Review

Nutriproteomics: Identifying the Molecular Targets of Nutritive and Non-nutritive Components of the Diet

Stephen Barnes[†,‡,§,#,††,]* and Helen Kim[†,‡,§,||]

[†]Departments of Pharmacology and Toxicology, [‡]UAB Center for Nutrient-Gene Interaction,
[§]Purdue-UAB Botanicals Center for Age-related Disease, [||]2D-Proteomics Laboratory, and
[#]Mass Spectrometry Shared Facility, Comprehensive Cancer Center,
[††]Department of Pharmacology and Toxicology, 452 McCallum Building, University of Alabama at Birmingham,
1918 University Boulevard, Birmingham, AL 35294, USA

The study of whole *patterns* of changes in protein expression and their modifications, or *proteomics*, presents both technological advances as well as formidable challenges to biological researchers. Nutrition research and the food sciences in general will be strongly influenced by the new knowledge generated by the proteomics approach. This review examines the different aspects of proteomics technologies, while emphasizing the value of consideration of "traditional" aspects of protein separation. These include the choice of the cell, the subcellular fraction, and the isolation and purification of the relevant protein fraction (if known) by protein chromatographic procedures. Qualitative and quantitative analyses of proteins and their peptides formed by proteolytic hydrolysis have been substantially enhanced by the development of mass spectrometry technologies in combination with nanoscale fluidics analysis. These are described, as are the pros and cons of each method in current use.

Keywords: High dimensional analysis, Organelle and protein fractionation, 2-Dimensional electrophoresis, Peptide mass fingerprinting, Tandem mass spectrometry, Post-translational modification

Introduction

For the better part of the 20th Century, research in nutrition focused on the isolation and analysis of foods for both the nutritive (carbohydrates, fats and proteins) and non-nutritive

components. Several of the latter became classified as vitamins. This was followed by the discovery of metabolic pathways and the relevant enzymes involved. Nutrition research was aided in the second half of the 20th Century by the introduction of chromatographic techniques (paper chromatography, thin-layer chromatography, gas-liquid chromatography, and high performance liquid chromatography). During this time, gas-liquid chromatography was linked to the emerging and powerful method of *mass spectrometry*, at least for those molecules (mostly low molecular weight compounds such as amino acids, fatty acids, some pathway intermediates and simple saccharides) that could be moved into the gas phase without damage. Although this did not include proteins, their analysis and separation was considerably improved in the 1970's with the introduction of first one dimensional sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) (Laemmli, 1970) and subsequently two dimensional-isoelectric focusing/SDS-PAGE analysis (O'Farrell, 1975), a way of providing high resolution of proteins not only by their molecular weight, but also by their isoelectric points. The latter enabled resolution of isoforms due to amino acid heterogeneity or posttranslational modifications. Until relatively recently, isolating and determining the properties of a single protein was a major exercise - for the graduate student, it was usually sufficient to earn a PhD thesis.

By 1980, this world of investigative science became overwhelmed by the rapidly developing technologies inherent in molecular biology. The quest for the *holy grail of science*, the chemistry of DNA and the sequence of the entire genome of each cell, attracted the minds of most scientists, graduate students, their faculty mentors, postdoctoral fellows, university administrators, large and small pharmaceutical companies, and finally governments. Enthusiasm for molecular biology by investigators in nutrition sciences was no different from that of their colleagues in other disciplines.

*To whom correspondence should be addressed.
Tel: 1-205-934-7117; Fax: 1-205-934-6944
E-mail: sbarnes@uab.edu

During the 1980's, determination of the complete amino acid sequences of proteins shifted from Edman degradation chemistry of trypsin-derived peptides to the use of molecular cloning methods to isolate cDNAs encoding the protein - the nucleotide sequence of the cloned cDNA thereby being translated to yield the amino sequence. The resulting excitement enabled a more rapid assembly of a protein's sequence and led to the bold step in 1990 to sequence the entire human genome, using methods no longer based on protein-driven research. DNA sequencing methods grew by leaps and bounds and the major goal - the human genome sequence, albeit consisting of only 30% of the expected number of genes - was accomplished two years ahead of schedule (Lander *et al*., 2001; Venter *et al*., 2001). Many other genomes have also been described.

Has having all the gene sequences in a genome really solved the major issues of life and disease? Not really. Nature turns out to be quite a bit more complex than previously thought. DNA microarray analysis is enabling investigators to look at changing expression of thousands and thousands of genes caused by diets and their food components. Young investigators engaged in this research endeavor, instead of seeing one target gene changing expression, are confronted by long lists of genes whose cryptic names make no sense, or whose encoded proteins were never really learned or even covered in graduate school. They lean heavily on the services of a new breed in the scientific community, *bioinformaticists,* and ponder experimental design and statistical analysis with statisticians who despair at trying to find statistically significant changes in these *high dimensional* systems (Page *et al*., 2003).

The universe as defined by molecular biology has undergone bigger cracks as it was "remembered" that almost all genes are only functional when their proteins are expressed. Also there is little or no correlation between the expression of a gene (as its mRNA) and its corresponding protein. Thus, for many situations, it is the events in the protein universe that are more closely related to processes of growth, development, disease and death.

So, in 2004 and onwards, many investigators find themselves in the world of protein analysis, biochemistry and chemistry. For some, this is a return journey, but for many, especially the younger investigator, it is a new experience. The purpose of this review is to introduce the terminologies of *proteomics* and *protein mass spectrometry*, and to outline the major analytical methods of this discipline. Our goal is to educate the nutrition science community and those from other disciplines whose research has led them into nutrition-related research.

The proteome, proteomics and where to start The *proteome* is analogous to the genome - it is the complement of proteins that are expressed from the mRNAs that comprise the *transcriptome*. Although the DNA content of each cell is the same, the proteome may be quite different from cell type to cell type. In fact, the proteome is both dynamic and highly positional in character (the proteins in a cell are arranged regionally and it matters which compartment they are in). The "disappointment" of only discovering 24,000 to 30,000 human genes (instead of the expected 80,000-100,000) is more than compensated by the plethora of protein forms resulting from mRNA splice variants, protein processing, and posttranslational modifications. Protein chemistry does not yet have the equivalent of the polymerase chain reaction to specifically amplify a protein once detected. Thus proteins still have to be separated from each other and then analyzed with ultrasensitive mass spectrometry instrumentation to both identify the protein and characterize any posttranslational modifications. The challenge facing the nutrition scientist is how best to utilize these proteomics technologies.

Cellular and subcellular proteomes Analyzing the proteome of a food or diet, of an organ, or even a single cell is a daunting task if the goal is to detect specific changes in response to a stimulus to a specific nutritive or non-nutritive food component. Initial applications of 2D-electrophoresis were, and still are, largely on unfractionated lysates. The proteomes of a food or a target organ (brain, heart, kidney, liver, etc.) are very complex. Although 2D-electrophoresis on large format (20×20 cm) gels can reproducibly display about 2,000 protein spots, this amounts to less than 5% of a typical proteome. While some initial hypotheses can be developed for proteins that are highly expressed (and therefore easy to detect), many interesting changes are simply not observable when attempting to examine the whole proteome.

One approach to increase the success of proteomics analysis of important but less abundantly expressed proteins is the use of cultured cells from an organ. This, of course, limits the analysis to one cell type. Moreover, cells grown in tissue culture behave differently from those actually in the tissue matrix. Three dimensional growth on materials such as Matrigel™ overcomes this problem to a degree, but this still does not guarantee that the proteome is the same as in tissue. So, the ideal way to proceed is to isolate a particular cell type from its natural tissue environment. One recent approach has been developed in which mouse germ-line cells are transfected with a gene encoding a protein for the cell of interest modified so that it includes in its open reading frame the green fluorescent protein (GFP) (Roper *et al*., 2003). Expression of this GFP-fusion protein can be targeted to particular cells by incorporating tissue-specific or cell-specific promoters Once a mouse has developed to a certain age or at particular timepoint in the nutrition experiment, the organ containing the cells of interest is perfused with collagenase to initially disperse the tissue and then trypsin-treated to yield a single cell population. These are subjected to cell-sorting techniques such as flow cytometry that sort cells containing the green fluorescence from the other cells (Roper *et al*., 2003).

Proteins are not uniformly distributed throughout cells. Certain proteins are localized to the plasma membrane, to
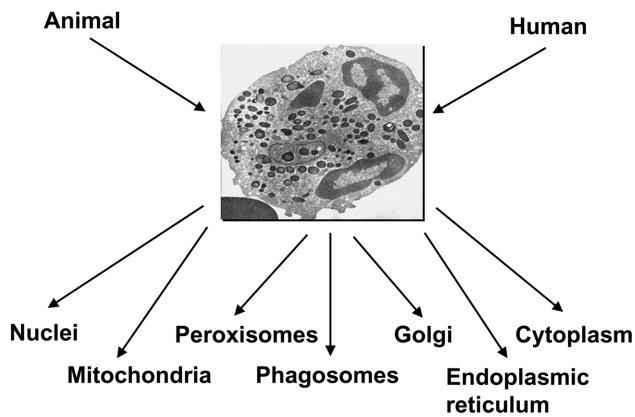
Fig. 1. Fractionation of a cell. In this example, individual neutrophils from different origins can be isolated and homogenized. The homogenate is subjected to centrifugation to separate it into many particulate fractions and the cytoplasm.

mitochondria, to lysosomes, to nuclei, etc. Historically, protein chemists and enzymologists systematically searched through subcellular fractions to identify the fraction with the highest amount or activity of the protein/enzyme of interest. Typically, differential ultracentrifugation has been used to accomplish this - for instance, nuclei will pellet when cell homogenates are centrifuged at $800 \times g$. While not pure, the nuclear fraction is enriched in nuclear proteins compared with the original homogenate. Similarly, each of the other particulate fractions and the soluble protein fraction (cytosol) can be obtained (Fig. 1). In recent years, much use has been made of Percoll density gradients for preparation of cellular fractions (Pertoft, 2000). These ultracentrifugation methods are old techniques that nonetheless work and should be re-visited by young investigators with access to new ideas and technologies in biochemical separations. These are needed to address for instance the question of whether an intracellular compartment or structure is a single population, or is instead a heterogeneous family with differing properties and functions.

One "new-old" approach in this challenging area is the use of free-flow electrophoresis to refine the separation of mitochondria (Pasquali *et al.*, 1999) - in this method the mitochondria flow in a liquid matrix (a neutral buffer) through a chamber. A large potential difference is maintained orthogonally across the chamber. The mitochondria take up position in the potential gradient according to their surface, not internal, charge and thereby emerge separated into groups (Fig. 2). The resulting proteomes in individual mitochondrial fractions can quite different from each other and may be altered by the nutritional stimulus applied to biological model.

