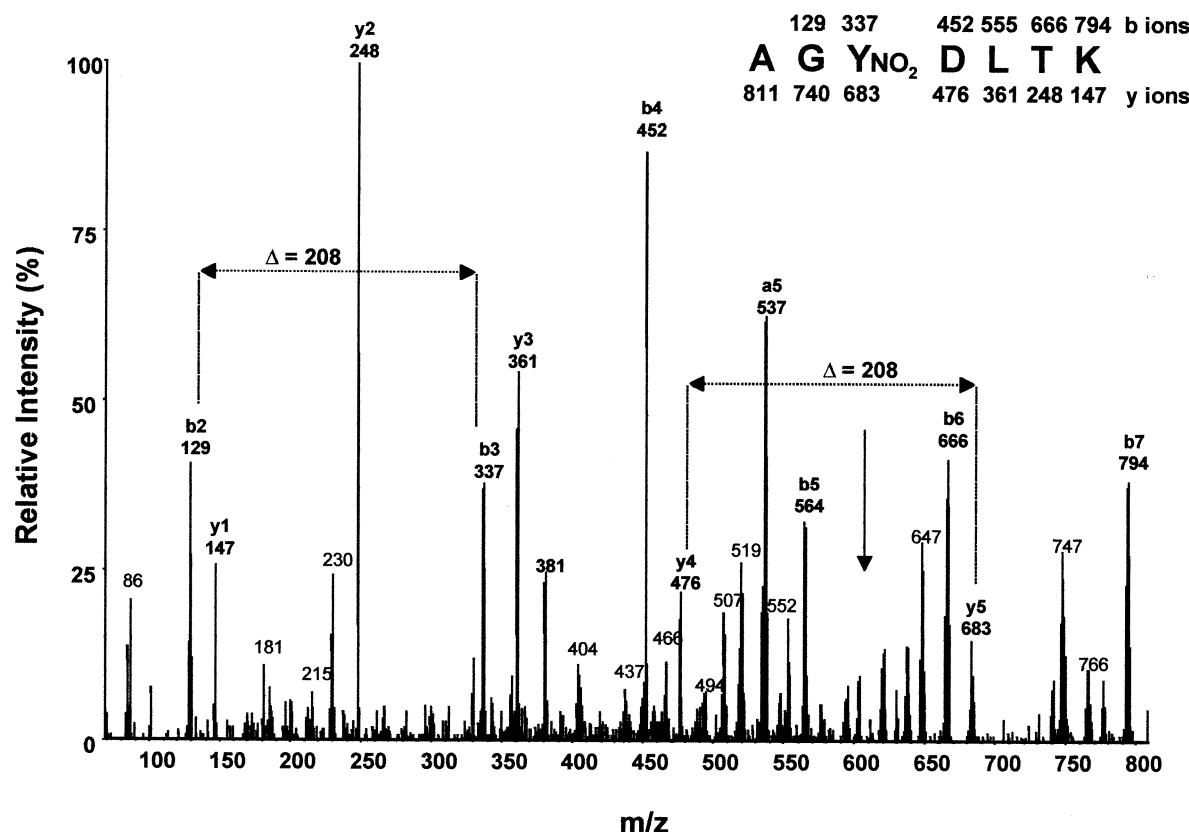


FIG. 8. Tandem mass spectrum of a nitrated tryptic peptide. Bacteriophage-P22-coated protein was treated with trypsin and the products were analyzed by nano-liquid chromatography, electrospray ionization, mass spectrum mass spectography on a Micromass quadrupole orthogonal time-of-flight instrument. Two series of fragment ions were produced by collision-induced dissociation with argon gas. The expected masses (to the nearest atomic mass units [amu]) of the "b" and "y" series are given above and below the sequence of the peptide, respectively. Note the  $m/z$  208 difference caused by the nitrated tyrosine group.