Techniques of protein separation Following the initial steps to enrich for the proteins of interest, biochemists have used multiple procedures based on the sizes of native proteins (possibly in complexes) by gel filtration, their charges (ion exchange and chromatofocusing), their hydrophobicity (reverse-phase chromatography), their differential binding to immobilized dyes (e.g., Blue Sepharose™), and their specific binding to affinity ligands (small molecules or antibodies) (Fig. 3). All of these methods are appropriate in 2004 in combination with high throughput and highly sensitive proteomics technologies. In the future, the older technologies will be modified to employ microfluidics approaches on etched chips (Kvasnicka, 2003). The proteome can be significantly enriched for low abundance proteins using one or a combination of these techniques. Using molecular biology techniques, specific immunologically detectable groups can be introduced into the expressed protein (Fig. 3). For instance, we have engineered a recombinant form of the human liver bile acid CoA:amino acid N-acyltransfease (hBAT) that contained a C-terminal region that was biotinylatable *in vivo* (Sfakianos *et al.*, 2002). The biotinylated protein was easily recovered from cells by passing cell lysates over strept-avidin affinity phase and then eluting the bound protein(s) with 5 mM biotin. If washing and elution steps are mild, this method is also a way to isolate proteins that are associated with the biotinylated protein.
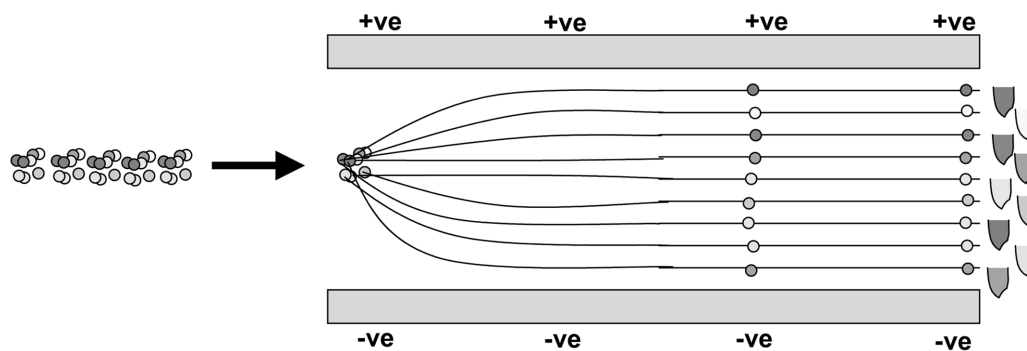

Fig. 2. Schematic of the principles of free-flow electrophoresis. The sample containing particulate matter (e.g., mitochondria) to be fractionated is introduced into a chamber along which it flows linearly. A voltage is applied across the chamber orthogonal to the direction of the flow. The particulate species, depending on their surface charge, move to a place on the potential gradient and emerge from the chamber in parallel streams. They can be collected in individual wells in a multi-well plate.
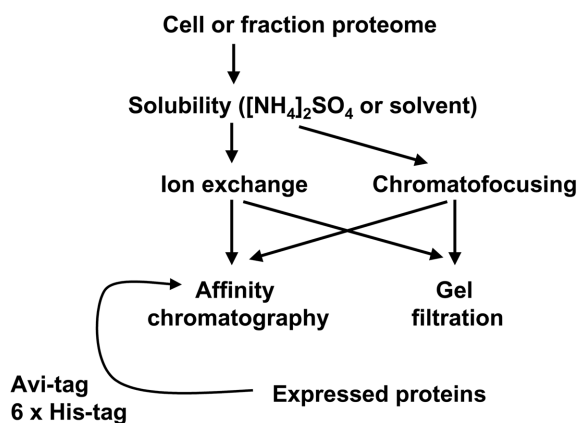
Fig. 3. Fractionation of the proteome. Proteins can be separated into groups using their chemical and physical properties (differential solubility, charge, hydrophobicity, size and affinity for particular reagents). In some cases, a particular affinity group can be added using molecular biology techniques.

Fig. 4. Principles of two-dimensional electrophoresis for the separation of proteins. Proteins solubilized under denaturing conditions are first resolved by isoelectric focusing (IEF) where they "focus" at their isoelectric points or pIs. The IEF strip is then laid on top of a SDS-PAGE gel and the proteins therein resolved in the second dimension according to differences in molecular weight. Proteins that have single amino acid differences or phosphorylated or otherwise modified, often form horizontal "strings" of spots, indicating charge heterogeneity.

Protein separation by electrophoresis and other 2D-methods As mentioned earlier, two-dimensional electrophoresis has been used for several decades to resolve complex mixtures of proteins and it remains a viable approach in proteomics (Anderson *et al.*, 2001). In this method, the proteins are denatured and dissociated from each other by homogenization of the biological sample in a solution containing high urea (usually 5-9 M), thiourea (usually 2 M), a detergent (4% CHAPS), and a reductant (5 mM tributyl phosphine). The individual polypeptides are resolved from each other in the first dimension (an immobilized pH gradient [IPG] strip) according to differences in charge, then in the second dimension according to differences in size. In this way, a two-dimensional "display" of the polypeptides is obtained, that by definition gives greater resolution of the polypeptides in a mixture than running either dimension alone (see Fig. 4). While this method is one of the most labor intensive of the several types of two-dimensional protein separation methodologies available at the "front end" of proteomics technology, it is one of the few that yields a physical separation of the intact (i.e. nonproteolytically-cleaved) polypeptides from each other in a display that itself is informative; that is to say, the position of a polypeptide spot in a 2D gel tells us what is its molecular weight, as well as its isoelectric point (pI), both parameters that can be used to narrow down or confirm the identification of the protein. Since the pI reflects the charged state of a protein, its observed pI and any heterogeneity in the spots at the same molecular weight can aid in determining whether and how the polypeptide is modified *in vivo*, and whether there are differently modified forms of the polypeptide. For example, if the predicted pI (once the protein is identified following mass spectrometry tryptic mass fingerprint analysis) for a protein is 6.5, and the one gel spot for the protein is found at a pI of 5.5, this would indicate t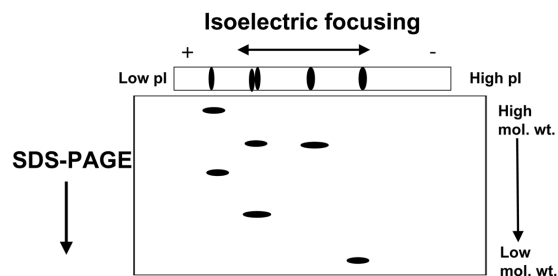hat the protein could have undergone modification by phosphorylation that would result in its having a more acidic pI than the predicted unmodified protein.

It should be noted that if the molecular weight and the pI of a protein(s) of interest are known beforehand, the 2D gel parameters can be selected to optimize for the separation of these proteins of interest. For example, if one wanted to determine if a protein of interest of molecular weight 60 kDa and pI 4.5 underwent phosphorylation in response to a stimulus such as a dietary component, a first dimension separation would probably be a pH gradient of 3-6, and the second dimension separation would probably be a medium range (10%) acrylamide gel. On the other hand, if the goal was to identify as many proteins as possible that interacted with the protein of interest, one would use antibody to affinity purify the protein and its associated proteins, then resolve the mixture on a 3-10 pH gradient, and probably a 10-20% acrylamide gradient; these are 2D gel conditions we commonly use to resolve complex mixtures of proteins of unknown molecular weights and pIs.

Once proteins have been separated by 2D-electrophoresis, images of the gels are acquired by either densitometry (of visibly stained gels) or by fluorescence scanning instrumentation. These images are then analyzed by software specific for this purpose to "detect" the spots. This assigns intensities to the spots/peaks, as well as x,y coordinates. These data can then be analyzed statistically using various algorithms such as provided by Statistical Analysis Systems (SAS), to determine which spots in the gels representing different samples in the experiment are statistically different in either intensity (indicating differential expression), or in horizontal position (indicating a difference in posttranslational modification). This area of statistical analysis of 2D gel images is probably the single largest bottleneck in 2D gel analysis, because while various software packages exist for 2D gel image analysis (such as PDQuest [BioRad Laboratories, Hercules, USA] and Progenesis Discovery

[NonLinear Dynamics, Inc., Newcastle-on-Tyne, UK]), *none* (to date) provides adequate statistical analytical capabilities; as such, the investigator must rely on assistance from biostatistician colleagues, who both understand the biological aspects of the experiment, as well as have the required statistical expertise.

A major objective of the statistical and image analysis is the identification of gel spots that are designated for "picking," that is to say, these spots will be excised out of the gel, for analysis by mass spectrometry to identify the polypeptide and/or its modifications. The image analysis software that generates the "picklist" also usually results in an "annotated" gel image, where each spot that is on the picklist is indicated on the gel by either a spot number, or letter. While gel spots can be excised manually, robotic spot picking that is driven by software is preferable because spots that were detected in fluorescently stained gels can be excised by the instrument, whereas the investigator would have to post-stain after Sypro Ruby with Coomassie Brilliant Blue, then re-align the gel images, to determine which spots to excise. On the other hand, manual spot picking allows picking as large a spot as desired, increasing the amount of protein for subsequent analysis. Most robotic pickers will excise a precise 1.0-1.2 mm diameter spot out of each spot, no matter how small or big the spot on the gel. This can be an issue if the amount of protein in that precise plug is barely above the threshold for detection by subsequent mass spectrometry analysis.

All in all, 2D gel image analysis and spot-picking technology have developed to such an extent that the rate limiting factor in extracting information out of an experiment is the speed with which one can repeat the experiment, or generate enough sample to have sufficient amounts in the lower abundance protein spots.

As well as ultimately identifying targets of dietary components (Kim *et al.*, 2002), proteomics technology can also be used to study the protein composition of foods themselves. Figure 5 shows a 2D gel separation of the soybean cotyledon proteome. In recent years, much attention has been drawn to the health benefits of the soybean and soy-derived foodstuffs, such as tofu and soy protein isolate (Barnes, 1998). In particular, the health benefits of the soy isoflavones, contained in unextracted soy protein, but often obtained as preparations extracted from soy protein, have been intensively studied in cell (Barnes, 1995) and animal models of disease (Barnes, 1997; Lamartiniere *et al.*, 2002) because of their purported estrogenic actions (Hsieh *et al.*, 1998), and their anti-oxidant activities (Boersma *et al.*, 2003). Many experiments have been carried out comparing the effects of isoflavone-free soy protein with intact soy protein, with the assumption that the protein component in both cases was the same. Since the extraction to remove the isoflavones involves several chemical steps, including a hot alcoholic extraction, it would not be shocking if the protein composition in the resulting isoflavone-depleted soy protein preparation was not the same as the starting material. Systematic analysis has not
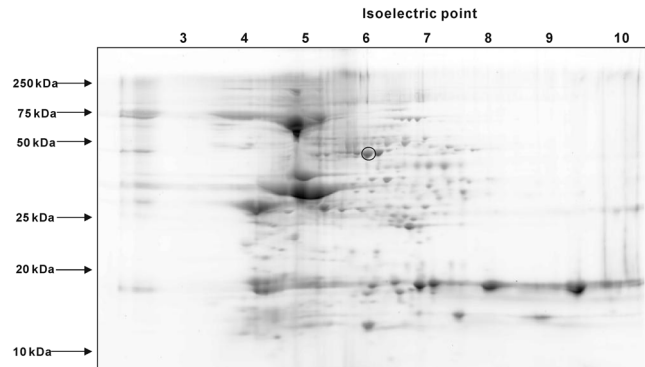

Fig. 5. Two-dimensional electrophoresis of soy cotyledon proteins. Soybeans were pulverized in liquid nitrogen and homogenized in 7 M urea: 2 M thiourea containing 5 mM tributyl phosphine. Particulate material was removed by ultracentrifugation. 11 cm pH 4-7 IPG strips were rehydrated overnight with 125 µg soluble soybean protein, and the proteins resolved by isoelectric focusing. Resolution in the second dimension was carried out on a 8-16% acrylamide gradient Criterion gel (Bio-Rad Laboratories, Richmond, USA). The gel was stained with the fluorescent dye Sypro Ruby Red and visualized by fluorescence imaging with a ProXpress imager (Perkin Elmer Corp., Wellesley, USA). (courtesy of Todd Sanderson, Stephen Barnes and Helen Kim, unpublished observations).

been carried out on the protein complement of soy protein versus "isoflavone-free" soy protein to determine whether both preparations actually contain the same complement of proteins, although preliminary data indicate differences (Gianazza *et al.*, 2003). The soybean cotyledon has a complex proteome as judged by 2D-electrophoresis (Fig. 5). This is a starting point for a systematic analysis of this proteome as the soybean is processed to yield the various protein-rich preparations that are used for foodstuffs and for experimental purposes.

In the next decade, some other physical separation technology other than "regular" 2D - electrophoresis will undoubtedly evolve, because of the need for a method where the same sample runs EXACTLY the same, where positional differences are true differences between samples, and not gel-to-gel variations, as is the case now. Even with commercially manufactured IPG strips, and commercially poured acrylamide gels, it is not uncommon for the same protein to run slightly differently on three replicate gels, under ostensibly the same pH and % acrylamide conditions. One 2D gel technology that minimizes gel-to-gel variation is "difference gel electrophoresis" (DIGE), (Unlu *et al.*, 1997), where two similar but nonidentical protein mixtures such as brain homogenates from two animals, one treated, one control, are each reacted with a fluorescent dye (cy-dyes, Amersham Biosciences, Piscataway, USA) such that the proteins in one sample fluoresce green, and the proteins in the other sample fluoresce red. The key to DIGE is that after reaction with the fluorescent dyes, both samples are MIXED,
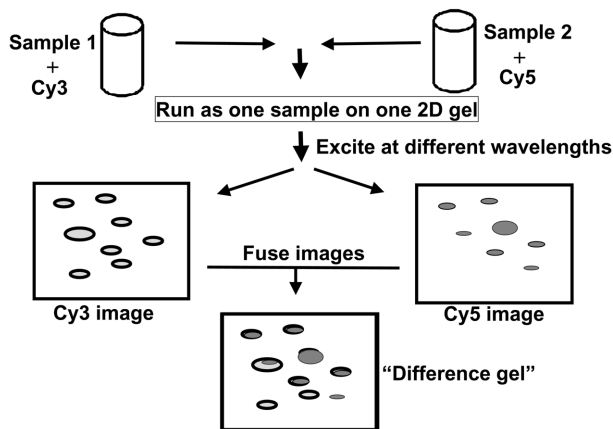
Fig. 6. Difference gel electrophoresis. This method is an adaptation of standard 2D-IEF/SDS-PAGE. Two samples to be compared are separately labeled with Cy-dyes that fluoresce at different wavelengths (green and red). The samples are then mixed and run on the same 2D gel. This ensures that each pair of protein mixtures is subject to the same analytical conditions. The proteins from the green sample can be viewed independently from the red sample, or fused into a single image, thereby enabling the detection of any differences in the proteome of each sample.

and co-electrophoresed on the same 2D gel, so that a given protein from the two different samples co-migrates with itself. Figure 6 schematically shows the elements of DIGE.

Yet, DIGE has caveats, too. For one thing, the current technology of labeling with the fluorescent dye involves "minimal" labeling, where only 1-3% of the lysines in a protein get labeled, because labeling beyond this extent modifies the mass of the protein. The result of this "minimal" labeling is that for lower molecular weight spots, the spot that is fluorescently labeled migrates at a slightly different position to the rest of the spot that is unlabeled, yet the latter is what is excised for mass spectrometry analysis, because it has the bulk of the protein. The end result: the cy-dye labeled gels must be post-stained, with either a straight fluorescent dye such as Sypro Ruby, or with Coomassie Brilliant Blue, to visualize the position of the rest of each spot that was not labeled with the cy-dyes. Secondly, with complex proteomes, if a spot shifts due to a modification, it is not inconceivable that it will be superimposed on top of a neighboring protein, resulting in the appearance of that spot having increased in intensity. Mass spectrometry analysis will detect this heterogeneity, but the quantitation will be skewed in such cases, requiring re-running of the samples separately, to determine which spot belongs to which protein.

Given the limitations of DIGE, and the other conventional 2D electrophoretic methods, 2D - electrophoresis is as stated earlier, the only method where intact related polypeptides are physically resolved from each other, and the chemical basis for their relatedness is often predicted directly from their gel positions.

One acknowledged limitation of "regular" 2D-electrophoresis as discussed above is that some categories of proteins, particularly the intrinsic membrane proteins such as the mitochondrial complexes, are poorly solubilized in conventional isoelectric focusing sample buffer. This results in their absence from the 2D gel. However, this difficulty has been overcome by developments in the "blue-native" gel electrophoresis methodology (Schagger and von Jagow, 1991). In this method, mitochondria are solubilized in a buffer containing a mild detergent (sodium lauryl maltoside) and a zwitterionic buffer (aminocaproic acid), along with low amounts of Coomassie Brilliant Blue. This sample is then electrophoresed in a low percent polyacrylamide gel. Coomassie Blue not only binds to the proteins, enabling visual detection during the electrophoresis (hence the nomenclature "blue-native"), but also results in all the proteins being charged similarly, so that they migrate in the same direction *down* the gel, when loaded at the top of the non-denaturing gel. After this first dimension separation of the native complexes (some of which are greater than 800 kDa molecular weight), the individual components of the complexes are resolved in the second dimension SDS-PAGE step forming a ladder in the gel much as if they had been loaded in a conventional well at the beginning of the electrophoresis (see Fig. 7, which illustrates the blue native electrophoresis of heart mitochondrial complexes). Although the protein spots are not nearly as tightly focused as in conventional 2D-IEF/SDS-PAGE, nonetheless each protein can be excised and identified by mass spectrometry (Brookes *et al.*, 2002). This technique is not limited to mitochondrial protein complexes, but in theory could be used for any intracellular protein complex (Brookes *et al.*, 2002; Camacho-Carvajal *et al.*, 2003).

Besides 2D-electrophoresis, proteins can be separated entirely in the liquid phase (2D-LC) using combination techniques. Recently, a commercial 2D-LC instrument has been introduced that combines chromatofocusing and hydrophobic chromatography (Chong *et al.*, 2001). Fractions eluting from the chromatofocusing column over a narrow pH range are further resolved on a reverse-phase column. The advantage of this method is that it isn't restricted to a particular mass range and further can be completely automated, thereby being very reproducible. The resolved proteins are hydrolyzed with proteases and subjected to mass spectrometry analysis (see below).

Mass spectrometry of proteins and peptides The introduction of two soft ionization techniques, electrospray ionization (ESI) (Fig. 8), and matrix-assisted laser desorption ionization (MALDI) (Fig. 9), has allowed protein chemists to readily analyze both peptides and whole proteins by mass spectrometry (Yates, 1998). Ions formed by ESI are multiply charged - the importance of this is explained later - whereas MALDI ions are singly charged. There are several types of mass spectrometer detectors that are used to observe the
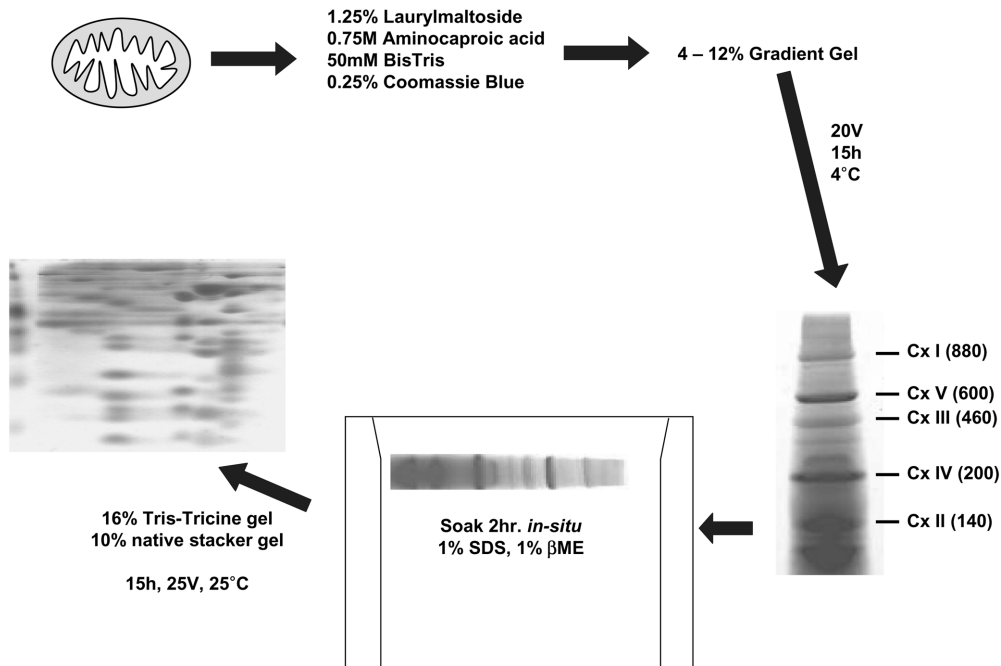
1.25% Laurylmaltoside
0.75M Aminocaproic acid
50mM BisTris
0.25% Coomassie Blue