intake of a food, or response to a stimulus that has relevance in nutrition, may be the levels of the substrates or products of the enzymes rather than the proteins themselves. These compounds range from the components of the major pathways of metabolism, e.g., glycolysis, Krebs' cycle, and biosynthesis of amino acids, nucleotides, lipids, steroids, etc. They have been studied by nuclear magnetic resonance (NMR), a method referred to as *metabonomics*. However, these compounds can also be easily analyzed with LC-MS techniques with far higher sensitivities. Studies thus far have emphasized the more hydrophilic metabolites and those with molecular weights below 1000 Da.<sup>49</sup> However, this is unnecessarily limiting. There have been several reports of applying this technique to urine samples, largely because they are easy to collect and are non-invasive.<sup>50</sup> However, it should be appreciated that the physiologic status of the kidney itself could be affected by the treatment, which in turn might bias the observed data considerably. The metabolome of the affected organ(s) is a far better site to investigate the power of this approach.

## SUMMARY

The past decade has seen phenomenal leaps in technology development and implementation for analysis of biological experiments. No sooner had various genomes been sequenced than we were in the postgenome era, with the understanding that knowing what genes are turned on or off is only a beginning. Nonetheless, protein sequence databases for different phyla exist only because of the genome databases. However, we should be mindful that these

genomic sequence databases are not writ in stone—that is, if a protein spot from a gel does not match a polypeptide in the human database with a sufficiently high Molecular Weight Search Engine (MOWSE) score, it could be that the predicted sequence is not correct. Given these frailties in the system, the technologies available to the enthusiastic researcher are such that the limiting factors are the goodness of the experiment and the amount of sample and not identifying the protein or cataloging the metabolites. The kinds of experiments that nutrition researchers carry out are fundamentally no different from those in cell biology, physiology, or biochemistry in terms of amenability for analysis by the modern technologies discussed in this paper. Because of the expense of the instrumentation and the high degree of specialization of training required to generate and interpret MS data, the successful nutrition researchers and researchers in the other disciplines will be those who collaborate with other scientists who have the appropriate expertise in protein separation, statistics, and MS and available instrumentation. The next decade should see an explosion of information in nutrition sciences as nutrition researchers exploit these technologies.

## ACKNOWLEDGMENTS

The authors thank Peter Prevelige, University of Alabama at Birmingham, for his permission to use the tandem mass spectrum shown in Figure 8.

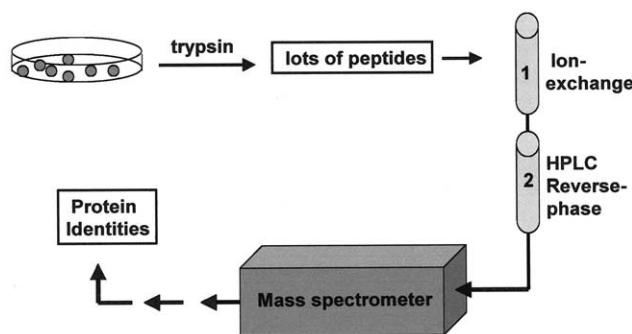


FIG. 9. Simple schematic of multidimensional protein information technology analysis. The biological fraction (a cell lysate or a subfraction of the cells) is trypsinized. The large number of peptides is first desalting (not shown) and then passed onto a strong cation exchange resin in the  $[H^+]$  form. Elution of the bound peptides is accomplished with step increases in the concentration of KCl (two-column system) or ammonium acetate (one-column system). The peptides from each elution step are captured on the reverse-phase packing and then eluted with a linear gradient (0% to 50% of acetonitrile in 0.1% formic acid). This process can be automated so that the fractions from the strong cation exchange column are analyzed independently. Tandem mass spectra are recorded by the mass spectrometer (a quadrupole ion trap or hybrid quadrupole orthogonal time-of-flight instrument) and the data are analyzed by automated software such as SEQUEST. HPLC, high-performance liquid chromatography.