4 – 12% Gradient Gel

20V
15h
4°C

— Cx I (880)
— Cx V (600)
— Cx III (460)
— Cx IV (200)
— Cx II (140)

16% Tris-Tricine gel
10% native stacker gel

15h, 25V, 25°C

Soak 2hr. *in-situ*
1% SDS, 1% βME

Fig. 7. Blue native gel electrophoresis of heart mitochondria. The mitochondrial preparation was electrophoresed in the first dimension on a 4-12% polyacrylamide gradient in the presence of a weak detergent (1.25% laurylmaltoside), an amphoteric buffer (0.75 M aminocaproic acid-50 mM Bis-Tris) and 0.25% Coomassie Blue. After this the gel piece was soaked in 1% SDS-1% β-mercaptoethanol for 2 h and placed on top of a 16% polyacrylamide-Tris-Tricine SDS-PAGE gel. It was stained with Coomassie Blue. These data were adapted from data presented by Brookes *et al.* (2002) (Brookes, unpublished data).
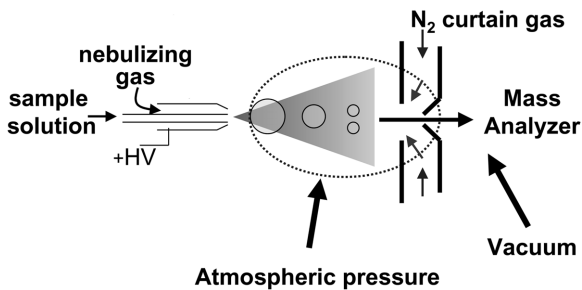

N₂ curtain gas

nebulizing gas

sample solution

+HV

Mass Analyzer

Vacuum

Atmospheric pressure

Fig. 8. Principles of electrospray ionization. Samples are forced through a narrow capillary to create a charged spray (the nebulizing gas assists spray formation). The finely divided fluid droplets undergo rapid evaporation (assisted by a countercurrent flow of warm dry N_2 curtain gas). The charged ions within the droplet exert an outward coulombic pressure that causes the droplets to burst, thereby releasing the ions into the gas phase. They then move orthogonally towards the orifice of the mass spectrometer.
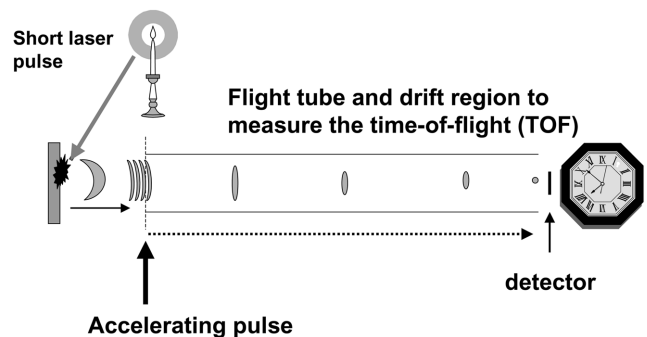

Short laser pulse

Flight tube and drift region to measure the time-of-flight (TOF)

detector

Accelerating pulse

Fig. 9. Matrix-assisted laser desorption ionization. Ions are formed by firing short ns 337 nm N_2 laser pulses at the crystalline matrix containing the peptide(s) or protein(s). Matrix and sample ions accumulate at a wire mesh and are then accelerated. The ions *drift* down the flight tube. Their *time-of-flight* is proportional to their *m/z* values.

peptide and protein ions. A quadrupole detector is a filter - ions pass down the long axis of four rods exhibiting a spiral motion. At one combination of applied Rf and voltage on the rods, a particular ion of the form $[M+nH]^{n+}$ exhibits a stable spiral, whereas ions at other mass-to-charge (*m/z*) ratios spiral outwards and strike the rods. This type of detector is very stable, but has low mass resolution and it takes 1-2 sec to obtain a full spectrum. A further disadvantage of the

quadrupole detector is that the quadrupole filter throws away >99% of the signal since it only measures one *m/z* value at any one moment.

This problem is overcome to a degree by using a time-of-flight (TOF) detector (Fig. 9). For TOFs, ions are injected into a drift region of the mass spectrometer (down a tube 1-2 m in length). The time taken for the ions to reach a photodiode detector at the end of the flight tube is accurately measured and converted to *m/z* values. In this method, all the ions in an

ion packet are detected and time-resolved. The duty cycle of this detector is such that MALDI-TOF mass spectrometry is much more sensitive than one based on ESI-quadrupole analysis.

Ions can also be captured in a "trap" and ejected out one at a time in order to be detected. This *ion trap* is relatively simple in construction and is the most economic choice for LC-ESI-mass spectrometry analysis of peptides. Since it can also be used to carry out multiple steps of fragmentation (MS^n), it is a very valuable all-round instrument in a nutrition research laboratory. The concept of using a trap for ions during mass spectrometry analysis has been extended to linear ion traps that accumulate ions formed after fragmentation of peptide molecular ions $[M+nH]^{n+}$. This technique permits peptide sequencing at the attomole level. In a related method (Fourier Transform-Ion Cyclotron Resonance-Mass spectrometry - FT-ICR-MS) (Marshall *et al.*, 1998), ions can be injected into a cylindrical cell that serves as an ion trap. The cell is located in the center of field of a superconducting magnet. Ions undergo circular motion within the cell caused by the magnet field. They can be excited by the application of ion cyclotron radiation. This is analogous to nuclear magnetic resonance spectroscopy. The ions decay back to their ground state once the cyclotron radiation is turned off and in doing so, they induce a current in solenoids surrounding the cell (free induction decay; FID). The FID is subjected to Fourier Transform processing, converting it from the time domain to the frequency domain and hence to give mass spectrum. FT-ICR-MS provides not only high sensitivity and mass resolution, but also very high mass accuracy (for a 9.4 Tesla magnet, this can be less than one part per million).

MALDI and peptide mass fingerprinting   When proteins (with a particular amino acid sequence) are incubated with proteases with good specificity, a predictable set of peptides is produced (Mortz *et al.*, 1994). Since these fragments in turn have known amino acid sequences, they also have expected and exact molecular weights. Thus each protein produces a *peptide mass fingerprint* that is particular for a given protease. For trypsin that hydrolyzes at arginine and lysine residues, a typical 500-residue protein may be converted to 50 or so peptides. Approximately one third to one half of these peptides have masses between 800 and 2500 Da. These peptides are readily detected using MALDI-TOF-MS. This method is readily applied to protein spots that were resolved by 1D-SDS-PAGE or (better) 2D-IEF/SDS-PAGE. The gel piece is washed with 50% aqueous acetonitrile to remove dye reagent, buffer salts and detergents. After drying, it is rehydrated in 25 mM ammonium bicarbonate, pH 8.0, and trypsin (1 : 20 by weight) added. Incubation is carried out at 55°C for 2 h or overnight at 37°C. Peptides are recovered by aspirating the buffer and then extracting the gel piece with 50% aqueous acetonitrile-5% formic acid. These are combined and evaporated to dryness. For MALDI-TOF
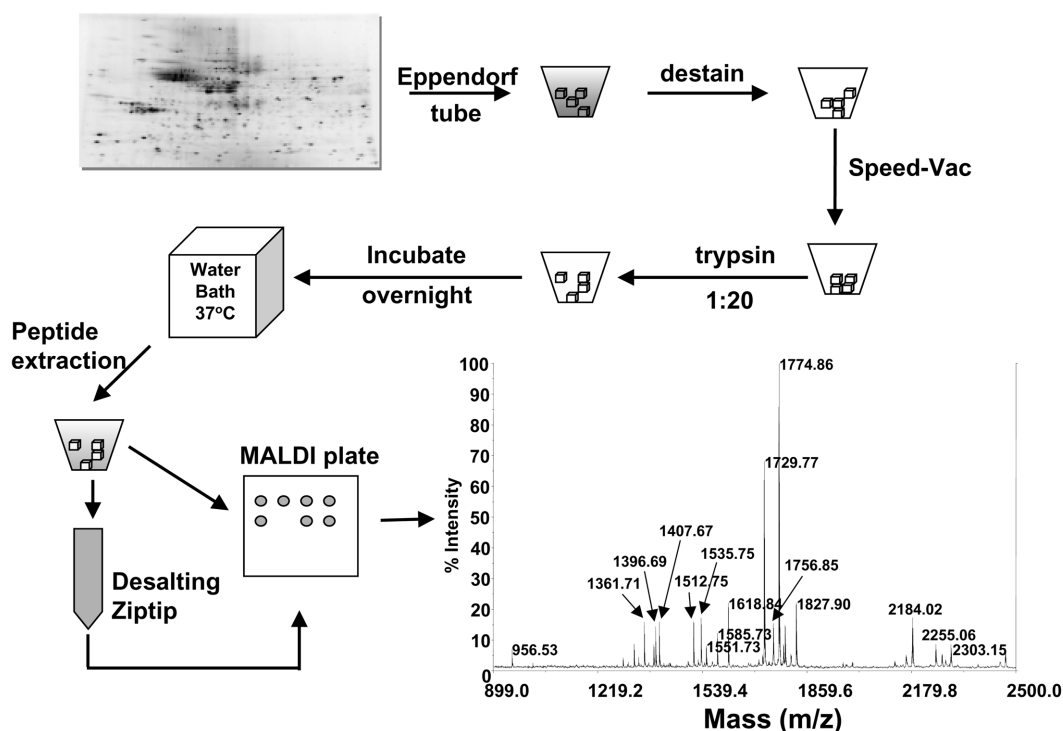

Fig. 10. Summary of peptide mass fingerprinting. Individual spots from a 1D- or 2D-gel are excised, washed to remove the dye and other reagents, dried and rehydrated in buffer containing trypsin. After incubation, the tryptic peptides are recovered and spotted onto a target plate prior to MALDI-TOF-MS analysis.
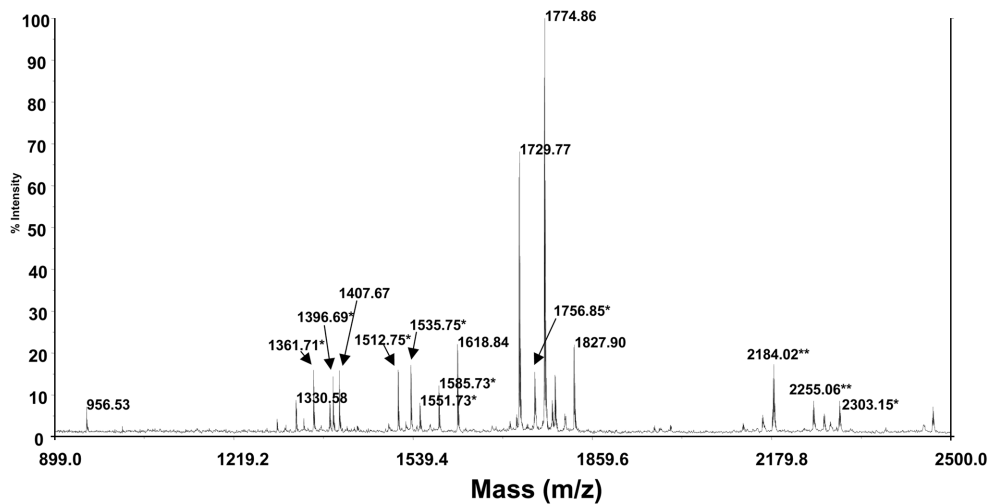
Fig. 11. Peptide mass fingerprint of soybean protein. MALDI-TOF mass spectrum of a trypsin-digested, protein spot from the 2D-electrophoresis map in Fig. 4. The best MOWSE score for these peptides was for the soybean beta-conglycinin beta subunit.

analysis, the peptides are dissolved in a small volume (10-20 μl) of 50% aqueous acetonitrile-0.1% formic acid and an aliquot (1 μl) mixed with a saturated solution of α-cyano-4-hydroxycinnamic in 50% aqueous acetonitrile. One microliter of this mixture is applied to a MALDI target plate and allowed to form crystals and then dry thoroughly. Depending on the application, 2,5-dihydroxybenzoic acid and sinapinic acid are also used as matrices (Katta *et al.*, 1998; Katayama *et al.*, 2001). Once in the mass spectrometer, each target in turn is subjected to short nanosecond laser pulses that "evaporate" a small portion of the crystal surface forming ions of the matrix and the peptides therein. This MALDI-TOF-MS analysis produces a tryptic *mass fingerprint* (Fig. 10). For modern MALDI-TOF instruments, mass accuracies better than 50 ppm are routine. After extracting the *m/z* values for the peptides, these are analyzed by freely available search engines (e.g., MASCOT at http://www.matrixscience.com and MS-Fit at http://prospector.ucsf.edu/ucsfhtml4.0/msfit.htm). The peptide pattern of the unknown protein is compared against the peptide patterns of all known proteins. A statistical scoring scheme (MOWSE when using MASCOT) is used to determine the proteins that most closely correspond to the peptides of the unknown protein. Typically, 4-8 peptides are sufficient for this purpose. In the example shown in Fig. 11, the protein with the highest MOWSE score was soybean beta-conglycinin beta subunit. Analysis of MALDI ions by FT-ICR-MS produces much better mass accuracy and thus substantially improves the statistical scores (Witt *et al.*, 2003). MASCOT and MS-Fit also permit the peptides to be modified, for instance acetylation at the N-terminus, oxidation of methionine residues, and phosphorylation at serine, threonine and tyrosine residues.

Interpretation of peptide fragmentation spectra While peptide mass fingerprinting using MALDI-TOF-MS has greatly

facilitated the MS analysis of proteins, it does not produce a high enough level of proof of identity. Proteins often exist in related families and the limited number of peptides that are observed may be in common between the family members. Further, as more genomes are sequenced, it becomes increasingly more difficult to be certain that the correct protein has been "identified". A much higher degree of certainty can be obtained by fragmenting the peptide molecular ion and analyzing the daughter ions (Fig. 12). The peptides are first resolved on an acetonitrile gradient (in 0.1% formic acid) by reverse-phase chromatography. Since ESI-MS is concentration-dependent, there is an advantage in decreasing the scale of the analysis; most analysts now use 75 μm i.d. or smaller columns (so-called *nanoHPLC*) (Shen *et al.*, 2002). Compared to analysis on a conventional 4.6 mm i.d. HPLC column, a sample analyzed by nanoHPLC will have an increased sensitivity of $(4.6/0.075)^2 = 3,762$ times. The nanocolumn operates at a flow rate of 200-250 nl/min; so, consequently there is no need to split the flow entering the ESI interface. Although there is great flexibility in carrying out tandem mass spectrometry using a triple quadrupole instrument (Fig. 13A), a more sensitive method is the hybrid orthogonal quadrupole-TOF (Qq-TOF) (Fig. 13B). The recently introduced Applied Biosystems Inc. Q-trap takes advantage of using a hexapole to accumulate daughter ion fragments prior to their analysis. As an alternative method, the column eluant can be laid down on a pre-coated MALDI target plate (Rejtar *et al.*, 2002). This method has the advantage that the eluant is captured on the plate and can be re-interrogated using MALDI-TOF-MS analysis as needed. This method can be further enhanced by using MALDI-TOF-TOF-MS. In this approach, the MALDI-generated peptide ions are individually selected by a highly time-resolved ion gate, collided with argon gas, and then the fragment ions re-accelerated for a second TOF analysis (Bienvenut *et al.*, 2002).

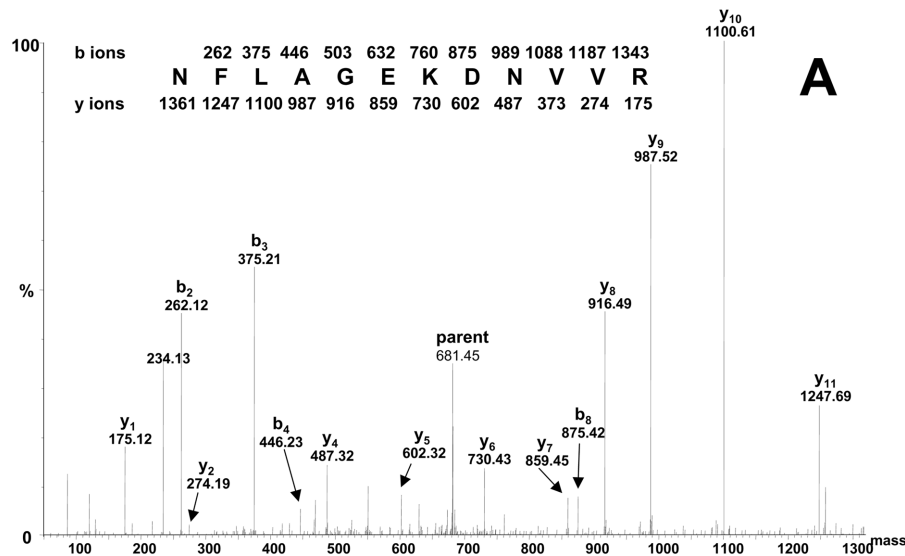Tryptic peptides form doubly charged ions in the ESI

NH_3^+-CHR^1-CO-NH-CHR^2-CO-NH-CHR^3-CO-NH-CHR^4-CO-NH-CH-COOH

$(CH_2)_4$

NH_3^+

NH_2-CHR^1-CO-NH-CHR^2-C≡O$^+$ b_2 ion

NH=CHR^3-CO-NH-CHR^4-CO-NH-CH-COOH y_3 ion

$(CH_2)_4$

NH_3^+

Fig. 12. Fragmentation of a doubly charged tryptic peptide. These peptides have a positive charge on the N-terminal amino acid and on the C-terminal arginine or lysine. The energy derived from collision with a neutral inert gas (usually Argon) causes cleavages at the peptide bonds. Ions that contain the N-terminal amino acid are termed "b" ions; those that contain the C-terminal arginine or lysine are "y" ions. They are formed by independent processes.

(A)

(B)

Fig. 13. Tandem mass spectrometry. In A (the triple quadrupole), electrospray peptide molecular ions are selected in the first quadrupole, collided with argon gas in the second quadrupole to produce daughter ion fragments, and analyzed in the third quadrupole. In B, a similar set of processes occurs, except that the fragment ions are injected orthogonally into a TOF detector. This provides not only increased sensitivity, but also mass accuracies for the fragment ions better than 20 ppm.

interface - they have a charged N-terminal amino group and a charged amino group on the arginine or lysine C-terminal residue. When they are collided with an inert gas (argon), they undergo fragmentation largely at their peptide bonds (Figs. 12 and 14). There are two major types of fragment ions; "b" ions that include the N-terminus and "y" ions that contain the C-terminus. They are produced by two independent mechanisms. Not all the "b" and "y" ions will be detected for a given peptide. As shown in Fig. 14, the daughter ions that are observed in the tandem mass spectrum of the soybean beta-conglycinin beta subunit tryptic peptide NH_2-NFLAGEKDNVVR-COOH (residues 371-382) include b_2, b_3 and b_4, and y_1, y_2, y_4, y_5, y_6, y_7, y_8, y_9, y_{10} and y_{11} ions. Additional types of fragmentation of peptides can occur -

these are well-described at http://www.matrixscience.com/help/fragmentation_help.html.

Peptide sequencing and posttranslational modifications
Discovery of posttranslational modifications can be carried out in several ways. In one type, a pan antibody to the protein of interest is used to immunoprecipitate all forms of the protein. This purified fraction is digested with several proteases and the resulting peptides closely examined for evidence of any posttranslational modification. An alternative approach is the use of an antibody to the modification. For instance, antibodies to nitrotyrosine, phosphoserine, phosphothreonine and phosphotyrosine have been utilized to isolate the nitroproteome (Aulak *et al*., 2001) and the

Fig. 14. Tandem mass spectrum of a tryptic peptide of soybean beta-conglycinin beta subunit. The doubly-charged *m/z* 681.45 tryptic peptide ion (residues 371-382) was collided with argon in a Qq-TOF mass spectrometer producing the daughter ion mass spectrum. The expected "b" and "y" ions are shown above and below the amino acid sequence.

phosphoproteome (Gronborg *et al.*, 2002; Imam-Sghiouar *et al.*, 2002). It is important to understand that these antibodies may be quite selective for the modification site and may be sequence-dependent and therefore may not provide a complete proteome. Some investigators combine several different forms of each antibody to enable a greater coverage of the proteome (Aulak *et al.*, 2001; Gronborg *et al.*, 2002). Lectins and antibodies to specific glycosyl groups have been used to isolate classes of glycoproteins (Kaji *et al.*, 2003). Of course, the materials isolated these ways are complex and investigators use both 2D-gel electrophoresis and 2D-LC methods to resolve the individual proteins or peptides prior to mass spectrometry analysis. Methods based on the chemical properties of the modification have been used. For instance, the nitro group in nitrotyrosine can be reduced to an amine (Viera *et al.*, 1999). This allows it to selectively react with a sulfonic acid derivative linked to a biotin group and thereby be recovered from a mixture using a Strept-Avidin column (Helbeck *et al.*, 2003). Glycosyl and phospho-groups linked to serine and threonine undergo beta-elimination under alkaline conditions (Greis *et al.*, 1996; Wells *et al.*, 2002). This reaction removes a molecule of water from these amino acids forming dehydroalanine and dehydroaminobutanoic acid. These amino acids can be recognized because of their unique residue molecular weights (69 and 83, respectively) (Greis *et al.*, 1996). Alternatively, the double bond can undergo a Michael reaction and thereby either introduce a sulfhydryl group (with ethanedithiol) (Adamczyk *et al.*, 2001) or an amino group (aminoethanethiol) (Zhou *et al.*, 2003). The former will react with the ICAT reagent (Fig. 15A) whereas the latter has created a new trypsin-sensitive site (Fig. 15B).