## REFERENCES

- Venter JC, Adams MD, Myers EW, et al. The sequence of the human genome. *Science* 2001;291:1304
- Lander ES, Linton LM, Birren B, et al. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* 2001;409:860
- Barnes S, Allison DB. Challenges in the biological credentialing of high dimensional data. *Nutr Today* 2003(in press)
- Hogenesch JB, Ching KA, Batalov S, et al. A comparison of the Celera and Ensembl predicted gene sets reveals little overlap in novel genes. *Cell* 2001;106:413
- Pennisi E. A low number wins the GeneSweep pool. *Science* 2003;300:1484
- Potter JD. At the interfaces of epidemiology, genetics, and genomics. *Nat Rev Genet* 2001;2:142
- Lee ML, Kuo FC, Whitmore GA, Sklar J. Importance of replication in microarray gene expression studies: statistical methods and evidence from repetitive cDNA hybridizations. *Proc Natl Acad Sci USA* 2000;97:9834
- Mirnics K. Microarrays in brain research: the good, the bad and the ugly. *Nat Rev Neurosci* 2001;2:444
- Boguski MS, McIntosh MW. Biomedical informatics for proteomics. *Nature* 2003;422:233
- Ulrich M, Morgan ME, Minden JS. Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. *Electrophoresis* 1997;18:2071
- Margolin J. From comparative and functional genomics to practical decisions in the clinic: a view from the trenches. *Genome Res* 2001;11:923
- Pan W, Lin J, Le CT. How many replicates of arrays are required to detect gene expression changes in microarray experiments? A mixture model approach. *Genome Biol* 2002;3:research0022
- Gadbury GL, Page GP, Edwards JW, et al. *Stat Methods Med Res* 2003(in press)
- Kerr MK, Churchill GA. Statistical design and the analysis of gene expression microarray data. *Gene Res* 2001;77:123
- Kerr MK, Martin M, Churchill GA. Analysis of variance for gene expression microarray data. *J Comput Biol* 2000;7:819
- Tamayo P, Slonim D, Mesirov J, et al. Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. *Proc Natl Acad Sci USA* 1999;96:2907
- Dudoit S, Fridlyand J, Speed TP. Comparison of discrimination methods for the classification of tumors using gene expression data. *J Am Statist Assoc* 2002;97:77
- Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* 1998;95:14863
- Brazma A, Vilo J. Gene expression data analysis. *FEBS Lett* 2000;480:17
- Kenyon GL, DeMarini DM, Fuchs E, et al. Defining the mandate of proteomics in the post-genomics era: workshop report. *Mol Cell Proteomics* 2002;1:763
- Heck S, Ermakova O, Iwasaki H, et al. Distinguishable live erythroid and myeloid cell s in beta-globin ECFP  $\times$  lysozyme EGFP mice. *Blood* 2003;101:903
- Egan MF, Kojima M, Callicott JH, et al. The BDNF val66met polymorphism affects activity-dependent secretion of BDNF and human memory and hippocampal function. *Cell* 2003;112:257
- Zischka H, Weber G, Weber PJA, et al. Improved proteome analysis of *Saccharomyces cerevisiae* mitochondria by free-flow electrophoresis. *Proteomics* 2003;3:906
- Schagger H, von Jagow G. Blue native electrophoresis for isolation of membrane protein complexes in enzymatically active form. *Anal Biochem* 1991;199:223
- Brookes PS, Pinner A, Ramachandran A, et al. High throughput 2D blue-native electrophoresis—a tool for functional proteomics of mitochondria and signaling complexes. *Proteomics* 2002;2:969
- O'Farrell PH. High resolution two-dimensional electrophoresis of proteins. *J Biol Chem* 1975;250:4007
- Deshane J, Oh J, Chaves L, et al. Proteomics identification of neurodegeneration-relevant proteins modulated by soy isoflavones in rodent brain. *J Nutr* 2003(in press)
- Rogers M, Graham J, Tong RP. Using statistical models for objective evaluation of 2-DE gel image analysis. *Proteomics* 2003;3:879
- Lopez MF, Berggren K, Chernokalskaya E, Lazarev A, Robinson M, Patton WF. A comparison of silver stain and SYPRO Ruby Protein Gel Stain with respect to protein detection in two-dimensional gels and identification by peptide mass profiling. *Electrophoresis* 2000;21:3673