Verification of the identity and position of the modified amino acid residue is carried out by tandem mass spectrometry. In some cases, the modified group remains intact during fragmentation of the peptide backbone by collisional dissociation. This is the case for nitrotyrosine (Crow *et al.*, 1997) and phosphotyrosine residues (Carr *et al.*, 1996), thereby easily pinpointing the modification site. However, O-glycosyl groups undergo decomposition prior to fragmentation of the peptide backbone (Medzihradszky *et al.*, 1990). In this case, electron capture dissociation in a FT-ICR-MS is an effective solution providing collisional processes are minimized. In this procedure, c ions containing the O-glycosyl group are prominent ions in the mass spectrum (Spear P, Barnes S, Hart, GW, unpublished observations).

2D-LC of peptides Another approach to a more automated analysis of the proteome is the application of Multi-Dimensional Protein Identification Technology (MuDPIT) (Washburn *et al.*, 2001). In this method, the proteome under investigation is hydrolyzed to its component peptides. These are fractionated in two ways. First, they are absorbed onto a strong cation exchanger on a silica base and then eluted with increasing concentrations of a counter cation in 5% acetonitrile. The peptides in each fraction are captured on a reverse-phase column and eluted with a linear gradient (5-50%) of acetonitrile in 0.1% formic acid. If the different phases are contained in a single, multi-layered column, then ammonium acetate is used for elution from the cation exchanger (Fig. 16A). For a two column system, KCl can be used since the peptides can be captured on an intermediate reverse-phase cartridge and the excess KCl that would otherwise interfere with the ESI process passed to waste (Fig. 16B) prior to beginning the acetonitrile gradient (Davis *et al.*, 2001). In both cases, the analysis of the peptides that undergo tandem MS can be carried out using a quadrupole ion trap -
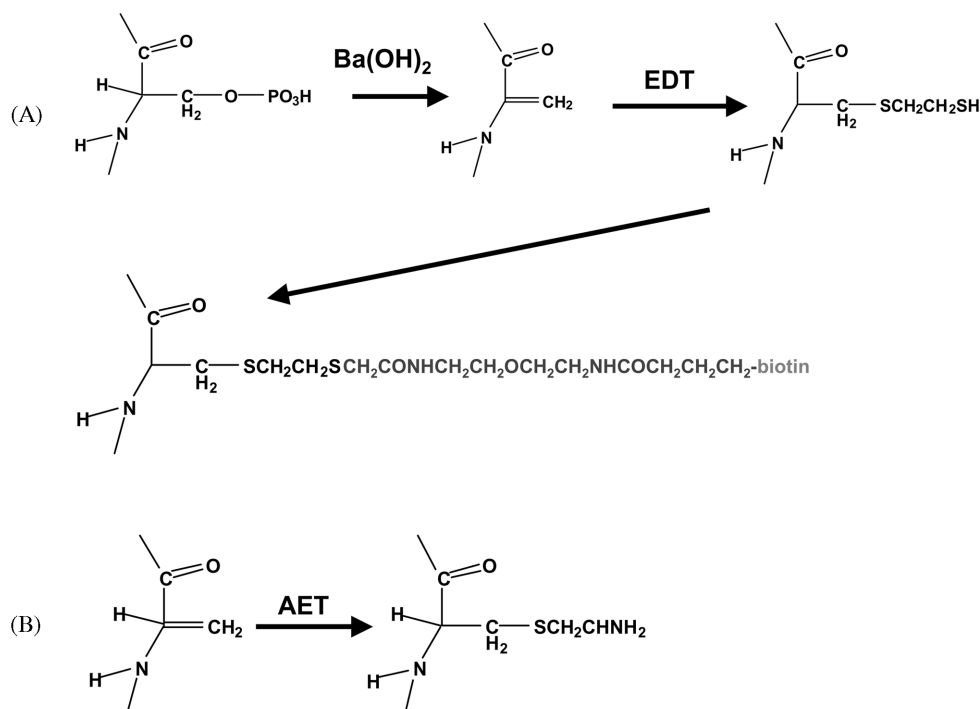
Fig. 15. Beta elimination of phosphate group from phosphoserine. The mild alkaline treatment causes elimination of phosphoric acid, creating the double bond in dehydroalanine. Michael addition of ethanedithiol (EDT) and aminoethanethiol (AET) to the double bond introduces a biotinylatable site (A) and a trypsin-sensitive site (B).

this is the most economic method - or with much greater mass accuracy on a Qq-TOF. For both these methods, the sizes of the tandem MS files are enormous and have a requirement for large scale computing. The software program SEQUEST (available from ThermoFinnigan) is used for the automated analysis of the tandem MS spectra. Similarly, the data files can be submitted (in the appropriate format) for analysis by MASCOT. For each observed peptide parent ion, these programs make a list of all known peptides that are within ± 1 Da. The observed tandem mass spectrum of the unknown peptide is compared against the predicted mass spectra of all the selected peptides. A statistical scoring scheme is used to determine the goodness of fit.

Quantitative and qualitative burrowing of the proteome
Isotope labeling is used to obtain quantitative proteomics information. This includes the use of $^{15}N/^{14}N$ (Wang *et al.*, 2002), $^{18}O/^{16}O$ (Reynolds *et al.*, 2002) and so-called ICAT (isotope-coded affinity technology) (Gygi *et al.*, 1999) to differentially label the control and treated sample proteomes. Analysis of the labeled peptides is carried out by LC-ESI-MS, in many cases with an ESI-TOF configuration. This places a heavy emphasis on successfully chromatographically resolving the peptides using the 2D-LC approach in MUDPIT analysis.

ICAT analysis involves a reagent with three different parts, an iodoacetamide group that reacts with peptide sulfhydryl groups, an intermediate bridge that contains the isotopic carbon or hydrogen atoms allowing for differential labeling, and a biotin group for affinity capture of the derivatized peptides onto a Strept-avidin column. In each of these applications, the LC-MS analysis is carried out on samples (control and treatment) that have been mixed thereby insuring that each sample is subject to the same analytical conditions. In the original version of the ICAT reagent, deuterium labeling of the bridge region was used. However, the d_0 and d_8-labeled peptides are chromatographically partially resolvable resulting in an inconsistent d_0/d_8 ratio across the peptide peak. This problem has been overcome by using ^{13}C labeling of the ICAT reagent - this isotopic label has no effect on the chromatographic mobility of the derivatized peptide. There are two other limitations of the ICAT reagent. First, it can only be applied to proteins that have a cysteine group (approximately 80% of the proteome) and even then may only provide information about one peptide from that protein. Second, elution of the biotinylated peptides from the Strept-Avidin column is not quantitative, particularly the low abundance proteins. As a result, an acid cleavable group has been included in the current ICAT reagent.

Another quantitative approach for the measurement of peptides has been termed AQUA (Absolute Quantitative Analysis) (Gerber *et al.*, 2003). This method is based on multiple reaction ion monitoring using a triple quadrupole instrument. This type of analysis has been extensively used for the analysis of small molecules such as isoflavones (Coward *et al.*, 1996). It is based on tandem mass spectrometry of the
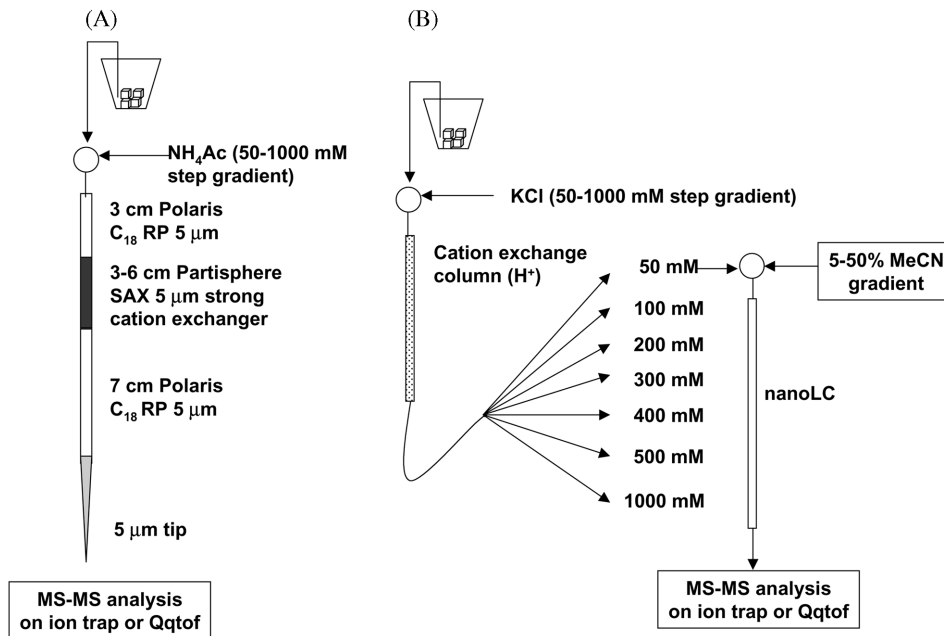
(A)   (B)

NH$_4$Ac (50-1000 mM step gradient)

3 cm Polaris C$_{18}$ RP 5 μm

3-6 cm Partisphere SAX 5 μm strong cation exchanger

7 cm Polaris C$_{18}$ RP 5 μm

5 μm tip

MS-MS analysis on ion trap or Qqtof

KCl (50-1000 mM step gradient)

Cation exchange column (H$^+$)

50 mM
100 mM
200 mM
300 mM
400 mM
500 mM
1000 mM

5-50% MeCN gradient

nanoLC

MS-MS analysis on ion trap or Qqtof

Fig. 16. Two dimensional liquid chromatography of peptides. The MUDPIT method comes in two forms. In A, the strong cation exchange phase and the hydrophobic reverse-phase are placed in the same column. This arrangement maximizes sensitivity by minimizing the dead volume of the system. A small amount of the reverse-phase material is placed on top of the column to initially capture the peptides. They are eluted with 80% aqueous acetonitrile and then absorbed by the strong cation exchanger. They are sequentially eluted with increasing concentrations of ammonium acetate (50-1000 mM) in 5% aqueous acetonitrile and are captured by the lower reverse-phase packing. They are subsequently eluted with a 5-50% linear gradient of acetonitrile. In B, two separate columns are used. This allows KCl to be used as the eluting agent for the strong cation exchanger.

peptide of interest. Instead of examining the entire daughter ion mass spectrum, a particular daughter ion is selected so that the combination of parent molecular ion and daughter ion is specific. Triple quadrupole instruments typically can have up to 8 channels in which different parent ion/daughter ion pairs can be monitored. If a ^{13}C-labeled form of the peptide is synthesized, it is possible to add it to a digest and thereby determine any losses during the processing of the sample, hence the term *Absolute Quantitative Analysis*. The obvious limitation of this method is the small number of peptides (8 or less) that can be measured in a single analysis.

As noted earlier, only a small proportion of the proteome in a cell has been investigated. There are advocates of 2D-gel electrophoresis (Anderson *et al.*, 2001), MUDPIT 2D-LC-tandem mass spectrometry (Washburn *et al.*, 2001), and the ICAT method (Gygi *et al.*, 1999). Several groups have recently reported on the comparative capabilities of these methods (Lee *et al.*, 2003; Schmidt *et al.*, 2003). What has become clear is that each approach samples a different region of the proteome. Although the MUDPIT approach generates many thousands of peptides, the data are both extensively degenerate (many peptides from the same protein) as well as having only one peptide from a given protein. Because of this, there is uncertainty as to the quality of the identification. In hydrolyzing proteins to their peptides, there is a considerable *loss* of information. 2D-gel electrophoresis, albeit slower and

difficult to automate, has the advantage that each protein form is gathered together at one location (pI and molecular weight). Thus at the present time, serious investigators must be capable of using all of the above techniques. In a recent study, the protein bands from 1D-SDS-PAGE analysis were then subjected to LC-tandem mass spectrometry analysis. This combination was superior with respect to protein coverage to either MUDPIT or 2D-electrophoresis (Lee *et al.*, 2003). Clearly, all of these approaches must be combined with other methods of pre-fractionating proteins described at the beginning of this article.

The scale of the problem that exists for the protein analyst, and that requires new methodologies, is seen in the case of the expression of two genes. The *Drosophila melanogaster* gene *Dscam* (Down syndrome cell adhesion molecule) has 100 alternative splicing sites and has over 38,000 putative protein products (Schmucker *et al.*, 2000), more than the total number of genes that have been identified in the human genome. This high degree of alternative splicing is not the case for the human homolog (Yamakawa *et al.*, 1998). The pattern of forms of *Dscam* fluctuates during development in these flies (Celotto *et al.*, 2001). Although *Dscam* is an extreme case, it highlights the need to identify the whole protein, not just a single peptide. The second example is the posttranslational modifications of the histones. These proteins have an important role in the regulation of gene expression. At least 20

sites of modification have been identified (Zhang et al., 2003; Zhang and Tang, 2003). Acetylation and mono-, di- and trimethylation of the lysine groups cause opposing effects on histone function (Geiman and Robertson, 2002). In addition, phosphorylation of serine residues in this region of the histone protein is a regulator of histone acetylation, and hence function (Zhang et al., 2003). The question is whether each molecule of the histones is modified similarly, or whether there are distinct, differentially modified histone molecules. Separating and identifying them may require an approach based on isolation of specific genes (with the histone attached).

Advanced applications of proteomics Besides the qualitative analysis of the content of a particular proteome, mass spectrometry is being used to address other major challenges of biochemistry and cell biology and will be of interest to the nutrition scientist. These include the structure of proteins in solution using hydrogen-deuterium exchange (Lanman et al., 2003), their quantitative interaction with ligands or substrates (enzymology) (Ge et al., 2001), and the nature of the protein-protein interaction networks (Gavin et al., 2002). The latter are considered crucial in understanding biological systems. In addition, clinical investigators are actively examining tissue and fluid proteomes (plasma/serum, urine, nipple aspirate, bronchial-alveolar lavage, prostatic fluid, etc.) for novel biomarkers for the detection and progression of disease (Noel-Georis et al., 2001; Spahr et al., 2001; Adkins et al., 2002; Varnum et al., 2003). The recent advances in tissue imaging of the proteome is providing important insights into the changing proteome in situ (Yanagisawa et al., 2003). Clearly, proteomics and protein mass spectrometry technologies can provide rich resources to investigators in nutrition and related areas of biomedical research. Those who are able to incorporate these technologies in their experiments will make the meaningful contributions over the next decade.

References

Adamczyk, M., Gebler, J. C. and Wu, J. (2001) Selective analysis of phosphopeptides within a protein mixture by chemical modification, reversible biotinylation and mass spectrometry. *Rapid Commun. Mass Spectrom.* **15**, 1481-1488.

Adkins, J. N., Varnum, S. M., Auberry, K. J., Moore, R. J., Angell, N. H., Smith, R. D., Springer, D. L. and Pounds, J. G. (2002) Toward a human blood serum proteome: analysis by multidimensional separation coupled with mass spectrometry. *Mol. Cell. Proteomics* **1**, 947-955.

Anderson, N. G., Matheson, A. and Anderson, N. L. (2001) Back to the future: the human protein index (HPI) and the agenda for post-proteomic biology. *Proteomics* **1**, 3-12.

Aulak, K. S., Miyagi, M., Yan, L., West, K. A., Massillon, D., Crabb, J. W. and Stuehr, D. J. (2001) Proteomic method identify proteins nitrated *in vivo* during inflammatory challenge. *Proc. Natl. Acad. Sci. USA* **98**, 12056-12061.

Barnes, S. (1995) Effects of genistein in *in vivo* and *in vitro* models of cancer growth. *J. Nutr.* **125**, 777-783.

Barnes, S. (1997) The chemopreventive properties of soy isoflavonoids in animal models of breast cancer. *Breast Cancer Res. Treatment* **46**, 169-179.

Barnes, S. (1998) Evolution of the history of soy and genistein. *Proc. Soc. Exp. Biol. Med.* **217**, 386-392.

Bienvenut, W. V., Deon, C., Pasquarello, C., Campbell, J. M., Sanchez, J. C., Vestal, M. L. and Hochstrasser, D. F. (2002) Matrix-assisted laser desorption/ionization-tandem mass spectrometry with high resolution and sensitivity for identification and characterization of proteins. *Proteomics* **2**, 868-876.

Boersma, B. J., Patel, R. P., Benton, M. R., Kirk, M., Wilson, L. S., Botting, N. P., Barnes, S. and Darley-Usmar, V. M. (2003) Neutrophil myeloperoxidase chlorinates soy isoflavones and enhances their antioxidant properties. *Free Radical Biol. Med.* **35**, 1417-1430.

Brookes, P. S., Pinner, A., Ramachandran, A., Coward, L., Barnes, S., Kim, H. and Darley-Usmar, V. M. (2002) High throughput 2D blue-native electrophoresis a tool for functional proteomics of mitochondria and signaling complexes. *Proteomics* **2**, 969-977.

Camacho-Carvajal, M. M., Wollscheid, B., Aebersold, R., Steimle, V. and Schamel, W. W. A. (2003) Two-dimensional blue native/SDS gel electrophoresis of multi-protein complexes from whole cellular lysates a proteomics approach. *Molecular Cellular Proteomics*, in press.

Carr, S. A., Huddleston, M. J. and Annan, R. S. (1996) Selective detection and sequencing of phosphopeptides at the femtomole level by mass spectrometry. *Anal. Biochem.* **239**, 180-192.

Celotto, A. M. and Graveley, B. R. (2001) Alternative splicing of the Drosophila Dscam pre-mRNA is both temporally and

spatially regulated. *Genetics* **159**, 599-608.

Chong, B. E., Yan, F., Lubman, D. M. and Miller, F. R. (2001) Chromatofocusing nonporous reversed-phase high-performance liquid chromatography/electrospray ionization time-of-flight mass spectrometry of protein s from human breast cancer whole cell lysates: a novel two -dimensional liquid chromatography/mass spectrometry method. *Rapid Commun. in Mass Spectrom.* **15**, 291-296.

Coward, L., Kirk, M, Albin, N. and Barnes, S. (1996) Analysis of Plasma Isoflavones by Reversed-Phase HPLC-Multiple Reaction Ion Monitoring-Mass Spectrometry. *Clin. Chim. Acta* **247**, 121-142.

Crow, J., Ye, Y.-Z., Strong, M., Barnes, S., Kirk, M. and Beckman, J. (1997) Superoxide dismutase catalyzes nitration of tyrosines by peroxynitrite in the rod and head of neurofilament-L. *J. Neurochem.* **69**, 1945-1953.

Davis, M. T., Beierle, J., Bures, E. T., McGinley, M. D., Mort, J., Robinson, J. H., Spahr, C. S., Yu, W., Luethy, R. and Patterson, S. D. (2001) Automated LC-LC-MS-MS platform using binary ion-exchange and gradient reversed-phase chromatography for improved proteomic analyses. *J. Chromatogr. B. Biomed. Sci. Appl.* **752**, 281-291.

For the latest information on the human genome, go to http://www.genome.gov/

Gavin, A. C., Bosche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J. M., Michon, A. M., Cruciat, C. M., *et al.* (2002) Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415**, 141-147.

Ge, X., Sirich, T. L., Beyer, M. K., Desaire, H. and Leary, J. A. (2001) A strategy for the determination of enzyme kinetics using electrospray ionization with an ion trap mass spectrometer. *Anal. Chem.* **73**, 5078-5082.

Geiman, T. M. and Robertson, K. D. (2002) Chromatin remodeling, histone modifications, and DNA methylation - how does it all fit together? *J. Cell Biochem.* **87**, 117-125.

Gerber, S. A., Rush, J., Stemman, O., Kirschner, M. W. and Gygi, S. P. (2003) Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc. Nat. Acad. Sci. USA* **100**, 6940-6945.

Gianazza, E., Eberini, I., Arnoldi, A., Wait, R. and Sirtori, C. R. (2003) A proteomic investigation of isolated soy proteins with variable effects in experimental and clinical studies. *J. Nutr.* **133**, 9-14.

Greis, K. D., Hayes, B. K., Comer, F., Hart, G. W., Kirk, M. and Barnes, S. (1996) Selective detection and site analysis of glycopeptides by capillary LC-ESMS. *Anal. Biochem.* **234**, 38-49.

Gronborg, M., Kristiansen, T. Z., Stensballe, A., Andersen, J. S., Ohara, O., Mann, M., Jensen, O. N. and Pandey, A. A. (2002) mass spectrometry-based proteomic approach for identification of serine/threonine-phosphorylated proteins by enrichment with phospho-specific antibodies: identification of a novel protein, Frigg, as a protein kinase A substrate. *Mol. Cell Proteomics* **1**, 517-527.

Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H. and Aebersold, R. (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature Biotech.* **17**, 994-999.

Helbeck, T., McComb, M. E., Lim, A., Schoneich, C., Cohen, R.