- Chong BE, Yan F, Lubman DM, Miller FR. Chromatofocusing nonporous reversed-phase high-performance liquid chromatography/electrospray ionization time-of-flight mass spectrometry of proteins from human breast cancer whole cell lysates: a novel two-dimensional liquid chromatography/mass spectrometry method. *Rapid Commun Mass Spectrom* 2001;15:291
- University of California-San Francisco Mass Spectrometry Facility. In-gel method. Available at: <http://donatello.ucsf.edu/ingel.html>
- Padliya ND, Wood TD. Optimizing MALDI matrix formulation: a strategy to improve protein identification via peptide mass fingerprinting. Paper presented at the 51st Conference of the North American Society for Mass Spectrometry; Montreal, Canada; June 9–12, 2003
- Mortz E, Vorm O, Mann M, Roepstorff P. Identification of proteins in polyacrylamide gels by mass spectrometric peptide mapping combined with database search. *Biol Mass Spectrom* 1994;23:249
- 2002 Announcement of funding by the National Human Genome Research Institute and several NIH Institutes and Centers for the Creation of the UniProt database. Available at: <http://www.nih.gov/news/pr/oct2002/nhgri-23.htm>
- University of California-San Francisco Mass Spectrometry Facility. MS-Digest, part of the Protein Prospector package. Available at: <http://prospector.ucsf.edu>
- Washburn MP, Wolters D, Yates JR, III. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotech* 2001;19:242
- MacCoss MJ, McDonald WH, Saraf A, et al. Shotgun identification of protein modifications from protein complexes and lens tissue. *Proc Natl Acad Sci USA* 2002;99:7900
- Zhang L, Eugeni EE, Parthun MR, Freitas MA. Identification of novel histone post-translational modifications by peptide mass fingerprinting. *Chromosoma* 2003;112:77
- Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotech* 1999;17:994
- Tao WA, Aebersold R. advances in quantitative proteomics via stable isotope tagging and mass spectrometry. *Curr Opin Biotech* 2003;14:110
- Yao X, Freas A, Ramirez J, Demirev PA, Fenselau C. Proteolytic  $^{18}\text{O}$  labeling for comparative proteomics: model studies with two serotypes of adenovirus. *Anal Chem* 2001;73:2836
- Wang YK, Ma Z, Quinn DF, Fu EW. Inverse  $^{15}\text{N}$ -metabolic labeling/mass spectrometry for comparative proteomics and rapid identification of protein markers/targets. *Rapid Commun Mass Spectrom* 2002;16:1389
- Coward L, Kirk M, Albin N, Barnes S. Analysis of plasma isoflavones by reversed-phase HPLC–multiple reaction ion monitoring–mass spectrometry. *Clin Chim Acta* 1996;247:121
- Gerber SA, Rush J, Stemman O, Kirschner MW, Gygi SP. Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc Natl Acad Sci USA* 2003;100:6940
- Horn DM, Zubarev RA, McLafferty FW. Automated de novo sequencing of proteins by tandem high-resolution mass spectrometry. *Proc Natl Acad Sci USA* 2000;97:10313
- Meng F, Cargile BJ, Patrie SM, Johnson JR, McLoughlin SM, Kelleher NL. Processing complex mixtures of intact proteins for direct analysis by mass spectrometry. *Anal Chem* 2002;74:2923

47. Resemann A, Suckau D. Terminus-specific fragmentation, a novel tool for the direct characterization of intact proteins. Paper presented at the 51st Conference of the North American Society for Mass Spectrometry; Montreal, Canada; June 11, 2003
48. Reid GE, McLuckey SA. 'Top down' protein characterization via tandem mass spectrometry. *J Mass Spectrom* 2002;37:663
49. Phelps TJ, Palumbo AV, Beliaev AS. Metabolomics and microarrays for improved understanding of phenotypic characteristics controlled by both genomics and environmental constraints. *Curr Opin Biotech* 2002;13:20
50. Plumb RS, Stumpf CL, Gorenstein MV, et al. Metabonomics: the use of electrospray mass spectrometry coupled to reversed-phase liquid chromatography shows potential for the screening of rat urine in drug development. *Rapid Commun Mass Spectrom* 2002;16:1991