A. and Costello, C. E. (2003) An affinity labeling method for the identification of nitrotyrosine containing proteins and peptides. 51st Annual Conference of the American Society for Mass spectrometry, Montreal, Canada.

Hsieh, C. Y., Santell, R. C., Haslam, S. Z. and Helferich, W. G. (1998) Estrogenic effects of genistein on the growth of estrogen receptor-positive human breast cancer (MCF-7) cells in vitro and in vivo. *Cancer Res.* **58**, 3833-3888.

Imam-Sghiouar, N., Laude-Lemaire, I., Labas, V., Pflieger, D., Le Caer, J. P. Caron, M. and Nabias, D. K. (2002) Joubert-Caron R. Subproteomics analysis of phosphorylated proteins: application to the study of B-lymphoblasts from a patient with Scott syndrome. *Proteomics* **2**, 828-838.

Kaji, H., Saito, H., Yamauchi, Y., Shinkawa, T., Taoka, M., Hirabayashi, J., Kasai, K., Takahashi, N. and Isobe, T. (2003) Lectin affinity capture, isotope-coded tagging and mass spectrometry to identify N-linked glycoproteins. *Nature Biotechnol.* **21**, 667-672.

Katayama, H., Nagasu, T. and Oda, Y. (2001) Improvement of in-gel digestion protocol for peptide mass fingerprinting bymatrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Rapid Commun. in Mass Spectrom.* **15**, 1416-1421.

Katta, V., Chow, D. T. and Rohde, M. F. (1998) Applications of in-source fragmentation of protein ions for direct sequence analysis by delayed extraction MALDI-TOF mass spectrometry. *Anal. Chem.* **70**, 4410-4416.

Kim, H., Chaves, L., Hall, P., Mills, R., DeSilva, T., Coward, L. and Barnes, S. (2002) The use of proteomics technology to study brain proteins affected by soy isoflavones; in Soy and Health 2002 - clinical evidence - dietetic applications. K. Descheemaeker and I. Debruyne (eds.), pp. 155-166. Garant, Antwerpen.

Kvasnicka, F. (2003) Proteomics: general strategies and application to nutritionally relevant proteins. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **787**, 77-89.

Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680-685.

Lamartiniere, C. A., Cotroneo, M. S., Fritz, W. A., Wang, J., Mentor-Marcel, R. and Elgavish, A. (2002) Genistein chemoprevention: timing and mechanisms of action in murine mammary and prostate. *J. Nutr.* **132**, *552-558.*

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., *et al.* (2001) International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921.

Lanman, J., Lam, T., Barnes, S., Sakalian, M., Emmett, M. R., Marshall, A. G. and Prevelige, P. E. Jr. (2003) Identification of novel interactions in HIV-1 capsid protein assembly by high resolution mass spectrometry. *J. Mol. Biol.* **325**, 759-772.

Lee, C. -L., Hsiao, H. -H., Lin, C. -W., Wu, S. -P., Huang, S. -Y., Wu, C. -Y., Wang, Ah. -J. and Khoo, K. -H. (2003) Strategic shotgun proteomics approach for efficient construction of an expression map of targeted protein families in hepatoma cell lines. *Proteomics* **3**, 2472-2486.

Marshall, A. G., Hendrickson, C. L. and Jackson, G. S. (1998) Fourier transform ion cyclotron resonance mass spectrometry: a primer. *Mass Spectrom. Rev.* **17**, 1-35.

Medzihradszky, K. F., Gillece-Castro, B. L., Settineri, C. A., Townsend, R. R., Masiarz, F. R. and Burlingame, A. L. (1990) Structure determination of O-linked glycopeptides by tandem mass spectrometry. *Biomed. Environ. Mass Spectrom.* **19**, 777-781.

Mortz, E., Vorm, O., Mann, M. and Roepstorff, P. (1994) Identification of proteins in polyacrylamide gels by mass spectrometric peptide mapping combined with database search. *Biol. Mass Spectrom.* **23**, 249-261.

Noel-Georis, I., Bernard, A., Falmagne, P. and Wattiez, R. (2001) Proteomics as the tool to search for lung disease markers in bronchoalveolar lavage. *Dis. Markers* **17**, 271-284.

O'Farrell, P. H. (1975) High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.* **250**, 4007-4021.

Page, G. P., Edwards, J. W., Barnes, S. and Allison, D. B. (2003) A Design & Statistical Perspective on Microarray Gene. Expression Studies in Nutrition the need for playful creativity and scientific hard-mindedness. *Nutrition* **19**, 997-1000.

Pasquali, C., Fialka, I. and Huber, L. A. (1999) Subcellular fractionation, electromigration analysis and mapping of organelles. *J. Chromatogr. B Biomed. Sci. Appl.* **722**, 89-102.

Pertoft, H. (2000) Fractionation of cells and subcellular particles with Percoll. *J. Biochem. Biophys. Methods* **44**, 1-30.

Rejtar, T., Hu, P., Juhasz, P., Campbell, J. M., Vestal, M. L., Preisler, J. and Karger, B. L. (2002) Off-line coupling of high-resolution capillary electrophoresis to MALDI-TOF and TOF/TOF MS. *J. Proteome Res.* **1**, 171-179.

Reynolds, K. J., Yao, X. and Fenselau, C. (2002) Proteolytic ^{18}O labeling for comparative proteomics: evaluation of endoprotease Glu-C as the catalytic agent. *J. Proteome Res.* **1**, 27-33.

Roper, J. M., Staversky, R. J., Finkelstein, J. N., Keng, P. C. and O'Reilly, M. A. (2003) Identification and isolation of mouse type II cells on the basis of intrinsic expression of enhanced green fluorescent protein. *Am. J. Physiol. Lung Cell Mol. Physiol.* **285**, 691-700.

Schagger, H. and von Jagow, G. (1991) Blue native electrophoresis for isolation of membrane protein complexes in enzymatically active form. *Anal. Biochem.* **199**, 223-231.

Schmidt, F., Donahoe, S., Hagens, K., Mattow, J., Schaible, U. E., Kaufmann, S. H., Aebersold, R. and Jungblut, P. (2003) Complementary analysis of the Mycobacterium tuberculosis proteome by two-dimensional electrophoresis and isotope coded affinity tag technology. *Mol. Cell Proteomics*, in press.

Schmucker, D., Clemens, J. C., Shu, H., Worby, C. A., Xiao, J., Muda, M., Dixon, J. E. and Zipursky, S. L. (2000) Drosophila Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. *Cell* **101**, 671-684.

Sfakianos, M., Wilson, L., Sakalian, M., Falany, C. N. and Barnes, S. (2002) Conserved Residues in the Putative Catalytic Triad of Human Bile Acid Coenzyme A: Amino Acid N-acyltransferase. *J. Biol. Chem.* **277**, 47270-47275.

Shen, Y., Zhao, R. and Berger, S. J. (2002) Anderson GA. Rodriguez N. Smith RD. High-efficiency nanoscale liquid chromatography coupled on-line with mass spectrometry using nanoelectrospray ionization for proteomics. *Anal. Chem.* **74**, 4235-4249.

Spahr, C. S., Davis, M. T., McGinley, M. D., Robinson, J. H., Bures, E. J., Beierle, J., Mort, J., Courchesne, P. L., Chen, K.,

Wahl, R. C., Yu, W., Luethy, R. and Patterson, S. D. (2001) Towards defining the urinary proteome using liquid chromatography-tandem mass spectrometry. I. Profiling an unfractionated tryptic digest. *Proteomics* **1**, 93-107.

Unlu, M., Morgan, M. E. and Minden, J. S. (1997) Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. *Electrophoresis* **18**, 2071-2077.

Varnum, S. M., Covington, C. C., Woodbury, R. L., Petritis, K., Kangas, L. J., Abdullah, M. S., Pounds, J. G., Smith, R. D. and Zangar, R. C. (2003) Proteomic characterization of nipple aspirate fluid: identification of potential biomarkers of breast cancer. *Breast Cancer Res. Treat.* **80**, 87-97.

Venter, J .C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., *et al.* (2001) The sequence of the human genome. *Science* **291**, 1304-1351.

Viera, L., Ye, Y. Z., Estevez, A. G. and Beckman, J. S. (1999) Immunohistochemical methods to detect nitrotyrosine. *Methods Enzymol.* **301**, 373-381.

Wang, Y. K., Ma, Z., Quinn, D. F. and Fu, E. W. (2002) Inverse ^{15}N-metabolic labeling /mass spectrometry for comparative proteomics and rapid identification of protein markers/targets. *Rapid Commun. Mass Spectrom.* **16**, 1389-1397.

Washburn, M. P., Wolters, D. and Yates, J. R. (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.* **19**, 242-247.

Wells, L., Vosseller, K., Cole, R. N., Cronshaw, J. M., Matunis, M. J. and Hart, G. W. (2002) Mapping sites of O-GlcNAc modification using affinity tags for serine and threonine post-translational modifications. *Mol. Cell Proteomics* **1**, 791-804.

Witt, M., Fuchser, J. and Baykut, G. (2003) Fourier transform ion cyclotron resonance mass spectrometry with NanoLC/microelectrospray ionization and matrix-assisted laser desorption/ionization: analytical performance in peptide mass fingerprint analysis. *J. Am. Soc. Mass Spectrom.* **14**, 553-561.

Yamakawa, K., Huot, Y. K., Haendelt, M. A., Hubert, R., Chen, X. N., Lyons, G. E. and Korenberg, J. R. (1998) DSCAM: a novel member of the immunoglobulin superfamily maps in a Down syndrome region and is involved in the development of the nervous system. *Hum. Mol. Genet.* **7**, 227-237.

Yanagisawa, K., Shyr, Y., Xu, B. J., Massion, P. P., Larsen, P. H., White, B. C., Roberts, J. R., Edgerton, M., Gonzalez, A., Nadaf, S., Moore, J. H., Caprioli, R. M. and Carbone, D. P. (2003) Proteomic patterns of tumour subsets in non-small-cell lung cancer. *Lancet.* **362**, 433-439.

Yates, J. R. (1998) Mass spectrometry and the age of the proteome. *J. Mass Spectrom.* **33**, 1-19.

Zhang, K. and Tang, H. (2003) Analysis of core histones by liquid chromatography-mass spectrometry and peptide mapping. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **783**, 173-179.

Zhang, L., Eugeni, E. E., Parthun, M. R. and Freitas, M. A. (2003) Identification of novel histone post-translational modifications by peptide mass fingerprinting. *Chromosma* **112**, 77-86.

Zhou, J., Rusnak, F. and Hathaway, G. (2003) A method for the identification of post-translationally modified peptides by chemical targeting. 51st Annual Conference of the American Society for Mass spectrometry, Montreal, Canada.